

**Cloning, Sequencing and Partial Characterization of
the Accessory Gene Region of Plasmid pTC-F14
isolated from the Biomining Bacterium
Acidithiobacillus caldus f.**

Gunther Karl Goldschmidt



**Dissertation presented for the degree of Master of Sciences at the University
of Stellenbosch**

Supervisor: Prof. D. E. Rawlings

April 2005

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously, in its entirety or part, submitted in to any university for a degree.

.....

.....

(Gunther K. Goldschmidt)



Abstract

Plasmid pTC-F14 is a 14.2kb promiscuous, broad-host range IncQ-like mobilizable plasmid isolated from *Acidithiobacillus caldus* f. *At. caldus* is a member of a consortium of bacteria (along with *Acidithiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*) that is used industrially for decomposing metal sulphide ores and concentrates at temperatures of 40°C or below which is now a well-established industrial process to recover metals from certain copper, uranium and gold-bearing minerals or mineral concentrates. These biomining microbes are usually obligately acidophilic, autotrophic, usually aerobic iron- or sulphur-oxidizing chemolithotrophic bacteria. Their remarkable physiology allows them to inhabit an ecological niche that is largely inorganic and differs from those environments populated by the more commonly studied non-acidophilic heterotrophic bacteria. *At. caldus*, is a moderately thermophilic (45 to 50°C), highly acidophilic (pH1.5 to 2.5) sulphur-oxidizing bacterium, and its role as one of the major players in the industrial decomposition of metal sulphide ores has become evident in recent years. *At. caldus* f from which pTC-F14 was isolated was found to be one of two dominant organisms in a bacterial consortium undergoing pilot-scale testing for the commercial extraction of nickel from ores.

The majority of the plasmids studied to date have been isolated from a clinical or a non-acidophilic heterotrophic background and little is known about plasmids from acidophilic autotrophic microorganisms. Work on the plasmid biology of these microbes is important to the understanding of the nature of plasmids, and the genes which participate in the horizontal gene pool of biomining bacteria. In this study it has been of particular interest to discover what accessory genes are carried on plasmid pTC-F14 and this question was addressed by the cloning, sequencing and analysis of this region.

The accessory gene region of pTC-F14 has been sequenced and sequence analysis revealed six open reading frames which gave putative translation products of 9 kDa or larger. Two of these open reading frames, ORF13 and ORF9.5 gave no meaningful similarity hits using the BLAST program. ORF33 gave relatively weak similarity and identity to approximately one third of the amino acid sequence of an aminotransferase. The remaining three open reading frames showed strong amino acid sequence similarities and identities to protein sequences already deposited in the NCBI database. ORF20.8 had the strongest match to an invertase or recombinase gene. ORF17.4 showed a close relationship to a hypothetical protein that is highly conserved in a

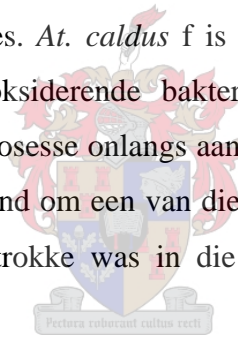
wide variety of bacteria. The remaining ORF had very high sequence identity to the transposase of an insertion sequence, *ISAfeI*, previously identified in a different but related biominer bacterium, *At. ferrooxidans*. Analysis of the nucleotide sequence of the regions upstream and downstream of this open reading frame revealed that it conformed to the criteria of a bacterial insertion sequence element and was designated here as *ISAtcI*.

In this study an attempt was made to partially characterize this insertion sequence. The study revealed that *ISAtcI* is 1,303bp in size with imperfectly conserved 26-bp terminal inverted repeats. Amino acid sequence comparisons revealed that it is 92% identical to *ISAfeI* and the analysis of the overall organization of *ISAtcI* showed that it is also a member of the ISL3 family of insertion sequences. We have shown that *ISAtcI* is present on the chromosome of three *At. caldus* strains isolated from South Africa but not present in three *At. caldus* strains from Europe or Australia. The presence of insertion sequences on both a plasmid and the chromosome allows for plasmids to integrate into the chromosome and provides an enhanced level of genome plasticity. We showed that *ISAtcI* is actively transposing in its natural *At. caldus* host. This study also revealed that *ISAtcI* has the ability to form cointegrate-like structures in an *Escherichia coli* background.



Opsomming

Die plasmied pTC-F14 is 'n 14.2-kb IncQ-tipe, mobiliseerbare plasmied geisoleer uit *Acidithiobacillus caldus* f. pTC-F14 kan in 'n wye reeks gashere repliseer en tesame met die vermoë om gemobiliseer te word is die plasmied 'n hoogs uitruilbare plasmied. *At. caldus* is 'n lid van 'n konsortium van bakterieë (tesame met *Acidithiobacillus ferrooxidans* en *Leptospirillum ferrooxidans*) wat industrieel aangewend word in die biologiese afbraak van metaal sulfied ertse by temperature van 40°C of laer. Die aanwending van hierdie mikrobies in die biologiese myning van sekere koper, uranium en goud-bevattende minerale of mineraal konsentrate is tans 'n goed-ontwikkelde kompeterende, maar steeds groeiende, industriële fermentasie proses. Hierdie organismes is gewoonlik verpligte asidofiliese, outotrofiese, aerobiese yster-en swawel-oksiderende chemolitotrofiese bakterieë. Hulle interessante en unieke fisiologie maak dit vir hulle moontlik om 'n hoogs anorganiese ekologiese nis te bevolk wat baie verskil van daardie omgewings wat bevolk word deur die meer algemene nie-asidofiliese heterotrofiese organismes. *At. caldus* f is 'n matigde termofiel (45-50°C), hoogs asidofiliese (pH1.5-2.5), swawel oksiderende bakterium en wie se rol as een van die domineerende organismes in hierdie prosesse onlangs aan die lig gekom het. Die ras *At. caldus* f, waaruit pTC-F14 geisoleer was gevind om een van die twee dominerende organismes te wees in 'n bakteriese konsortium wat betrokke was in die kommersiële ekstraksie van nikel uit mineraal ertse.



Die meerderheid van plasmied studies tot datum was gedoen op plasmiede wat geisoleer was vanuit 'n kliniese of 'nie-asidofiliese heterotrofiese gashere en min kennis aangaande plasmiede afkomstig vanuit asidofiliese mikrobies. Biologiese plasmied studies van hierdie mikroorganismes is belangrik vir die begrip van die aard en gedrag van plasmiede, en die gene wat bydrae en deelneem aan die horisontale geen poel van die hierdie lae pH, hoogs anorganiese bakterieë. Dit was spesifiek interessant om te ontdek watter bykomstige gene gedra word op die plasmied pTC-F14 en hierdie vraag is aangespreek deur die klonering, DNA volgordebepaling en analise van hierdie area.

Die bykomstige geen area van pTC-F14 se DNS volgorde was bepaal en die DNS volgorde is ge-analiseer. Ses ooplesrame was geïdentifiseer met waarskynlike translasië produkte van 9 kDa of groter. Twee van hierdie ooplesrame, ORF13 en ORF9.5 het geen rekenbare ooreenkoms en identiteit getoon tydens 'n aminosuurvolgorde vergelyking met die BLAST

program. ORF33 het 'n relatiewe swak ooreenkoms en identiteit getoon met gemiddeld een derde van die aminosuurvolgordes van 'n aminotransferase. Die oorblywende drie ooppleesrame het sterk aminosuurvolgorde ooreenkomste en identiteite getoon met aminosuurvolgordes wat alreeds gedeponeer was in die NCBI databasis. ORF20.8 het die grootste ooreenkoms getoon met 'n invertase of rekombinase geen. ORF17.4 het 'n baie nou verhouding getoon met 'n hipotetiese protien wat hoogs gekonserveerd is in 'n wye reeks van bakterieë. Die oorblywende ooppleesraam het hoë volgorde identiteit en ooreenkoms op die aminosuur en DNS vlak getoon met 'n transposase geen van 'n eenvoudige transposon, *ISAFE1*, voorheen geïdentifiseer in 'n verskillende maar naverwante organisme, *At. ferrooxidans*. Analise van die nukleotied volgordes van die areas opstroom en afstroom van hierdie ooppleesraam het getoon dat dit voldoen aan die kriteria van 'n eenvoudige bakteriese transposon en is hier genoem *ISATC1*.

In hierdie studie was dit aangewend om hierdie eenvoudige transposon gedeeltelik te karakteriseer. Dit is bevestig dat *ISATC1* 1303bp in grootte is met 26-bp onperfekte gekonserveerde terminale herhalings. Aminosuur volgorde vergelykings het getoon dat *ISATC1* is 92% identities aan *ISAFE1* is en analise van die algemene organisasie van *ISATC1* het getoon dat 'n lid is van die ISL3 familie van eenvoudige bakteriese transposons. Die studie het getoon dat *ISATC1* is teenwoordig op die chromosoom van drie *At. caldus* rasse geïsoleer vanuit Suid-Afrika maar is nie teenwoordig op die chromosoom van drie *At. caldus* rasse van Europa of Australia. Die teenwoordigheid van hierdie element op beide die chromosoom en die plasmied pTC-F14 skep 'n ideale toestand vir plasmiede om te integreer in die chrosomale DNS en verhoog die plastisiteit van die genoom. In die studie is ook getoon dat *ISACT1* funksioneel is binne die natuurlike gasheer naamlik *At. caldus* f asook binne 'n *Escherichia coli* gasheer.

Acknowledgements

I wish to thank the National Research Foundation (Pretoria, South Africa) and BHP Billiton Minerals Technology (Randburg, South Africa) whom has funded this work. Thanks are also given to the University of Stellenbosch whom has financially supported me in the form of the Stellenbosch 2000 merit bursary.

My deepest gratitude goes out to the following people:

Professor Douglas Rawlings for his unswerving patience, encouragement and concern at difficult times and who always had time to give advice when needed. His relentless experience, skill, determination and his quenchless passion for Science made it an honor to study under his supervision.

Dr. Shelly Deane for her support and ever-willingness to help. Thank you for helping me take that first few steps into the research world.

I wish to thank the rest of the BRG (Biomining Research Group) lab members and members and friends in the Department of Microbiology for their loyal friendship and support.

A special thank you to my beloved family, especially my parents, brother and sister, for their love, constant support and patience throughout my years of study especially during the time I was writing this thesis. Thank you for always giving me the comfort that everything is possible through GOD.

I wish to thank friends outside the Department of Microbiology for their loyalty, righteousness and for the true friendships that have evolved over the years.

In fear of that I might have forgotten to thank someone, I wish to thank all who have assisted me in some way or the other in completing this work.

Most importantly I want to acknowledge that all the support I was given throughout the years and all of my achievements could not have been accomplished without GOD, the divine force that guides my life. I wish to thank all my Blessing Bearers for their support, love, prayers and spiritual guidance throughout my life.

Dedication

This work is dedicated to my family, my parents, my brother and sister. GOD could not have given me a greater gift and this is one way to say thank you and that I appreciate everything you have done for me throughout my life.



Table of Contents

1. Chapter 1: Literature review	2
1.1. Microbial Mining: A General Introduction	3
1.2. Historical background.....	5
1.3. Mechanisms of leaching	6
1.4. Current commercial applications of biomining	13
1.5. Biomining Microbes	18
2. Bacterial Insertion Sequences	21
2.1. Introduction	21
2.2. General Structure and Functional Properties of IS elements.....	24
2.3. Components and Mechanisms involved in Transposition.....	26
2.4. Control of Transposition Activity.....	48
2.5. Transposition Immunity	57
2.6. Target Specificity	58
2.7. Genetic effects of the IS elements	60
2.7.1. IS-mediated DNA rearrangements	60
2.7.2. Agents of mutations and effects on transcription	61
2.8. The IS families.....	63
2.9. Significance and evolutionary implications of IS elements	65
2.10. The Aim of the study.....	69
2. Chapter Two: The cloning, sequencing and analysis of the accessory gene region of the plasmid pTC-F14, isolated from <i>Acidithiobacillus caldus</i> f.	70
2.1. Introduction	71
2.2. Material and Methods	73
2.3. Results	76
2.4. Discussion.....	89
3. Chapter 3: Partial characterization of ISAtcI from plasmid pTC-F14, a new member of the ISL3 family of insertion sequences	91
3.1. Introduction	92
3.2. Materials and Methods	93
3.3. Results	97
3.4. Discussion.....	117
4. Chapter 4: General Discussion and Perspectives	121
4.1. General Introduction.....	122
4.2. Plasmid pTC-F14 in perspective	123
Appendix One: Annotated sequence of the accessory gene region	128
Literature cited	137

1. Chapter 1: Literature review

Table of Contents

1.1. Microbial Mining: A General Introduction	3
1.2. Historical background.....	5
1.3. Mechanisms of leaching	6
1.4. Current commercial applications of biomining.....	13
1.4.1. Irrigation type processes	14
1.4.2. Stirred tank processes	16
1.5. Biomining Microbes	18
1.5.1. Microbial dominance in biomining processes.....	20
2. Bacterial Insertion Sequences	21
2.1. Introduction	21
2.2. General Structure and Functional Properties of IS elements.....	24
2.3. Components and Mechanisms involved in Transposition.....	26
2.3.1. Conservative Transposition	27
2.3.2. Replicative Transposition	30
2.3.3. The Reaction Mechanisms of Transposition	31
2.3.3.1. First strand Donor Cleavage.....	32
2.3.3.2. Second Strand Donor Cleavage.....	34
2.3.3.3. The Strand Transfer Reaction.....	35
2.3.3.4. The Post Transfer Processing Reaction.....	36
2.3.4. The Role of Divalent Metal Ions.....	36
2.3.5. The Transposase	37
2.3.5.1. DNA binding domain	39
2.3.5.2. The Catalytic Domain.....	39
2.3.5.3. Multimerization	44
2.3.6. The Synaptic Complex (The Transposome).....	45
2.4. Control of Transposition Activity.....	48
2.4.1. Transposase Expression and Activity.....	49
2.4.1.1. Production of Translational Inhibitors.....	49
2.4.1.2. Impinging Transcription	49
2.4.1.3. Programmed Translational Frameshifting.....	50
2.4.1.4. Transposase Stability	51
2.4.1.5. Translation Termination	52
2.4.1.6. Activity in cis.....	52
2.4.2. Host Factors in Transposition.....	52
2.4.2.1. Histone-like proteins.....	53
2.4.2.2. Dam Methylase.....	54
2.4.2.3. DNA polymerase I.....	55
2.4.2.4. DNA gyrase	56
2.4.2.5. DnaA protein	56
2.4.2.6. RecA protein.....	56
2.4.2.7. The ClpX, ClpP and Lon proteases	57
2.5. Transposition Immunity	57

2.6. Target Specificity	58
2.7. Genetic effects of the IS elements	60
2.7.1. IS-mediated DNA rearrangements	60
2.7.2. Agents of mutations and effects on transcription	61
2.8. The IS families.....	63
2.9. Significance and evolutionary implications of IS elements	65
3. The Aim of the study	69

1.1. Microbial Mining: A General Introduction

Microbial encounters with metals and metalloids are unavoidable in the environment; it is therefore not surprising that they should interact. These interactions are sometimes for their benefit and sometimes to their detriment (Ehrlich, 1997). Metals of particular interest include the base metals vanadium, chromium, manganese, iron, cobalt, nickel, copper, zinc, molybdenum, cadmium and lead and the precious metals gold and silver. Metalloids include arsenic, selenium and antimony. Whether prokaryotic or eukaryotic, all microbes require metals that play a role in structural and/or catalytic functions. The levels of interaction with metals depends in part whether the organism is prokaryotic or eukaryotic. Some metals can exist in more than one oxidation state and several prokaryotes can employ them as electron donors or acceptors in their energy metabolism. Such metals include Fe, Co, Cu, As and Se. When used for this purpose, sufficiently high concentrations are needed to meet the demand of the organisms, resulting in a noticeable impact on the metal distribution in the environment. It is these interactions that have a potential role in biotechnological applications and are commercially exploited. One of the most important biotechnological applications is biomining, which is the main focus of this part of this chapter.

Biomining is a general term that refers to bioleaching and biooxidation processes. Bioleaching or bacterial leaching, as it is commonly also referred to, is generally accepted as the conversion of an insoluble metal, usually a metal sulphide, into its water soluble form, usually a sulphate. The metal is extracted into water and the metal can be said to have been bioleached (Kelly *et al*, 1979). Biooxidation describes the microbial decomposition of the host minerals that contain the metal compound of interest, but the metal is not solubilized. As a result, the metal values remain in the solid residues in a more accessible form (Rawlings, 2002). An example is in the recovery of gold from refractory arsenopyrite ores where the gold remains in the mineral after

biooxidation and is extracted subsequently by cyanidation. All these processes are oxidations performed by specialized chemolithotrophic organisms.

Prior to using bioleaching technologies in the recovery of metals, conventional methods such as roasting or pressure leaching of sulphide minerals were used. The cost of these operations was high and the pollution levels were unacceptable. The use of microbially-assisted mining has grown in recent years (Rawlings and Silver, 1995). The growing interest in such processes is attributed to the following:

- a) In many places, high-grade mineral deposits resources have been depleted resulting in the tendency for mining to be extended deeper underground and miners being forced to mine lower-grade surface deposits (Brierley, 1982; Rawlings and Silver, 1995). The recovery of metals from low-grade ores using the conventional physico-chemical methods is uneconomic and bioleaching provides a more economical method for the recovery of metals from these low-grade mineral deposits. Bacterial leaching can be as effective in removing all the metal from a 0.3% ore as from a 0.03% ore (Rawlings and Silver, 1995). Leaching of metals from deep or low-grade ores could also be carried out *in situ* thereby saving the cost of bringing the waste rock to the surface.
- b) Biological processes, such as bioleaching, are generally less energy intensive in comparison to the traditional methods used in the treatment of recalcitrant ores. An example is the consumption of large quantities of energy metabolism when using traditional roasting methods to pre-treat recalcitrant gold-bearing arsenopyrite ores (Ehrlich and Brierley, 1990).
- c) Bioleaching procedures are environmentally less hazardous and also have potential environmental benefits over physico-chemical treatment methods. The production of sulphur dioxide and other environmentally harmful gaseous emissions (e.g. arsenic-laden flames) as experienced during roasting of ores does not occur during microbial extraction. Mine tailings produced from physicochemical processes are more chemically active and when exposed to rain and air may be biologically leached, resulting in the formation of acid mine drainage and unwanted metal pollution (Schippers *et al*, 1996). Tailings resulting from biomining operations are less chemically active and their potential to pollute is reduced by the extent to which they have already been leached in a contained process.

It is evident from the brief survey presented that there is a strong case for the increase in implementation of microbial mining processes where suitable minerals exist. Biomining offers

distinct advantages that include operational simplicity, low capital and operating costs and shorter construction time which no other alternative process can provide.

1.2. Historical background

The Roman writer Gauss Plinius Secundus (23-79AD) gave one of the first reports where leaching might have been involved in the mobilization of metals (reviewed in Brandl, 2001). He described in his work on natural sciences how copper minerals were obtained using a leaching process. The Rio Tinto mines are usually considered as the cradle of biohydrometallurgy, which includes the field of bioleaching. These mines have been exploited since pre-Roman times for their copper, gold and silver values. Bioleaching of copper from ores was practised for many centuries before the discovery of bacteria. The mining processes were purely empirical then, without recognizing the participation of microbes in the leaching process. Documentation of commercial copper leaching from partially roasted ores at the Rio Tinto mines dates back to 1752 (Rossi, 1990), but it is unclear to what extent bacteria were involved in this process.

The first bacterium isolated from a coal mine drainage was, *Acidithiobacillus ferrooxidans* (previously, *Ferrobacillus ferrooxidans* and later *Thiobacillus ferrooxidans*), and its discovery was reported in 1947 (Colmer and Hinkle, 1947). The genus, *Thiobacillus*, was recently subdivided and a new genus, *Acidithiobacillus* was created, to accommodate the highly acidophilic members of the former genus *Thiobacillus* (Kelly and Wood, 2000). The biotechnology of microbial mining has grown significantly since the discovery of this bacterium, with the first demonstration of microbial involvement in copper leaching reported in the 1950s, following reports on the role of bacteria in the formation of acid mine drainage from bituminous coal deposits (Ehrlich, 1997). Since these recordings many research groups demonstrated microbial assisted leaching of metal sulfides which include ZnS, NiS, and PbS as well as molybdenum (Bryner and Anderson, 1957; Bryner and Jameson, 1958).

Intensive research and development by a variety of scientists from a variety of disciplines since the 1960's, yielded much information regarding microbial mining. It is impossible to mention all the contributions made by the scientists in this short review. In general it includes the following: identification of microbes capable of growing in biomining environments and phenotypic characterization of leaching bacteria; describing microbial-mineral interactions during bioleaching processes, elucidation of physicochemical and microbiological parameters

influencing bacterial mineral biooxidation, much genetic research has been done on *Acidithiobacillus ferrooxidans*, which has been considered for a long time to be the major role player in bioleaching processes; the elucidation of copper dump leach operations and the improvement of mineral sulphide oxidation by employing stirred tank reactors for bioleaching. Commercial bioleaching of copper began in the 1950's with dump leaching. Today heap and dump leaching remains a vital process for the copper industry. These processes are discussed below in more detail.

It is apparent from the brief foregoing survey that the science of biohydrometallurgy using acidophilic autotrophs has significantly progressed in its development as a commercial technology for the processing of sulphide ores. Elucidation of microbial ecology studies of extremely acidic, metal rich environments has laid the foundation for the development for mineral technologies, which is now well established in the mining industry.

1.3. Mechanisms of leaching

The iron- and sulphur-oxidising bacterium *At. ferrooxidans* was the first bacterium associated with the dissolution of metals from ores (Bryner and Anderson, 1957; Colmer and Hinkle, 1947). The chemical oxidation of ferrous iron by dissolved oxygen occurs very slowly. *At. ferrooxidans* has the ability to oxidize ferrous iron at a rate 500,000 times as fast as would occur in their absence (abiotic) (Lacey and Lawson, 1970). The ferric iron produced, as a result of the oxidation of ferrous iron, is one of the least expensive and most effective natural metal-oxidizing agents (Brierley, 1978) and is capable of solubilizing metal sulphides.

The mechanism of microbially assisted biooxidation of sulphide minerals has interested researchers for a long time (Rossi, 1990; Sand *et al* 1995). Most work regarding the mechanism of dissolution has been done with *At. ferrooxidans* and almost since the time of its discovery two mechanisms have been debated: direct and indirect. There have been heated discussions on the role microorganisms play in the leaching of minerals, and whether the so-called direct or indirect mechanism is used (Lundgren and Silver, 1980). Much of the disagreement has been caused by a lack of clarity as to what is meant by direct and indirect. Especially the concept of the direct mechanism has been to some extent imprecise and equivocal. Many researchers have reported experiments either confirming or rejecting the hypothesis (Sand *et al*, 2001). As a result, the final outcome of this discussion is still pending.

The indirect mechanism is generally accepted as the chemical attack (oxidation) by ferric ions or protons on a metal sulphide. In the course of the chemical reaction ferrous irons and various

forms of sulphur are formed. Ferrous iron serves as an electron donor for iron-oxidizing microbes in their energy metabolism, which re-oxidise it to ferric ions thus restoring the mineral oxidant. Because of this shuttle mechanism direct contact between the bacterium and the sulphide mineral is not necessary. The definition of the direct mechanism is not always clear. An assumption is that the microbes must be attached to a metal sulphide but it is not always clear whether it is only the requirement for attachment that is referred to as the direct mechanism or whether the oxidation of the metal is due to direct enzymatic attack on the metal and sulphide moiety of the mineral (Sand *et al*, 1995). In its strict sense, the direct mechanism refers to a direct enzyme-facilitated attack on a mineral.

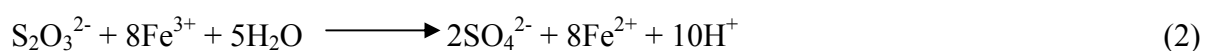
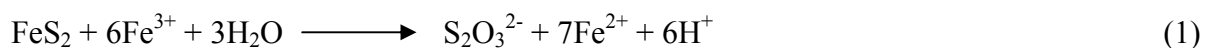
Studies on the attachment of *At. ferrooxidans* to mineral samples indicate that the association is quite tenacious and that bacteria are found in the pitted and eroded surfaces of mineral particles (Brierley, 1978). The attachment of leaching bacteria to the mineral surface is thought to enhance dissolution (Schippers *et al*, 1996). It is mainly for this reason that the direct mechanism was proposed. Several inconsistencies arose from experimental data that shed serious doubt on whether the direct mechanism existed (reviewed in Sand *et al*, 1995). This problem has been addressed by scientists from a variety of disciplines. However, in recent years, new insights have been derived by combining knowledge obtained from areas like sulphur chemistry, mineralogy and solid state physics with evidence obtained from analysis of degradation products occurring during metal sulphide dissolution and the analysis of extracellular polymeric substances (EPS layers) (Sand *et al*, 2001). The focus of this survey is to briefly cover the most important contributions made by researchers in elucidating the mechanisms of mineral dissolution. Helpful contributions were made from the laboratories of Wolfgang Sand, Frank Crundwell and Helmut Tributsch.

Sand *et al* (1995) proposed that bacterial leaching of metal sulphides proceeds strictly via the indirect mechanism, which is initiated by ferric iron. In their report they presented experimental data that places a considerable amount of doubt on the existence of the direct mechanism. In their hypothesis of the indirect action they emphasize the important role played by Fe^{3+} (ferric iron) bound in the EPS layers of microbial cells and the essential role of these layers in the attachment of microbes to the mineral surface (Gherke *et al*, 1998). This will be returned to later.

Using molecular orbital and valence band theories, Sand and co-workers realised that metal sulphides have different types of crystal structures and they observed that the oxidation of

different metal sulphides proceeds via different intermediates. Based on these observations they proposed two indirect oxidation mechanisms for mineral sulphide dissolution (Schippers and Sand, 1999; Sand *et al* 2001).

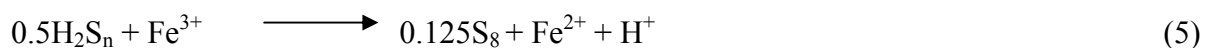
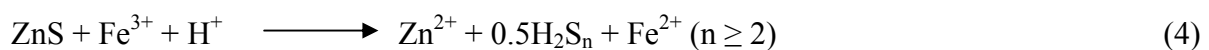
They proposed the thiosulphate mechanism for the oxidation of acid-insoluble metal sulphides such as pyrite (FeS₂), molybdenite (MoS₂) and tungstenite (WS₂). Degradation studies on these sulphides indicated they are degraded by an oxidizing attack by ferric iron with thiosulphate being the main intermediate and sulphate the main end product. Using Pyrite as an example the reactions proposed by Schippers and Sand (1999) are:



Thiosulphate is supposedly formed from the disulphide in the crystal (equation 3):



The dissolution of acid-soluble metal sulphides is proposed to occur by the polysulphide mechanism. Acid soluble metal sulphides include sphalerite (ZnS), galena (PbS), realgar (As₄S₄) and chalcopyrite (CuFeS₂). Their dissolution occurs via combined attack by ferric iron (Fe³⁺, for a oxidation attack) and protons (H⁺, for a hydrolysis attack). Elemental sulphur is the main intermediate with the possible formation of polysulphides. Elemental sulphur is stable over a variety of conditions but may be biologically oxidized to sulphuric acid by sulphur oxidizing bacteria. Using zinc as an example of an acid soluble mineral, the following equations are proposed by Schippers and Sand (1999):



Schippers and Sand (1999) also point out that these two proposed mechanisms explain why the sulphur oxidizing bacterium, *Acidithiobacillus thiooxidans* is able to solubilize some metal sulphides, the acid soluble sulphides, that are amenable to a proton attack (hydrolysis), but not acid-insoluble minerals. An overview of these two mechanisms is depicted in Figure 1.1. Included in the scheme are both abiotic (chemical) and biotic-catalyzed parts. The finding that

only iron oxidizing bacteria are able to dissolve pyrite emphasises the importance of ferric irons as the principal pyrite attacking agent and therefore, supports the hypothesis of the indirect leaching mechanism as the basic mechanism for the biooxidation of mineral sulphides.

It is evident from the data above that Sand and co-workers regard ferric iron (Fe^{3+}) and protons (H^+) as the only (chemical) agents involved in dissolving a metal sulphide and the mechanism involved is *sensu strictu* an indirect one. Therefore, the role of microorganisms is only to regenerate sulphuric acid for a proton hydrolysis attack and to oxidize ferrous iron (Fe^{2+}) to ferric irons (Fe^{3+}) for an oxidative attack on the sulphide mineral (Schipers and Sand, 1999). These reactions are purely chemical and do not require direct contact/attachment of the

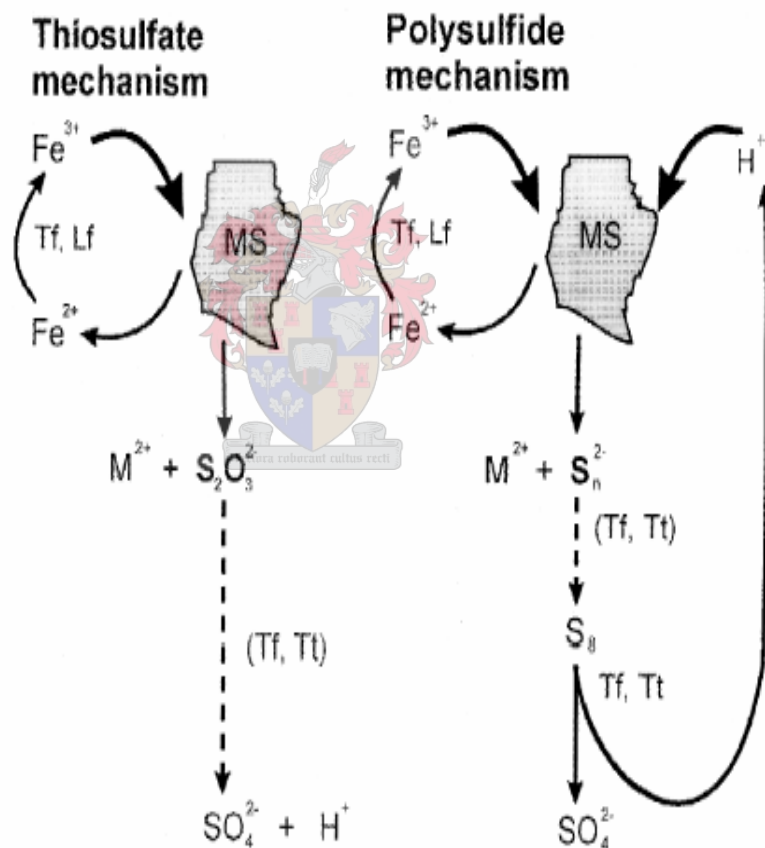


Figure 1.1: Bioleaching of metal sulphides (MS) via the thiosulphate and polysulphide mechanisms (Reprinted from Sand *et al*, 2001).

microbes to the mineral surface for dissolution . Mineral attachment may enhance the rate of mineral leaching as it brings the means of ferric iron and acid production closer to the mineral increasing the concentration of reactants at the surface of the mineral. However, the mechanism of mineral solubilization remains chemical and therefore indirect (Fig 1.2) (Rawlings, 2002;

Tributsch, 2001). Blake *et al* (1994) demonstrated that ferric irons are needed to overcome the repellent effect that exists between the negatively charged sulphide minerals and bacterial cells. Ferric ions are thus of crucial importance for cell attachment.

Besides the production of ferric iron and acid, a second function of the bacteria is to concentrate the ferric irons and protons at the mineral/water or mineral/bacterial cell interface in order to enhance the dissolution of mineral sulphide. A critical factor here is the attachment of the bacteria to the mineral surface which is mediated by the extracellular polymeric substances (EPS layers) (Schippers *et al*, 1996). Sand and Tributsch have stressed the role played by the EPS layers produced by *At. ferrooxidans* (Gherke *et al*, 1998) and *L. ferrooxidans* (Tributsch, 2001) when attached to the mineral. The EPS layers assist the bacteria to attach to the mineral and form a matrix wherein the cells divide and eventually form a biofilm (Rawlings, 2002). The analysis of the chemical composition of these EPS layers showed that it is primarily made up of lipopolysaccharides complexed with ferric irons (Sand *et al*, 2001). The iron species are presumably bound to the glucuronic acid subunits of the carbohydrate moiety. The EPS layers contain a considerable amount of iron, estimated to be ~53 g/l which is a considerable higher concentration than in the non-EPS phase (Schippers and Sand, 1999). Therefore, ferric ion-impregnated EPS layers serves as a reaction space in which ferric ions carry out the primary steps in the degradation of metal sulphides. Ferric iron is reduced to ferrous iron, which in turn is re-oxidized to ferric irons by the specialized iron-oxidizing bacteria:



EPS layers may also be interpreted as a compartment where other special conditions prevail that are different from the bulk phase, e.g. altered pH and redox potential. An interesting paper by Fowler *et al* (1999) reports that the dissolution of pyrite at fixed concentrations of ferrous and ferric irons was faster in the presence of *At. ferrooxidans* because of an increase in the localized rise in pH caused by bacterial attachment. The question of why the rate of mineral oxidation is enhanced by attachment if mineral attack is purely chemical becomes explainable when considering the concentration of the degradative agents at the mineral interface within the EPS. The role of the EPS also explains the localized etching that occurs on the sulphide surface, once taken as support for the direct attack (Bennet and Tributsch, 1978; Edwards *et al*, 1999).

Comparative leaching experiments performed with 16 different synthetic metal sulphides (Tributsch and Bennet, 1981a; Tributsch and Bennet, 1981b) revealed that the bacterial activity of *At. ferrooxidans* and thus the rate of bacterial dissolution, is approximately proportional to the solubility product of the sulphide concerned (Tributsch, 2001). The solubility product of sulphides describes the reactivity of protons with a sulphide resulting in the disintegration of the sulphide into SH^- and metal ions. Sulphur oxidizing bacteria can use SH^- as an energy source. In addition to this mechanism, he argues for four mechanisms that can increase sulphide dissolution. These processes include: (a) extraction of electrons by Fe^{3+} from the sulphide valence band resulting in the disruption of the interfacial bonds of the sulphide crystal and thus liberating metal ions and sulphur; (b) broken chemical bonds already present in the sulphide (p-type conduction, their presence in sulphides leads to the higher rate of interfacial dissolution); (c) reaction with a polysulphide or metal complex-forming agent; or (d) the electrochemical dissolution that results from multiple electron extraction and depolarization of the pyrite, which occurs at high concentrations of ferric irons.

Interfacial reactions of sulphides are determined by the co-ordination chemistry of the metal. Based on this theory Tributsch (2001) explains why pyrite, FeS_2 , is so stable and why electron extraction does not directly lead to the disintegration of the sulphides. Pyrite is the most abundant sulphide in nature and bacteria had to evolve specialized mechanisms to disintegrate this sulphide for energy resources. These workers propose that *At. ferrooxidans* has acquired the ability to use an unidentified carrier that results in a polysulphide intermediate (mechanism in “c”). This carrier molecule is able to disrupt pyrite and it has been found that this carrier molecule works using a thiol-group (SH^-) provided by the amino acid cysteine. To support this view workers have shown that cysteine on its own can oxidize pyrite in the absence of oxygen or bacteria (Rojas-Chapana and Tributsch, 2001). As many proteins contain cysteine, it raises the possibility of a direct protein attack on the mineral. Although thiol-dependant mechanisms may contribute a small amount to overall leaching, as a sole mechanism, it is insufficient to oxidize pyrite at the rates observed (Rawlings, 2002).

L. ferrooxidans is an organism that can only oxidize ferrous iron and not sulphur. Nevertheless, it has acquired the ability to dissolve pyrite electrochemically (mechanism in “d”). As mentioned above, electron extraction alone does not break chemical bonds in pyrite, but when many electrons are extracted the electrical potential of pyrite becomes so positive that the electrochemical formation of thiosulphate and sulphide occurs (Tributsch, 2001). This means that *L. ferrooxidans* has learned to use a sufficiently large concentration of Fe^{3+} in an

interfacial reaction with pyrite converting interfacial S_2^- into sulphate. *L. ferrooxidans* is able to gain redox energy from Fe^{2+} at a very positive redox potential, meaning that their ability to oxidize ferrous iron is not inhibited by ferric iron, unlike is the case with *At. ferrooxidans*.

It is clear that the mechanisms involving strictly H^+ and Fe^{3+} (proton interaction with the sulphide; electron extraction by Fe^{3+} and p-type conduction as described above) can be considered as indirect leaching (non-contact leaching) since the attachment of bacterial cells is not necessary, but only to recycle the chemical agents. On the other hand, the specialized mechanism employed by *At. ferrooxidans* (sulphur carrier) and *L. ferrooxidans* (electrochemical dissolution of pyrite), requires close proximity to the mineral surface. They must be able to produce a sulphur carrier molecule and artificially increase the electron extracting agent (Fe^{3+}), close to the pyrite surface. The mechanisms would not be as effective if they relied on carriers or chemicals available in the bulk liquid. An artificially controlled reaction zone between the bacterial membrane and the mineral surface is needed. Such a reaction zone is provided by the EPS layers. The cysteine-based sulphur carrier, in *At. ferrooxidans*, is carried through the EPS layers to the pyrite-sulphur and forming polysulphide. Colloidal sulphide is stored in the EPS layers and as a temporary energy reservoir (Rojas *et al*, 1995). A sufficiently high Fe^{3+} concentration is accumulated in the EPS layer in the case of *L. ferrooxidans* in order to extract electrons so that a sufficiently positive redox potential is created which induces the electrochemical formation of thiosulphate and sulphate. This means that a close contact across the EPS layer with the mineral surface is essential to dissolve it. Previously direct leaching implied a direct interaction between the bacterial membrane and the sulphide surface, using an enzyme system to dissolve the mineral surface. Electron microscopy does not confirm this and it has been suggested that direct leaching should be renamed contact leaching (Fig 1.2) because of the confusion around the term direct leaching. Contact leaching can thus be defined as the process where the bacterium attaches itself (EPS layers) to the mineral surface with the purpose to create special conditions which can facilitate the disintegration of the mineral sulphide. This definition does not need the presence of an enzyme system, which up to now has not yet been found, although a small contribution to mineral solubilization by cysteine containing proteins may occur.

Recently it has been shown that cooperative leaching (Fig 1.2) is possible (Rojas-Chapana *et al*, 1998), during which there is cooperation between the attached and free bacteria. *At. ferrooxidans* and *L. ferrooxidans* leach pyrite releasing sulphur compounds and small pyrite

crystals while being attached to the mineral surface and thereby liberating free energy carrying species which other non-attached bacteria feed on.

In conclusion, it is evident from the data presented above that there is a growing consensus view that mineral bioleaching occurs primarily via the indirect mechanism.

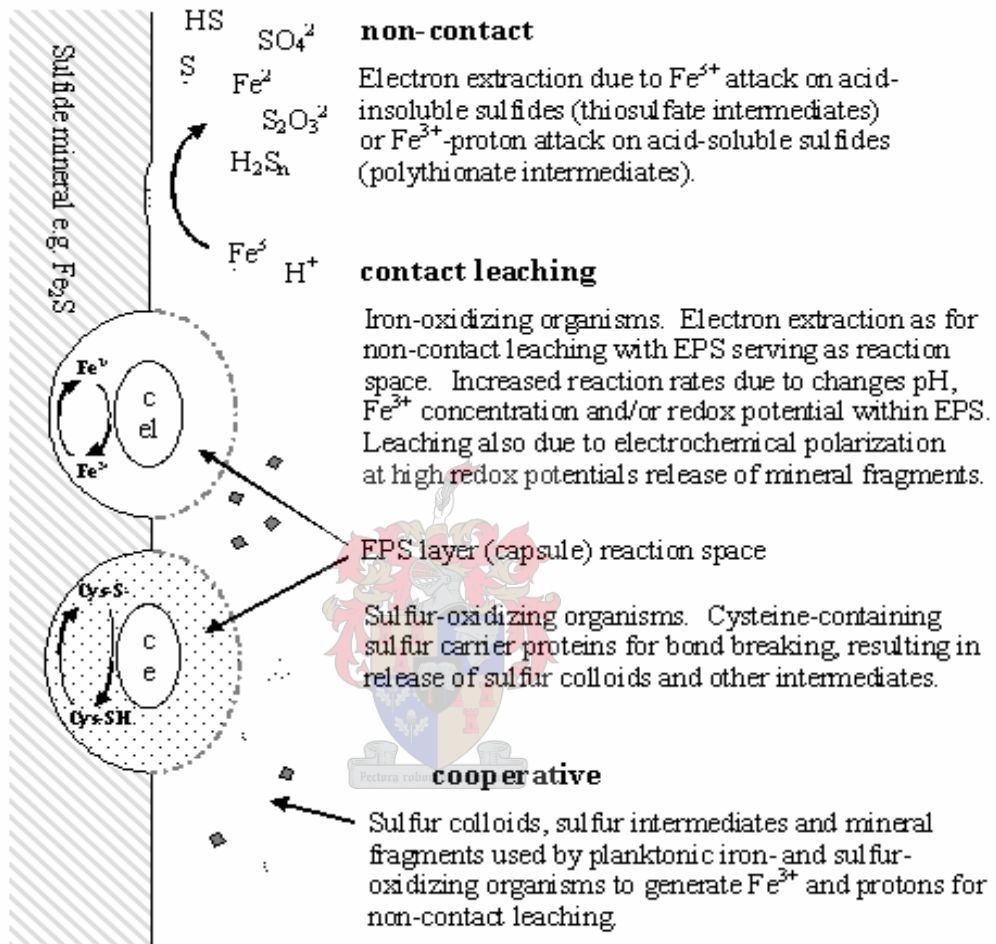


Figure 1.2: Diagrammatic illustration of the proposed mechanisms of pyrite biooxidation (Rawlings, 2002).

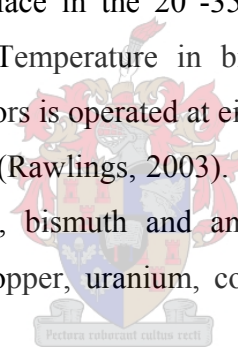
1.4. Current commercial applications of biomining

As mentioned above, biomining includes bioleaching and biooxidation. These processes make use of a consortia of bacteria which include the mesophiles *At. ferrooxidans*, *At. thiooxidans* and *L. ferrooxidans* (Kelly and Harrison, 1989; Markosyan, 1972) and the moderate thermophiles *At. caldus* (Hallberg and Lindström, 1994), *Acidomicrobium ferrooxidans* and

Sulfobacillus species as well as the thermophiles *Acidianus*, *Metallosphaera* and *Sulfolobus* (Brierley, 1997). The types of bacteria involved in a given process are highly dependent on the temperature of the process and to a lesser extent by the mineral being oxidized. Commercial bioleaching became a reality in the 1950's with the dump leaching of copper from submarginal-grade, run-of-the mine material.

Currently two main types of commercial biomining processes are employed. The irrigation-type and the stirred-tank type processes (Rawlings, 2002). The first type describes the percolation of leaching solutions through crushed ore or concentrates that have been stacked in columns, heaps or dumps (Brierley, 1978; Schnell, 1997). It includes the *in situ* irrigation of an ore body, without bringing the ore to the surface. The second type employs continuously operating, highly aerated, stirred tank reactors (Rawlings and Silver, 1995; Rawlings, 2002).

Mineral disintegration can take place at a variety of temperatures. In the case of heap and dump leaching most biooxidation takes place in the 20°-35°C range although higher temperature processes are being investigated. Temperature in bioreactors is controlled and currently biooxidation in the stirred-tank reactors is operated at either 40° or 50°C. Processes that operate at 75°-78°C are under development (Rawlings, 2003). Although several metal containing ores (such as zinc, lead, cobalt, nickel, bismuth and antimony) are amenable to bioleaching technology, the leaching of only copper, uranium, cobalt and gold bearing ores have been commercially exploited.



1.4.1. Irrigation type processes

In terms of tonnages, copper is the most extensively bioleached metal. Current physical methods commercially employed for the copper recovery are of irrigation type, dump, *in situ* and heap. One of best-known dump leaching operations is located at the Kennecott Copper Mine in Bingham Canyon, Utah (Brierley and Brierley, 1999). The Bola Ley plant of Chuquicamata division of Codella in Chile is another example of a more recently constructed dump operation (Schnell, 1997).

Dumps consist of very large quantities of uncrushed, untreated low grade ore, piled to depths of up to 350 meters. Dump leaching is a mechanically simple process that is subjected to cycles of preconditioning, irrigation, rest, conditioning and washing, each of which may be for extended periods of over a year (Rawlings, 2002). The leach dumps are irrigated with raffinate, an iron- and sulphate rich recycled wastewater and is introduced by spraying, flooding or injection

through vertical pipes (Lundgren and Silver, 1980). Leach dumps are usually not inoculated with specialized leaching bacteria. The leaching microorganisms are ubiquitous and when the correct conditions prevail they will proliferate and catalyze the chemical reactions that result in the solubilization of copper sulphides as the leach solution percolates through the dump. Insoluble copper sulphides are converted to the soluble copper sulphate. The metal laden leach solution is removed from the bottom of the dump and copper is recovered by solvent extraction and electrowinning (Schnell, 1997).

Heap leaching of copper is more efficient than dump leaching and is used to extract metals from a higher-grade ore to those subjected to dump leaching. It uses finer crushed ores which are mixed with sulphuric acid in an agglomerating device to bind fine material to coarser ore particles and precondition the ore for bacterial development. The agglomerate is stacked in heaps of 2-10m high onto irrigation pads lined with highly density polythelene to avoid the loss of the leaching solution (Rawlings, 2002). Aeration is enhanced by including aeration pipes which forces air through the heap from the bottom and thus speeding up the bioleaching process (Fig 1.3). Small amounts of inorganic nutrients in the form of fertilizer grade

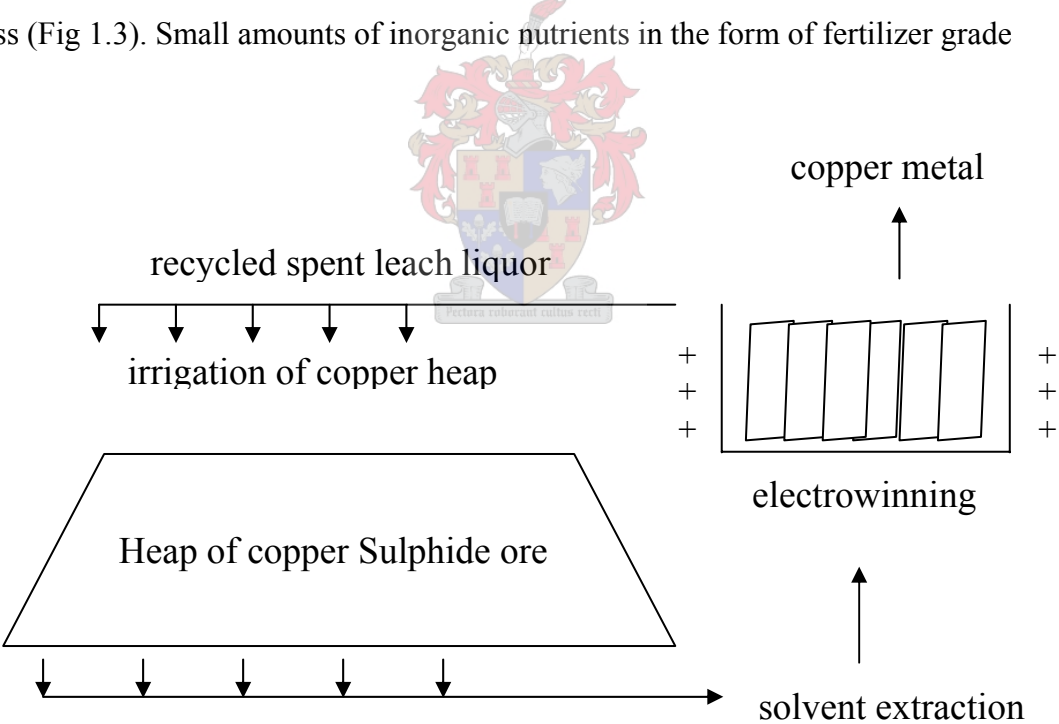
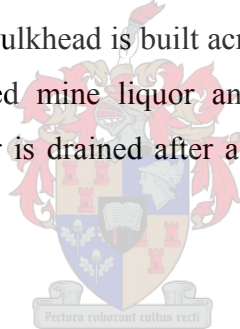


Figure 1.3: Illustration of heap leaching of a copper-containing ore. Refer to literature for an explanation (Reprinted from Rawlings, 2002).

ammonium sulphate and potassium phosphate are frequently added to the raffinate prior to the irrigation, through drip lines placed on the surface of the heap. The solution percolates through the heap and bacteria growing on the surface of the ore and in solution catalyze the release of copper. In contrast to dump leaching, heap bioleaching processes are completed in months

rather than years. Similar process have been applied to refractory gold bearing ores at the Carlin Trend Newmont Mine (reviewed in Rawlings, 2002). The Quebrada Blanca plant in northern Chile is an example of a modern heap bioleaching operation. It is estimated that 75 000 tons Cu/annum are produced from a chalcopyrite ore containing 1.3% copper (Schnell, 1997).

Examples of bioleaching operations that employ irrigation type processes for *in situ* metal extraction includes operations at San Manuel near Tuscon, Arizona (Schnell, 1997) and the Gunpowder's Mammoth mine in Queensland, Australia (Brierley, 1997). *In situ* bioleaching has been used for nearly 30 years to extract uranium and copper from the depleted underground operations as well as from the new mines. Several mines within the Elliot Lake district of northern Ontario employed *in situ* bioleaching for uranium recovery. During 1988 approximately 300 tons of uranium with a value of over US\$ 25 million was recovered from a single mine, the Dennison mine in the Lake Elliot district. This mine has stopped production in recent years because of the reduction in demand for uranium. Typically the underground ores are blasted to fracture the ore and a bulkhead is built across the opening of the stope. The ore is subsequently irrigated with acidified mine liquor and aerated by passing compressed air through perforated pipes. The liquor is drained after a period of 3 weeks and pumped to the surface where uranium is extracted.



1.4.2. Stirred tank processes

Due to capital and operating costs of highly aerated, stirred tank bioreactors their use is generally reserved for high value ores and concentrates. The use of bioreactors increases the rate and efficiency of mineral biooxidation immensely in comparison to the irrigation-type processes. The process is usually comprised of a number of reaction vessels arranged in series and operating in continuous flow mode, with feed added to the first tank and overflowing from tank to tank until the biooxidation of the mineral ore is sufficiently complete (Rawlings, 2002). The primary bioreactor tanks operate in parallel in order to achieve a longer retention time for the ore to allow a stable microbial population to establish (Fig 1.4). The feed comprises a mineral concentrate suspended in water to which small amounts of nutrients, $(\text{NH}_4)_2\text{SO}_4$ and KH_2PO_4 have been added. Mineral biooxidation is an exothermic process and in order to maintain an optimum temperature for microbial activity large volumes of air are blown through each reactor. In addition cooling water may be sprayed over the outside of the tanks or passed through coils inside the tanks. Large agitators ensure that the solids remain even in suspension

and are carried into the next tank. The overall residence time for biooxidation typically varies between four and six days.

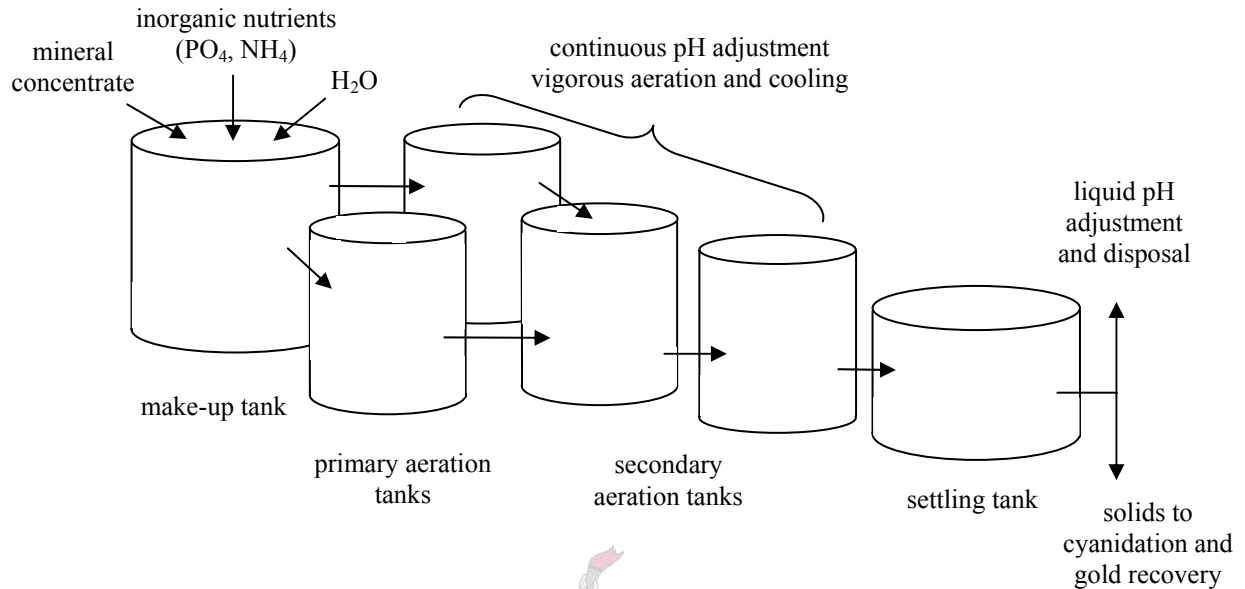


Figure 1.4: Illustration of a typical continuous-flow biooxidation facility used in the pretreatment of a gold-bearing arsenopyrite concentrate (Reprinted from Rawlings, 2002).

Gold bearing concentrates are valuable substrates compared with copper and uranium ores and therefore biooxidation of refractory gold-bearing arsenopyrite ores is carried out in a more efficient and controlled manner in stirred-tank bioreactors. Biooxidation is used as a pre-treatment in the recovery of gold. The gold is embedded in a matrix of pyrite/arsenopyrite and cannot be easily solubilized by the usual process of cyanidation. The arsenopyrite is decomposed during this process in order for the cyanide to make contact with the gold. Gold-bearing mineral usually makes up a small fraction of the mined ore and the ore is crushed and the gold containing concentrate is prepared by flotation (Rawlings, 2002). Microbes that decompose these ores release ferric iron and sulphate and create a highly acidic environment (pH1.5-1.6). Without pre-treatment to expose the gold, usually less than 50% of the gold is recovered by cyanidation in contrast with the 95% recovery after biooxidation. After pretreatment the gold is recovered by cyanidation (Rawlings and Silver, 1995).

The potential of microbes to assist in the extraction of gold from refractory ores was first realized in the early 1980's. Much of the credit is owed to the late Eric Livesey-Goldblatt, who

at the time was the director of the Gencor Process Research Laboratory in Krugersdorp, South Africa (Livesey-Goldblatt *et al*, 1983). This technology was developed into a fully commercial process in 1986 by Gencor at their Fairview mine at Barberton, South Africa. Since then, additional gold-recovery plants based on Gencor biooxidation technology have been commissioned. These plants include Bento (Brazil), Harbour Lights, Wiluna and Youanmi (Australia), Ashanti and Sansu (Ghana) and Tamboraque (Peru) (Rawlings, 2002). The largest plant is at Sansu, Ghana, commissioned in 1994 and is probably the largest fermentation process in the world.

As explained earlier, the use of microorganisms in the extraction of the metal from mineral ores is more efficient in both an economical and an environmental sense in comparison to conventional physicochemical methods for certain iron- and sulphur-containing minerals. Biomining is now an established process and has certainly made its mark as an important industrial process. In order to advance the biomining technology, for example to expand the use of the process to a broader spectrum of base and precious metal ores, we must understand the microbiology of mineral biooxidation.



1.5. Biomining Microbes

An exhaustive report on biomining bacteria is not presented but only a brief overview on the physiology of the more important, commonly isolated biomining microbes. A variety of microbes have been isolated from sites of natural mineral oxidation or from industrial leaching operations (Kelly, 1988; Harrison, 1984). As mentioned earlier, the solubilization of metals from sulphide minerals is catalyzed by a microbial consortium, consisting of complex mixtures of autotrophic and heterotrophic bacterial strains. Their unique physiology permits their growth and reproduction in these low pH, metal-rich, inorganic environments. The primary biomining microbes are all chemolithotrophic and are capable of oxidizing iron- and/or sulphur containing minerals, using ferrous iron and/or reduced inorganic sulphur sources as electron donors. Because of their metal rich niche, biomining bacteria are commonly resistant to a range of metal cations such as Cu^{2+} , Zn^{2+} , Al^{3+} , Ni^{2+} , Hg^{2+} , Ag^{2+} and As^{3+} . The level of resistance varies with different strains. All of the important biomining bacteria are obligately acidophilic, thriving within the pH range 1.5–2.0 (Rawlings, 2002). The production of sulphate during the oxidation of sulphur containing minerals results in the accumulation of sulphuric acid. The acidophily is also very important to biomining organisms capable of only oxidizing iron, which may grow together with sulphur-oxidizing bacteria or generate acid through a

chemical reaction in which ferric iron is the oxidant. Their unique physiology means that biomining bacteria are able to grow in very nutrient poor solutions. Air provides the source of carbon (CO₂) and although other electron acceptors (e.g. ferric iron) can be used, oxygen (O₂) is the primary electron acceptor. The electron donor is supplied by the mineral ore in the form of ferrous iron and/or reduced inorganic sulphur, with water being the growth medium. Essential trace elements are provided as impurities in the ore or water.

Although being regarded for many years as the principal biological catalyst in bioleaching processes, *At. ferrooxidans* shares its environmental niche with other acidophilic bacteria with a similar physiology (Harrison, 1984). The most commonly encountered bacteria primarily involved in mineral decomposition have been characterized as *At. ferrooxidans*, *At. thiooxidans*, *L. ferrooxidans* (Kelly and Harrison, 1989) and *At. caldus* (Hallberg and Lindström, 1994). Acidophilic microorganisms are subdivided on the basis of their preferred temperatures of growth (Norris and Johnson, 1998). Three groups have been recognized: the mesophiles (T_{opt} ca. 20–40 °C), moderate thermophiles (T_{opt} ca. 40–60°C) and extreme thermophiles (T_{opt} > 60°C).

At. ferrooxidans is a mesophilic, Gram-negative, rod-shaped bacterium approximately 0.5µm in diameter and 1 to 1.5µm in length (Norris, 1990). This microbe is capable of oxidizing both ferrous iron and reduced inorganic sulphur compounds, which serves as electron donors during its energy metabolism with oxygen being the preferred electron acceptor. In anaerobic conditions this bacterium can use ferric ions as an electron acceptor provided a reduced inorganic sulphur compound will serve as an electron donor (Pronk *et al*, 1991). *At. ferrooxidans* tolerates high concentrations of sulphuric acid and is well adapted to grow optimally in the range pH 1.8–2.0 (Rawlings, 2002) and are favoured in the temperature range 20–35°C. All *At. ferrooxidans* strains tested have *nif* genes and therefore have the ability to fix atmospheric nitrogen. The nitrogen requirement for cell mass is met by reducing and incorporating atmospheric N₂ (Rawlings and Silver, 1995).

The mesophilic *At. thiooxidans* was first characterized by Waksman and Joffe in 1922. This Gram-negative rod shaped microbe is motile by means of a polar flagellum (Doetsch *et al*, 1967). It is unable to oxidize iron but produces metabolically useful energy from the oxidation of sulphur or sulphide minerals. *At. thiooxidans* is more acid tolerant than *At. ferrooxidans*, and grows in the range pH 0.5–5.5.

L. ferrooxidans is a mesophilic, highly acid-tolerant (pH optimum ~1.5-1.8) Gram-negative bacterium capable of only using ferrous iron as an electron donor. These vibroid cells are much more motile than *At. ferrooxidans* by means of a long polar flagellum (Norris, 1990). Unlike *At. ferrooxidans*, *Leptospirilli* have a high affinity for ferrous iron thus tolerating high concentrations of these ions. Their ability to oxidize ferrous iron is not inhibited by the surrounding ferric iron concentration. *L. ferrooxidans* appears to be more tolerant of high temperatures and less tolerant of low temperatures, than *At. ferrooxidans*. It has been reported to have an upper limit of about 45°C and a lower limit of about 20°C (Rawlings, 1997).

At. caldus is closely related to *At. thiooxidans* and for several years *At. caldus* was mistaken for *At. thiooxidans*. Hallberg and Lindström (1994) described *At. caldus* recently as a separate species. These bacteria are short, Gram-negative motile rods. They are moderately thermophilic bacteria and grows optimally at 45°C and in the pH range 2.0–2.5. They do not oxidize ferrous iron and is only capable of oxidizing reduced sulphur sources. Some strains of *At. caldus* are able to grow mixotrophically using yeast extract or glucose (Rawlings, 2002).

1.5.1. Microbial dominance in biomining processes.

Since its discovery in 1947, all the attention was focused on *At. ferrooxidans*. As mentioned earlier, it was considered for many years to be the main player in bioleaching processes that operated at 40°C or less and as a result much of the bulk of research in this field has been focused on elucidating the genetics, biochemistry and physiology as well as the microbial ecology of this organism (Brierley, 1978; Ingledew, 1982; Nicholaidis, 1987; Southam and Beveridge, 1993). *At. ferrooxidans* shares its highly selective niche with a plethora of strains capable of growth in these highly inorganic, low pH metal rich environments. There is a rapid growth in the number of the variety of microbes isolated and identified from these environments. This is due to an increase in new techniques available to screen for the presence of these organisms and because of an increase in the number of environments being screened.

Recent microbial diversity studies of commercial biomining operations have shown that most of these processes are in fact dominated by “*Leptospirillum*”- like species in combination with *At. thiooxidans* or *At. caldus* (Rawlings *et al*, 1999a; Rawlings *et al* 1999b). Rawlings (2002) sites a few references in support of these findings (Espejo and Romero, 1997; Pizarro *et al*, 1996; Vasquez and Espejo, 1997; Rawlings, 1995; Lawson, 1997; Gardner and Rawlings, 2000).

The molecular genetics of other biomining microorganisms including “*Leptospirillum*”, *At. caldus* and *At. thiooxidans* are in its infancy. Studies of the molecular genetics of biomining bacteria and of their plasmids are mostly initiated with a long-term view for the development of the improved strains for bacterial leaching.

In our analyses of the accessory gene region of plasmid pTC-F14, isolated from *At. caldus* f, we detected an open reading frame that conformed to the criteria of a bacterial insertion sequence. Because of the remarkable nature of insertion sequences, a partial characterization has been done on the element and it is therefore that a section on bacterial insertion sequences has been included to provide insight into the biological nature and importance of these elements.

2. Bacterial Insertion Sequences

2.1. Introduction

Since the discovery of mobile DNA elements by Barbara McClintock in the 1940s, the concept of the chromosome as being an invariable entity of genetic information has been abandoned for a more dynamic view (Iida *et al*, 1983). McClintock first suggested mobile DNA when she discovered that there were genetic determinants in maize chromosomes that caused chromosomal breaks and called them “dissociaters” (Ds). She found that these elements could move from place to place within the genome (Craig, 2002) and discovered a second determinant called “activator” (Ac) which can control the activity of the Ds elements. These Ac elements were also capable of moving from place to place within the genome. A segment of DNA that has the ability to translocate or reconfigure is called a mobile genetic element. Bacterial mobile DNA elements are extremely diverse and new types are continuously being discovered (Toussaint and Merlin, 2002). They are traditionally classified as bacteriophages or plasmids or transposable elements.

Transposable elements are discreet segments of DNA which are characterized by their ability to translocate from one nonhomologous locus to another in their host genome or between different genomes. Such elements are widespread in nature having been identified in all three biological kingdoms. The major distinguishing factor between transposable elements is whether their

mechanism of transposition is dependent on exclusively DNA intermediates or whether it includes an RNA stage (Haren *et al*, 1999). Transposable elements which have an RNA step in their transposition cycle include the retroviruses and retrotransposons and are restricted to eukaryotic organisms. Transposable elements depending strictly on DNA intermediates for their transposition include the genetically complex bacteriophage Mu, the large and complex transposons of the Tn554-type, Tn7, the conjugative transposons, the Tn3-like elements and the genetically elementary insertion sequences (ISs). These elements are found in both eukaryotes and prokaryotes. This review focuses on bacterial insertion sequences.

IS elements are the simplest transposable DNA elements. They generally encode only one protein, a transposase, which catalyzes their transposition. Bacterial IS elements were originally discovered during the investigation of the molecular genetics of gene expression in *Escherichia coli* and bacteriophage lambda. They were identified as causative agents of highly polar and unstable mutations in the galactose and lactose operons of *E. coli* and in the early genes of lambda. They reduced the expression of the genes downstream of their insertion points (Kallastu *et al*, 1998; Galas and Chandler, 1989). Hybridization and heteroduplex analysis showed that these mutations were insertions of the same 0.8–1.5 kilobase pairs (Kbp) long discrete segments of DNA, in different positions and orientations. These segments were repeatedly isolated as insertion mutations and became known as insertion sequences and the observed insertion mutations were evidence for their transposition to new sites in the genome of *E. coli*. They were similar to the genetic elements discovered by McClintock.

The acquisition of point mutations provides bacteria with a capacity to adapt to an ever-changing environment. A more drastic modification is the exchange of segments of DNA between bacterial cells or the reshuffling of genetic information within a given genome. The former process, called horizontal gene transfer, is a key contributor to evolutionary change. Much of the speciation of bacteria has probably been mediated by lateral DNA transfer events (Bushman, 2002; Toussaint and Merlin, 2002; Merlin *et al*, 2000; Syvanen, 1994). Mobile elements play a crucial role in mobilizing genes within a given genome or between cells in a bacterial community. Copies of an IS element flanking a DNA segment can act in concert and render the interstitial region mobile. These structures are called composite or compound transposons. These interstitial DNA (central parts of compound transposons) of naturally occurring transposons include a wide variety of genes, usually genes coding for antibiotic resistance or catabolic functions. Accordingly, IS have been implicated in the acquisition and dissemination of accessory genes involved in antibiotic resistance, pathogenicity, virulence and

detoxifying catabolic pathways (e.g. those allowing for the degradation of manmade recalcitrant molecules) (Galas and Chandler, 1989; Mahillon and Chandler, 1998; Merlin *et al*, 2000). IS elements on their own cannot move through bacterial populations horizontally, but by hitchhiking (associating) on intercellular mobile elements such as plasmids or bacteriophages, they help provide a bacterial community with a pool of mobile genes. In addition to the assembly of accessory functions in bacteria, they have shown to cause various genome rearrangements and in some organisms a substantial fraction of spontaneous mutations derive from their movement (transposition) (Craig, 2002). They are found in the genome of most bacteria at multiplicities of between a few and a few hundred copies per genome. By their ability to transpose, they cause genome rearrangements such as deletions, inversions and replicon fusions (Craig, 1996a; Galas and Chandler, 1989). Multiple copies of the elements can also serve as substrates for homologous recombination, which can lead to alterations in genome structure and activity.

Due to recent advances in bacterial genomics the number of identifiable and sequenced ISs continues to increase, with more than 19 families containing more than 800 ISs. The purpose of this part of the chapter is to provide a general overview on the present state of the understanding of the key properties, genetic activities and occurrence of bacterial insertion sequences.

Several systems of nomenclature appear in literature. One such system, initiated in 1978 attributes a single number to an IS element (e.g. IS1) (Lederberg, 1981). This system does not include sufficient information about the source of the element and it becomes an inadequate system with the large numbers of elements today. A second system, includes some information concerning the source of the element, and includes the initials of the bacterial species from which it was isolated; e.g. ISRm1 for *Rhizobium meliloti*. Yet other system, specific names have been reserved for a specific species; e.g. ISRm and ISC for *Sulfolobus*. The most simplified convention for nomenclature suggested by Mahillon and Chandler (1998), the first letter of the genus followed by the first two letters of the species and a number is used; e.g. ISBce1 for *Bacillus cereus*. In the case the following problem is encountered; e.g. *Bacillus cereus* and *Burkholderia cepacia*, the first two letters of both genus and species should be included; e.g. ISBuce1 for *Burkholderia cepacia*. The current database (<http://www-isbio-toul.fr>.) contains a list of bacterial species and assigned names.

2.2. General Structure and Functional Properties of IS elements

IS elements are small, compact and have a rather elementary genetic organization (Fig 1.5). They range in length between 750 and 3200bp (Merlin *et al*, 2000). IS elements encode only functions involved in their mobility and confer no obvious phenotype (selective advantage) on the host. The functional organization of IS element varies. The majority of ISs contain a large open reading frame (ORF) which extends the entire length of the element. This ORF encodes an enzyme, the transposase (Tpase), which catalyze the breaking and joining reactions that mediate transposition. Certain IS elements, encode a second protein from this ORF, which is a truncated version of the Tpase and acts as an inhibitor of transposition. In some elements such as IS1 and IS3, the Tpase is a fusion protein generated via a translational frameshifting between two overlapping ORFs (see section 2.4, “Control of Transposition activity”) (Galas and Chandler, 1989; Merlin *et al*, 2000).

The termini of the majority of known IS elements carry short inverted repeats of about 10–40bp. These ends are specifically recognized, bound and processed by the Tpase. Alteration within these IRs often affects transposition, reflecting its importance in transposition (Iida *et al*, 1983; Surette *et al*, 1991). The repeats are in inverted orientation with respect to the IS element DNA sequence. However, the absence of IRs in the IS911 and IS110 families of ISs as well as in the IS200/605 complex, shows that the presence of IRs is not a strict criteria for IS elements (Chandler and Mahillon, 2002). These families might employ a different strategy for their transposition than those used by ISs carrying IRs. These IRs are unique to each type of element and the two IR sequences of an IS element are identical or related. Mutational analysis and deletion studies on the IR, showed that the terminal repeats can be subdivided into two distinct functional domains (Fig 1.5) (Mizuuchi, 1992a; Chandler and Mahillon, 2002; Galas and Chandler, 1989). Domain I includes the two to four terminal base pairs which fit into the active site of the Tpase and are involved in the DNA cleavage and joining reactions. This terminal domain is generally identical at both ends of the element and tends to be conserved between related elements. The most common terminal end is the 5'-CA-3' dinucleotide found at the ends of Tn7, IS30, Tn552, the IS3 family members and the bacteriophage Mu elements (Haren *et al*, 1999). Domain II is located within the IR and acts as the sequence specific binding site for the Tpase. Members of the IS21 family carry multiple repeated sequences at both ends. The function of the IRs is to symmetrically position the Tpase at both ends to so that identical reactions are catalyzed at each end.

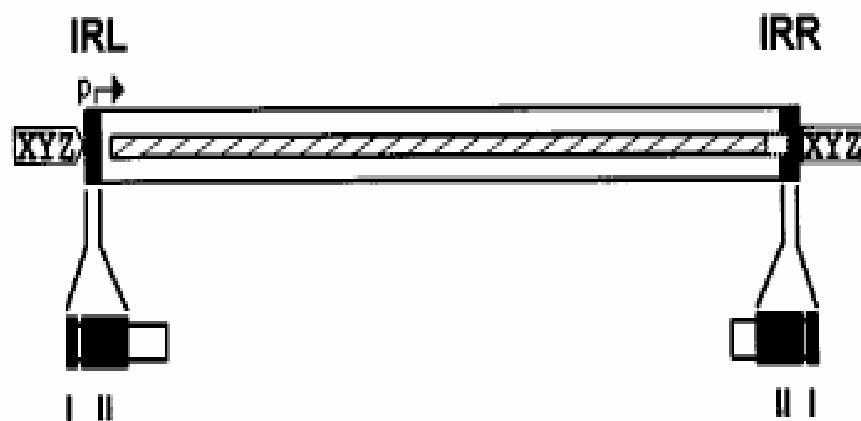


Figure 1.5: Functional organization of a typical insertion sequence. The IS is represented as an open box in which the terminal IRs are shown as gray boxes labeled IRL (left inverted repeat) and IRR (right inverted repeat). The single open-reading frame is indicated as a hatched box stretching the entire length of the IS and extending to within the IRR sequences. Direct target repeats generated in the target DNA during transposition is indicated as XYZ enclosed in a pointed box. The Tpsase promoter, p, partially localized in IRL is shown by a horizontal arrow. The domain structure of (gray boxes) of the IRs is indicated beneath. Refer to text for details. (Reprinted from Chandler and Mahillon, 2002).

In addition to sites primarily concerned with Tpsase recognition and binding, the terminal IRs may carry an array of binding sites for other proteins such as host factors. These proteins may play a role in controlling Tpsase expression or in transposition activity by influencing the assembly of the stable synaptic complex that controls the transposition reactions (see section 2.4, “Control of transposition activity”). The cognate IS promoters are located upstream of the Tpsase, partially within the left inverted repeat (IRL) (Fig 1.5). Coupling the binding sites for the Tpsase and RNA polymerase may provide a method for autoregulating Tpsase synthesis through Tpsase binding.

Another feature common to the majority of known IS elements, is that they generate small, directly repeated duplications of the target DNA at the point of insertion (Fig 1.5). This duplication is a result due to the staggered cleavage of the target DNA and subsequent joining of donor element during the transposition process. After a transposition event the transposed element is flanked by directly repeated duplications of the target DNA. The length of the direct repeat (DR) is usually between 2 and 14 bp and is a characteristic of each type of element. The nucleotide sequence of the repeat is not specific to a type of element and is merely dependent on the sequence of the target site. Although many elements induce a duplication of a fixed number of base pairs, exceptions have been reported for IS elements generating DRs of atypical length at low frequencies (Galas and Chandler, 1989). This may be caused by small variations

in the geometry of the transposition complex (Mahillon and Chandler, 1998). Exceptions exist where insertion lacks the generation of duplicated target sequences and IS91 is an example of such an exception (Diaz-Aroca *et al*, 1987). The absence of DRs may indicate that restructuring may have occurred subsequent to the initial transposition event. This may involve a homologous inter-or intramolecular recombination between two IS elements, each with a different DR. The result would be an IS element carrying one DR of each parent element.

2.3. Components and Mechanisms involved in Transposition

Presently it is known that a wide variety of different types of DNA elements (plasmids, viruses, transposable elements) move from place to place within and between genomes and these recombination reactions underlie many different types of biological transactions. The recombination reactions discussed here involve the actions of a specialized element-specific DNA recombinase, the T_pase that mediates the transposition process. Biological transactions that make use of this process include the acquisition of bacterial genes, replication of certain bacteriophages, the integration of retroviruses and the intracellular movement of retroviral-like elements (Craig, 1995). Discussion will be limited to a particular group of T_pases that uses a conserved triad of acidic amino acids (DDE) to catalyze the DNA transactions necessary for the mobility of IS elements. The T_pases and the integrases (responsible for the integration of viruses and retrovirus-like elements) form the second major class of recombinases and catalyze a set of mechanistically related reactions essential for their mobility (Mizuuchi and Baker, 2002). It should be noted that much of the fundamental insights into the understanding of the nature of IS elements resulted from parallel progress in analyzing related recombinase proteins that each function in a distinct biological niche. This includes recombinases from mobile elements such as bacteriophage Mu, the Tn7, Tn554 and Tn3-like transposons, the retroviruses and retrotransposons. These elements share significant functional similarities that provide insight into the mechanism of transposition and where useful references will be made to these.

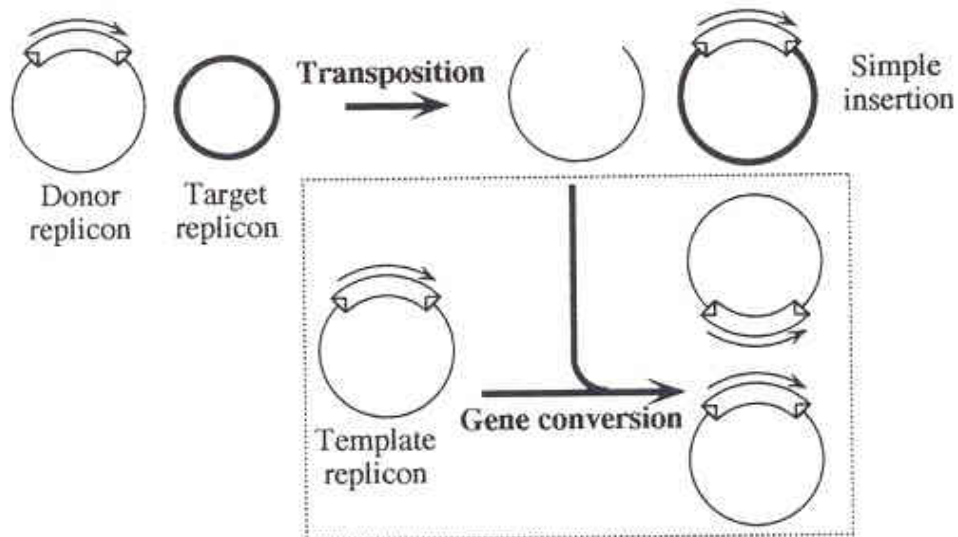
Transposition is the T_pase-mediated recombination reaction which catalyzes the movement of discrete DNA segments between distinct nonhomologous sites. This process is generally, with a few exceptions, independent of the host's sequence homology-requiring recombination machinery because the recombination events do not require homology between the donor and target DNAs. A second reason is that transposition appears to occur with equal facility in wild-type cells and cells deficient in *recA* (Craig, 2002; Hirschel *et al*, 1982; Berg, 1983). Transposition involves the hydrolysis of phosphodiester bonds at the ends of the IS element

(IRs) and the joining (transfer) of these ends to the target DNA molecule. These breakage and joining reactions are well coordinated and controlled and proceed through the formation of a stable highly organized nucleoprotein synaptic complex, the transposome (see section 2.3.6, “The Synaptic Complex”). The nature of the DNA cleavage reactions has a profound influence on the transposition products. Whether just the 3′ ends of the IS elements or both the 3′ ends and 5′ ends are liberated by the T_pase, dictates which mode of transposition will be followed. Transposition mechanisms have historically been separated into two broad alternative modes, conservative (nonreplicative) and replicative transposition (Fig 1.6).

2.3.1. Conservative Transposition

Conservative transposition proceeds through a cut-and-paste mechanism where the element is completely excised from the donor sites and via a mini-circle intermediate reinserted into a target site without the replication of the element (Figs 1.6b, and 1.7a). In other words, both DNA strands at each end of the element in the donor DNA are cleaved and rejoined to a target site to generate a simple insertion. A potentially lethal double-strand break is introduced into the donor molecule during the excision. Most bacteria do not tolerate double-nicked DNA and as a result of a cellular response to transposition, the donor molecule is either lost (degraded by nucleases) or rescued through double strand break repair or by gene conversion using an intact copy of the donor replicon as a template (Coen *et al*, 1989; Engels *et al*; 1990; Hageman and Craig, 1993; Merlin *et al*, 2000). The donor replicon from which excision occurred will be repaired, gaining a copy of the element, so that the cell that underwent transposition will have two copies of the element, one in the donor and one in the target molecule from which excision occurred (Fig 1.6b). Rejoining of the flanking DNA of gaps can also occur but it is thought to be present in non-bacterial systems (Craig, 1997). Examples of elements using the conservative strategy for transposition includes IS10 and IS50 IS elements (Goryshin and Reznikoff, 1998; Kleckner *et al*, 1996) and the transposon Tn7 (Craig, 1996b). In some cases, passive replication of the IS element by the host replication machinery activates transposition. The element is replicated before transposition is initiated. The formation of dimers of the element in the donor replicon is formed as seen in IS10 and members of the IS4/IS5 group, so that after a conservative transposition event both the donor and target replicons have a copy of the element.

B - Conservative transposition



C - Replicative transposition

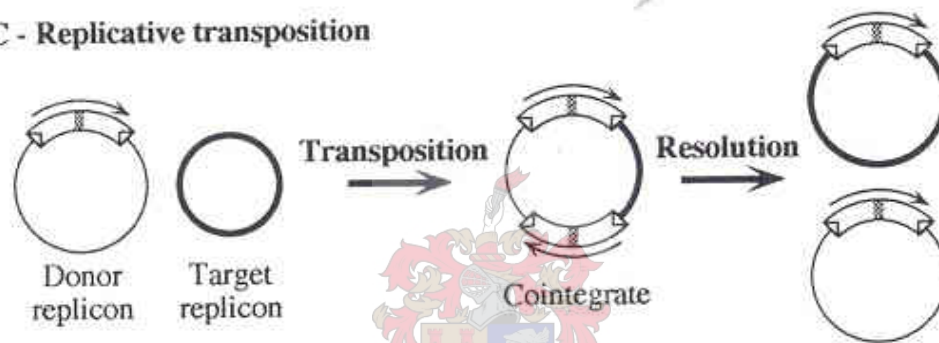


Figure 1.6: The different modes of translocation. **B.** Conservative Transposition. **C.** Replicative transposition. The transposable element is represented by an open box flanked by open triangles (inverted repeats). The thin and bold line circles represent the donor and target replicons respectively (see “Conservative Transposition” and “Replicative Transposition” for details) (Modified from Merlin *et al*, 2000).

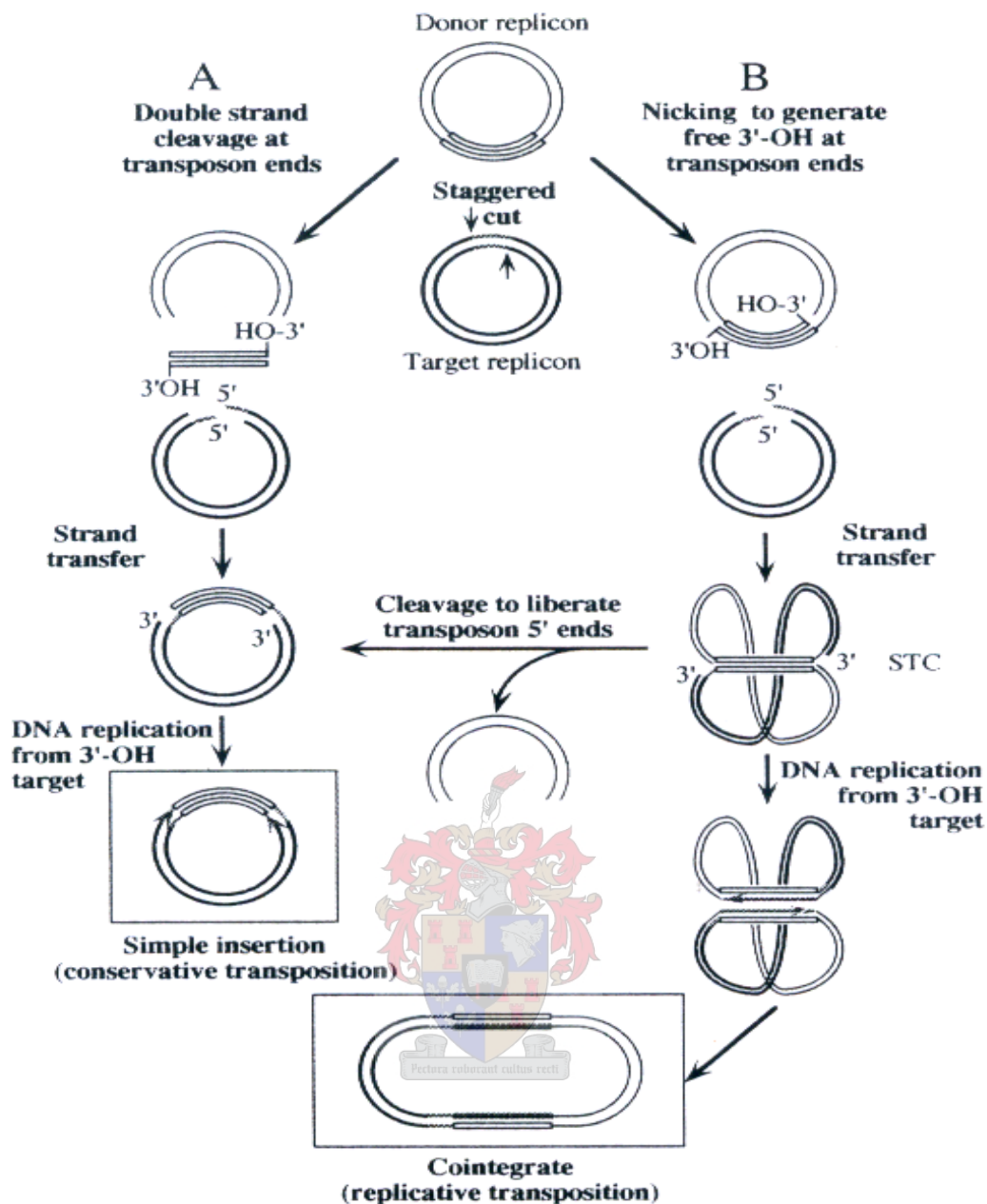


Figure 1.7: Conservative and replicative transposition. The IS elements (or Transposable element) are represented by open boxes. Thin and bold lines indicate the donor and target replicons respectively. **A: Conservative transposition.** The T_pase cleaves both strands at each end of the element (element-donor DNA junctions) completely liberating the element, which ends in free 3'-OH groups. The 3'-OH groups launch a staggered nucleophilic attack and ligation to the target replicon. Replication fills the gaps flanking the reinserted element and generates the short DRs which flanks the element at the new location. **B: Replicative transposition.** T_pase introduce nicks only at the 3' ends of the element which also attacks the target DNA in a staggered manner resulting in the ligation of these ends to the target DNA to form the Strand Transfer Complex (STC). Semi-conservative replication is initiated at the target 3'-OH ends and proceeds along the element to generate a cointegrate. DRs are also generated during replication as a result of the staggered cleavage of target DNA. Resolution of the cointegrate would yield a donor and target replicons, each with a copy of the element (see Fig 1.6C) (reprinted from Merlin *et al*, 2000).

2.3.2. Replicative Transposition

In some bacterial systems only single strand nicks at the 3' ends of the element are introduced during a replicative transposition process, in contrast with the double strand breaks associated with conservative transposition. The IS element is not completely excised from the donor DNA molecule and the exposed 3'-OH tips are joined to the target DNA molecule. A fusion product is formed in which the IS element is covalently linked to both the donor backbone and the target DNA, i.e. the 3' ends of the element are joined to the target DNA while its 5' ends remain attached to the donor DNA (Figs 1.6b and 1.7b). (Craigie and Mizuuchi, 1985). At this stage the process can go in one of two possible ways. Subsequent 5' strand cleavage at the IS element-donor DNA junction, will separate the element from the donor backbone and will also result in a simple insertion. Secondly, the fork-like structures generated leaves a 3'-OH in the target DNA molecule which serves as primers for replication along the element and generates a cointegrate, which contains two copies of the element and one copy each of the donor backbone and target DNA (Fig 1.6c, 1.7b). This mechanism was first predicted by Shapiro (1979) to explain the formation of cointegrates during replicative transposition. The donor and target DNA backbone are separated by a directly repeated copy of the IS element at each junction (Fig 1.6c). The cointegrate structure can be resolved by subsequent recombination between the two element copies to generate two separate replicons, one target replicon carrying a copy of the element and the original donor replicon (Merlin *et al*, 2000). In the more complex transposons, e.g. the Tn3 family (Sherrat, 1989), a second recombinase (a resolvase) is encoded that ensures the recombination between the two IS element copies by acting at the cognate *res* sites (recombination sites). Cointegrate resolution can also be effectively catalyzed by the host homologous recombination machinery, as seen in the IS6 family of ISs (Mahillon and Chandler, 1998). Elements that transpose using the replicative mode include bacteriophage Mu (Mizuuchi, 1992a; Lavoie and Chaconas, 1996), members of the Tn3 family of bacterial transposons (Sherrat, 1989) and the IS6 family of ISs (Mahillon and Chandler, 1998). Care should be taken when using the presence of a cointegrate structure as evidence for a replicative transposition process. A nonreplicative process can also lead to the formation of cointegrates from a plasmid dimer (Fig 1.8) (Lichens-Park and Syvanen, 1988; Berg, 1983) or from tandemly repeated copies of an IS element (Chandler and Mahillon, 2002 and references therein). In other words, a cointegrate could be generated after a simple insertion process (nonreplicative), between a target (already containing a copy of the element) and donor replicon. This reaction is dependent on the host's *rec* enzymes. Here the replication of the element does not form an intimate step during the transposition process and cannot be taken as evidence for replicative transposition.

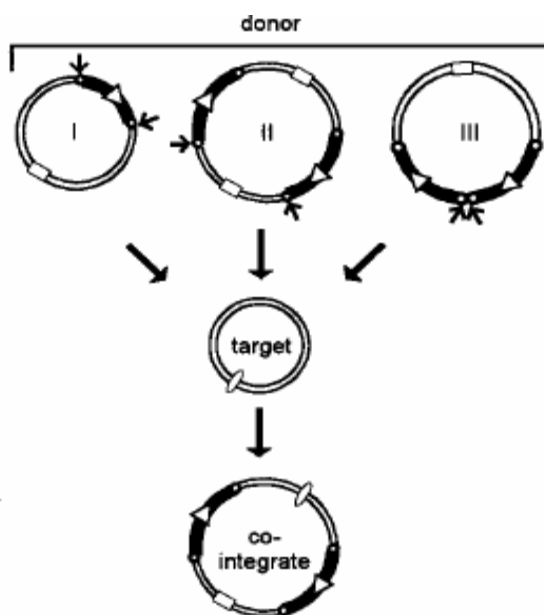
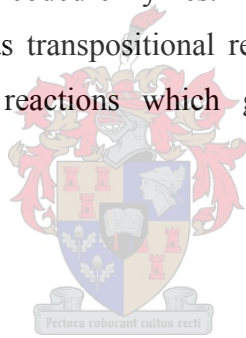


Figure 1.8: Replicative and nonreplicative transposition as mechanisms leading to cointegrate formation. Three pathways are shown which generate cointegrates. (i): replicative transposition; (ii) simple insertions from a dimeric form of the donor molecule; (iii) simple insertion from a donor carrying tandem copies of the IS element (Transposable element in general). The transposable element is indicated in the bold lines with IR represented as small open circles. The orientation of the elements is indicated by the open arrowheads. The square and oval symbols represent the donor and target molecules respectively (reprinted from Chandler and Mahillon, 2002).

2.3.3. The Reaction Mechanisms of Transposition

Although only a handful of transposable elements have been analyzed in biochemical detail, all these elements share similar reaction steps. There are important variations in the reaction pathway of different elements, but the basic mechanistic aspects are shared among a diverse variety of transposable elements. The development of defined *in vitro* systems made valuable research on the elucidation of the reaction steps involved in transposition possible. A great part of the present understanding of the mechanism of transposition is based on *in vitro* studies done on bacteriophage Mu (Mizuuchi; 1992a,b; Haniford and Chaconas, 1992). Defined *in vitro* systems have since been developed for several retroviruses [human immunodeficiency virus (HIV), avian sarcoma virus (ASV) and the murine leukemia virus], the P element in *Drosophila*, members of the *Tc/mariner* group of eukaryotic ISs, the prokaryotic elements Tn7 and the bacterial ISs of the IS4 (IS10, IS50) and the IS3 (IS911) families. Reference to these reports are made in Haren *et al* (1999) (Beall and Rio, 1997; Brown, 1997; Coen *et al*, 1989; Craig, 1996b; Engels *et al*, 1990; Goryshin and Reznikoff, 1998; Hageman and Craig, 1993; Kleckner *et al*, 1996; Ton-Hoang *et al*, 1998; Kaufman and Rio, 1992). Although an examination of the reaction mechanisms is critical for an understanding of the behaviour of

transposable elements, a detailed report on the mechanisms is not pertinent to this study and only a brief description of these reaction steps will be included. Detailed examinations of the reaction steps of transposition has been subject of the following reviews: Mizuuchi (1992a,b), Craig (1995), Craig (1996a,b), Haren *et al* (1999), Chandler and Mahillon (2002), Craig (2002) and Mizuuchi and Baker (2002). Transposition is accomplished through a series of DNA breakage and joining reactions. It is these cleavage and strand-joining reactions that are remarkably similar amongst the wide variety of transposable elements which entails a series of consecutive T_pase-catalyzed endonucleolytic cleavage reactions (hydrolysis) of the phosphodiester bonds at the ends of the transposable elements and the transfer of these ends into a target DNA molecule (transesterification reactions). These reactions have been shown to require no external energy source. After the elaboration and stabilization of the synaptic complex the transposition is accomplished in three distinct steps: the liberation of the transposable element from its donor site by single-or double-strand cleavage at the ends; transfer of the cleaved element ends to a specific or random target site and the processing of the transposition products by the host-encoded enzymes. The first two steps are catalysed by the T_pase and are collectively known as transpositional recombination that involve two distinct types of polynucleotidyl transfer reactions which generate the strand transfer product (transposition intermediate).



2.3.3.1. First strand Donor Cleavage

This is the first critical chemical step in transposition which initiates the separation of the transposable element from the donor DNA. T_pases belonging to the DDE group catalyse single strand cleavages at the 3' ends of the IRs of the elements by an attacking nucleophile to liberate a free 3'-OH group at each end of the element. This is a hydrolysis reaction (also referred to as a nuclease reaction) in which H₂O serve as the attacking nucleophile to generate a free 3'-OH group. That is, it separates the 3'-OH group of the IS element ends from the 5' phosphoryl ends of the adjoining host DNA. This is illustrated in Fig 1.9A. Elements that transpose via the replicative mode of transposition makes use of only single strand cleavages at the 3' ends of their IRs.

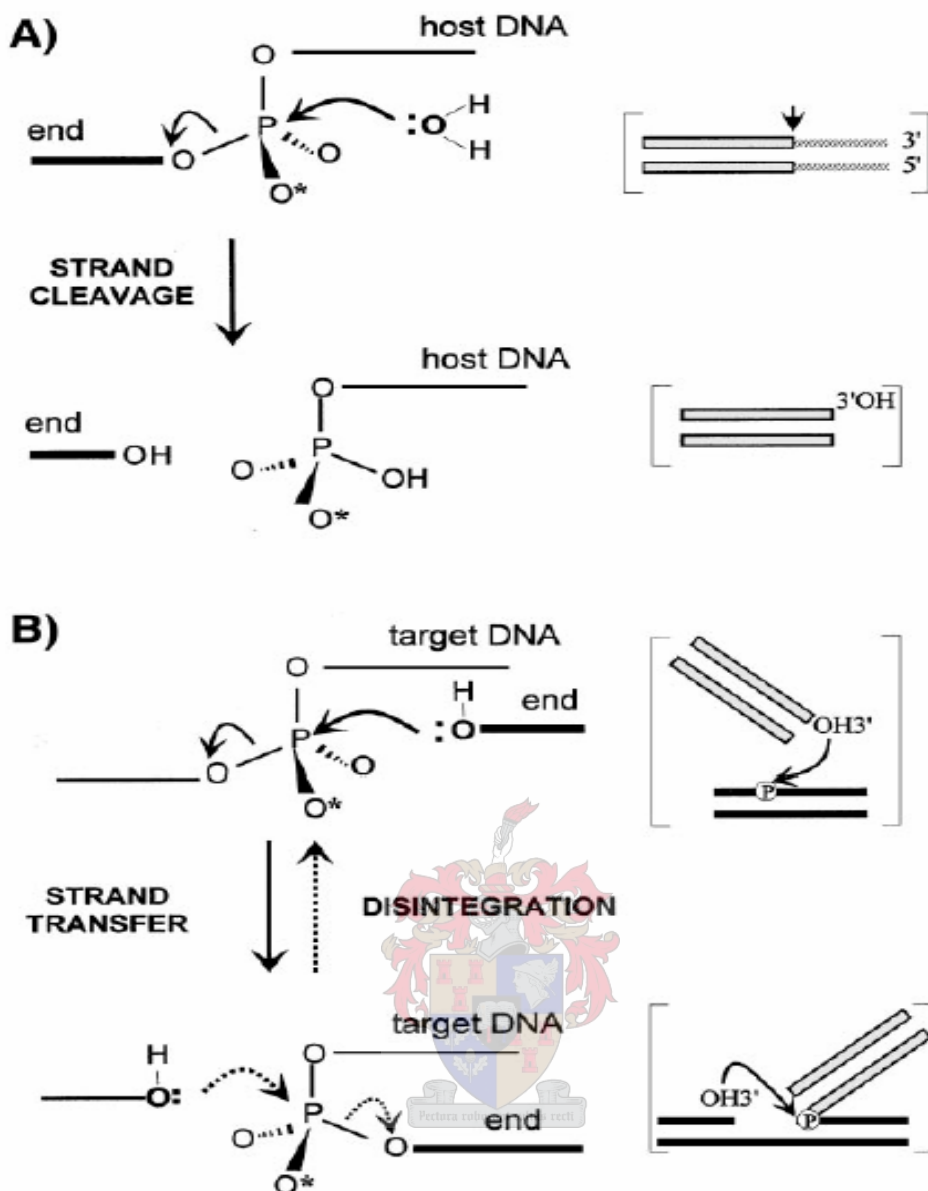


Figure 1.9: The chemistry of the donor cleavage and transfer reactions. **(A)** strand cleavage, **(B)** strand transfer and disintegration. Only one of the two relevant DNA strands is shown in each case. Transposable element DNA is represented as a bold line. The nucleophilic attack by H₂O in donor cleavage reaction or a 3' hydroxyl (OH⁻) from the exposed element ends in the strand transfer reaction or from the exposed 3'-hydroxyl in the target DNA for disintegration, is shown from the right and indicated by curved arrows. The target phosphodiester bonds are shown as a chiral form. The phosphate is usually achiral but by replacing one of the nonbridging oxygens (O*) with a sulphur it traps the phosphate in one of its two chiral forms. The hydrolysis (A) and strand transfer (B) reactions result in an inversion of chirality. The steps involving double strand breaks in the donor and target DNA are shown on the right of this figure (Reprinted from Haren *et al.*, 1999).

2.3.3.2. Second Strand Donor Cleavage

Elements which do not transpose via the replicative mode of transposition require a second donor strand cleavage reaction. This means that in addition to the cleavage at the 3' ends of the element, a second cleavage reaction of the complementary strand (5' end) must also take place which completely excises the element from the donor DNA molecule. This second cleavage reaction is not common amongst all transposable elements and is observed in elements which transposes via the cut-and-paste mechanism. DDE Tpsases are capable of catalyzing cleavages only at the 3'-ends of the element. Different elements have developed different strategies to cleave the second strand. Three such mechanisms are well understood (Fig 1.10) (Turlan and Chandler, 2000). In the first strategy a second element-encoded protein catalyst other than the Tpsase are used for cleavages at the 5' ends. Tn7 encodes a second protein in addition to the Tpsase (TnsB). It encodes a type II restriction endonuclease-like enzyme, TnsA which is dedicated to cleavages at the 5' ends. Inactivation of the catalytic domain of TnsA abolishes 5' cleavage activity and the element transpose via the replicative mode (Hickman *et al*, 2000; Sarnovsky *et al*, 1996). In the other strategies the Tpsase promotes second strand cleavage. The exposed 3'-OH groups generated during the donor cleavage reaction launches an "intratransposon" attack on the opposite, complementary DNA strand. This generates a double strand break at each end, completely excising the element from the flanking donor. The reaction generates a hairpin-like DNA structure at the ends of the element. The ends of the element are opened by a second hydrolysis reaction to generate the 3'-OH and 5'phosphate ends, before the element can be joined to the target DNA. This strategy is used by members of the IS4 family of bacterial insertions sequences such as IS10 and IS50 (which are part of the Tn10 and T5 composite transposons respectively) (Bhasin *et al*, 1999; Kennedy *et al*, 1998). This type of strand transfer reaction can also occur on the same strand, between the two ends of the element. Here the liberated 3'-OH ends attacks the same DNA strand on the opposite end of the element. This results in a figure-eight shaped intermediate and is processed into transposon circles. The ends of the transposon circles undergo additional hydrolysis to generate the excised transposon which then undergoes integration. IS3 family members such as IS911 use this mechanism (Turlan and Chandler, 2000). This type of reaction intermediate may also be generated by family members of IS21, IS30, IS110 and IS256 (http://www-is.biotoul.fr/is/IS_infos/is_general.html).

2.3.3.3. The Strand Transfer Reaction

Once donor strand cleavage has occurred, the exposed ends are transferred into the target DNA. The strand transfer reaction step of transposition does not proceed via the cleavage of the target DNA followed by the ligation of the exposed elements ends into this cleaved target DNA. It has been shown through stereochemistry studies on the target phosphodiester bond in the target DNA that strand transfer occurs via a one-step transesterification reaction. In this reaction the exposed free 3'-OH element ends generated during the donor strand cleavage reactions, acts as the attacking nucleophiles which attacks the 5' phosphate groups in the target DNA and creates a new element-target DNA junction with the production of an exposed 3'-OH in the cleaved target DNA. It is this direct attack of the 3' element ends that couples the element to the target DNA without prior cleavage. This creates a new phosphodiester bond between the elements and target DNA and leaves a 3'-OH group in the target DNA at the point of insertion. It is thus the donor strand itself that performs the cleavage-ligation step in the target DNA. The strand transfer reaction is illustrated in Fig 1.9b.

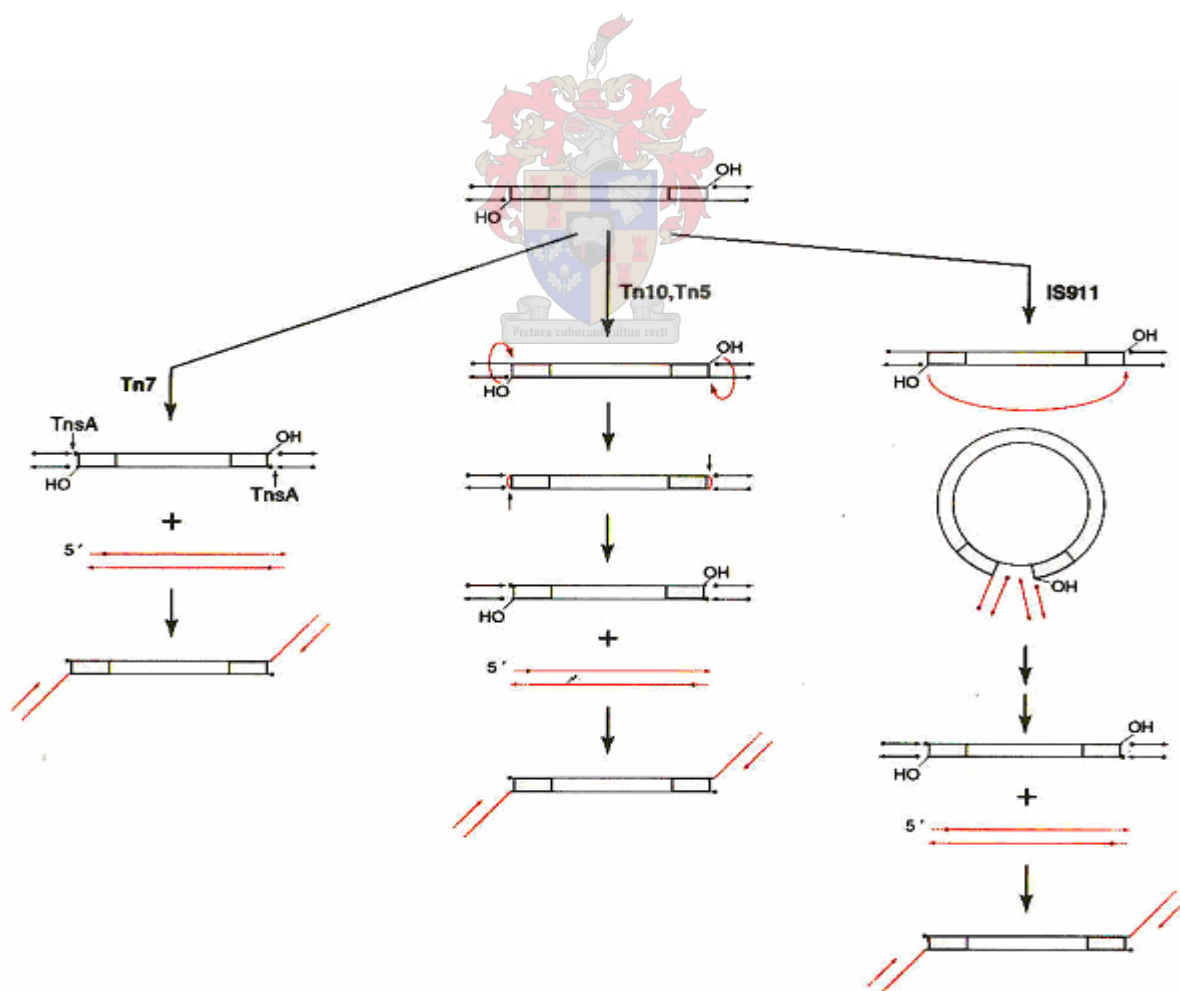


Figure 1.10: Multiple mechanisms for 5'-end cleavage (nontransferred strand cleavage) for element excision during conservative transposition (Reprinted from Mizuuchi and Baker, 2002).

2.3.3.4. The Post Transfer Processing Reaction

The concerted insertion of each 3' end of the element into the target generally occurs in a staggered manner and results in the creation of short, complementary single-strand regions flanking the inserted element. The gap that is produced in this way, i.e. the regions that where the element is separated from the target DNA is repaired by the host replication/repair machinery to complete the various transposition products and the concomitant production of the short direct target repeats. These target repeats are a characteristic of most insertion events. This step of transposition usually requires the dissociation of the transposome from the element to allow the host enzymatic machinery to complete the end transposition products.

2.3.4. The Role of Divalent Metal Ions

Electron starved entities such as positively charged amino acids and metal cations can play the principal role as Lewis acids in stabilizing the transition state. The divalent metal cation, Mg^{2+} , is a perfect candidate as an active site component in phosphoryl transfer, especially when water is the nucleophile (acceptor of transfer e.g. during hydrolysis). The ions coordinate the hydroxide anion for nucleophilic attack on the target DNA phosphodiester. Metal ions have the ability to lower the pKa values of water to an extent that a metal-bound hydroxyl ion is a potent nucleophile. In effect, the metal-bound water can function as a general acid/base and therefore metal cations can form part of an acid/base catalysis mechanism. On the other side of the reaction center, the leaving group side, the leaving oxyanion needs stabilization. This is also accomplished by a metal cation or any other electron starved entity instead of by proton donation from a general acid. Mechanisms using metal cations often involve additional functional groups to stabilize the transition state. Indeed, based on a mechanism proposed for the transesterification reactions catalyzed by Tpsases and integrases, it was suggested that the acidic triad of aa (DDE; see 2.3.5 below, "The Transposase") coordinate divalent metal ions essential for activity (Yang and Steitz, 1995; Khan *et al*, 1991). In this two-metal ion model (Fig 1.11) the ions are positioned specifically to assist substrate engagement and stabilizing the proper active site configuration. In addition to acting as a general acid and general base, the two metal ions, make contact with one of the nonbridging oxygens. This interaction polarizes the P-O bond, i.e. it displaces the negative charge (O^-) away from the positive phosphorus center. This makes the phosphorus center a better target for a nucleophilic attack and contributes to the stabilization of the transition state, and may increase the reaction rate (Mizuuchi and Baker, 2002). The rate of deprotonation of the nucleophile and protonation of the leaving oxygen depends on how efficient the transition state is stabilized. Attractive forces between the

substrate and enzyme surface play a role in stabilizing the intermediate state, in the immediate vicinity of the reaction center. Positively charged functional groups including ions are frequently found within the active centers of proteins that catalyze phosphoryl transfers.

In summary, the two-metal ion model: one metal ion is proposed to function as a general base which increases the potential negative charge of the incoming nucleophile by deprotonation. The second ion assists the leaving 3' oxyanion by acting as a general acid. Further, these ions interact with one of the three nonbridging oxygens (O) to displace the negative O^- away from the positive phosphorous center and thus making it susceptible for a nucleophilic attack by the incoming group.

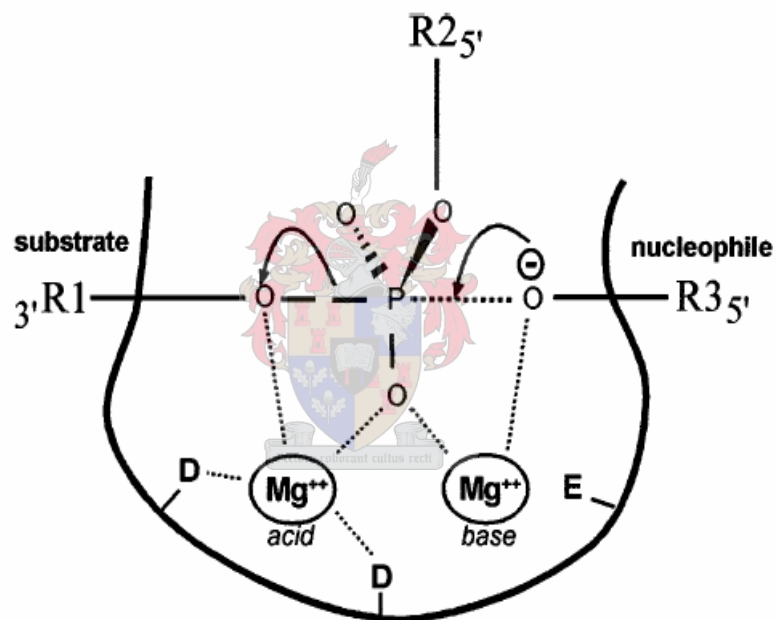


Figure 1.11: Schematic illustration of the two metal-ion model for catalysis of phosphoryl transfer reactions. The figure illustrates the catalytic pocket of the T_pase containing the pentacoordinated phosphate in the (hypothetical) transition state. One Mg^{2+} ion coordinated by two aspartate residues (D) interacts with and stabilizes the leaving group oxyanion (R1) and with the nonbridging oxygen. The second Mg^{2+} also interacts with the same nonbridging oxygen in helping to establish a polarized P-O bond and acts as a base for the deprotonation of the incoming group (nucleophile). The substrate is 3'R1-R25', and the product is 3'R3-R25'

2.3.5. The Transposase

Insights into the mechanism of transposition have come from biochemical and structural studies of multiple members of this protein family. Since transposition involves the breaking and joining of the hereditary material of a cell, it must be carried out in a well-coordinated and

controlled process. The T_pase plays an intimate role as a multifunctional protein in this complex multistep process. To catalyze the reactions described above the T_pase must be able to (i) locate, recognize and bind the correct pair of ends of an element; (ii) elaborate a synaptic complex which involves the terminal ends and one or more host proteins; (iii) recognize the correct donor phosphodiester bonds to be cleaved; (iv) ensure target capture and incorporation at an appropriate stage in the process; (v) strand transfer of the cleaved donor DNA molecule into the target; (vi) dissociate/disassemble the synaptic complex to relinquish their place for the host cellular enzymes to complete the transposition intermediate. These functions require the following abilities: specific and nonspecific binding of DNA, catalysis, the ability to oligomerize (multimerize) and the capacity to interact with accessory proteins. While the primary sequences of many T_pases are generated rapidly because of an increase in genome sequencing projects, only a handful of these proteins have been analyzed in biochemical and structural detail to determine the functional organization, structure and catalytic properties of this protein family (Chandler and Mahillon, 2002; Merlin *et al*, 2000). Partial proteolysis *in vitro* and deletion analysis *in vitro* and *in vivo* are employed to dissect the proteins into topological independent domains. Partial proteolysis generally yields an N-terminal region, a central core region and a C-terminal region. It is convenient to analyze proteins in terms of domain structures because a given domain often exhibits a given function. It should be noted however, that a specific function may not necessarily be carried out by a single domain, but through several different regions of a single polypeptide or by more than one polypeptide in a multimeric complex (Haren *et al*, 1999). From the limited T_pases being studied in detail a general pattern seems to be emerging. The sequence specific DNA binding activities are generally located in the N-terminal region and the catalytic domain is often localized towards the C-terminal end (Fig 1.11). This functional organization has been observed for a few elements. This has been reviewed and referenced in Haren *et al*, 1999 for the following IS elements: IS1 (Machida and Machida, 1989; Zerbib *et al*, 1985), IS30 (Stalder *et al*, 1990), Mu (Lavoie and Chaconas, 1996), Tn3 (Maekawa *et al*, 1993), IS50 (Wiegand and Reznikoff, 1994), IS903 (Tavakoli *et al*, 1997), IS911 (Polard *et al*, 1996; Haren *et al*, 1998). As mentioned later, in some cases the functional domains of a T_pase are the result of a fusion protein generated via translational frameshifting (see section 2.4.1.3, “Programmed Translational Frameshifting”).

2.3.5.1. DNA binding domain

The DNA recognition abilities of the T_pase allows it to specifically recognize their cognate mobile elements (IRs), thus discriminating between the ends of the element and “non-transposable element” DNA. DNA binding and recognition also plays a role in the elaboration of the synaptic complex, positioning of the element substrate DNA in the catalytic pocket and in target capture (target selection). These reactions involve specific and non-specific binding to DNA. The ability of the T_pase to specifically recognize the correct pair of ends of an element is easily imagined when looking at the bifunctional organization of both the IRs and the T_pase. As mentioned above, the terminal ends are composed of two distinct functional domains. Domain I includes the terminal 2-4 bp which are essential in the cleavage and strand transfer reactions. Domain II is located internally to the IRs and is specifically bound by the T_pase. They primarily function to symmetrically position the T_pase on the ends. The T_pase sequence specific-DNA (recognition) domain is located in the N-terminal region and is topologically independent of the C-terminal localized catalytic domain. This functional organization allows the T_pase to simultaneously bind the internal domain (domain II) with its N-terminal region and the terminal bps (domain I) with its catalytic domain. A helix-turn-helix motif (HTH) is generally found in the sequence-specific DNA binding domain and reflects a capacity for sequence specificity to the ends (Fig 1.12) (Polard and Chandler, 1995). Binding of the N-terminal T_pase domain to the IRs has been observed for IS1, IS30, IS903, IS50, IS911 and Tn3 (Haren *et al*, 1999 and references therein). The interaction of the T_pase with the target molecule forms an integral part of the transposition process. In general, the target DNA sequence does not show significant homology with the ends of the element. Thus, a region(s) of the T_pase must be able to bind DNA non-specifically during target site selection (Bender and Kleckner, 1992). These regions may contain a high proportion of basic amino acids, which will allow for non-specific DNA binding. This has been observed in MuA (Mu element T_pase) which contains a string of amino acids RRRQK (Fig 1.12) (Wu and Chaconas, 1995).

2.3.5.2. The Catalytic Domain

The earliest comparisons of the structural domains of the Mu T_pase, several integrases (protein catalysts that mediate retroviral integration) and the Tn5 T_pase revealed that proteins forming part of the T_pase/Integrase family have highly related catalytic domains (Mizuuchi and Baker, 2002; Haren *et al*, 1999, Polard and Chandler, 1995; Chandler and Mahillon, 2002), so much so that the structures surrounding the active centers of these proteins can be nearly superimposed. Another important observation was that these enzymes also exhibit a relationship, undetectable

by primary sequence comparison, to other enzymes that mediate phosphoryl transfer reactions, notably RNaseH and the Holliday junction resolvase RuvC. This has led to the idea that these enzymes are members of a superfamily of polynucleotidyltransferases (Dyda *et al*, 1994; Rice and Mizuuchi, 1995). The catalytic domain is approximately 160 aa residues and has a mixed three-dimensional α/β -fold structure, containing a five stranded β -sheet which is flanked by α -helices (Mizuuchi and Baker, 2002). The domain carries the highly conserved triad of acidic amino acids with, the two Aspartate (D) residues separated from a Glutamate (E) residue by about 35 aa and is commonly known as the DDE motif. This motif is present in all the enzymes mentioned above and a recent compilation includes the majority of known ISs (Chandler and Mahillon, 2002). Because of the characteristic spacing between the second D residue and the E residue, this triad is known as the DD(35)E motif in the IS3 and IS6 families of bacterial ISs and retroviral integrases (Fig 1.14) (Fayet *et al*, 1990; Katzman *et al*, 1991; Kulkosky *et al*, 1992). Additional alignments of several Tpsases revealed four relatively conserved regions designated N1, N2, N3 and C1 which encompass the D (N2), D (N3) and the E (C1) regions of the DDE motif respectively (Fig 1.14) (Rezsohazy *et al*, 1993).

The C1 region is the most defined and conserved region in the catalytic core, and forms part of an α helix. It carries several additional conserved residues in particular a K (Lysine) or R (Arginine) residue, seven aa residues downstream of the E residue (Fig 1.13) (Polard and Chandler, 1995 and reviewed in Chandler and Mahillon, 2002; Haren *et al*, 1999; Doak *et al*, 1994, Jenkins *et al*, 1997). Additional relatively well conserved residues observed include a basic residue four residues (approximately one helical turn) upstream and four residues downstream from the final E residue and a W or other hydrophobic amino acid three residues upstream from the first D, and a G two residues downstream from the second D. The IS4 family of ISs has additional conserved residues in their active sites, which has been implicated in catalysis. This family transpose via the conservative mode by making use of the DNA hairpin mechanism for second strand cleavage at the 5' ends to completely liberate the element from the donor DNA (see above, "Mechanism of the Donor DNA Cleavage Reaction"). The conserved residues have a "YREK" signature, which conforms to the sequence pattern $YX_2RX_3EX_6K$ where E is the catalytic E residue in the DDE motif. These residues has been implicated in hairpin formation (Mizuuchi and Baker, 2002). The N2 and N3 regions consist of one β sheet each.

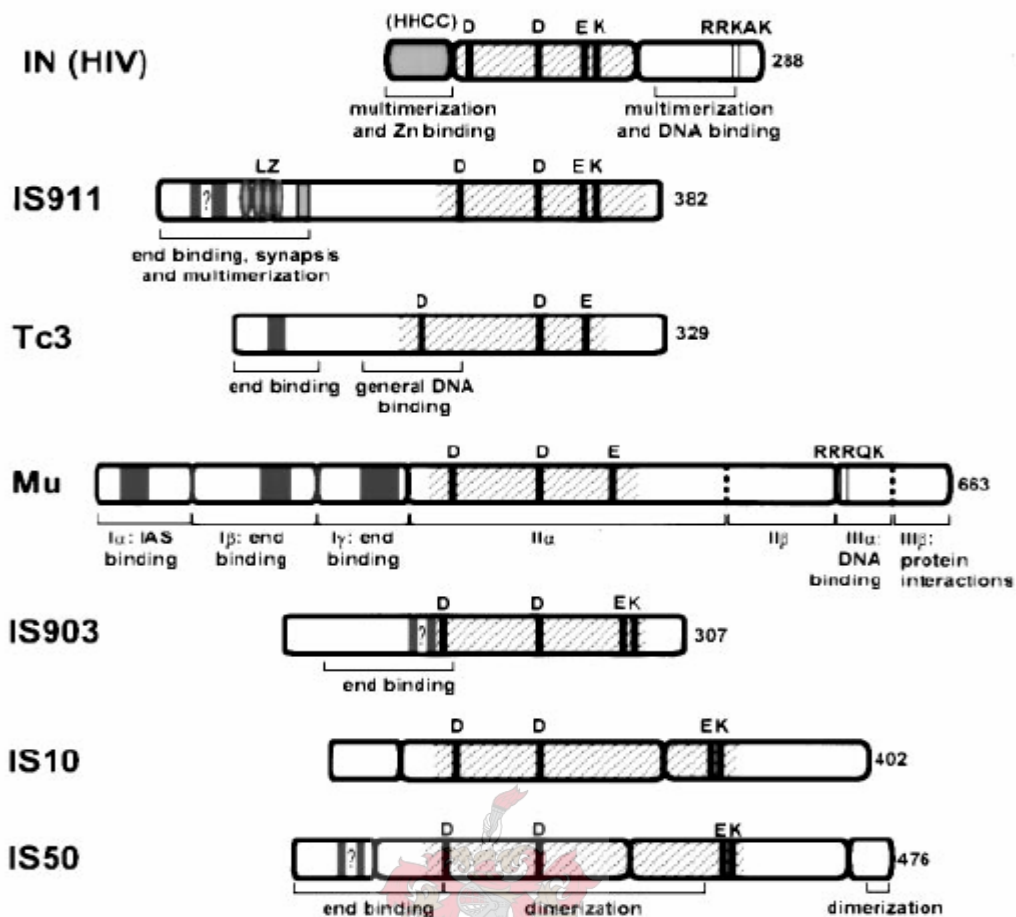


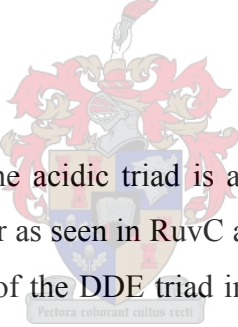
Figure 1.12: The functional organization of several Tpsases. The protease sensitive sites are delimited by the open boxes. The dark grey boxes represent potential or real helix-turn-helix (HTH) motifs. Potential HTH motifs are indicated by “?”. The catalytic core is indicated by pale grey and carries the DDE motif. These residues together with a number of others mentioned in the text are indicated in uppercase letters above each Tpsase molecule. LZ indicates the leucine zipper motif with the four repeating heptads observed in the *IS911* Tpsase. A second region involved in multimerization is shown slightly downstream. In cases investigated, the catalytic core is also capable of promoting multimerization. Known functions of the different regions are indicated below. The Tpsase alignments are centered in the second aspartate residue. The length of each protein in aa is indicated at the right (Reprinted from Haren *et al*, 1999).

Selective site-directed mutagenesis of these enzymes has revealed the importance of the DDE motif residues in the catalysis of both the donor DNA cleavage and DNA strand transfer steps of transposition. Selective mutation of these residues either leads to the abolishment of the donor cleavage or the joining reactions or both or has been shown for Mu Tpsase, HIV-1 integrase, *IS50*, *IS10* and *IS903* Tpsases (reviewed in Haren *et al*, 1999; Chandler and Mahillon, 2002).

	64	116	152
HIV-1 (IN)	wql D cth (51)	vht D ngsnf (35)	ynpqs Q gvi E sm N Kel K
	269	336	392
Mu (MuA)	ing D gyl (66)	iti D ntrga (55)	kgwgqa K pv E rafgvg
	28	114	149
Tn7 (TnsA)	(hgz D yip) (85)	mst D flvdc (34)	ertle K lel E rrywqq K
	273	361	396
Tn7 (TnsB)	yei D ati (87)	lla D rgelm (34)	rrfda K giv E stfRtl
	166	240	276
Tn552	wqa D htl (73)	fyf D hgscf (35)	gvprg R gki E rff Q tv
	689	765	895
Tn3	asa D gmr (75)	imt D tagas (129)	riltql N rg E sr H ava R
	207	287	323
IS911 (IS3)	wcg D vty (59)	fhs D qgshy (35)	gnwd N spm E rff R sl K
	97	161	292
IS10 (IS4)	vlv D wsd (63)	ivs D agfkv (130)	niysk R mqi E etf R dl K
	119	188	326
IS50 (IS4)	siq D ksr (67)	avc D readi (136)	diyth R wri E ef H Kaw K
	121	193	259
IS903 (IS5)	lvi D stg (71)	asa D gaydt (65)	tdynr R sia E tamyrv K
	78	138	173
IS26 (IS6)	whm D ety (59)	int D kapay (36)	qikyr N Nvi E cd H gkl K
	237	293	327
IS30	weg D lvs (55)	ltw D rgmel (33)	qspwq R gtn E nt N gli R
	122	184	230
IS21 (IstA)	lqh D wge (61)	vlv D nqkaa (46)	rrart K gkv E rmv K yl K
	181	261	297
IS630	fye D evd (80)	liv D nyiih (35)	vyspww N hv E rlw Q al H
	112	192	237
IS982	sii D sfp (79)	vlg D mgylg (45)	nfskr R Kvi E rvfsfl
	167	233	341
IS256	lmt D vly (65)	vis D ahkgl (107)	nrlkst N li E rl N Qev R
	86	177	286
Tc1	iws D esk (90)	fqq D ndpkh (108)	sqspdl N pi E hmweele R
	N2	N3	C1

Figure 1.13: The DDE consensus in Tpsases and Integrases (INs). Individual representative members of each family are shown. Where the enzymes shown are not the founding members of a IS family, this is included in brackets. The residues forming part of the DDE motif are shown in large bold letters. Uppercase letters indicate conservation within a family and lowercase letters indicate that the particular amino acid is predominant. The numbers in parentheses show the distance in amino acids between the amino acids of the conserved motif. The N2, N3 and C1 region are those defined by Reszhohazy *et al* (1993).

There is also growing evidence that the conserved K and R residues 7 aa downstream of the final E residue, in IS10 and IS50 Tpsases forms part of the catalytic pocket and is essential for catalysis (Bolland and Kleckner, 1996; Reznikoff, 1993). This implies the possibility that the protein might undergo a conformational change to displace these residues before catalytic activity. Additional partial conserved regions are also implicated in catalytic activities of Integrases and Tpsases of various ISs (Haren *et al*, 1999). These results provide strong evidence that the DD(35)E triad lies at the heart of the active site and that these proteins employ one catalytic center to catalyze the multiple reaction steps in transposition. This shows a remarkable flexibility in both the substrate and nucleophile binding sites. The active center that binds and stabilizes the intermediates during the reaction steps are occupied by three different phosphate groups during transposition namely, the first one from the transferred DNA strand, then one from the nontransferred strand and a third from the target DNA. Further, when a particular phosphate group has been processed it does not fully dissociate from the active site but remains in the synaptic complex during the next reaction step. This requires a series of conformational changes to rearrange the strands between the sequential transposition steps (Mizuuchi and Baker, 2002).



Structural analysis has shown that the acidic triad is arranged close together in the catalytic pocket, in a three dimensional manner as seen in RuvC and RnaseH (Rice *et al*, 1996; Rice and Mizuuchi, 1995). One essential role of the DDE triad in catalysis is to coordinate the divalent cations, Mg^{2+} or Mn^{2+} (see 2.3.4, “The Role of Divalent Metal Ions), necessary for catalysis of both DNA cleavage and DNA strand transfer which has been supported by structural studies (Mizuuchi and Baker, 2002; Haren *et al*, 1999). In the two metal-ion model for catalysis, the DDE traps the two Mg^{2+} ions that make the phosphodiester bond at the 3' ends of the element fragile and susceptible to a nucleophilic attack by a water molecule. It is however, not clear how these ions are being positioned in the catalytic center and the exact functions they perform in the catalytic mechanism. Recent studies on the Tn5-DNA cocrystal structure revealed one Mg^{2+} -ion bound to each active site, which is coordinated by the two DDE-motif residues (Asp-97 and Glu-326) and the 3'-OH of the transferred strand (Mizuuchi and Baker, 2002). Although one metal is observed in this Tn5 structure, many enzymes that catalyze polynucleotidyltransfers, use a two-metal ion mechanism. It is thus at present unclear whether a one or two metal-ion model best describes catalysis by Tpsases. Studies aimed to elucidate the structure-function relationships of the catalytic domain and their role in target sequestration and positioning are still required.

The limited information available for the greater part of the IS families, serves to generalize the importance of the presence of the DDE triad. The DDE motif is not defined in all ISs and implies that these elements may employ different mechanisms for transposition. Indeed, the IS91 family shows significant similarities with enzymes associated with replicons that use a rolling-circle replication mechanism and this family has adopted a rolling-circle transposition mechanism (Galas and Chandler, 1981; Mendiola *et al*, 1994; Garcillan-Barcia *et al*, 2002). IS110 family members have been thought to encode a novel type of site-specific recombinase although recent analysis has suggested the presence of an atypical DDE motif (Lenich and Glasgow, 1994; http://www-is.biotoul.fr/is/IS_infos/is_general.html).

2.3.5.3. Multimerization

Multimerization is another fundamental property of many, if not all, Tpsases of both prokaryotic and eukaryotic origin (Haren *et al*, 1999). This has been physically demonstrated for prokaryotic elements such as Mu (Mizuuchi, 1992a; Lavoie *et al*, 1991), IS50 (Weinreich *et al*, 1994; Braam and Reznikoff, 1998) and IS911 (Haren *et al*, 1998) and eukaryotic elements such as retroviruses (Andrake and Skalka, 1996). Transposases function primarily as multimeric proteins for example the IS4 family Tpsases work as dimers (Davies *et al*, 2000), the Mu Tpsase as a tetramer (Lavoie *et al*, 1991) and the HIV-1 integrase has been proposed function as a tetramer (Zheng *et al*, 1996) or an octamer (Heuer and Brown, 1998). The nature of the oligomerization properties of Tpsases is complex and poorly understood because the proteins undergo a series of conformational changes during the transposition process, which involves protein-protein interactions.

The regions involved in multimerization have been probed for only a limited number of proteins. Integrases carry determinants in all three domains (N-terminal, core and C-terminal) for self-associating properties (Fig 1.12). The IS911 Tpsase carries three self-interaction regions implicated in multimerization (Fig 1.12). The leucine zipper motif promotes multimerization as does the catalytic core. This allows binding to the IRs where the adjacent region is apparently involved in multimerization and finally leads to synapsis (reviewed in Haren *et al*, 1999). Two regions capable of self-interaction with a second Tpsase have been uncovered in the IS50 Tpsases (Fig 1.12).

2.3.6. The Synaptic Complex (The Transposome)

As in the case with many other DNA transactions (Echols, 1990) and RNA reactions, the chemical steps of transposition occur within high order protein-nucleic acid complexes. These elaborate synaptic complexes, promoted by the Tpsases, contain oligomers of the Tpsase, the DNA substrates and in some cases accessory proteins as well. The assembly of these structures to form an active complex as well as the chemical steps of transposition involves a large number of protein-protein and protein-DNA interactions (Mizuuchi, 1992b). The complex pathway of the formation of the transposome as well as the multiple protein-protein and protein-DNA interactions appear to be designed to accomplish coordination of the chemical reactions at the two terminal IR termini of the element. The requirement for proper assembly of such a complex prior to activation is a key regulatory step in transposition and seems to accommodate a number of control points (Craig, 1997). This complexity ensures that the appropriate substrates, i.e. the proper combination of IRs and sometimes the target as well, are present and precisely coordinates the timing of these reaction steps (Referenced in Mizuuchi, 1992b; Mizuuchi, 1992a; Mizuuchi *et al*, 1992; Baker and Mizuuchi, 1992 and referenced in Craig, 1997; Bainton *et al*, 1993; Chaconas *et al*, 1996 and Craigie; 1996).

The initial reactions of transposition start out as reversible interactions between the Tpsase and DNA sites. It is during these reactions that the Tpsase discriminates between possible terminal ends, leading to the recognition and binding of the correct pair of terminal IRs. All bacterial elements examined exhibit end-pair selection mechanisms and all these elements require a certain degree of supercoiled DNA substrates. After end synapsis has been established transposition proceeds through a series of ordered steps showing consecutive increases in protein-DNA complex stability to temperature and protein denaturing agents (Sakai *et al*, 1995; Surette *et al*, 1987). The correct DNA substrates and correct pair of terminal IRs are identified by protein-DNA interactions and this selection is achieved through the interaction between the sequence-specific binding domain of the Tpsase and the IRs. As mentioned above, Tpsases are multimeric proteins in which the active sites for the DNA cutting and joining reactions are formed by collaboration between domains from different Tpsase molecules (Chaconas *et al*, 1996; Craigie, 1996). Only once the Tpsases are positioned at the correct pair of termini and the stable synaptic complex is assembled, do they become active catalysts which then execute the cutting and joining reactions (Davies *et al*, 2000). The Mu Tpsase functions as a tetramer and it has been reported that the individual protein subunits within the synaptic complex recognize one terminal end via their sequence-specific DNA binding domain and donates its catalytic domain to the opposite terminal end (Mizuuchi and Baker, 2002). This observed interlinking

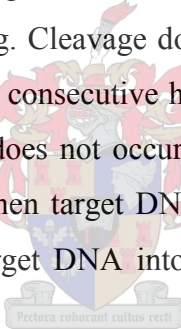
subunit rearrangement is a possible explanation for why the complex assembly is a prerequisite for catalytic activity. It is also believed that additional conformational changes in the T_pase proteins occurs during the assembly of the synaptic complex and will prove to be essential for the activation of the protein's catalytic function (Mizuuchi and Baker, 2002).

It is evident that the initial reactions mentioned above, are key regulatory steps in transpositional recombination. DNA rearrangements are potentially dangerous for the cell. DNA damage caused by recombination using the incorrect pair of IR ends or the premature dissociation of the synaptic complex before completion of the transposition process, would be deleterious and non-productive and could lead to cell death. By demanding the formation of the elaboration of the synaptic complex, these deleterious effects are minimized (Mizuuchi, 1992a; Mizuuchi and Baker, 2002; Haren *et al*, 1999). The product-protein complexes are unlikely to dissociate spontaneously within the time frame needed for the completion of the process and require an external energy source to achieve this.

The assembly of the synaptic complex is understood in some detail and has been demonstrated for IS10/Tn10 (Sakai *et al*, 1995; Kleckner *et al*, 1996). Tn10 is a composite transposon consisting of two copies of IS10 elements flanking genes coding for tetracycline resistance. Tn10's IRs are defined as outside and inside ends (OE and IE) with respect to their relationship within the parental transposon Tn10. These IRs are distinguished by their different localized positions of *dam* methylation sites and the presence of an IHF binding site proximal to OE. Synapsis involves accessory factors and a possible two-step mechanism leading to the activation of the T_pase multimer. In the first step a transiently constrained complex is formed. In the second step, this complex undergoes conformational changes during which the accessory factors are discharged and the active sites of the T_pase oligomers are brought closer to the reactive phosphodiesterbonds (Haren *et al*, 1999).

A precleavage synaptic complex has been identified for Tn10. This transitory complex is composed of the two terminal ends bridged by the T_pase and is dependent on IHF but not divalent metal ions. IHF creates a bend at each transposon end sequence which allows the T_pase to contact each terminal end and also helps to stabilize this sequence in a folded conformation. It is thought that this complex is transitory because titration of IHF from this complex results in conformational changes and renders the complex competent for subsequent

target capture and strand transfer (Kleckner *et al*, 1996). In the presence of Mg^{2+} or Mn^{2+} , the precleavage synaptic complex is assembled *in vitro*, undergoes double-strand cleavage to generate complexes with single- or double-end breaks. Kinetic data has consistently shown that the double-end breaks are generated from single-end breaks. Cleavage of each strand at the terminal ends occurs in a sequential manner with the transferred strand being cleaved first followed by cleavage of the complementary strand (Bolland and Kleckner, 1995). As mentioned above (see section 2.3.3.2, “Second Strand Donor Cleavage”), the liberated 3'OH ends of the transferred strand cleave the complementary strand. The double-end break complexes are competent for strand transfer, which occurs in a coordinated way, and resembles that of an excised transposon fragment. This is observed both *in vivo* and *in vitro* and the entire element is precisely excised from the donor backbone in which the two terminal ends are held together by the Tase (Chalmers and Kleckner, 1994; Haniford *et al*, 1991). The donor DNA that flanked Tn10, which is cut away by double-strand cleavages, is relaxed and not stably maintained within the complex. The excised transposon fragment, held together by the tightly bound Tase within the complex, is supercoiled because the Tase that bridges the two ends prevents the DNA ends from swiveling. Cleavage does not occur simultaneously and a single catalytic site is repeatedly used for the consecutive hydrolysis and transesterification reactions. Because strand cleavage at each end does not occur simultaneously (not concerted cleavage) this could be potentially dangerous when target DNA is immediately available and therefore IS10 allows only late entry of the target DNA into the synaptic complex subsequent to IR cleavage.



In order for target capture to occur the transposon ends must be unfolded. It is thought that one of the transposon arms blocks the entry of an intermolecular target DNA into the transposome. If IHF is not discharged, the transposon arm is not unfolded and only intramolecular strand transfer events occur. The intramolecular transposition product has been observed to be an unknotted inversion circle. Another histone-like protein, H-NS, plays an important role in Tn10 transposition. H-NS has a high affinity for bent DNA and is naturally attracted to the folded form the transposome which acts in opposition to IHF, which upon binding to the transposome induces transposome unfolding. This unfolding facilitates intermolecular transposition. The mechanism by which this unfolding is stimulated by H-NS is currently being investigated. Synaptic complexes including target DNA have been detected only with double end complexes, indicating that prior cleavage at both ends is a requirement for target capture (Sakai and Kleckner, 1997). The introduction of target DNA into the complex leads to the formation of the

strand transfer product, which exists as a stable protein complex. This stable protein-DNA architecture must be disassembled first to assist the host machinery in the processing of the DNA into a simple, physiologically tolerable structure. The disassembly process and the processing of the strand transfer product have not been studied in detail but is also thought to occur in a coordinated manner.

Not much is known concerning the steps involved in the progression of the synaptic complexes of other IS elements.

2.4. Control of Transposition Activity

The multiple copies of ISs present in the genome of bacteria places an extra burden on the host cell to replicate this extra DNA content. In addition, the multiple copies serve as substrates for recombination which can lead to lethal genome rearrangements and cell death (Doolittle *et al*, 1984). Through their ability to transpose, these elements are a major source of mutations and responsible for disrupting genes. In the light of these effects of ISs on the host, it is not difficult to imagine that transposition of these elements at high frequencies would be detrimental to the host cell. Indeed, all known ISs maintain relatively low frequencies of transposition depending on the genotype and the physiological state of the host organism. Different IS elements transpose with different frequencies, which varies between 10^{-3} or more usually 10^{-5} to 10^{-9} per cell division. However, transposition is also essential for the propagation and survival of ISs in a bacterial population. Thus, ISs and transposable elements in general, must have co-evolved with their hosts to invent mechanisms that facilitate the establishment of a long-term stable relationship between transposable element and its host. It is therefore not surprising that transposition is regulated not only by the element itself, but also by host factors especially those involved in signaling the physiological state of the host cell (Chandler and Mahillon, 2002). It is often seen that transposition activity increases when the host is stressed, where the transposable element is “called on” to stimulate evolution (Shapiro, 1999).

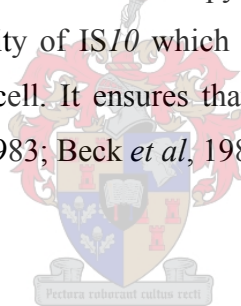
As mentioned above, the cognate T_pase promoters are located partially within the left inverted repeat. This arrangement may allow a degree of autoregulation by coupling T_pase binding (and thus transposition activity) with T_pase expression. In addition, these promoters are generally weak and facilitate weak binding of the RNA polymerase and yields low transcript levels

during transcription, and subsequent low expression of the T_pase gene. IS elements have adopted a variety of strategies to regulate T_pase activity. This has been reviewed in Galas and Chandler (1989) and more recently in Chandler and Mahillon (2002) and a summary will be given of the 2002 review as well as references made herein.

2.4.1. Transposase Expression and Activity

2.4.1.1. Production of Translational Inhibitors

The earliest studies on IS10 revealed that it employs one of the classical mechanisms of controlling gene expression. IS10 encodes an outward directed transcript that acts as an RNA repressor of translation. The antisense RNA transcript has a much longer half-life than the T_pase transcript so there is always more of the antisense transcript than the sense transcript. The antisense transcript binds to T_pase mRNA and inhibits the expression of this gene. It has been suggested that this strategy is to limit multicopy transposition in the host cell. In other words it reduces transposition activity of IS10 which is stimulated by the presence of many copies of the element in the same cell. It ensures that a fix copy number of the element is maintained (Simons and Kleckner, 1983; Beck *et al.*, 1980).



2.4.1.2. Impinging Transcription

There is *a priori* reason to expect that insertion of an IS element into an actively transcribed host gene would result into high expression levels of IS-encoded genes, which is driven by external transcription (transcription from the strong host gene promoter). No evidence has however been reported for this expected outcome. On the contrary, a mechanism observed in IS10 and IS50 prevents the high expression of the T_pase gene in this situation. The ribosome binding site or the translational initiation codon of the T_pase is also located close to or partially within the left inverted repeat (Krebs and Reznikoff, 1986; Davies *et al.*, 1985). Transcription from the host gene promoter would yield transcripts in which the ORFs as well as the inverted repeats of the IS element have been transcribed into mRNA. The inverted repeats are related by a consensus sequence and act as substrates for the generation of secondary structures in the mRNA molecule. This would render the translation initiation codons ineffective which results in no expression of the T_pase gene and no transposition activity. Transcripts from the

indigenous T_pase promoters would only contain the right inverted repeat and no secondary structure would be observed.

In some cases it has been observed that transposition activity is reduced by transcription across the end of element (IR). T_pases bind these ends specifically during transposition and transcription across the end would disrupt the essential complexes (transposomes) formed between the T_pase and the ends (Galas and Chandler, 1989; Goosen and van de Putte, 1986).

2.4.1.3. Programmed Translational Frameshifting

Members of some IS families of both eukaryotic and prokaryotic origin, use translational frameshifting to produce an active T_pase as a transframe protein from two consecutive out-of-phase reading frames (Ohtsubo and Sekine, 1996). In other words, a level of control is exercised by low-frequency frameshifting between the first and the second ORF to generate a fusion protein. These frames usually overlap in the –1 reading frame and frameshifting occurs in the –1 direction. The control of transposition activity through translational frameshifting has been demonstrated for IS1 (Sekine and Ohtsubo, 1989; Luthi *et al*, 1990; Escoubas *et al*, 1991) and IS3 (Sekine *et al*, 1994; Chandler and Fayet, 1993; Polard *et al*, 1992). The upstream frame contains the DNA binding domain and the downstream frame the catalytic domain. The upstream frame protein product can act as modulator of transposition activity by binding the IRs of the element and therefore prevents T_pase binding to these ends. Through translation frameshifting these two domains are fused into one protein, the T_pase.

Frameshifting occurs in a region called the slippery codons. This is a heptanucleotide sequence of the type Y YYX XXZ in phase 0 (triplet bases are paired with the anticodon) (Chandler and Fayet, 1993; Farabaugh, 1997; Gesteland and Atkins, 1996). When the sliding (translating) ribosomes reach these codons, they slide one base upstream and resumes translation in the –1 direction (phase). The slippery codons are now read as YYY XXX Z. An example of this phenomenon is shown in Fig 1.14. The most common example of a set of slippery codons is the A AAA AAG heptanucleotide. Several structures in the mRNA elevate the change of reading frames and obstruct the progression of the translating ribosomes in the phase 0 frame. These obstructions are usually potential ribosome binding sites upstream of the slippery codons or stem-loop structures (pseudoknots) downstream of the codons (Fig 1.14). It has been suggested that translational frameshifting may be influenced by the physiological state of the cell. However, this is not yet clear.

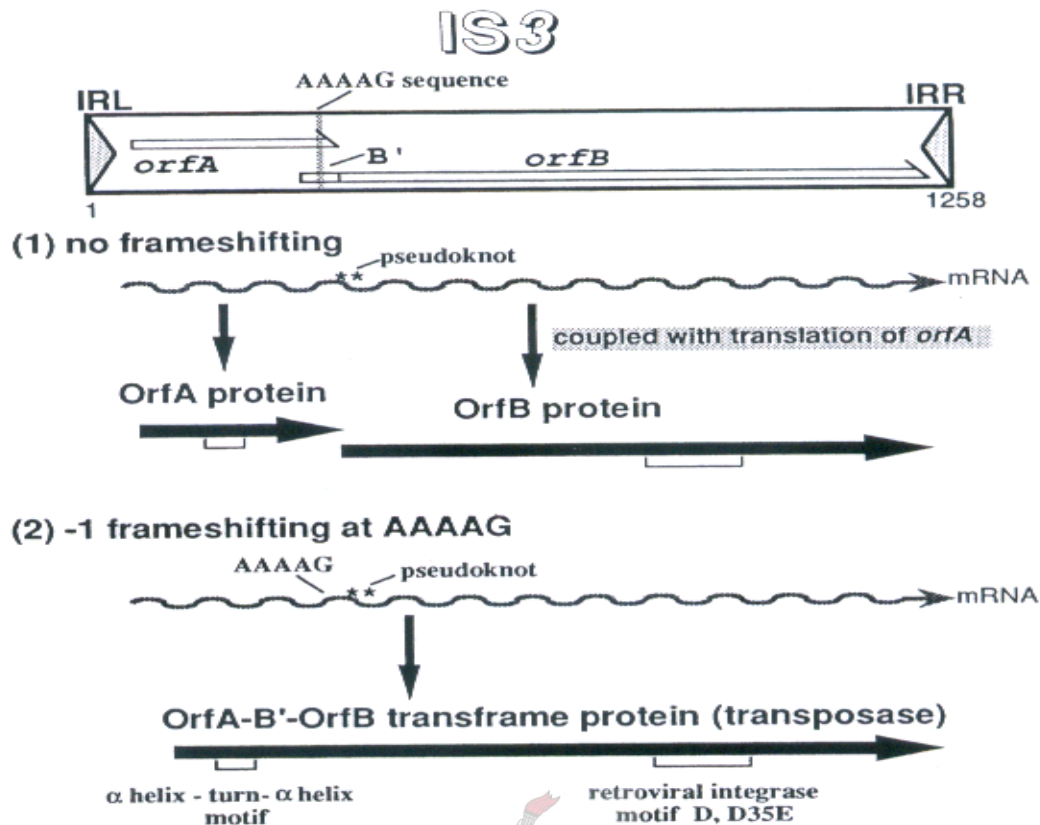


Figure 1.14: Structure and expression of IS3. IS3 codes for two ORFs, ORFA and ORFB. Reading frame, B', extending upstream of from the initiation codon ATG of ORFB overlaps ORFA in the -1 frame. The Tpsase is produced by -1 translational framehsifting at an AAAAG sequence (slippery codons) in the overlapping region between the two ORFs, inferred by the pseudoknots downstream of the slippery codons. Unlike most ISs, both ORFA and ORFB are produced unless frameshifting occurs (Reprinted from Ohtsubo en Sekine, 1996)

2.4.1.4. Transposase Stability

Like for any other protein, its stability would have an affect on its activity and it has been shown that stability of the Tpsase contributes to the control of transposition activity. Tpsases are very unstable so it is not normally accumulated to high concentration in the cell, which would lead to high transposition frequencies. Factors reported that affect Tpsase stability include the *E. coli* Lon protease which limits the activity of the Tpsase of IS903 temporally and spatially. This obervation has been suggested as a reason why several Tpsase preferentially acts in *cis*. Indeed, when mutants of IS903 have been derived that showed more resistance to the proteolytic action of the Lon protease, it was discovered that this mutant exhibited an increased tendency to function in *trans*. (Derbyshire *et al*, 1996; Derbyshire *et al*, 1990). Observations suggest that the efficiency of transposition activity, and thus Tpsase stability, may be dependent on temperature as seen in Tn3 and IS911 transposition (Haren *et al*, 1997; Kretschner and Cohen, 1979).

2.4.1.5. Translation Termination

In certain IS elements the translation termination codons of the Tpsases are located in the IR sequences. As mentioned, the IRs are bound by the Tpsase during transposition and this organization allows an autoregulation mechanism in which Tpsase binding and activity are coupled with translation termination. Several elements (Chandler and Mahillon, 2002) do not carry any translation termination codons but transposition into specific target sites generates a translation termination codon in the target DNA duplications, created upon insertion.

2.4.1.6. Activity in cis

A remarkable property of IS Tpsases is that they act efficiently only in *cis*, in other words on the same DNA molecule from which the Tpsase gene has been transcribed, and show a strong preference for closely linked ends (Grindley and Reed, 1985). Tpsases in general are quite basic (Galas and Chandler, 1989), a common feature of sequence-specific DNA binding proteins, and may play some role in the *cis*-action of Tpsases. Complementation of an IS element with a defective Tpsase, in *trans* occurs with an efficiency of about 1% or less. Chandler and Mahillon (2002) proposed that the observed *cis* effect of the Tpsase might be forced upon the element by the structural organization of the Tpsase. The N-terminal region carries the DNA binding domain and if this domain were capable of undergoing its three-dimensional folding independently of the C-terminal region, it would allow the DNA binding domain upon synthesis to bind neighboring DNA binding sites. The N-terminal region generally shows the highest affinity for DNA binding than any other part of the Tpsase and is theoretically capable of binding IRs of a different copy of the same IS element. It has been shown for both IS50 and IS10 that the presence of the C-terminal region reduces the binding activity of the Tpsase to the terminal ends (Jain and Kleckner, 1993; Weinreich *et al*, 1994). Accordingly it has been suggested that the C-terminal region masks the DNA binding activity of the N-terminal region to the IRs and thus restricts it to binding only its cognate terminal repeats.

2.4.2. Host Factors in Transposition

As described in the above report, IS elements are simple genetic entities and it is not surprising that certain cellular factors may play some role in stimulating or modulating transposition or possibly a structural role in the process. The earliest detailed studies on different ISs demonstrated that several *E. coli* chromosomal mutants effect IS activities and that the ends of

these elements carry putative binding sites for an assortment of host-encoded factors. It is the discovery of these mutants and the presence of these candidate binding sites that initially suggested the possible involvement of host factors in transposition (Syvanen, 1988; Galas and Chandler, 1989). These binding sites are usually located close to or within the terminal repeats. The T_{ps} promoter region is located within the left inverted repeat and together with the location of these binding sites this allows for a strategy to reduce the amount of active T_{ps} produced and to modulate transposition activity. The effect a host factor has on a certain element is specific for that element and might have a different effect on a different element and can intervene at different stages of the transposition process. A recent review on the effects of host-factors is presented by Chandler and Mahillon (2002). The role of host factors is in most cases poorly understood and only the factors understood to some degree will be included in the discussion here.

2.4.2.1. Histone-like proteins

Histone-like proteins form a group of small host encoded DNA-binding proteins which contribute to the organization of the bacterial nucleoid (nucleoid-associated proteins). They tend to introduce significant bends to DNA upon binding and play a role in the control of gene expression as well as participating in DNA transactions such as recombination and DNA replication (Dorman and Deighan; 2003). Various histone-like proteins (or DNA chaperones as they are more commonly known as) have been implicated in playing some role in mediating the transposition of certain transposable elements. IHF, HU, HNS and FIS are the three most intensively studied members of this protein group. There is a general notion that the DNA chaperones may play an engineering role, ensuring the correct three-dimensional nature of the synaptic complex prior to the DNA cutting and joining reactions. IHF (Integration Host Factor) was the first histone-like protein speculated to be involved in transposition when putative binding sites for this protein was discovered in both ends on *IS1* and subsequently in *IS10* and *Tn1000* (Galas and Chandler, 1989 and references therein). Candidate binding sites for IHF have been found in the terminal ends of several elements. IHF is a member of a family of small, heterodimeric, heat stable DNA binding proteins and was originally identified as a host protein required for the integration of bacteriophage lambda DNA into a host genome (Mizuuchi, 1992a). This host protein binds DNA sites with a consensus sequence of C/T AANNNTTGAT A/T. Putative IHF binding sites occur between the -35 and -10 T_{ps} promoter regions, which is located within the left inverted repeat, as seen in *IS1* and *IS903* (Chandler and Mahillon, 2002). Although *IS1* was the first element in which IHF binding sites

were identified, no convincing results have been reported which show a clear effect of IHF on *IS1* transposition. In *IS10* the IHF binding site is located outside the *Tpase* promoter, 43bp away, and the binding of IHF to this site appears to influence the nature of the transposition products. Morisato and Kleckner (1987) reported that IHF greatly stimulates, but is not absolutely required, for the *IS10* *Tpase*-dependent cutting reaction. IHF binding in *Tn1000* enhances *Tpase* binding to its cognate ends (Wiater and Grindley, 1988).

Hu is a small heterodimeric protein which is structurally related to IHF but binds DNA non-specifically. Its role in transposition was first recognized when it was discovered that it is required for the formation of the *Mu* transposition intermediate (Craigie and Mizuuchi, 1985). It plays an architectural role during the elaboration of the synaptic complex where it ensures that the correct pair of inverted repeats is bound by the *Tpase* and brought together in a stable protein-nucleic acid complex (Mizuuchi, 1992a). HU have been reported to stimulate *IS10* movement (Morisato and Kleckner, 1987).

H-NS has been reported to stimulate *IS1* transposition by promoting the formation of the *IS1* synaptic complex in which the cutting reactions are initiated (Shiga *et al*, 2001). FIS (factor for inversion stimulation) and H-NS have both been implicated in the *Mu* transposition process (Chandler and Mahillon, 2002). The precise role of these host factors in transposition is at present not clear. Further development of efficient *in vitro* transposition reactions for other transposable elements would lead to a better assessment of the role of these types of host proteins.

2.4.2.2. Dam Methylase

In *E. coli*, 5'-GATC-3' sequences are substrates for the Dam methylase (D-adenine methylase) which attaches methyl groups to the adenine (A) residues of these target sequences on each strand. The dam methylase was first discovered to play a role in *IS10* and *IS50* transposition where it reduces the transposition frequencies of these two elements by methylating the adenines in the multiple GATC sequences present in these elements (Syvanen, 1988). These sites are located in the extremities of these elements and the reduced transposition frequencies could be the result of two mechanisms: (1) fully methylated GATC sites in the terminal ends are not successfully recognized by the *Tpase* which results in poor binding of the *Tpase*; thus the activity of the ends are modulated by its methylation status; (ii) fully methylated GATC sites in the *Tpase* promoter regions results in a weak binding of the RNA polymerase and thus

reduces the amount of active T_pase expressed (referenced and reviewed Syvanen, 1988; Roberts *et al*, 1985; Yin and Reznikoff, 1987). The opposite effect was shown in *E. coli dam* mutants where increased transposition frequencies were seen for IS10 and IS50 than in wild-type *E. coli* strains. This *dam* control of expression of IS encoded genes has also been observed for IS903 where T_pase promoter activity is increased in a *dam*⁻ host (Chandler and Mahillon, 2002). In the light of the presented data, it appears that the intrinsic activity of these terminal ends is determined by its methylation status and these ends are more active when these sites are unmethylated or hemimethylated (newly replicated DNA) than when fully methylated. It is well known that parental DNA is always uniformly methylated providing the cell a means to distinguish between parental DNA strands and daughter strands during DNA replication and this plays an important role in the fidelity of the replication process. It is during the passage of the replication fork during DNA replication, where transiently hemimethylated DNA is found. Although the newly synthesized daughter strand is quickly methylated, this short hemimethylation state favours transcription and transposition and thereby coupling T_pase production and transposition to passage of the replication fork (Bushman, 2002). (It is interesting to note that this strategy can be used by transposable elements that transpose via the cut-and-paste mechanism to increase in copy number. In other words, it moves from replicated DNA to unreplicated DNA by excising after the passage of the replication fork and integrating ahead of the fork. This also ensures that a second copy of the donor DNA is available for the subsequent repair of the gap from which the element transposed from. The gap can be repaired by gene conversion using the second copy of the donor element, and in this way the cut-and-paste mechanism can also result in a net increase in copy number). Most of the IS families contain members which have similar methylation sites within the first 50 bp of one or both IRs. It should be however noted that the probability of a 100bp sequence carrying a GATC site is about 40% and in most of the ISs these sites may have no biological significance.

2.4.2.3. DNA polymerase I

DNA polymerase I have been directly implicated in both IS50 and IS10 transposition (Sasakawa *et al*, 1981; Syvanen *et al*, 1982; Syvanen, 1988). Host cells with mutations in *polA*, the gene coding for DNA polymerase I, and defective in both the polymerase and 5'-3' exonuclease activities are deficient in transposition. It is currently unknown which of the functions of DNA polymerase is needed for transposition, but it is thought that the transposition process leaves the DNA nicked or gapped and DNA polymerase I is required to fill and repair these staggered gaps leading to the formation of the DRs. The notion of transposition leaving a

gap in the donor DNA is further supported by the evidence presented by Little and Mount (1982), which showed that the host SOS functions are induced in the event of transposition. The host SOS function is a cellular response to prevent and limit DNA damage, for example during transposition.

2.4.2.4. DNA gyrase

Sternglanz *et al* (1981) first reported that mutations in *top*, the structural gene for DNA topoisomerase I affect the transposition of some transposable elements such as Tn5, Tn10 and Tn9, but not others which includes Tn3. These mutations were found to be in the DNA gyrase subunit genes, *gyrA* and *gyrB*, which results in the reduction of the degree of supercoiling of DNA. This affect can be explained by the fact that certain transposable elements are sensitive to the degree of supercoiling of donor or recipient DNA or it might require a certain DNA topology for transposition. It was also proposed that the energy of supercoiling might drive one or more of the reaction steps of transposition (Galas and Chandler, 1989).

2.4.2.5. DnaA protein

DnaA binds DNA specifically and is primarily involved in the initiation of chromosome replication. A few insertion elements have been identified that carry a consensus sequence for possible DnaA protein-binding sites (Galas and Chandler, 1989; Chandler and Mahillon, 2002). IS50 carries a binding site for DnaA in one of its terminal ends and it has been shown that this end is recognized and bound by DnaA *in vitro*, and appears to affect the activity of this end in transposition *in vivo* (reviewed and referenced in Galas and Chandler, 1989).

2.4.2.6. RecA protein

RecA is the most intensively studied recombination gene which directly and indirectly regulates recombination. *RecA* null mutations almost completely loose their recombination activities in most *E. coli* mutations. Transposition is documented to occur independently of the host's recombination machinery and that could explain to a degree the DNA rearrangements seen in *recA*⁻ mutants. As explained above (see section 23.2, “Replicative Transposition”), cointegrates could be formed via two pathways, either as an intermediate during replicative transposition or as a product of rearrangement activities which is RecA-dependent. It appears

that RecA has no direct effect on transposition. However, as a result of contradicting findings, it is still uncertain what role RecA plays in Tn5 (IS50) transposition (reviewed and referenced in Chandler and Mahillon, 2002).

2.4.2.7. The ClpX, ClpP and Lon proteases

Although the involvement of these proteases has not been widely documented, it has been implicated in the transposition of Mu and IS903. ClpX is an unfolding chaperone and is a member of the Clp/Hsp100 ATPase family and is involved in the disassembling of the Mu synaptic complex to make way for the host replication machinery to complete the process. ClpP plays a role in the degradation of the Mu repressor (Chandler and Mahillon, 2002) and thereby increasing transposition activity. It was mentioned above that the Lon protease degrades the IS903 T_pase and reduces its transposition frequency.

2.5. Transposition Immunity

Transposition immunity (target immunity) is a phenomenon by which the presence of a transposable element in a target molecule renders the nearby DNA region a poor target for insertion of a second copy of the same element (Mizuuchi, 1992a; Chandler and Mahillon, 2002). It reduces the affinity of the target molecule for an additional insertion of the same type of element so that the elements do not insert into DNA already containing a copy of the element. Transposition immunity is a *cis*-acting phenomenon effective only on potential target DNAs already containing a copy of the element. It does not act to serve as a global inhibition of transposition within a host but only to regulate, to some degree, the transposition activity of some elements once they are established within a host genome. This property is usually displayed by transposons that transpose at high frequencies such as the Tn3 family of transposons (Lee *et al*, 1983; Wiater and Grindley, 1990), Tn7 (Arciszewska *et al*, 1989; Hauer and Shapiro; 1984) and Mu (Darzins *et al*, 1988; Reyes *et al*, 1987; Adzuma and Mizuuchi, 1988). At present it is believed that IS elements have not adopted this strategy although recent evidence suggests that IS21 may display some degree of transposition immunity (Berger and Haas, 2001). In order to determine at which distance immunity can act, studies on Tn7 target immunity in the *E. coli* chromosome were done and revealed that immunity was exercised effectively over distances of at least 190kb (Craig, 1997; DeBoy and Craig, 1996a). That is, Tn7 at one position can prevent insertion of another copy in a position closer than 190kb.

Transposition immunity was first discovered for Tn3 and it has been demonstrated that a single copy of the terminal IR sequence of Tn3 imposes target immunity on a plasmid molecule (Mizuuchi, 1992a) and this has been also demonstrated for bacteriophage Mu. Transposition immunity may provide an element with a barrier for preventing self-destruction where it helps transposable elements to avoid another copy of the element to integrate into it or prevents integrating itself into another copy of an element. Although such events (intramolecular insertions) may lead to self-destruction of the element via deletions, Tn10 channels these intramolecular insertions to create inversions, and exploits this kind of recombination reaction for element evolution (Craig, 1997). Transposition immunity is also a means by which the insertion of multiple copies of an element in close proximity on a chromosome can be prevented, which may otherwise cause great genome instability, as homologous recombination would yield potentially dangerous gene rearrangements. Composite transposons are generally composed of two IS elements, in close proximity to each other, flanking and rendering bacterial genes (interstitial DNA) mobile. Therefore, transposition immunity is probably not practiced by IS elements and this allows the formation of composite transposons.

In both bacteriophage Mu (Berns and Linden, 1995; Lavoie and Chaconas, 1996; Adzuma and Mizuuchi, 1988) and Tn7 (Bainton *et al.*, 1993; and referenced in http://www-is.biotoul.fr/is/IS_infos/is_general.html), transposition is catalyzed by two proteins, the T_pase which catalyzes the breaking and joining reactions of transposition, and a second ATP-dependent nonspecific DNA-binding protein involved in binding target DNA that provides the T_pase with a preferential target during the strand transfer step. It is this second protein involved in transposition that selects a target that carries no copy of the specific element over a certain length of area and ensures a degree of transposition immunity. Tn3 target immunity is less well understood and the major difference between this system and systems from Mu and Tn7 is that Tn3 target immunity is not mediated by a second protein, but by the Tn3 T_pase (Sheratt, 1989).

2.6. Target Specificity

Target site selection differs from element to element and the specific target site preference is a characteristic unique to a specific type of element. While some transposable elements can insert into many different sites, target sites is usually not random. Some elements prefer certain sites or regions for insertion while others may avoid certain sites or regions, thus displaying some degree of target site selectivity (Craig, 1997). This skill is likely to ensure element propagation while maintaining and establishing an optimized element-host relationship. A good element-host relationship would be favored if transposable elements avoid insertion into essential host

genes but rather into sites where insertion would not have an unfavorable effect on the host organism. Some elements usually insert specifically upstream of promoters and therefore avoid disrupting the expression of a gene. A second strategy for avoiding insertional inactivation of essential genes would be to insert into genes that have multiple copies within the genome. Some mobile elements are also introns (Craig, 1997) and when these elements insert into essential genes they can be removed from the mRNA transcript through RNA splicing. Despite these rather clever strategies, some elements do insert into essential genes.

Several mechanisms exist by which target sites may be selected. They could be selected through the direct interaction of the T_pase and the target DNA molecule during the transposition process (before or after the donor cleavage reaction). The interaction between the T_pase and element- and/or host-encoded accessory proteins, may also be a way of selecting a specific or suitable target site. The spatial location as well as the state (global features) of the target DNA could play a role in the accessibility of a target site for insertion. Some elements avoid transcribed regions. Selection of a target site (target capture) can occur at different stages of transpositional recombination: (i) during the formation of the synaptic complex; (ii) after the stabilization of the transposome and just before the donor cleavage reaction; or (iii) after the donor cleavage reactions.

Some transposable elements display a high degree of sequence-specific insertion, i.e. a marked preference for target sites that share a common consensus sequence. This stringency differs from element to element. Tn7 can use one of two kinds of target sites: (i) it can insert at high frequencies into a specific site, *attTn7*, in the *E. coli* chromosome of many bacteria; and (ii) it can also insert at lower frequencies into non-*attTn7* sites (Craig, 1996b). IS91 also exhibits a high degree of target site selection, which requires a GAAC/CAAG target consensus sequence for insertion (Mendiola *et al.*, 1989). Some elements exhibit less-strict sequence specific insertions and these include members of the IS630 and *mariner*/Tc families, which both require a TA dinucleotide in the target. The insertion into a low-specificity consensus sequence has also been observed in IS10, IS50 and IS231. IS10 inserts preferentially into the consensus target sequence 5'-NGCTNGACN-3' and insertion results in the duplication of these nine base pairs (Halling and Kleckner, 1982; Lee *et al.*, 1987). IS50 transposes into 5'-AGNTYWRANCT-3' (Goryshin *et al.*, 1998). IS231 shows a preference for 5'-GGG(N₅)CCC-3' (Hallet *et al.*, 1994). Mu shows a preference for 5'-NYG/CRN-3' (Mizuuchi and Mizuuchi, 1993). One of the possible reasons for elements recognizing a particular sequence for insertion is that the specific sequence might provide an optimal geometric feature for the T_pase-target

interaction. It can also provide the element with a strategy to avoid insertion into itself by preferentially inserting into sites that is not contained within the element allowing its dispersal and propagation.

Some elements prefer insertion into certain regions for example a GC- or AT-rich segments such as *IS186* (Sengstang *et al* 1986) and *IS1* (Zerbib *et al*, 1985, Chandler and Mahillon, 2002). The geometry of the DNA substrates at these regions may be ideal for an optimal T_pase-target interaction. Some elements prefer insertion into bent DNA, for example when DNA is assembled with histones into nucleosomes or when the bend is caused by an interaction of the DNA with DNA-binding proteins. *IS231*, *IS4* and *IS10* are examples of elements preferentially inserting into bent DNA. Other factors documented by Chandler and Mahillon (2002) and Craig (1997) to have an effect on target specificity include the following: (i) the degree of supercoiling of target DNA as in the case with *IS50*; (ii) some elements are attracted to replicating DNA and conjugal replicating DNA, e.g. *IS102* and Tn7; (iii) transcription of a potential target site is reported to reduce transposition frequency and may serve as a strategy to direct the element away from the most essential genes (those being actively transcribed); e.g. *IS102*, Tn7, Tn5 and Tn10.

A few IS elements have been shown to display a preference for insertion into other elements. *IS231* targets the terminal 38bp of Tn4430 for insertion. *IS21*, *IS30* and *IS911* show a preference for insertion close to the end of a second copy of itself (Chandler and Mahillon, 2002).

2.7. Genetic effects of the IS elements

Only a brief discussion on the genetic effects of IS elements on the host genome is presented.

2.7.1. IS-mediated DNA rearrangements

In addition to simple insertions, IS elements have been widely documented for their ability to mediate a variety of DNA rearrangements such as deletions, replicon fusions and inversions (Iida *et al*, 1983; Galas and Chandler, 1989; Syvanen, 1988; Craig, 1996a; Merlin *et al*, 2000; Chandler and Mahillon, 2002). Certain IS elements transpose via the replicative mode of transposition. This mechanism can be used as a powerful tool for rearranging DNA and does not require the need for any sequence homology. As explained above, intermolecular transposition of an IS element stimulates the formation of replicon fusions (cointegrates). If

intramolecular replicative transposition occurs, the process can result in adjacent deletions and inversions (see Fig 1.15 for an explanation). It should be noted however that not all elements are found to exhibit all these restructuring capacities and that the relative frequencies of occurrence of these rearrangements varies widely among the ISs (Galas and Chandler, 1989).

As mentioned above, the multiple copies of IS elements in the host genome can serve as substrates for homologous recombination which depends on the host recombination machinery, thus forming new joints between previously non-homologous DNA. The importance of IS elements in homologous recombination was initially recognized during the behavioral studies of the sex plasmids (F plasmids) and the multi-drug resistance plasmids (R-plasmids) in various bacterial hosts. Both these plasmids carry a cluster of IS elements (Galas and Chandler, 1989). Studies revealed that, homologous recombination between these elements located on the plasmids and copies on the *E. coli* chromosome resulted in the integration of these elements into the host chromosome. Integration of F plasmids into the host chromosome leads to the formation of Hfr (high frequency of recombination) male strains. It is also reported that IS elements are responsible for the excision of the integrated F plasmid together with flanking host DNA to produce F-prime plasmids. Merlin *et al* (2000) also proposed that R-prime plasmids can be formed either via a replicative transposition event of an IS element on the R plasmid or through homologous recombination between two appropriately located IS sequences, one on the plasmid and the second on the host chromosome. R-prime plasmids play an intimate role in mobilizing chromosomal genes between bacterial populations, which can be stably incorporated by these cells. Homologous recombination using the IS elements as regions of homology, can result in the formation of duplications, inversions, deletions and replicon fusions.

2.7.2. Agents of mutations and effects on transcription

In addition to disrupting gene function through insertional inactivation and the generation of target repeats, IS elements have been reported to be one of the major sources of spontaneous mutations. Several insertion elements were initially discovered because of their ability to exert strong polar effects on the expression of downstream genes (Shapiro, 1969; Malamy, 1970; Jordan *et al*, 1968). Polarity refers to a decrease in expression of downstream genes. One explanation for this type of behaviour is that IS elements carry internal transcription or translation termination sites. If transcription enters the element from flanking donor DNA these sites would induce premature termination of transcription so that the downstream genes are not

transcribed. This effect has been reported for a number of IS elements such as IS1, IS2, IS4 and IS5 and appears to be a common property of these elements (Iida *et al*, 1983; Galas and Chandler, 1989). One example of these transcription termination sites is the Rho-dependent termination sites. Rho is an oligomeric protein which binds weakly to DNA and tightly to RNA. The binding of Rho to these sites would cause the RNA polymerase to pause and transcription to terminate. Polarity can be partially suppressed by mutations in the *rho* gene or by using antitermination proteins. A second proposal is that certain elements carry nonsense codons in all three reading frames in their primary sequence which would result in transcription being terminated (reviewed in Iida *et al*, 1983) or by producing nonsense mutations upon insertion (Galas and Chandler, 1989).

Most, but not all mutations caused by ISs revert back to the wild-type phenotype and often the wild type genotype. The ability to restore gene function inactivated by IS insertion or the relief of polarity on genes distal from the insertion relies on precise and precise or imprecise excision of the IS element, respectively. Precise excision involves the deletion of the IS element including one copy of the target repeat, which was generated upon insertion. This results in the restoration of both the phenotype and the genotype. Precise excision appears not to be Rec-dependent and seems to be independent of the T_pase functions as in the case of IS50 (Tn5) and IS10 (Tn10) but it is possible that for other IS elements the T_pase would be essential for transposition (Galas and Chandler, 1989 and references therein). It is generally proposed that excision is the result of intra-strand pairing between the inverted repeats, leading to the “bulging out” of the IS element and its deletion by subsequent replication “slippage” across the small directly repeated target sequence.

In contrast to inducing polar mutations some elements can activate the expression of neighboring downstream genes and has been observed for several elements, including elements reported to exert polar effects on distal genes (Chandler and Mahillon, 2002 and references therein). Many IS elements carry outwardly directed -35 promoter hexamers in their IRs and by placing this region at the correct distance from a resident -10 hexamer upon insertion (usually in AT-rich regions), would result in the formation of a functional hybrid promoter. This has been documented for numerous IS elements (extensively referenced in Chandler and Mahillon, 2002; and Galas and Chandler, 1989). How a certain element can both activate or decrease expression of distal genes upon insertion is explained by the fact that the potential -35 promoter region is located only at one end of the element, and thus the orientation of the insertion of the element, determines the effect of the IS element on neighbouring genes.

Although poorly understood, another genetic effect mediated by IS elements involves the activation of the *bgl* operon of *E. coli*. This operon is usually cryptic but can be activated by insertion of either IS1 or IS5 (referenced and reviewed in Galas and Chandler, 1989). The activation is not a result of transcription driven from the internal or hybrid (outwardly directed) promoters of the element and activation is independent of the orientation of the insertion of the elements (activation occurs in both orientations) and has been observed to occur over a 47bp region (between -77 and -124bp). Transcription always initiates at the same site in this operon, i.e. proximal to the beginning of the first gene and stimulation appears to occur at a distance. A possible explanation of this enhancer-like effect is that the presence of the element enhances binding of the cyclic AMP-binding protein (CAP) which is involved in the *bgl* expression. Point mutations can also result in increased binding of CAP and thus expression of the *bgl* operon. Mutations in the gyrase subunits are known to activate the *bgl* operon and thus conformational changes induced directly by the insertion or indirectly as a result of proteins that bind to the IS element, have also been implicated in this phenomena.

2.8. The IS families

ISs are a large group of intrachromosomal genetic elements that have been found in virtually all organisms examined. IS1 was the first IS element to be completely sequenced in the 1970s (Ohtsubo and Ohtsubo, 1978). Since then the number and diversity of these elements rapidly increased to a vast amount due to an increase in genomic sequencing projects. The current database, <http://www-IS.biotoul.fr>, contains more than 800 ISs isolated from 86 genera representing 196 bacterial species of both eubacteria and the archaea of which 523 have not yet been assigned to a family (Chandler and Mahillon, 2002). ISs are very versatile and promiscuous because a given IS element is often found flanking different bacterial genes in composite transposons (Reviewed in Galas and Chandler, 1989). The great number and diversity of ISs, which will certainly increase, provides some idea on the impact these transposable elements have on bacterial populations. At present, ISs are classified into 19 families. The families are based on the following criteria: (i) similarities in genetic organization, i.e. the arrangement of the open reading frames, (ii) common domains or motifs in their T_pase, (iii) similar features in the IRs, (iii) length of the target sequence duplications. A detailed discussion of the different families will not be given, but a summary of the most important characteristics of the families (Table 1.1). The family of interest is the ISL3 family, but a discussion on this family will be given in Chapter 3.

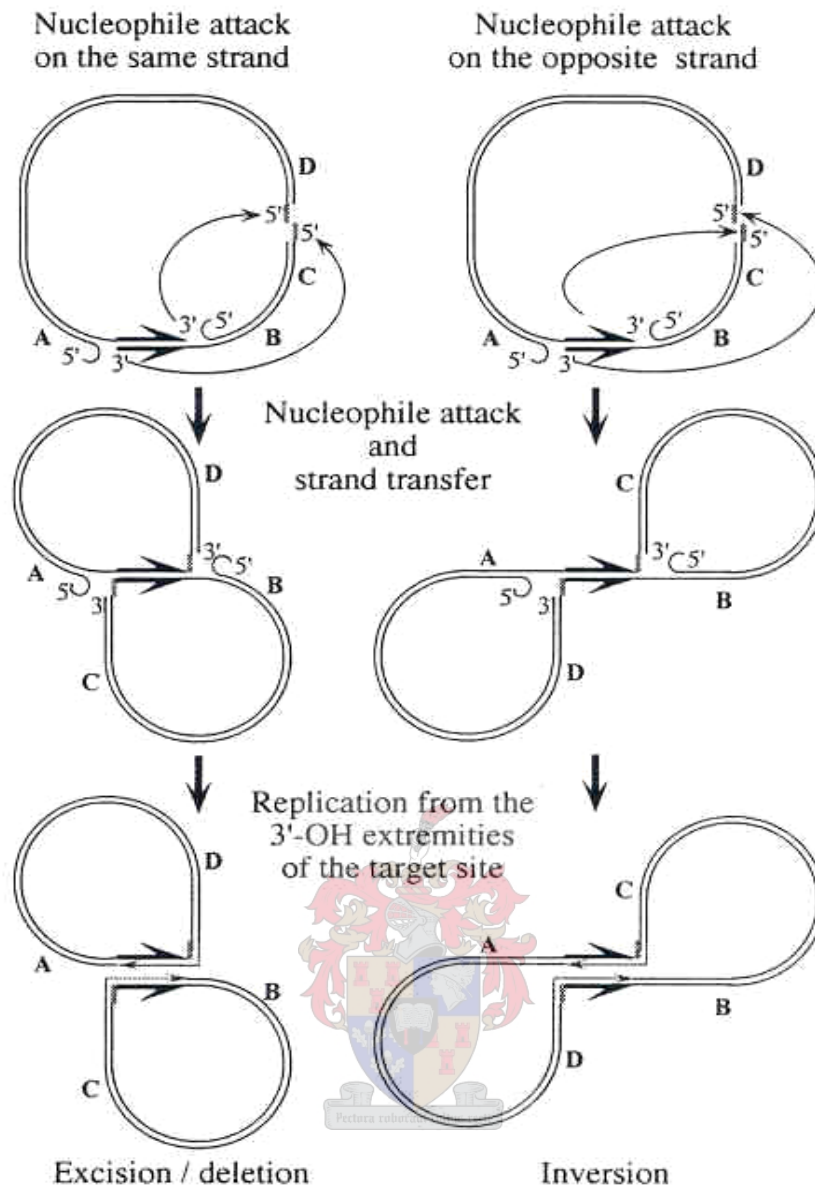


Fig 1.15: Intramolecular replicative transposition. The transposable element is represented by a double black arrow and the basic mechanism of transposition is the same as explained above (“Replicative Transposition”). Transfer of the free 3'-OH element ends generated by the T_pase to the target site can occur on either the same strand (pathway shown on the left) or on the opposite strand (pathway on the right). In both pathways free 3'-OH ends are generated and DNA replication is initiated at these ends which proceeds through the element. In the excision/deletion pathway (shown on the left), the intermediate structures are resolved into two covalently closed circles, each carrying a copy of the transposable elements. The deleted circular fragment is lost unless it carries an origin of replication to persist as an extrachromosomal element. In the inversion pathway on the right, a new circular DNA molecule is formed in which a section adjacent to the original transposable element copy has been inverted vs. its original position and is now flanked by two copies of the transposable element which are in opposite orientation (Reprinted from Merlin *et al*, 2000)

Table1.1.Major features of the IS families (Reprinted from Chandler and Mahillon, 2002).

Family	Group	Size (bp) ^a	DR ^b	Ends ^c	IR ^d	ORF ^e	Recombinase ^f
IS1		770	9 (8-11)	GGT	Y	2	Phage integrase?
IS3	IS2	1,300-1,350	5	TGA	Y	2	DD(35)E
	IS3	1,200-1,300	3 (4)			2	
	IS51	1,300-1,400	3 (4)			2	
	IS150	1,400-1,550	3-5			2	
	IS407	1,200-1,250	4			2	
IS4		1,300-1,950	9-12	C (A)	Y	1	DD-E
IS5	IS5	1,100-1,350	4	GG	Y	1	DD-E
	IS427	800-1,000	2-3	Ga/g		-2	
	IS903	1,000-1,100	9	GGC		1	
	IS1031	850-950	3	GAG		1	
	ISH1	900-1,150	8			1	
	ISL2	800-1,100	2-3			1	
IS6		750-900	8	GG	Y	1	DD(34)E
IS21		1,950-2,500	4 (5,8)	TG	Y	2	DD-E
IS30		1,000-1,250	2-3		Y	1	DD(33)E
IS66		2,500-2,700	8	GTA	Y	3	
IS91		1,500-1,850	0		N	1	Rolling circle transposition
IS110		1,200-1,550	0		N	1	Site-specific recombinase
IS200/IS605		700-2,000	0		N	1 (2)	
IS256		1,300-1,500	8-9	Gg/a	Y	1	DD-E
IS481		950-1,100	5-6	TGT (a/g)	Y	1	DD-E
IS630		1,100-1,200	2		Y	1	DD-E, eukaryote relatives
IS982		1,000	ND	AC	Y	1	DD-E
IS1380		1,650	4	Cc/g	Y	1	DD-E
ISAS1		1,200-1,350	8	C	Y	1	DD-E
ISL3		1,300-1,550	8	GG	Y	1	DD-E
Tn3		>3,000	5	GGGG	Y	1 (2)	DD-E

^aLength range in base pairs (bp) represents the typical range of each group

^bDirect target repeats in bp; ND, not determined. Less frequently observed lengths are included in parentheses

^cThe most conserved nucleotides present at the ends of each IS family are indicated, including some variations.

^dPresence (Y) or absence (N) of terminal inverted repeats

^eNumbers in parentheses denote variations

^fDDE represents the common acidic triad presumed to be part of the active site of several Tpsases.

2.9. Significance and evolutionary implications of IS elements

Transposable elements in general have proven to be a matter of interest to geneticists, genome annotators and structural and evolutionary biologists (Holmes, 2002). Geneticists mainly

exploit these elements as experimental tools. This will be returned to later (Chapter 3). Due to the multiple copies as well as fragments of these elements within genomes, they may comprise much of the repetitive sequence content within these genomes. As a result, genome annotators regard them as a nuisance and as “junk DNA”. They can profoundly confuse the outcome of statistical analysis such as the Expectation values (E values) by programs like BLAST (Altschul *et al*, 1990). However, these repetitive sequences are useful in their analysis of genome sequences because previously characterized elements can help them to identify and mask out “junk DNA” prior to DNA analysis. The very repetitive nature of these elements can also be used to identify them. Transposable elements have been documented to have many homologies to more complex systems like the replication machinery of many viruses, and also to transcription factors and other specific DNA- and RNA-binding proteins. Transposition involves many specialized tasks which include specific and nonspecific sequence recognition, DNA and RNA processing, host defenses and auto-regulation. Thus, to molecular biologists interested in protein structure and function, transposable elements are linked to a variety of interesting cellular processes.

Transposable elements arouse the most interest among the evolutionary biologists. Ever since the discovery that segments of DNA can translocate within genomes, has their functional significance and whether these elements are DNA sequences beyond mere genome parasites, been a matter of debate among molecular and evolutionary biologists (Labrador and Corces, 2002).

Many discussions have been documented focusing on this issue and the content hereof is too large to deal with comprehensively here. How IS elements (transposable elements) and host evolution are linked is an important and rational question. One view was that IS elements are genomic parasites, supported by the “Selfish DNA” hypothesis (Sapienza and Doolittle, 1981, Orgel and Crick, 1980; Orgel *et al*, 1980). Selfish DNA has been defined as sequences that make no positive contribution to the phenotype of the host, but instead place an extra burden on the host by forming additional copies, and thereby ensuring their persistence within the host. These sequences are parasites because they benefit (by being propagated) but do not positively contribute to the host (due to the lack of any measurable phenotype); instead they impose on the welfare of the host cell because of their replicative load and mutator activity. In support of this hypothesis it is believed that if these parasites contribute to adaptive changes it is their effect and not their function (Maynard-Smith, 1982). Another hypothesis which has received

much attention is the “Selection hypothesis”, which has more of a mutualistic view of transposable elements. Because IS elements confer no obvious phenotype on the host, except for their mutator activity, it was proposed that IS elements are sustained within genomes by direct selection acting on genetic variability (Syvanen, 1988 and references therein). In other words, these elements are selected for each time a mutation induced by these elements is selected for. The hypothesis claims that these elements have contributed to adaptive evolution and may still induce mutations to produce potential variants that can be fixed in the future. This hypothesis has been criticized as a teleological explanation of how seemingly functionless DNA is stored up in order to have it available in future times of need (Mayr, 1982). In the “Selection Hypothesis” it is explained that a species that has given rise to successful variants at the highest rate will be favored and it is reasonable to assume that the species will carry with it the mutational mechanism that gave birth to these variants. This mutational mechanism may give rise to future successful variants in the future when needed. Thus, if the benefit of carrying these mutators outweighs the disadvantage of carrying these mutators, a simple nonteleological explanation for the maintenance of the IS elements is apparent, based on their function as mutators. This holds true for the hospital environment which is experiencing recurrent antimicrobial threats (Rice, 2002), due to the emergence and spread of antimicrobial resistance in bacteria. One has the perception that in the absence of an immediate selection (stress) that a host will lose the resistance determinants because this extra DNA content confers a replicative cost to the microorganism. This cost is perceived as reasonable in an environment under selective pressure but becomes a burden when this selective pressure is removed. But it seems reasonable that a microorganism’s long term interest would be to maintain these resistance determinants, in the absence of any immediate threat, which could lead to resistance variants in the future. Instead of deciding whether the maintenance of these determinants would be advantageous, microorganisms develop ways of how to best maintain and transmit these determinants while conferring minimal costs on itself.

Today the idea of IS elements, or transposable elements in general, as being parasites or selfish DNA is fleshed out with many published examples of the contribution of these elements to the adaptive evolution not just of bacterial genomes but also in higher organisms including humans (Bushman, 2002; Bennett, 2004). It is firmly established that transposable elements are ubiquitous components of all living organisms, which in their coexistence with their respective hosts have played an important role in the evolution of complex genetic works (Miller and Capy, 2004). Transposable elements are regarded as natural genetic engineering systems capable of acting genome-wide and are not restricted to one site. In this chapter it was referred

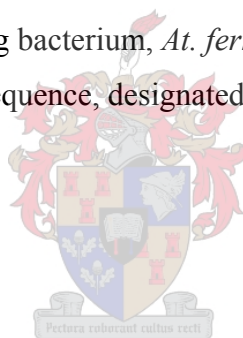
to the many genetic effects IS elements confer on the host as well as the properties of these elements which can play a role in the genome evolution of its host. Compound transposons, mentioned in the introduction, haven't been dealt with here in great detail. Antibiotic resistance genes are usually associated with compound transposons, but theoretically depending on the frequency of insertion and target specificity, two IS elements can mobilize any genes from one replicon to another, including chromosome-plasmid translocation of the interstitial DNA and hereby also creating conditions for lateral transfer between species. Both chemical and genetic forces can alter the genetic content of DNA by changing its nucleotide sequence. Mutations induced by IS elements and those of other transposable elements are more structured than those mutations caused by chemical toxicity, by rearranging rather than corrupting the host genome, more readily generating a neutral or advantageous phenotype (Kidwell and Lisch, 2001). An attractive result reported is that transposition activity increases when the host is stressed (Capy *et al*, 2000), leading to a mutualistic rather than a parasitic view of the host-transposable elements relationship. It is thought that these natural genetic engineers are called in to stimulate evolution at times of stress (Shapiro, 1999).

The subject of evolution is quite immense and it is almost impossible to be exhaustive on all the viewpoints reported on the role of IS elements in adaptive evolution, in the limited amount of space available here, many aspects haven't been dealt with. I think one important aspect would be the evolution of IS elements (and other transposable elements) as entities. Why and how these elements evolved. IS elements could have been an inevitable product of the replication machinery of the cell (Labrador and Corces, 2002) or even an inevitable by-product of the evolution of sequence-specific endonucleases (Federoff, 2002). In the light of this, one would think that transposable elements could be parasitic or mutualistic in nature, depending on the time frame in which the question is asked. They could have been parasites early in their evolution and could be parasitic in nature when there is no immediate reason for assisting the cell in evolution. However, because of the co-existence and co-evolving of these elements with its host, well developed systems arose which allowed a reasonable mutualistic transposable element-host relationship. Some of these mechanisms have been described in "Control of Transposition Activity". The biological importance of IS elements are evident from the short report presented here.

3. The Aim of the study

The importance of biomining as an industrial process is evident from the short report given on microbial mining. In order to advance the biotechnology of biomining an attempt has to be made to study and understand the molecular biology of the most important role-playing biomining microorganisms. Not only are studies on the mechanisms of the expression and regulation of important chromosomal genes necessary but also an understanding of the nature of plasmids that play a role as vectors of mobile genes within the horizontal gene pool of the bioleaching environment. Development of suitable cloning vectors would also be beneficial to the genetic engineering of these microbes.

This study focused on a 5.6kb possible transposon region of pTC-F14, isolated from *At. caldus* f. The first aim of the study was to determine what accessory genes, if any, are carried by the plasmid. This involved the mapping, cloning, sequencing and sequence analysis of this region. Sequence analysis of the region of interest yielded an open reading frame, which showed high sequence homology to IS*AfeI*, a member of the ISL3 family of insertion sequences, isolated from a different but related biomining bacterium, *At. ferrooxidans*. Thus, our second aim was to partially characterize this insertion sequence, designated IS*AtcI*.



2. Chapter Two: The cloning, sequencing and analysis of the accessory gene region of the plasmid pTC-F14, isolated from *Acidithiobacillus caldus* f. *

Table of Contents

2.1. Introduction	71
2.2. Material and Methods	73
2.3. Results	76
2.3.1. Restriction mapping and cloning of the accessory gene region	76
2.3.2. Sequence analysis of the accessory gene area	77
2.4. Discussion.....	89

*These results have been published (Goldschmidt *et al*, 2003). Since the publication the sequence analyses and text in this chapter have been updated.

2.1. Introduction

In addition to the chromosome, bacterial cells often contain plasmids which form a very important part of the genome. Plasmids have been found in essentially all types of bacteria. Plasmids species are extremely diverse in terms of size (1.5 to > 600kb), copy number (one to several hundred per cell) and in terms of the phenotypic traits they confer on the host bacterium (Osborn *et al*, 2000). Plasmids are most often circular DNA molecules but some bacteria contain linear plasmids. They also differ in terms of their host range and their stability. Plasmids do not encode functions essential to bacterial growth and usually encode two general classes of traits: (1) the essential plasmid functions, also called the plasmid-selfish characters, allows the plasmid to replicate and to be maintained within the cell and to be transferred between cells. These functions form the so-called plasmid “backbone” and involve mobilization, conjugation, inhibition of bacterial cell division and plasmid exclusion and partitioning functions. They promote the survival and replication of the plasmid as a unit. Although these characters are regarded as essential functions, it should be noted however that not all plasmids carry genes encoding for their transfer and maintenance. In other words, not all plasmids are mobile; (2) besides the machinery for their own maintenance and transfer, most plasmids carry genes coding for accessory functions. These functions hitchhike on the plasmid backbone and may confer a specific phenotype on the host and are not essential for host survival and growth but may enhance cell survival in certain hostile environments in which the host would otherwise not persist. The types of phenotypic traits found on plasmids will be returned to later. Due to the mobile character and maintenance functions of some plasmids they represent a fluid part of the chromosome and can move horizontally across species boundaries allowing for a certain degree of gene exchange between species and thus play a key role in bacterial adaptation and evolution. Plasmids are often cited as an important component of bacterial diversity (Smalla *et al*, 2000). Besides the evolutionary implications of plasmids, fully characterized plasmids also serve as important tools in molecular biology studies.

In the light of the above, plasmids can generally be defined as extrachromosomal genetic elements capable of replicating independently of the host chromosome and which may contain supplementary genes not essential for host survival, but enhance the proliferation and survival of the host in certain environments (Coplin, 1989). Plasmids have been documented to code for a variety of accessory phenotypic traits. This includes resistance to antibiotics, heavy metals, UV irradiation and other poisons; genes for the degradation of unusual toxic and/or xenobiotic organic compounds especially manmade compounds; bacteriocins; determinants for pathogenicity; phage resistance, genes coding for proteins involved in plant-host interactions

and nitrogen fixation (Eberhard, 1989 and references therein; Top *et al*, 2000). The increased use of antibiotics and industrial pollution with heavy metals and toxic compounds have resulted in the horizontal spread of these resistant genes between sometimes distantly related bacteria. The accessory functions located on plasmids may be acquired from the chromosome of an organism and the possibilities of mechanisms to achieve this are vast. This may be a result of IS elements, transposons, integrons, gene cassettes and conjugative plasmid activities (Osborn *et al*, 2000). It is probable that chromosomes and plasmids share the same gene pool, since the elements mentioned can move genes from the chromosome to the plasmid and from the plasmid to the chromosome. Why plasmids carry certain genes but avoid carrying housekeeping genes has been addressed by Eberhard (1989). He argues that genes that confer adaptations to restrictive environments (i.e. allowing survival in a larger variety of specialized ecological niches or specific set of environmental conditions) are more favorable to have on plasmids in contrast to “housekeeping genes” which are best located on the chromosome.

The majority of plasmids being studied and characterized originate from clinical bacteria where plasmids are responsible for the spread of antibiotic resistance or virulence determinants (Smalla *et al*, 2000). As a result only a limited amount of knowledge on the prevalence and the diversity of plasmids in bacteria from nonclinical environments, for example soil bacteria including those from the mining environment, are available. This study has focused on the plasmid pTC-F14, isolated from *Acidithiobacillus caldus* strain f. *At. caldus* is a sulphur-oxidizing, chemolithotrophic, obligately acidophilic (pH1.5-2.5) and moderately thermophilic (45°-50°) bacterium (Hallberg and Lindstrom, 1994). It has been reported that *At. caldus* strain f, together with the strictly iron-oxidizing bacterium of the genus *Leptospirillum*, have been found to be the dominant organisms present in the industrial continuous-flow tank reactors used in arsenopyrite oxidation as a pre-treatment step in the recovery of gold (Rawlings *et al*, 1999b). As mentioned before (Chapter 1), *At. caldus* inhabits a specialized ecological niche, i.e. nutrient poor, low pH inorganic environments and which requires in addition to the mineral, only air, water and trace elements to grow and reproduce. The isolation of plasmid pTC-F14 was previously reported (Gardner *et al*, 2001) and was found to be a 14.2-kb, broad-host range IncQ-like plasmid. We were particularly interested in investigating this plasmid in attempt to identify the accessory genes pTC-F14 might possess. It is of particular interest to discover what types of genes are present within the mobile gene pool of bacteria growing in an inorganic, low pH biomining environments and whether these biomining bacteria are part of an isolated gene pool or whether they share a gene pool with other types of bacteria. Although accessory genes located on plasmids are of limited value in the characterization of plasmids (Smalla *et al*,

2000), studies on the accessory genes carried by plasmids may help contribute to elucidating the ecological role of the plasmid as well as their epidemiology (occurrence, distribution and the factors governing their spread) in response to environmental stresses.

In this study, the 5.8-kb region of pTC-F14 thought to carry possible accessory genes was characterized using the techniques of restriction mapping, cloning in *Escherichia coli* and subsequent DNA sequencing and analysis.

2.2. Material and Methods

Bacterial strains, plasmids, media and growth conditions

The bacterial strains and plasmids used in this study are summarized in Table 2.1. The *E. coli* K-12 derivative, DH5 α was used for maintaining all cloning vectors used in this study as well as for the transformation, amplification and maintenance of subclones constructed in this study. *E. coli* strains were cultured in either Luria-Bertani broth or on LA plates (Sambrook *et al*, 1989) at 37°C and where appropriate ampicillin (Amp) was added at a concentration of 100 μ g/ml when required. A pUCBM21 construct containing the 5.6-kb *Bam*HI-*Sph*I accessory gene area was previously cloned into the *E. coli* vector pUCBM21 (Dr. Shelly Deane) and designated pTC-F14#1. Subcloning of this region for subsequent sequencing experiments was carried out using pUCBM21 as cloning vector.

Restriction mapping, construction of clones and general DNA techniques

Initial restriction mapping experiments were performed using restriction endonucleases according to the specifications of the manufacturer (Boehringer-Mannheim) and as described by Sambrook *et al* (1989). The cloning strategy used in this study is illustrated in Figure 2.1 and a description of the different clones constructed is summarized in Table 2.2. Large and small scale plasmid preparations, gel electrophoresis, making of competent *E. coli* DH5 α cells, transformations and cloning procedures were carried out by standard molecular genetics techniques compiled by Sambrook *et al*, (1989).

DNA sequencing and bioinformatics

Sequencing was done by the dideoxy-chain termination method, with an ABI PRISMTM 377 automated DNA sequencer. The sequence was analyzed using a variety of software

programmes but mainly the personal-computer based DNAMAN (version 4.1) from Lynnon BioSoft. Sequence comparison searches were done using the gapped-BLAST program of the National Center for Biotechnology Information (Altshul *et al*, 1997). Mostly the BLASTX service was used and in cases where gapped-BLAST searches did not yield any significant results, several web-based bioinformatics tools and databases were used to analyze the protein sequence of the relevant open reading frames. The sequence was mainly scanned for conserved domain regions, motifs related to a certain protein family, secondary structure predictions and transmembrane regions. General features such as pI, MW and amino acid composition have been inspected using DNAMAN but several web-based services are also available. A list of the bioinformatics tools and Databases used in these analyses is listed in Table 2.3 and various bioinformatics tools have been reviewed recently in Rehm (2001). Multiple alignments were done using DNAMAN and the web-based CLUSTALW software (Table 2.3).

Nucleotide sequence accession number

The nucleotide sequence of the accessory gene region has been submitted as part of the complete sequence of pTC-F14 to the EMBL-GenBank database under accession no. NC_004734.



Table 2.1: Strains, plasmids used in this study

Strain or plasmid	Genotype or description	Source or reference
Strains		
<i>E. coli</i> DH5 α	F'/ <i>endA1 hsdR17</i> ($\Gamma_k^- m_k^+$) <i>supE44 thi-1 recA1 gyrA</i> (Nal ^r) <i>relA1</i> Δ (<i>lacZYA-argF</i>) <i>U169</i> (ϕ 80 <i>dlac</i> Δ (<i>lacZ</i>) <i>M15</i>)	Promega Corp., Madison, Wisconsin, USA
Plasmids		
pUCBM21	Amp ^R <i>lacZ'</i> ; ColE1 replicon, cloning vector	Roche Molecular Biochemicals
pTC-F14#1	Amp ^R ; 5.6-kb <i>Bam</i> HI/ <i>Sph</i> I fragment containing the accessory gene region cloned into pUCBM21	Construct made by Dr. Shelly Deane, this laboratory

Table 2.2: Subclones constructed and sequenced in this study

Subclones	Description
pΔSmaI	Amp ^R , <i>SmaI</i> deletion of pTC-F14#1
pΔEcoRI	Amp ^R , <i>EcoRI</i> deletion of pTC-F14#1
pSH1.8	Amp ^R , 1.8-kb <i>SphI-HindIII</i> fragment of pΔEcoRI cloned into pUCBM21 digested with <i>SphI</i> and <i>HindIII</i>
pSS0.3	Amp ^R , 0.3-kb <i>SphI-StuI</i> fragment of pΔEcoRI cloned into pUCBM21 digested with <i>SphI</i> and <i>SmaI</i>
pHE1.2	Amp ^R , 1.2-kb <i>HindIII-EcoRV</i> fragment of pΔEcoRI cloned into pUCBM21 digested with <i>HindIII</i> and <i>EcoRV</i>
pSE0.6	Amp ^R , 0.6-kb <i>SphI-EcoRV</i> fragment of pΔEcoRI cloned into pUCBM21 digested with <i>SphI</i> and <i>SmaI</i>
pES1.0	Amp ^R , 1.0-kb <i>EcoRV-StuI</i> fragment of pΔEcoRI cloned into pUCBM21 digested with <i>EcoRV</i>
pHB2.3	Amp ^R , 2.3-kb <i>HindIII-BglII</i> fragment of pTC-F14#1 cloned into pUCBM21 digested with <i>HindIII</i> and <i>BamHI</i>
pEE1.1	Amp ^R , 1.1-kb <i>EcoRI-EcoRI</i> fragment of pHB2.3 cloned into pUCBM21 digested with <i>EcoRI</i>
pEVE0.6	Amp ^R , 0.6-kb <i>EcoRV-EcoRI</i> fragment of pHB2.3 cloned into pUCBm21 digested with <i>EcoRV</i> and <i>EcoRI</i>
pES0.6	Amp ^R , 0.6-kb <i>EcoRI-StuI</i> fragment of pHB2.3 cloned into pUCBM21 digested with <i>EcoRI</i> and <i>EcoRV</i>
pSB2.1	Amp ^R , 2.1-kb <i>StuI-BamHI</i> fragment of pTC-F14#1 cloned into pUCBM21 digested with <i>EcoRV</i> and <i>BamHI</i>
pHB0.6	Amp ^R , 0.6-kb <i>HindIII-BamHI</i> fragment of pSB2.1 cloned into pUCBM21 and digested with <i>HindIII</i> and <i>BamHI</i>
pBH1.1	Amp ^R , 1.1-kb <i>BglII-HindIII</i> fragment of pSB2.1 cloned into pUCBM21 digested with <i>BamHI</i> and <i>HindIII</i>
pSB0.9	Amp ^R , 0.9-kb <i>SphI-BamHI</i> fragment of pSB2.1 cloned into pUCBM21 digested with <i>SphI</i> and <i>BamHI</i>
pBS0.8	Amp ^R , 0.4-kb <i>BglII-SphI</i> fragment of pSB2.1 cloned into pUCBM21 digested with <i>BamHI</i> and <i>SphI</i>

Table 2.3: Web-based bioinformatics tools used in sequence analysis (Rehm, 2001)

Search engine/Database/Software	http site
NCBI BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
SwissProt	http://kr.expasy.org/sprot/
NCBI-RPS BLAST	http://www.ncbi.nlm.nih.gov/BLAST/
PROSITE pattern scan	http://kr.expasy.org/tools.scnpsit2.html
PROSITE profile scan	http://www.isrec.isb-sib.ch/software/PFSCAN_form.html
Pfam HMM scan	http://www.sanger.ac.uk/SOFTWARE/Pfam/
Expasy Tools	http://www.expasy.ch/tools
Jpred	http://jura.ebi.ac.uk:8888/ (also available on Expasy)
Propsearch	http://www.embl-heidelberg.de/prs.html
Protein analysis Tools	http://www.bioscience.org.urllists/protanal.htm
CLUSTALW	http://www2.ebi.ac.uk/clustalw

2.3. Results

2.3.1. Restriction mapping and cloning of the accessory gene region

pTC-F14#1 has previously been partially characterized for the following restriction endonuclease sites: an unique *Bam*HI and *Sma*I site and two sites each for *Hind*III and *Sph*I. Further single restriction digest and gel electrophoresis analysis revealed that this region contains two additional unique sites for *Bgl*II and *Sal*I, two sites for *Eco*RI, four sites for *Eco*RV and 6 sites for *Nco*I. Double restriction endonuclease experiments were then carried out to determine the relative positions for the *Bgl*II, *Sal*I, *Eco*RI, *Eco*RV and *Nco*I sites located on this region. It was problematic to accurately determine the location of the *Eco*RV and *Nco*I sites because of the multiple cleavage sites for these enzymes within this region. The initially subcloning strategy was by simply digesting pTC-F14#1 using a combination of restriction enzymes to yield a desired banding pattern, determined from the map, followed by excising the band of interest under a 360nm wavelength UV lightbox, purifying the band and cloning into the vector. In some cases the same bands were repeatedly sequenced. This confusion resulted from the specific distribution of restriction sites within this region where different single or double enzyme digests yielded identical sized smaller fragments which were ideal for sequencing. To rule out this confusion and ambiguity we decided to clone larger fragments of pTC-F14# into pUCBM21. This allowed us to confirm that our initial restriction map of pTC-F14#1 was accurate by individually mapping each construct and it allowed us to map the *Eco*RV sites. These larger fragments were then subcloned into smaller pieces of sizes ca. 600 bp so that all DNA could be sequenced in both directions with high fidelity. The restriction endonuclease map of pTC-F14#1 is shown in Fig 2.1 as are the constructs made and the cloning strategy used. More information on the clones is given in Table 2.2.

These constructs made the cloning process easier as these constructs were used to subclone the accessory gene area. The constructs made and the cloning strategy is depicted in Fig 2.1 More information of the clones is given in Table 2.2. The DNA was digested to give fragment sizes of 600-bp to 1.2-kbp and cloned into pUCBM21. It was desired to have cloned fragments with sizes of not more than 600-bp, if it was essential for the sequence in both directions were to be confirmed, since only the first 600-bp of a sequencing reaction shows high fidelity. Prior to the sequencing reactions restriction digests as positive controls was performed on the possible cloned fragments to double check clones for both the correct size and the correct internal sites, where possible (data not shown).

2.3.2. Sequence analysis of the accessory gene area

In our attempt to discover what accessory genes are carried by pTC-F14, the insert in pTC-F14#1, from the *SphI* site to the *BamHI* site (Fig 2.1) was completely sequenced on both strands and found to be 5.772-kb (GenBank Accession no. NC_004734). For a fully annotated sequence refer to Appendix One. After careful inspection of the DNA sequence mainly using DNAMAN, six complete ORFs were identified. One of these ORFs starts in a region adjacent to the *SphI* site of the 5.8-kb *BamHI/SphI* accessory gene fragment on pTC-F14. To assist in presentation, this 340-bp region of previously sequenced DNA has been joined to the original fragment carried in pTC-F14#1. The positions of the restriction endonuclease cleavage sites, the ORFs and their orientations relative to pTC-F14 plasmid are depicted in Fig 2.2. The six ORFs detected gave putative translation products of 9kDa and larger and were preceded by putative ribosome binding sites. The general characteristics of these ORFs are summarized in Table 2.4.

Sequence database matching is a powerful and easy method of finding the function of an unknown gene whose sequence has become available. If significant sequence similarity to one or more database sequences whose function is already known is obtained, the unknown protein might have the same function, biochemical activity or structure similar to the protein(s) in the database. In other words, much can be deduced about an unknown protein when significant sequence similarity is detected with a well studied protein. DNA sequences are variable in the third base position and in order to enable detection of more distant homologs, similarity searches were performed preferably at the protein sequence level. The predicted amino acid (aa) sequences of the ORFs were initially compared with sequences in the non-redundant database GenBank database (hosted at the National Center for Biotechnology, Table 2.3), using the BLASTX program. A BLASTX search performs a comparison at the protein level by comparing your transcribed DNA sequence with the amino acid sequences of proteins deposited in the database. In the worse case scenario these BLAST searches will yield no significant matches, or the search will result in a list of partial matches to an assortment of proteins with most of them being uncharacterized. This was the case with some of the ORFs detected in this study. In this case one has to rely on bioinformatics approaches to analyze the unknown protein of interest which may suggest a direction for subsequent laboratory investigations. Alternatively one might do BLAST searches regularly as the databases are updated with new sequences at a breathtaking speed, in the hope that its ortholog might be discovered.

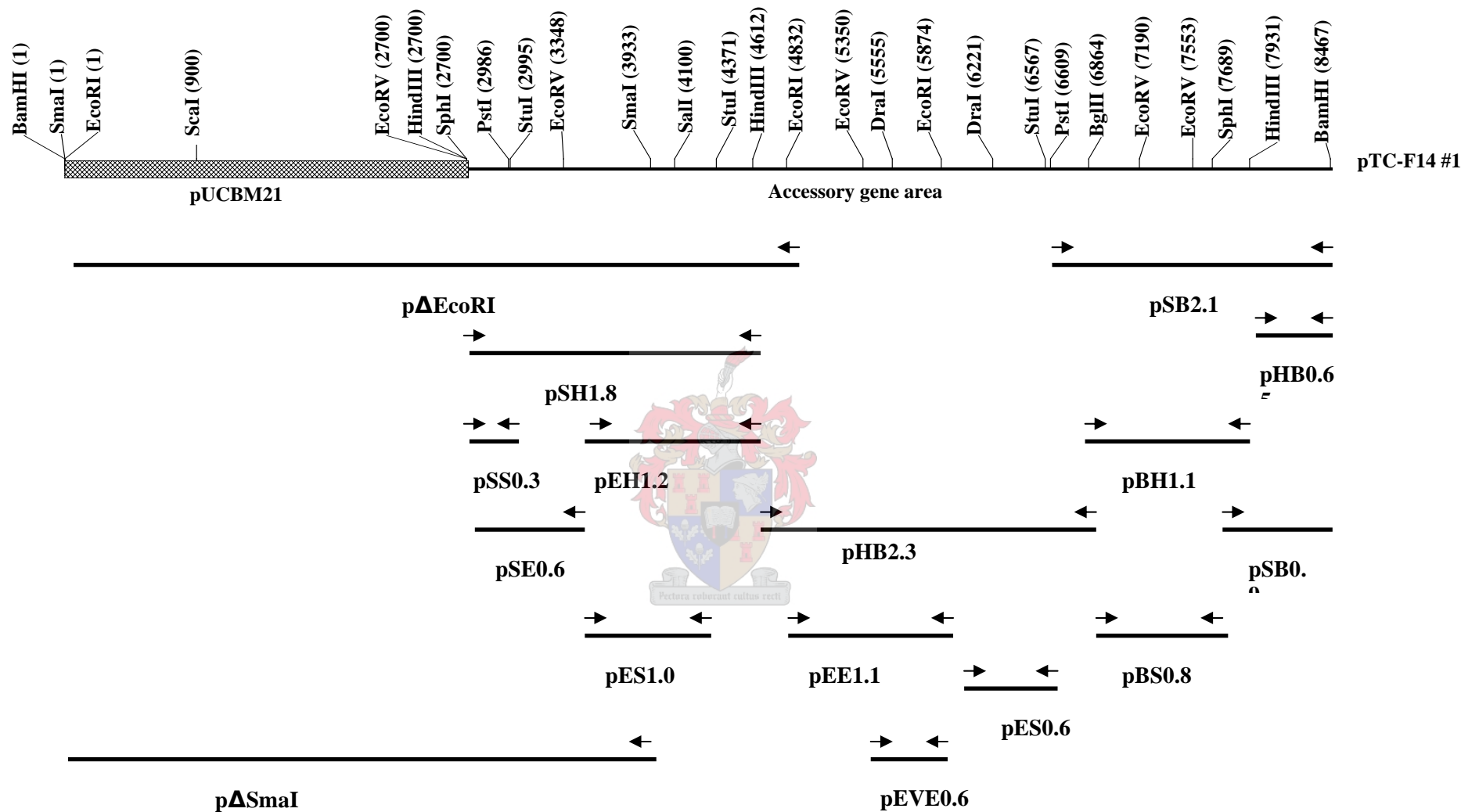


Figure 2.1: Cloning strategy of pTC-F14#1. Subclones are represented as thick black bars beneath the map. More information concerning the subclones and cloning vector is given Table 2.2. Thin arrows represent the direction of sequencing reactions. Map and subclones are drawn to scale.

One bioinformatics approach is to discover what conserved regions, motifs or domains the unknown protein might contain. This might help in placing the unknown protein in a family of specific proteins and it might help discover what domains it contains. A list of web-based bioinformatics sites used in this study is given in Table 2.3. Many more software and bioinformatics tools are reviewed in Rehm (2001).

Table 2.4: Accessory proteins on pTC-F14

Putative protein or ORF	Predicted number of aa (Da)	Putative ribosomal binding sites	Most related protein and proposed function and predicted size	% identity/similarity (part of protein) ^a	BLAST E value	Reference NCBI accession no.
ORF13	124 (13008)	AGGAGA	no meaningful BLAST hits	NA	NA	NA
ORF20.8	189 (20795)	AGGCGA	invertase/recombinase protein, <i>Aeromonas punctata</i> , 188aa	64/73 (181)	2e ⁻⁵⁸	CAG15093.1
ORF17.4	153 (17405)	AGGAG	conserved hypothetical protein, <i>Chromobacterium violaceum</i> ATCC 12472, 154aa <i>Pseudomonas fluorescens</i> PfO-1, 146aa	55/73 (149) 53/70 (141)	8e ⁻⁴³ 2e ⁻³⁷	AAQ59194.1 ZP_00267167.1
ORF33	286 (33169)	AGGAG	aminotransferase, <i>Bacillus halodurans</i> , 397aa	27/48 (133)	6.8e ⁻²	NP_244179
ISAtc1	404 (46188)	AGGAG	transposase from IS <i>Afe</i> 1, <i>Acidithiobacillus ferrooxidans</i> , 404aa	92/95 (404)	0.0	AAB07489
ORF9.5	86 (9485)	AGGAG	no meaningful BLAST hits	NA	NA	NA

^a number in brackets indicates the number of amino acids over which the percentage identity/similarity was obtained.

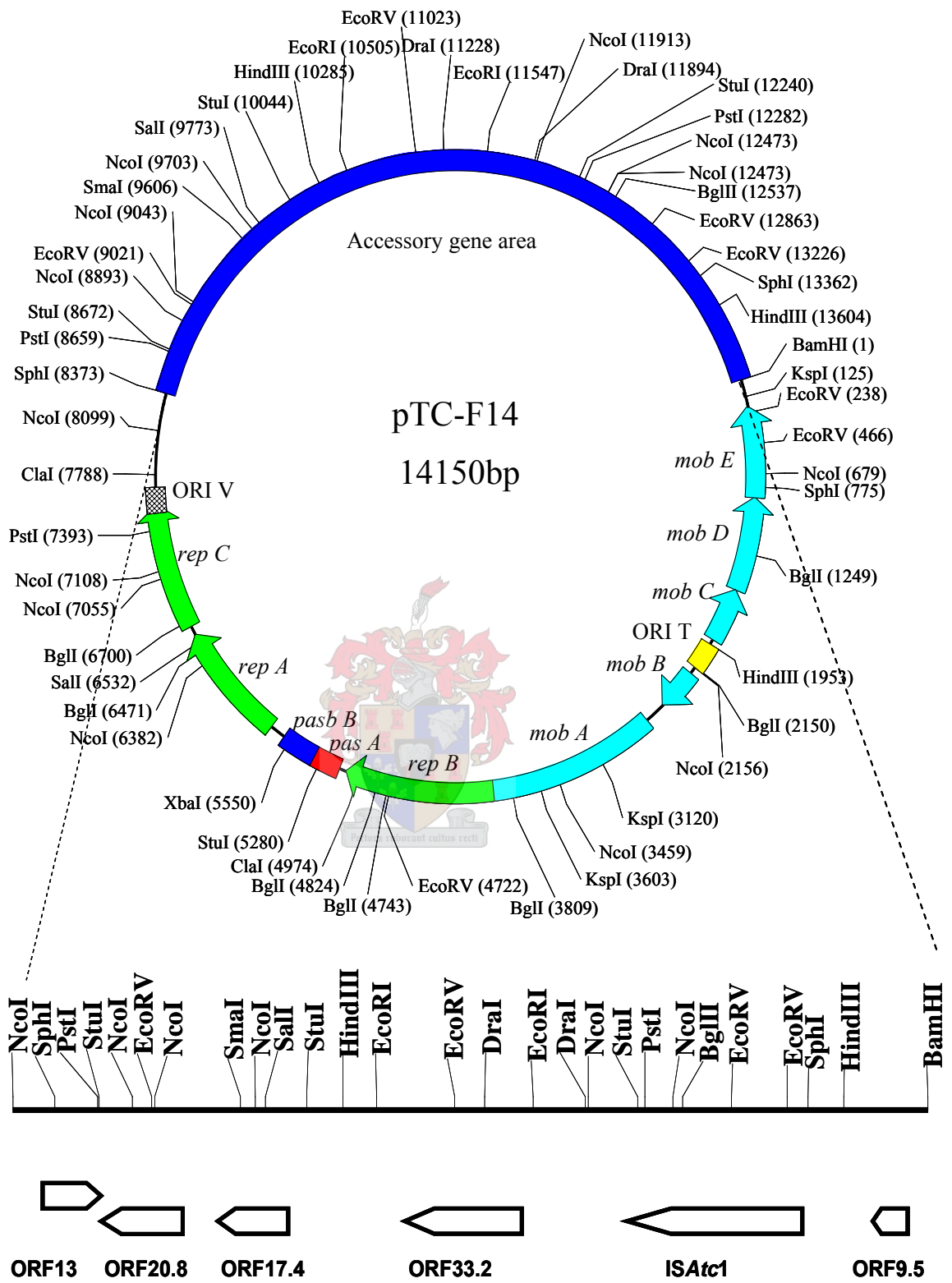


Figure 2.2: Restriction enzyme cleavage and genetic map of pTC-F14 showing the relative position and genetic map of the region carrying the accessory genes. Included are the ORFs identified and their relative orientations. The subclones used for the sequencing reactions are explained in Fig 2.1.

The first ORF (closest to the *NcoI* site, Fig 2.2) consisted of 124 amino acids with a molecular weight of 13008 Da and a theoretical pI (iso-electric point, pH where the net charge of the protein is zero) value of 11.42. This ORF was named ORF13. This ORF gave no meaningful similarity hits using the BLAST program. The secondary structure prediction was done using the SOPMA Secondary Prediction Software hosted on the Expasy website (Fig 2.3). No transmembrane regions could be detected during analysis with the PredictProtein web-based software utility at Expasy. The ProtParam tool hosted at the Expasy website predicts that this protein is unstable. In order to determine whether this protein contains conserved domains or functional regions or motifs similar to characterized proteins or to determine structural and/or functional homologs, the PROSITE pattern scan, RPS-BLAST, PropSearch and the Pfam utilities was employed for these purposes. The PROSITE pattern scan utility only tries to identify the functionally most important residue combinations and does not attempt to describe a complete domain or even protein (Reviewed and referenced in Rehm, 2001; Bairoch, 1991; Bairoch *et al*, 1997). The PROSITE utility yielded 5 sites of functional importance with 100% similarity to a consensus sequence. These included a site for N-glycosylation, 4 sites for protein kinase phosphorylation, a Casein kinase II phosphorylation site and 4 sites for N-myristoylation (mostly found in eukaryotes where a myristate, a C14 fatty acid is linked to the N-terminal glycine residue) and an ATP/GTP-binding site motif. Their relative positions and sequence are also indicated (data not shown). The PropSearch software attempts to find the putative protein family of an unknown sequence when the initial alignment methods such as the BLAST programs, did not yield significant similarity to already characterized proteins. PropSearch characterization is calculated from features such as amino acid composition, molecular weight, and content of bulky residues, content of small residues, average hydrophobicity, average charge and the content of selected di-peptide groups. Using this program in an attempt to find the putative protein family of the protein of ORF13, yielded what appears to be a false positive result. It placed the protein in a family together with Cu/Zn superoxide dismutase proteins, with a 80% chance of being in the same protein family as the superoxide dismutases from the eukaryote, *Mesembryanthemum crystallinum* (common ice plant) and a 68% chance of being in the same family of proteins as the tomato plant (*Lycopersicon esculentum*) superoxide dismutase, An *E. coli* super-oxide dismutase was also in the list of hits but only showed a 53% chance of being in the same family as this protein. Cu/Zn superoxide dismutase proteins is one of the three possible forms of this enzyme which catalyses the dismutation of superoxide radicals (Smith and Doolittle, 1992). By doing a CLUSTALW alignment of the protein sequences above 50% reliability, no significant conserved residues between the ORF13 amino acid sequence and the super-oxide

A.

```

      10      20      30      40      50      60      70
      |       |       |       |       |       |
MEAGPITQQDVAAAADQLGGPANTNAAKIREILGRGSLATIQRHLQALRDAQRAPELPEAVQTIPTPPEA
eccccchhhhhhhhhhtccccchhhhhhhhtttchhhhhhhhhhhhhhtccccchhhccccch
VSEASRGIWAAAWAMAEQRHAESLAHAPCWVITGITFAYSGNGCQSLSPNTLL
hhhhhhhhhhhhhhhhhhhhhhhhcccccccccccccccccttcteee

```

SOPMA :

Alpha helix (Hh) : 61 is 49.19%
 3₁₀ helix (Gg) : 0 is 0.00%
 Pi helix (Ii) : 0 is 0.00%
 Beta bridge (Bb) : 0 is 0.00%
 Extended strand (Ee) : 17 is 13.71%
 Beta turn (Tt) : 8 is 6.45%
 Bend region (Ss) : 0 is 0.00%
 Random coil (Cc) : 38 is 30.65%
 Ambiguous states (?) : 0 is 0.00%
 Other states : 0 is 0.00%

B.

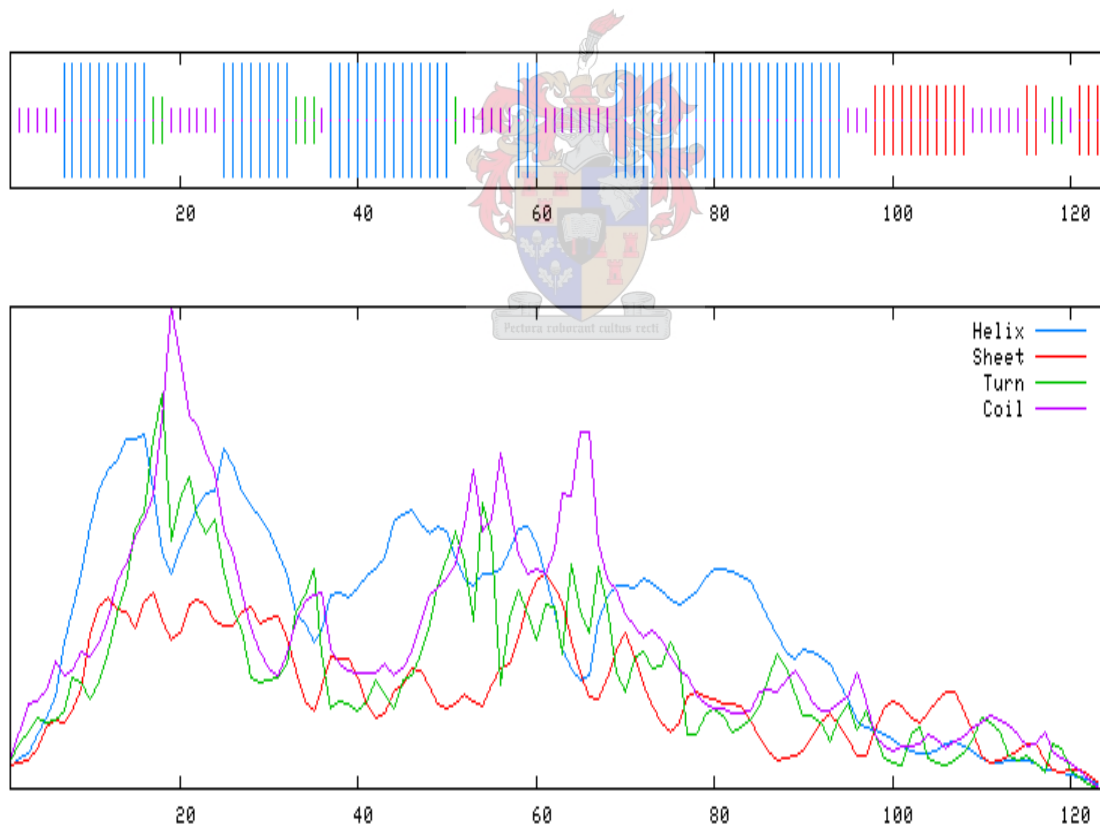


Figure 2.3: Illustration of the secondary protein structure prediction of ORF13 using the web-based SOPMA Secondary Prediction Software at the Expasy website (Geourjon and Deléage, 1995). (A) Statistical representation, (B) Diagrammatic representation.

dismutase amino acid sequences could be detected. Therefore this result is likely to be a false positive classification of the ORF13 protein into this family of proteins. All the other utility sites mentioned above, failed to give any meaningful results.

Adjacent to ORF13 was another complete ORF consisting of 189 amino acids, with a molecular weight of 20795 Da. This ORF named ORF20.8 showed clear amino acid sequence homology to various invertase/recombinase like proteins. The predicted amino acid sequence of ORF20.8 was most closely related to a DNA recombinase from *Aeromonas punctata* with 64%/73% amino acid identity/similarity (Table 2.4). An alignment of the amino acid sequence of ORF20.8 to members of this class of enzymes was performed using the DNAMAN program and is shown in Fig 2.4. The members include proteins from different bacterial strains, which showed the highest amino acid sequence similarity during the BLASTX database search. The function of these enzymes is the rearrangement of DNA within a sequence or the exchange of DNA between sequences.

Next to ORF20.8 was a region that contained another complete ORF, ORF17.4. This ORF was found to be consisting of 153 amino acids with a molecular weight of 17405 Da. The results of a BLASTX searched revealed that the amino acid sequence of ORF17.4 is related to a hypothetical protein that is highly conserved in a wide variety of bacteria. Only the two highest matches are given here (Table 2.4). It was most closely related to *Chromobacterium violaceum* ATCC 12427 showing 55%/73% amino acid identity/similarity. In an attempt to discover if this protein of ORF17.4 contains any conserved domains or motifs similar to any other characterized proteins several motif scans were performed using the various tools at the ExPasy website. All these searches failed in giving any significant results. However, performing PropSearch in the hope of finding a possible protein family for the predicted protein of ORF17.4, it was interesting to discover that this search utility yielded two distinct possible families in which our query could be placed. The first of these families, with a 94% chance of the ORF17.4 protein falling into this family of proteins, was the cyanate lyase (also known as cyanate hydratase) family of proteins. These enzymes typically catalyse the reaction of cyanate with bicarbonate to produce ammonia and carbon dioxide. The second family included the dissimilatory-type alpha subunit of a sulfite reductase. PropSearch gave an 87% chance of ORF17.4 being part of this family. These classes of enzymes are involved in catalyzing the reduction of sulfite to sulfide. This is the terminal reduction reaction in sulfate respiration, a process catalyzed by the sulfate-reducing bacteria. As *At. caldus* is a sulphur-oxidising bacterium it is not clear why a sulfite reductase gene should be present on pTC-F14

ORF20.8:M	1
A:M	1
B:	MSVKRSLVGHAGHCPAKLLHNGLSDIKIGGRNDGQERRQM	40
C:M	1
D:M	1
ORF20.8:	AQIGYARISTGDQNLDLQRDALNHAGCEHTFEDVASGAKA	41
A:	ALIGYARVSTAEQDTALQTDALRKAGCERVFEDTASGAKA	41
B:	ALIGYARVSTAEQDTALQTDALRKAGCERVFEDTASGAKA	80
C:	VLIGYARVSTAEQDTALQTDALRKAGCERVFEDTASGAKA	41
D:	ALIGYARVSTAEQDTALQTDALRKAGCERVFEDTASGAKA	41
ORF20.8:	ERTGLVQAMAFLRPGDVLTVWRLDRLARSLPHLIEVLNDL	81
A:	DRPGLADALAYLRDGDVLAVWRLDRLGRSMPHLIETIGAL	81
B:	DRPGLADALAYLRDGDVLAVWRLDRLGRSMPHLIETIGAL	120
C:	DRPGLADALAYLRAGDVLAVWRLDRLGRSMQHLIETIAAL	81
D:	DRPGLADALAYLRNGDVLAVWRLDRLGRSMPHLIETIGAL	81
ORF20.8:	EQRCCGFRSLTESIDTTTPAGRLIFHIFGALGQFERDLIR	121
A:	EARGVGFRSLTEAIDTTTPGGRLIFHIFGALGQFERDLIR	121
B:	EARGVGFRSLTEAIDTTTPGGRLIFHIFGALGQFERDLIR	160
C:	EARGVGFRSLTESIDTTTPGGRLIFHIFGALGQFERDLIR	121
D:	EARGVGFRSLTEAIDTTTPGGRLIFHIFGALGQFERDLIR	121
ORF20.8:	ERTLAGLAAAVARGRKGGRKPVMTPEKQQRQQQLIAQGLS	161
A:	ERTKAGLTAAAARGRKGGRKPVVTADKLQRAREHIANGLN	161
B:	ERTKAGLTAAAARGRKGGRKPVVTADKLQRAREHIANGLN	200
C:	ERTKAGLTAAAARGRKGGRKPVVTADKLQRAREHIANGLN	161
D:	ERTKAGLSAAAARGRKGGRKPVITADKLQQARKHIANGIN	161
ORF20.8:	VREAATRLKISKTVLYRELQRKRGGNGS	189
A:	VREAATRLKVSKTALYTALQSTSAADS.	188
B:	VREAATRLKVSKTALYTALQSTSAADS.	227
C:	VREAARKLKVSKTALYAALQSTSAANF.	188
D:	VREAATRLKVSKTALYAALQSTSAAGDLQ	189

Figure 2.4: Multiple sequence alignment of the amino acid sequence of invertase/recombinase-like ORF20.8 and homologs (invertase/recombinase like proteins) obtained during a BLASTX database similarity search. **A.** *Xanthomonas axonopodis* pv. *citri* str. 306 (NP_644789.1); **B.** *Aeromonas punctata* (CAG15093.1); **C.** *Pseudomonas aeruginosa* (AAP22617.1); **D.** *Xanthomonas campestris* pv. *campestris* str. ATCC 33913 (NP_637000.1). Accession numbers are given in brackets.

or on its chromosome. It may be that ORF17.4 protein might have a similar protein domain, motif or functional region to the proteins in the sulphate reductase family. To inspect the integrity of the PropSearch results of both searches the amino acid sequences of various family members of both families were obtained from the SwissProt database and used in two separate multiple alignment operations between the amino acid sequence of ORF17.4, one for each possible family. An alignment between members of the reductase family of enzymes and ORF17.4 did not reveal any significant relationships (homology) to support the results of the PropSearch (data not shown). However, the alignment using members of cyanate hydratase family yielded a result where more identity between these family members and the ORF17.4 protein was visible. The alignment is shown in Fig 2.5. A secondary structure prediction was done using SOPMA and is shown in Fig 2.6.

Adjacent to ORF17.4 we detected a complete ORF of 286 amino acids. This ORF was named ORF33.2 and has a predicted molecular weight of 33169 Da. The predicted protein sequence of ORF33.2 gave relatively weak similarity and identity to approximately one third of the amino acid sequence of an aminotransferase. This level of similarity was considered to be below the threshold (Rost *et al*, 1997). It was therefore considered to be insufficient to assign this a likely function of the putative protein. The ORF33 protein has been placed in putative family of aminoglycoside 3'-phosphotransferase. The PropSearch algorithm predicts a 94% chance of the ORF33 protein belonging to this family of aminoglycoside transferases. This family of proteins provides resistance to the kanamycin and structurally related aminoglycosides, by inactivating kanamycin through phosphorylation in the presence of ATP. An amino acid sequence alignment between ORF33.2 and several members of this protein family were made and although the identity was weak, a few conserved amino acids could be detected. The level of significance hereof is unclear and the data is not presented here. A comparison of the predicted secondary structures of ORF33 protein and the (259 amino acid) neomycin-kanamycin phosphotransferase, type VI of *Acinetobacter baumannii* are depicted in Fig 2.7. Although there are minor regional similarities between the two proteins, it is clear that these two proteins are not related. This is confirmed through the weak amino acid similarity obtained during the multiple alignments between ORF33.2 and members of the aminoglycoside transferases.

ORF 17.4:	RDLDPRYSGAPTRFAIALRNWANI	AVVAL	ITLDVLEKEEFY	40
A:	MAGTEISA	ITTKLLEAKKAK	20
B:	MSIPEIT	IQTLLQAKKDK	17
C:	MTSAIT	TEQLLKAKKAK	16
D:	MTVMSL	TATLMAAKKAK	17
E:	MTLP	SLTKLLLLKAKKEK	17
F:	MLGRM	TRNQLTEQIVVARLAK	21
ORF 17.4:	RRVNEYL	ADSVGED	IRTIADTWALAASERESFFPAERWAA	80
A:	GITFADL	EQLLGRDE	VWIAAVIYRQASASVDEAEKLLHCL	60
B:	GLSFADL	EATLGRDE	VCI AALFYRQASASEEEEAKLLVEAL	57
C:	GITFTE	LEQLLGRDE	VWIASV FYRQSTASPEEAELLLTAL	56
D:	GMSFADL	EAAIGLDE	VWVASL FYGQATASKEEAELLAGLL	57
E:	NLSFEDL	GNLINRDE	VWVASL FYGQATASEEEEATSLIAAL	57
F:	GLTWQEL	LADAI	GRPLLWTT SALLGQHP IPAELGRILVDKL	61
ORF 17.4:	LAYLRQEM	EGRARAEKSH	MDICFACGNLISARLKVLELLA	120
A:	GLSDDL	VP	ELTAPSVKGLG.PVVPTDPLIYRFYEIMQVYG	99
B:	GLDSSYI	KHLTEY	VPKGLG.PVVPTDPLIYRFYEIMQVYG	96
C:	GLDLAL	ADEL	TTPPVKGCLEPVIPTDPLIYRFYEIMQVYG	96
D:	SLDAET	TAALQ	EYPTKGSLEPVIPTDPLIYRFYEIMQVYG	97
E:	DLTSDL	KEDL	STPPVKGCLEPVIPTDPLIYRFYEIMQVYG	97
F:	GLDES	AVP	VLAAPP MRGGLPTAVPTDPTIYRFYEALQVYG	101
ORF 17.4:	PELNSGH	LYTELD	RLSRIGNPSVLGYDVRKEMM.....	153
A:	MPMKE	VIHEK	FGDGIMS AIDFTLDIEKEADPKGDRVKVTM	139
B:	FPIKE	VIQEK	FGDGIMS AIDFTLDVEKEADPKGDRVKITM	136
C:	LPLK	DVIQEK	FGDGIMS AIDFTLDVDKVEDPKGDRVKVTM	136
D:	MPLK	DVIQEH	FGDGIMS AIDFTIDVDKVEDPKGDRVKIVM	137
E:	LPMK	DVIQEK	FGDGIMS AIDFSIEVDKVEDPKGDRVLVKM	137
F:	GALKE	VIAEQ	FGDGIMS AINF SVDLQKKPHPSGDRVVVTF	141
ORF 17.4:			153
A:	NGKFL	PYKKW	...	149
B:	SGKFL	PYKKW	...	146
C:	CGKFL	AYKKW	...	146
D:	CGKFL	PYKKW	...	147
E:	CGKFL	PYKKW	...	147
F:	DGKFL	PYQ	WVSSE	154

Figure 2.5: Multiple alignment between the ORF17.4 protein and members of the cyanate hydratase family of proteins. **A.** *Synechocystis* sp.strain PCC6803 (Q55367); **B.** *Synechococcus* sp strain PCC7942 (Q59948); **C.** *Anabaena* sp strain PCC7120 (P58703); **D.** *Synechococcus* sp strain WH8102 (Q7U3E2); **E.** *Prochlorococcus marinus* subsp pastoris strain CCMP1378 (Q7V2UO); **F.** *Mycobacterium paratuberculosis* (P61192). Swiss-Prot accession numbers are given in brackets.

A.

```

      10      20      30      40      50      60      70
RLLDPRYSGAPTRFAIALRNWANIAVVALTLDVLEKEEFYRRVNEYLADSVGEDIRTIADTWALAASERE
cccccccccccccheeeehtttthheeehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
SFFPAERWAALAYLRQEMEGRARAEKSHMDICFACGNLISARLKVLELLAPELNSGHLYTELDRLSRIGN
cccchhhhhhhhhhhhhhhhhhhhhhhhhhhhheeehtthhhhhhhhhhhhhcttccttceehhhhhhtcc
PSVLGYDVRKEMM
ceeeeeehhhhhh

```

Alpha helix (Hh) : 100 is 65.36%
 3₁₀ helix (Gg) : 0 is 0.00%
 Pi helix (Ii) : 0 is 0.00%
 Beta bridge (Bb) : 0 is 0.00%
 Extended strand (Ee) : 17 is 11.11%
 Beta turn (Tt) : 11 is 7.19%
 Bend region (Ss) : 0 is 0.00%
 Random coil (Cc) : 25 is 16.34%
 Ambiguous states (?) : 0 is 0.00%
 Other states : 0 is 0.00%

B.

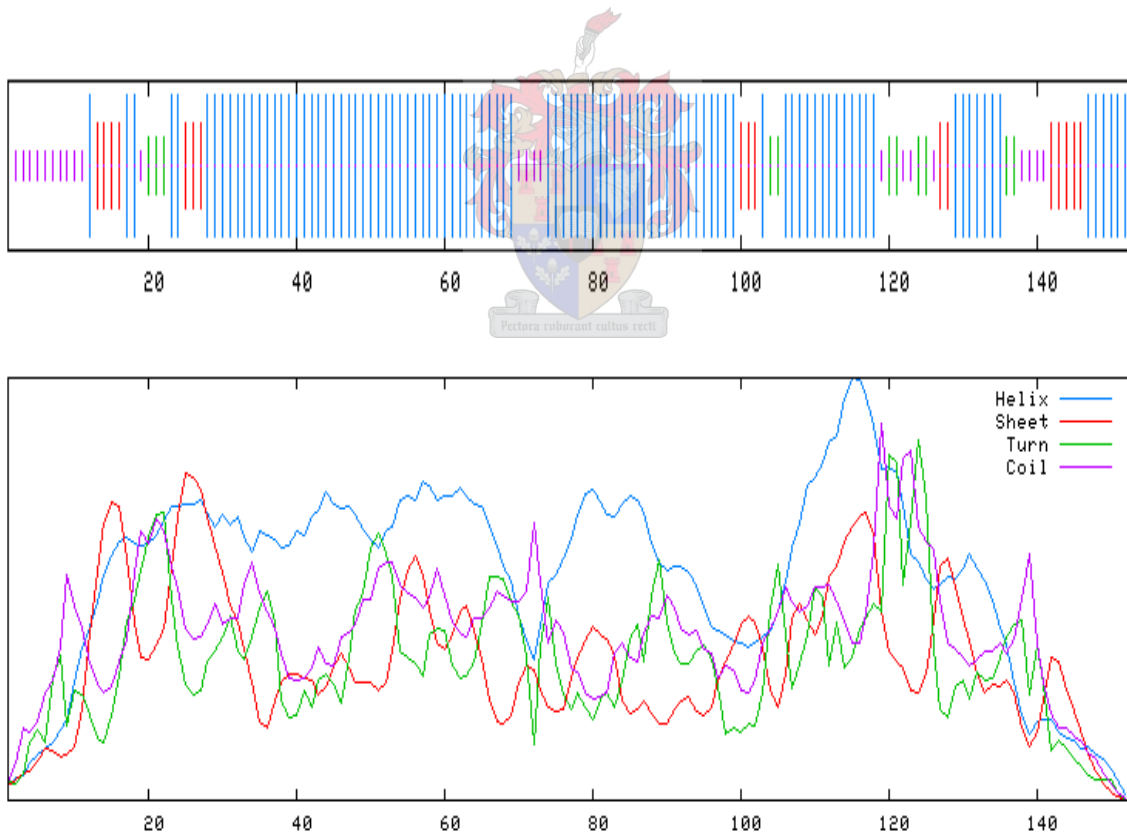
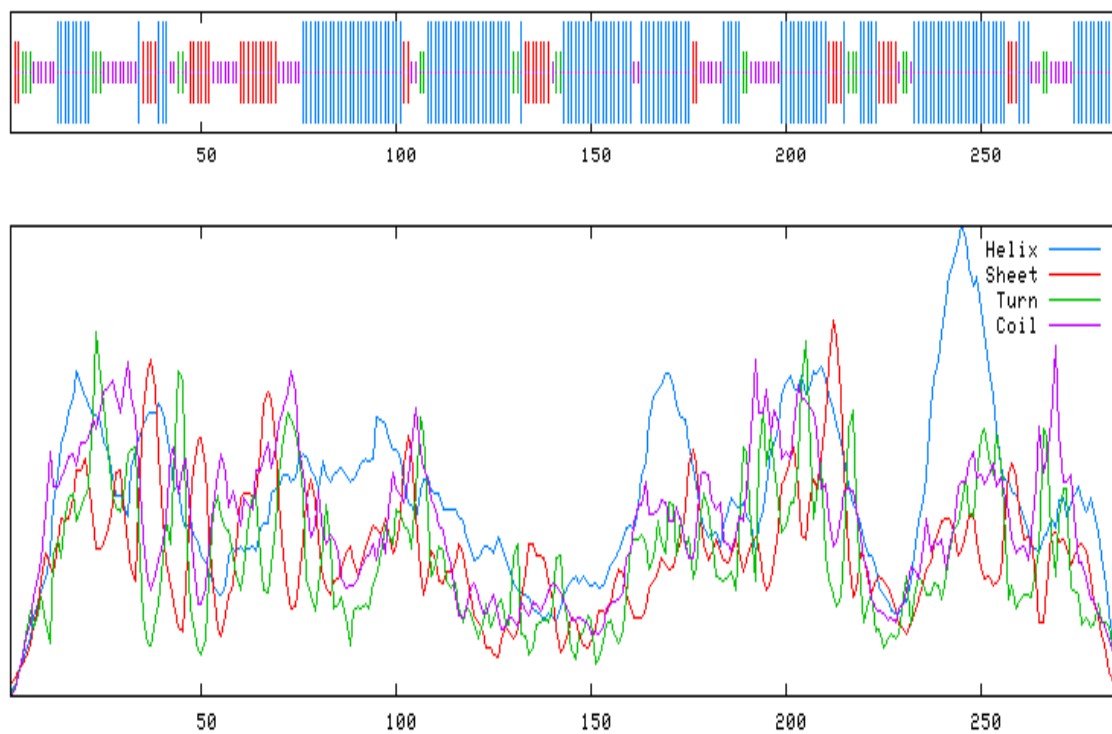


Figure 2.6: Secondary structure prediction of the predicted protein of ORF17.4. **A.** Statistical representation; **B.** Diagrammatic illustration

A.



B.

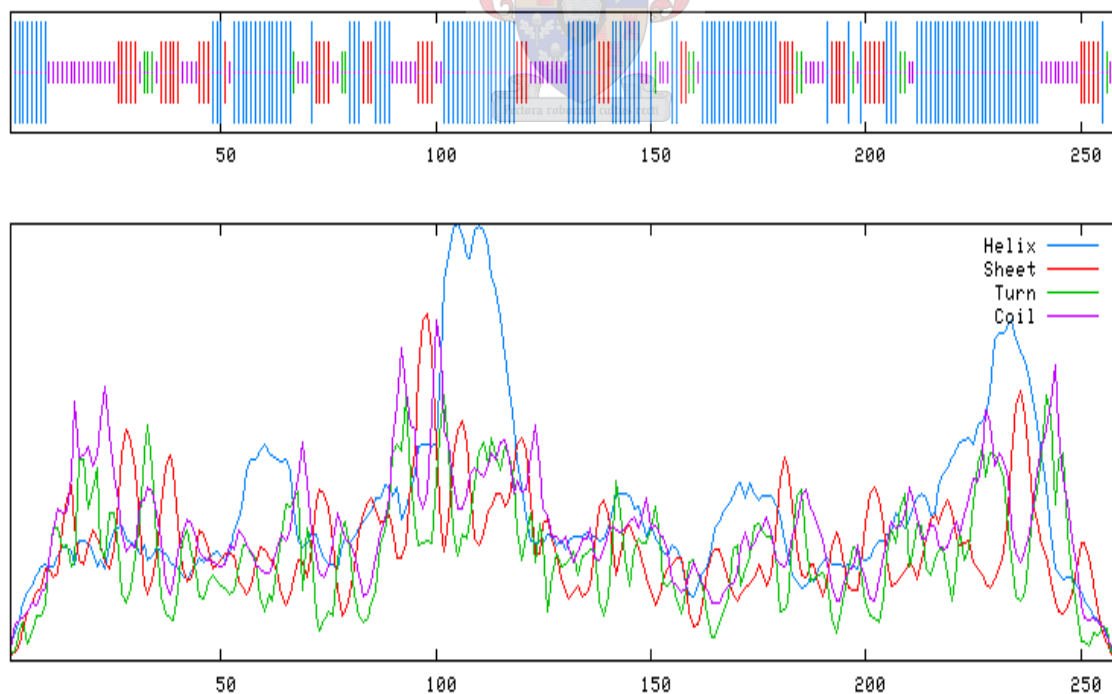


Figure 2.7: Comparison of the predicted secondary protein structure (SOPMA) of ORF30.2, shown in A, and the predicted neomycin-kanamycin phosphotransferase type VI of *Acinetobacter baumannii*, shown in B.

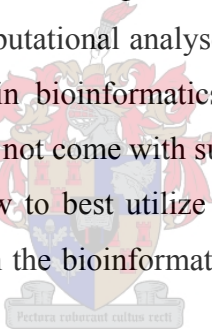
Downstream of ORF33.2 was an ORF of 404 amino acids (46188 Da). This ORF had highest sequence identity and similarity to the transposase of an insertion sequence, IS*Afe1*, previously identified in *At. ferroxidans* (Holmes *et al*, 2001). The ORf has subsequently been named IS*Atc1*. The high sequence similarity and identity observed (Table 2.3) has triggered further research into the biology of this mobile element. This is the focus of research of the next chapter (Chapter 3) and data of the analysis of this IS*Atc1* is presented there.

Adjacent to IS*Atc1* we detected a small ORF of 86 amino acids with a molecular weight of 9485 Da. This ORF, ORF9.5 gave no meaningful similarity using the BLASTX program and a PropSearch. No significant motifs, conserved domains or regions could be detected.

2.4. Discussion

In nature, bacteria are often confronted with environmental conditions they can only survive with the help of plasmids. Numerous plasmids encode accessory genes that allow cell adaptations to variations in environmental conditions and subsequent survival and propagation. These phenotypic traits conferred by the plasmids may be selected for by these environmental conditions and in the absence of these selective pressures the plasmids may be subsequently lost, because of the metabolic burden plasmids place on its host. These traits are thought of being accessory and allow cell survival only under certain hostile environments but are otherwise not essential for cell growth. These phenotypic traits can be transferred horizontally to a related or different species and thus cohabit cells with different chromosomal and plasmid genes (Eberhard, 1989). Movement of genes between chromosomes and plasmids has been documented to be mediated by transposons, insertion elements, bacteriophages and also integrating plasmids. Thus, bacteria have an access to collection of disposable genes, which is called the horizontal gene pool. Housekeeping genes have been argued to be more advantageous to carry on the chromosomal DNA. All the same it would not be surprising to discover genes on plasmids which would otherwise not been expected to be found on plasmids. Plasmids studied and characterized in the greatest detail have largely come from organic environments. Molecular genetic studies on microbes from low pH, inorganic environments are a relatively young field of study. These microbes occupy a highly specialized ecological niche (i.e. a highly acid, mineral rich, inorganic environment) so it is of particular interest to discover the mobile genes participating in the horizontal gene pool of this environment and whether it is an isolated gene pool. We specifically attempted to discover what accessory genes are carried on pTC-F14.

In this study no accessory genes conferring a known phenotype were detected. In this respect the analysis of the accessory gene area has been disappointing. Despite considerable computational efforts to characterize or find structural or functional regions, the possible functions for the protein products of ORF9.5, ORF13, ORF33.2 and ORF17.4 remains elusive. These proteins could possibly be advantageous to the cell but the function of the putative proteins remains to be discovered. One of these ORFs, ORF17.4 has shown significant similarity to a wide range of hypothetical proteins found in a variety of bacteria. This may indicate the likelihood of the discovery of the function of the ORF17.4 protein. Where no significant BLAST hits were obtained as was the case with these ORFS a bioinformatics approach was applied. These analyses were not very successful in gaining more significant information on the proteins in question. Bioinformatics are documented to be a powerful tool for analyzing protein and DNA sequence, for the functional assignment of genes and protein classification. The main problem with bioinformatics is the uncertainty of which software to use and how to obtain scientifically valid answers. However, considering the great amount of time and effort that is committed into the cloning and sequencing of foreign genes, it might be worth spending some time on the computational analyses of macromolecular sequences. There is much confusion surrounding certain bioinformatics tools as some of the free available software on the World Wide Web does not come with sufficient documentation or guidelines as to how to interpret your result, or how to best utilize the software to obtain a valid answer. Nothing concrete can be deduced from the bioinformatics analyses done on ORF17.4, ORF13 and ORF33.2.



Based on the database similarity search results it is clear that ORF20.8 codes for an invertase/recombinase like protein. These enzymes catalyses specialized recombination reactions, resulting in the alternating inversion of a single segment of DNA, which can generate one of two possible phenotypic traits (Johnson, 2002). These DNA inversions often function to help adapt a population or portion of a population to a sudden change in environmental conditions. These reactions have been termed programmed DNA rearrangements because they benefit the host under certain conditions. These inversion systems control gene expression by switching either coding sequences, the orientation of a promoter, or both. Similar enzymes are found encoded by DNA from a wide variety of organisms and this property is not specifically associated with the environment of *At. caldus*. An analysis of *ISAtc1* is the subject of Chapter 3.

3. Chapter 3: Partial characterization of ISAtc1 from plasmid pTC-F14, a new member of the ISL3 family of insertion sequences*

Table of Contents

3.1. Introduction	92
3.2. Materials and Methods	93
3.3. Results	97
3.3.1. Nucleotide sequence analysis of ISAtc1	97
3.3.2. Homology of ISAtc1 Tpnase to Tpnases of other IS elements	99
3.3.3. ISAtc1 Tpnase structure analysis	102
3.3.4. Genomic distribution of ISAtc1 in six <i>At. caldus</i> strains	107
3.3.5. Mobility of ISAtc1 in <i>At. caldus</i>	109
3.3.6. Transposition of ISAtc1 in <i>E. coli</i>	110
3.4. Discussion.....	117

*These results have been published (Goldschmidt *et al*, 2003). Since the publication the computational and genetic analysis of ISAtc1 have been updated

3.1. Introduction

The presence of insertion sequence (IS) elements as natural components of bacterial genomes has been firmly established in biology. IS elements are the simplest transposable elements which generally encode only protein(s) involved in their transposition, e.g. the transposase (Tpase). They form a diverse group of small, compact mobile units of DNA with the ability to translocate from one non-homologous site to another within a DNA molecule (intra-molecular transposition) or between different DNA molecules (inter-molecular transposition) with various degrees of specificity. These elements have long been recognized for their ability to mediate a variety of genetic effects on the host chromosome. These include chromosomal rearrangements such as deletions, duplications, inversions and replicon fusions (Chapter 1). In addition to these rearrangement abilities they have also been recognized for the ability to influence gene expression, either by reducing the expression or inactivating the genes downstream of their insertion points (polar mutations or insertional inactivations) or by providing promoters that increase the transcription of the downstream genes (activation of silent genes) (Galas and Chandler, 1989). Some IS elements can form composite transposons, where two copies of an IS element flank a DNA region and by jumping in unison the interstitial DNA is rendered mobile. Various genes have been identified that are associated with composite transposons, e.g. resistance, catabolic or pathogenic genes. By hitchhiking on other mobile elements such as a prophage or transmissible plasmids they can move horizontally between species and play an important role in the dissemination and acquisition of accessory genes. This allows IS elements to become distributed in a wide range of bacteria and it is these properties that indicate that IS elements could play a fundamental role in the adaptive evolution of bacteria by facilitating in horizontal gene transfer and in internal genetic rearrangements in the genome. It would be of special interest to discover whether the behavior of IS elements in the bacterial genome has contributed to the adaptation of bacteria to extreme environments such as the low pH, inorganic biomining environments. Besides the evolutionary interest in studying IS elements, they are also important to the molecular biologist and genome annotators. Several Biotechnology companies feature mobile-DNA based products in their catalogs (Boeke, 2002). The composite transposons Tn5 and Tn10 are currently being commercially exploited as sequencing tools. IS elements are also employed to knock out gene function by causing insertional inactivations. The analyses of the phenotypic effects of these insertional inactivations in a well characterized host-background could lead to a possible discovery of the function of the inactivated gene.

As a result of recent advancements in DNA sequence and analysis technologies there has been a burst in the increase in numbers of IS elements discovered. To date, the current IS database

(<http://www-is.biotoul.fr>) recognize 19 families of IS elements based on differences within the conserved features of these elements. These criteria have been mentioned in Chapter 1 and will be briefly mentioned here. They include the following: IS open reading frame (ORF) organization; conserved signature motifs among Tpsases; similarities of inverted repeats (IRs) and length of target site duplications (Mahillon and Chandler, 1998). It should be noted however, that a target-sequence-dependent site preference has not been demonstrated for some members, while other members appear to transpose in a random manner. Each family is named after the prototype member of the group. These 19 families contain more than 800 insertion sequences isolated from 86 genera representing 196 bacterial species of both eubacteria and archaea. Despite the considerable diversity amongst IS elements they share several structural features. They are usually between 800 and 2500 base pairs (bp) long and contain a gene encoding for the Tpsase, which is flanked by terminal IRs. These IRs are between 10 to 40 bp and although they serve as sites for recognition and cleavage by the Tpsase in transposition reactions, some IS elements have been identified that lack these terminal structures. During the transposition process most IS elements generate short directly repeats of the target DNA at their point of insertions. The chemical reactions, the mechanisms of transposition and the enzymology of the transposition process have all been described in Chapter 1.

In this study the identification and partial characterization of ISAtcI present on pTC-F14 isolated from *At. caldus* is reported. A computational analysis of ISAtcI and the proposal that it is a new member of the ISL3 family of IS elements is presented. Also demonstrated is that ISAtcI is present in multiple copies in the genome in three local *At. caldus* isolates but not in three foreign strains. An attempt was made to show that ISAtcI has the ability form a cointegrate-like structure within *E. coli*.

3.2. Materials and Methods

Bacterial strains, plasmids, media and growth conditions

The bacterial strains and plasmids used in this study are shown in Table 3.1. *E. coli* strains were grown in either Luria-Bertani broth or on LA plates (Sambrook *et al*, 1989) at 37°C and supplemented with antibiotics at the following concentrations as required: ampicillin (Amp), 100 µg/ml; chloramphenicol (Cm), 34 µg/ml; streptomycin (Str), 50 µg/ml; and nalidixic acid (Nal), 50 µg/ml. *At. caldus* strain f was cultured in tetrathionate medium which was made from a mineral salts solution containing the following (grams per liter): (NH₄)₂SO₄, 3.0; KCl, 0.1;

K_2HPO_4 , 0.5; $MgSO_4 \cdot 7H_2O$, 0.5; $Ca(NO_3)_2 \cdot 4H_2O$, 0.014; Na_2SO_4 , 1.45. The pH was adjusted to pH 2.5 with H_2SO_4 and the mixture was autoclaved. The trace element solution used contained the following (milligrams per liter): $ZnSO_4 \cdot 7H_2O$, 10.0; $CuSO_4 \cdot 5H_2O$, 1.0; $MnSO_4 \cdot 4H_2O$, 1.0; $CoCl_2 \cdot 6H_2O$, 0.5; $Cr_2(SO_4)_3 \cdot 15H_2O$, 0.5; $Na_2B_4O_7 \cdot 10H_2O$, 0.5; $NaMoO_4 \cdot 2H_2O$, 0.5. This mixture was also autoclaved and 1ml was added per liter. Filter-sterilized $K_2S_2O_6$ (tetrathionate) was added to a final concentration of 10mM, and the pH was adjusted to 2.5 with H_2SO_4 . The final growth medium was sparged with CO_2 -enriched (2%, vol/vol) air. *At. caldus* was grown at 37°C with constant shaking.

DNA techniques

Plasmid preparation, restriction endonuclease digestions, gel electrophoresis, ligations, bacterial transformations and hybridization experiments were carried out by standard methods (Sambrook *et al*, 1989). The labeling of probes, hybridization and detection was performed by using the dioxigenin-dUTP non-radioactive DNA labeling and detection system (Roche Molecular Biochemicals). Hybridization was done at 40°C in Easy Hyb (Roche Molecular Biochemicals) followed by two non-stringent washes at 25°C (in 2 X SSC, 0.1%SDS) and two stringent washes at 65°C (0.1 X SSC, 0.1% SDS).

Harvesting of *At. caldus* cells and preparation of chromosomal DNA

Cells were recovered by centrifugation and washed three times in acidified water (pH1.8). Each wash step was followed by a process of low- and high-speed centrifugation. This was to remove all the extraneous sulphur species attached to the cells. Washed cell pellets were resuspended in TE (0.01M Tris, 0.001M EDTA)-0.15M NaCl (pH7.6) buffer. Cells resuspended in TE-NaCl (pH7.6) buffer were used for the preparation of chromosomal DNA as described by Breed *et al* (1999). Chromosomal DNA was derived from six different strains of *At. caldus* (Table 3.1).

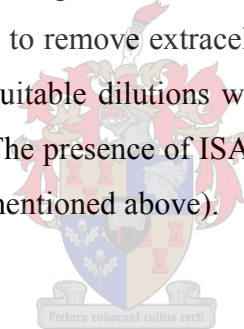
Construction of plasmids

An 1810-bp *EcoRI-SphI* fragment containing *ISAtc1* (Fig 2.2, Chapter 2 and Appendix one) was cloned into the *E. coli* non-mobilizable plasmid pUC19 and the resulting plasmid was designated pUC19-*ISAtc1*. An internal 300-bp *StuI-BglIII* fragment of *ISAtc1* was subsequently

used as probe in Southern Blot experiments for the detection of the presence of ISAtcI in transconjugants and in *At. caldus* chromosomal DNA.

Conjugation experiments

Transposition insertions into target plasmids were obtained by a mating-out assay using *E. coli* S17.1 and CSH56 as the donor and recipient strains respectively. To ensure that all insertion events were unique, each mating experiment was carried out with a single colony from independent transformations of the donor strain. The donor and recipient cells were cultured separately overnight with appropriate antibiotic selection. Cells were washed four times with 0.85% (wt/vol) NaCl solution and mixed in a donor/recipient ratio of 1:10. An LA plate without antibiotics was spotted with 100 µl of this mixture and incubated at 37°C for 2 hrs without agitation. The agar plug containing the mating mixture, was excised, suspended in 5ml of 0.85% NaCl solution, and vigorously shaken to dislodge mating cells. Cells were pelleted by a 2-min spin in a micro-centrifuge, resuspended and washed in 1ml 0.85% NaCl solution. The washing steps were repeated 5 times to remove extracellular β-lactamase and prevent growth of satellite colonies at low dilutions. Suitable dilutions were plated onto media that selected for the donor and transconjugant cells. The presence of ISAtcI in the transconjugants was detected through Southern hybridization (as mentioned above).



Computational analysis

The nucleotide sequence of ISAtcI was inspected using mainly the PC-based DNAMAN program (version 4.1) from Lynnon BioSoft. Putative promoter regions were detected through visual inspection and by using the web-based promoter prediction software (http://www.fruitfly.org/seq_tools/promotor.html). The BLASTP software (Altschul *et al*, 1997) was used to compare the deduced protein sequences with the non-redundant database entries at the server of the National Center of Biotechnology Information. A BLASTP search was also performed to compare the deduced amino acid sequence of ISAtcI with the sequences deposited at the Insertion Sequence database hosted at <http://www-is.biotoul.fr>. Phylogenetic trees were constructed first by extracting the various IS T_{pase} sequences, identified by BLAST, from the general IS database and subsequently performing a multiple sequence alignment using the ClustalW software at the EMBL-EBI server and selecting for the PHYLO data output format. By using the PHYLODRAW program (version 0.8) this data was used to construct a cladogram. The data of the multiple alignment performed using the ClustalW method were

manually assessed and shaded using the Genedoc Multiple Sequence Alignment Editor and Shading Utility (version 2.6.002) by Karl Nicholas. The alignment of the inverted repeats nucleotide sequence and the helix-turn-helix amino acid sequences were done using DNAMAN. Inspection of T_{pase} sequences for the potential to form HTH motifs were predicted by using the NPSA prediction service (<http://npsa-pbil.ibcp.fr/>). Secondary structures of T_{pase} proteins were analyzed using the PSIPRED software (<http://bioinf.cs.ucl.ac.uk/psipred/>).

Nucleotide sequence accession number

The sequence reported here has been deposited in the EMBL-GenBank database under accession no. NC_004734.

Table 3.1: Bacterial strains and plasmids used in this study

Bacterial strain or plasmid	Genotype or description	Source or reference
Strains		
<i>E. coli</i>		
DH5a	F ['] <i>endA1 hsdR17</i> (r _k ⁻ m _k ⁺) <i>supE44 thi-1 recA1 gyrA</i> (Nal ^r) <i>relA1 Δ(lacZYA-argF)U169</i> (φ80d <i>lacΔ(lacZ)M15</i>)	Promega Corp., Madison, Wisconsin, USA
S17.1	<i>recA pro hsdR</i> (RP4-2 Tc::Mu Km::Tn7)	Simon <i>et al</i> , 1983
CSH56	F ⁻ <i>ara Δ(lac pro) supD nalA thi</i>	Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
<i>At. caldus</i>		
'f'	Nickel pilot plant, Billiton, Randburg, South Africa	Own laboratory
#6	Fairview mine, Barberton, South Africa	Own laboratory
MNG	Arsenopyrite pilot plant, UCT	Murray Gardner
C-SH12	Continuous bioreactor, Brisbane, Australia	Kevin Hallberg
BC13	Birch Coppice, Warwickshire, UK	Hallberg and Lindström, 1994
KU	Kingsbury coal spoil, UK	Hallberg and Lindström, 1994
DSM8584		
Plasmids		
pUC19	Amp ^r <i>lacZ'</i> ; ColE1 replicon, cloning vector	Yanisch-Perron <i>et al</i> , 1985
pUC19-ISA _{tc1}	Amp ^r <i>lacZ'</i> ; A 2557-bp <i>Bam</i> HI/ <i>Eco</i> RI fragment containing ISA _{tc1} from pTC-F14 cloned into pUC19 digested with <i>Bam</i> HI and <i>Eco</i> RI	This study
pSUP106	Cm ^r Tc ^r , mobilizable	Prierer <i>et al</i> , 1985

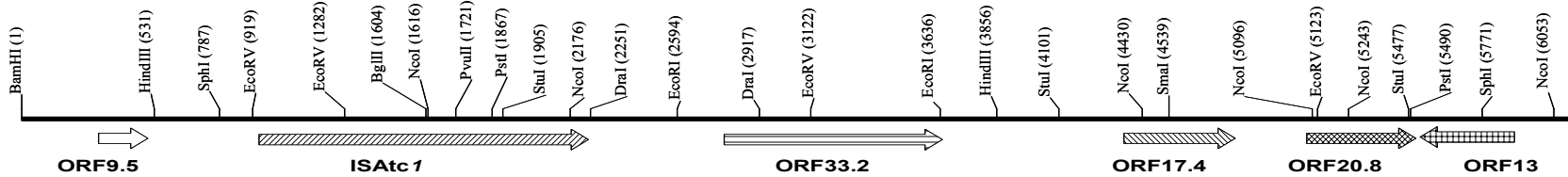
3.3. Results

3.3.1. Nucleotide sequence analysis of ISAtcI

During the nucleotide sequence analysis of the accessory gene region of pTC-F14 (Chapter 2) an open reading frame (ORF) with significant identities to those encoding transposases (Tpase) was detected. Through visual inspection it was discovered that the sequence surrounding the Tpase gene was found to fulfill the criteria required for bacterial insertion sequence (IS) elements. We designated the IS element discovered here, ISAtcI.

A fully annotated nucleotide sequence and the predicted amino acid (aa) sequence of ISAtcI is presented in Appendix One as well as in Fig 3.1. It should be noted however, to assist in presentation here, the sequence in Fig 3.1 is the reverse complementary sequence of ISAtcI relative to the sequence in Appendix one. A reversed physical map of accessory gene region containing ISAtcI is also included in Fig 3.1 for this purpose. ISAtcI is 1,303 base pairs (bp) long and is flanked on both sides by imperfectly conserved 26 bp terminal inverted repeats (IRs). The left and right inverted repeats (IRL and IRR) differ by 6 bp. The G+C content of ISAtcI is 56% which is fairly low compared to the 63.9% G+C content of *Acidithiobacillus caldus* (Hallberg and Lindström, 1994). This implies that ISAtcI is a recent addition to the *At. caldus* genome. Only one ORF, encoding for the Tpase, could be detected using DNAMAN. The Tpase ORF is 1,215 bps long and spans almost the entire length of ISAtcI. It starts with an ATG codon 74 nucleotides (nt) downstream of the IRL and ends with a TGA stop codon at position 1289, 10 bps inside the IRR. Putative promoter sequences were detected through visual inspection and resembled that of the RNA polymerase σ^{70} -type consensus sequence TTGACAN₁₆₋₁₈TATAAT. The -35 promoter hexamer 5'-TCGTCA-3' is located 7 bps within the IRL and is separated by a 18-nt-long spacer sequence from the -10 promoter sequence 5'-TAGAAT-3'. A putative ribosomal binding site resembling a typical Shine-Dalgarno sequence, 5'-AGGAG-3' is located 7 bps upstream of the Tpase ATG start codon. A potential outwardly directed -35 promoter sequence 5'-ATGAA-3' (4/6 consensus), identified on the basis of similarity to ISafeI of *Acidithiobacillus ferrooxidans*, lies at position 1285, 12 bps from the 3' end of IRR (Holmes *et al.*, 2001). Visual inspection of the sequence both upstream and downstream of the terminal IRs did not reveal the presence of an obvious direct target repeat which is a characteristic generated upon insertion during many transposition processes.

A.



B.

Left Inverted Repeat (IRL) Putative RBS Start of Tpmase ORF

1 **GGCTCTTCGTCAGATTGAGTGGGTAGGGGTAGAAAT**GCTGTGGAGCGGTCTGCCCTGAAT**AGGAG**TTAGCTG**ATGG**TCCCTGAAGAGCTGTTTTCTCTCGCGTTAGGATTGGTTCCGCC
 -35 -10 M V P E E L F S L A L G L V P P

121 GTGGTTGGTGGATCATGTGACTTTTCACGGTGGAGGAGAAACGCTGGATCTGCACATCAACTTTCCCAAAGGTAGCTCGTTGCTCCGCTGTGGTGAGGAGTGTCCGGTACATGA
 W L V D H V T F T V E E K R L D L H I N F P K G S R F A C S V C G E E C P V H D

241 TACCCGTGACCATAACCTGGCGGCACATGGATTTCTCCAGCATGAAGCCTATCTCCATGCCCGTGTACCTCGTGTGAAGTGCCAGGAGCATGGAGTGCATCA**GATATC**TGTTCCCTGGGC
 T R D H T W R H M D F F Q H E A Y L H A R V P R V K C Q E H G V H Q I S V P W A

361 GCGGGAAGGCTCGCGTTTCACCTGCTCTTTGAAGCGTGTATGACCTGGTGGCGGAGATGCCGGTATTGACGGCAGCTCGCATGGTGGTGAGACCGACAAGCTCCTGTGGCGAGT
 R E G S R F T L L F E A L I M T L V R E M P V L T A A R M V G E T D K L L W R V

481 GATTGACCATATGTGCCGAAGCTCGTGTGCGGTGGATATGGCCAATGTCCATGCCGTGGCGTTCGATGAAACCAGCAGTCCGGCATGGACATGACTACATCACGCTCTTCGTGGATCT
 I D H Y V P E A R A A V D M A N V H A V G V D E E T S S R H G H D Y I T L F V D L

601 GAATGCCCGGCGACTCTTGTTCGCTACTCCCGCAAGGATGCCAAGACCTTTGAGAAATTCTCCGG**AGATCT**TTCAGG**CCCATGG**TGGTAGCGCGGAAGCGATCACCGATGTGAGCATGGA
 N A R R L L F A T P G K D A K T F E K F S G D L Q A H G G S A E A I T D V S M D

721 CCTCTCGCGCGCTTCCAGAAAGGGGCTGCCGAGCACCTGCCAATGCCGAGATCACTTT**CGATCG**TTTTACCTCATGAAGCTCGTCAACGAGGCCGTAGACGACGTGCGCAAGGGGGA
 L S P A F Q K G A A E H L P N A E I T F D R F H L M K L V N E A V D D V R K G E

841 AGTCTCACCCAGCCAAATCTCAAAAAGACCCGCTGGCTTTGGCTCAAGAACGATTGCAACCTCAAAATGAAGCAGAAAGAAAAG**CTGCAG**GGAATTGCTCAAAGACCAGAACCTCAAGAC
 V L T Q P N L K K T R W L W L K N D C N L K M K Q K E K L Q E L L K D Q N L K T

961 GGCAGGCGCTACAGTTCCGCTGACCTTTACAGGACATCTTACGATCAAGAATCGCCACCAGGGGCTACCTCTTGAAAGCCTGGTTGGAAAACGCCAGAACCAGCGATCTGCCGCC
 A Q A Y Q F R L T F Q D I F T I K N R H C Q G A T T L L K A A W L E N A R T T S D L P P

1081 TATCGTCAGGGTGCCTACACCTCATGAACATTGGGACGGGGTCCCTCCGCTGGTTCGAGAGCCAAGTCAGCAATGGCATCCTGGAGGGCTTCAACAGCCTCCTCCAGTCCGCCAAGGC
 I V R V A Y T L M N H W D G V L R W F E S Q V S N G I L E G F N S L L Q S A K A

1201 GAAAGCCCGTGGCTACCGCACCCACAAGAACTTCATCG**CCATGG**CTTACCTCATCCTGGGCAAGCTGGATCTGAGGCT**TACCCACATGAAA**CGT**CGAAGAGCC**
 K A R G Y R T H K N F I A M A Y L I L G K L D L R L P T *

Stop of Tpmase ORF -35

Figure 3.1: A. Physical map of the accessory gene region of pTC-F14 coding for ISAtc1. **B.** Annotated sequence of ISAtc1 (1,303 bp). The amino acid sequence of the Tpmase encoded by ISAtc1 is shown below the nucleotide sequence. The right and left terminal inverted repeats are labeled and indicated in bold. The putative ribosomal binding site (RBS) is shaded and promoter regions are boxed. The stop and start codons of ISAtc1 Tpmase are labeled and shaded.

3.3.2. Homology of ISAtcI Tpmase to Tpmases of other IS elements

A preliminary BLASTP search of the non-redundant GenBank-EMBL database with the predicted amino acid sequence of the Tpmase of ISAtcI revealed a number of highly similar transposases of bacterial IS elements and Tpmases present on the genome of various bacteria. Table 3.2 lists the highest BLASTP results. The most closely related Tpmase are from the bacterial insertion element ISafeI isolated from a related biomining bacterium *Acidithiobacillus ferrooxidans*. Some of the chromosomally located Tpmases showing homology to the ISActI Tpmase have not been formally documented to be part of a simple bacterial insertion sequence. ISafeI from *At. ferrooxidans* has been documented to be a member of the ISL3 family of bacterial insertion sequences (Holmes *et al*, 2001). From the BLAST results it is clear that the putative Tpmase coded by ISAtcI, found on pTC-F14, is clearly a close relative of ISafeI. These two insertion sequences show high homology not only between their respective Tpmases on both amino acid and nucleotide level, but also in terms of their terminal inverted repeats, as will be seen below. However, these elements are not identical diverging by more than 5% and more than 10% in their amino acid and nucleotide sequences respectively (Chandler and Mahillon, 2002). They also show distinct differences in their respective restriction endonuclease maps (data not shown) as well as in the iso-electric points (pI) of their Tpmases (see below, section 3.3.3). As expected from the high similarity and identity to ISafeI, ISAtcI is also a member of the ISL3 family of insertion sequences. The IS element ISAeI present on the genome of *Ralstonia eutrophus* (previously known as *Alcaligenes eutrophus*) is also classified as an ISL3 family member. This further supports the idea that ISAtcI is an ISL3 member.

This idea was further strengthened when a BLASTP search of the current general IS database using the ISAtcI Tpmase as the query sequence revealed homology only to members of the ISL3 family of IS elements. ISafeI sequence was very recently deposited at this database and again shared the highest identity and similarity similar to that obtained from a BLASTP search of GenBank-EMBL database (Table 3.2). ISAeI showed the second highest identity/similarity. Only the top ten highest BLASTP results of the IS Database search are summarized in Table 3.3. ISAtcI also displayed motifs similar to the Tpmase sequences of rest of the ISL3 family members. A list includes the following, with the host (origin) and sequence accession numbers or Pubmed reference given in parentheses. (It should be noted that in cases where no accession numbers are available, these sequences have not been deposited in the GenBank/EMBL/DDBJ databases but have been directly submitted to the IS database): IS1096 (*Mycobacterium smegmatis* ATCC607; M76495), IS1411 (*Pseudomonas* sp. EST1001; M57500) , ISPsp2

(*Pseudomonas* sp. EST1001; M57500), IS651 (*Bacillus halodurans* C-125; NC_002570), IS1207 (*Corynebacterium glutamicum* B115; X96962), IS31831 (*Corynebacterium glutamicum* ATCC31831; D17429), IS2001 (*Bifidobacterium lactis*; AJ243948), IS1165 (*Leuconostoc mesenteroides* subsp. *cremoris* DB1165; X62617), ISBli3 (*Brevibacterium linens*), IS204 (*Nocardia asteroides* YP21; U10634), ISPst2 (*Pseudomonas stutzeri* M1; AJ012352), ISLdl1 (*Lactobacillus delbrueckii* subsp. *lactis*; AJ302652), IS1476 (*Enterococcus faecium*; U63997), ISPPu12 (*Pseudomonas putida*; AY128707), IS1396 (*Serratia marcescens*; AF027768), IS13869 (*Brevibacterium lactofermentum* ATCC13869; Z66534), IS1193D (*Streptococcus thermophilus* CNRZ368; Y13713), IS1193 (*Streptococcus thermophilus*; Y13713), ISL3 (*Lactobacillus delbrueckii* subsp. *bulgaricus* ATCC11842; X79114), ISSth1 (*Streptococcus thermophilus*; AY376237), ISPP2 (*Pseudomonas putida* ML2; U25434), IS1489v1 (*Pseudomonas putida* ML2; AF148496), IS466A (*Streptomyces coelicolor* A3(2) M145; AB032065), IS1181 (*Staphylococcus aureus* BM3121; L14544), IS1167 (*Streptococcus pneumoniae* CP1200; M36180), IS1167A (*Streptococcus pneumoniae* SP-VA96; AF030364), IS469 (*Streptomyces coelicolor* A3(2) M145; AB032065), ISP1 (*Lactobacillus plantarum*; NC_004567), ISSg1 (*Streptococcus gordonii* M5), ISCx1 (*Corynebacterium xerosis* M82B; U21300), IS1251 (*Enterococcus faecium* GUC; L34675). The identity of the aa sequence of ISA_{AtcI} Tpase with the Tpases of these IS elements ranged from 20 to 25%, with some having high sequence identity in only the C-terminal regions of these proteins. This will be returned to later.

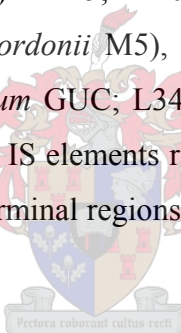


Table 3.2: Homology of ISA_{AtcI} Tpase with the most closely related sequences in the GenBank-EMBL databases

Most closely related Tpase	Length (aa)	No. of aa acids compared	% Identity/ Similarity	BLAST E value	Reference/NCBI Accession no.
ISAFel from <i>Acidithiobacillus ferrooxidans</i>	404	404	92/95	0.0	AAB07489
<i>Methanosarcina acetivorans</i> strain C2A	403	399	47/66	e ⁻¹¹²	NP_617312.1
<i>Methanosarcina acetivorans</i> strain C2A	394	388	47/66	e ⁻¹⁰⁹	NP_618857.1
<i>Methanosarcina barkeri</i> str. fusaro	371	351	45/65	2e ⁻⁹¹	ZP_00296546.1
<i>Methanococcoides burtonii</i> DSM 6242	402	398	39/63	9e ⁻⁹¹	ZP_00149031.2
<i>Ideonella dechloratans</i>	440	394	43/58	2e ⁻⁸⁶	CAD97446.1
ISAE1 from <i>Ralstonia eutropha</i>	408	395	41/59	2e ⁻⁸³	AAP85854.1
ISDvu5 from <i>Desulfovibrio vulgaris</i> subsp. <i>vulgaris</i> str. Hildenborough	413	378	37/57	6e ⁻⁷⁴	YP_011233.1
<i>Chlorobium tepidum</i> TLS	311	298	41/55	5e ⁻⁶⁴	NP_663030.1
<i>Chlorobium tepidum</i> TLS	408	394	30/51	2e ⁻⁵⁰	NP_661418.1
<i>Methanococcoides burtonii</i> DSM 6242	297	295	35/58	7e ⁻⁴⁹	ZP_00149383.2

Table 3.3: Comparison of ISAtcI with some insertion elements reported to be members of the ISL3 family elements.

IS element	Origin	Pubmed Reference/ EMBL Accession no.	Length of Tase (aa)	(%) Identity/similarity (part of protein) ^b	Blast E Value	Length of IR (bp)
ISAFE1	<i>Acidithiobacillus ferrooxidans</i> ATCC 19859	U66426	404	92/95 (404)	0.0	26
ISAE1	<i>Ralstonia eutropha</i> H1-4	M86608	408	41/59 (395)	2e ⁻⁸³	24
ISB106	<i>Bifidobacterium longum</i> NCC2705	NC_004307	411	42/57 (399)	1e ⁻⁸⁵	42
IST445 ^a _aa2	<i>Acidithiobacillus ferrooxidans</i>	Y18309	177	60/62 (133)	8e ⁻³²	23
ISBma1	<i>Burkholderia mallei</i>	AF285635	386	27/44 (363)	1e ⁻³¹	23
ISRso15	<i>Ralstonia solanacearum</i>	NC_003295	406	26/43 (380)	9e ⁻³¹	19
IS1001	<i>Bordetella parapertussis</i> B24	X66858	406	26/43 (349)	8e ⁻²⁹	26
IST445_aa3	<i>Acidithiobacillus ferrooxidans</i>	Y18309	130	45/59 (130)	1e ⁻²³	23
ISBli1	<i>Brevibacterium linens</i>	AF052055	436	26/40 (378)	5e ⁻¹⁹	25
ISAsp1	<i>Anabaena</i> sp. PCC7120	U13767	407	25/48 (263)	2e ⁻¹⁸	22
IST445_aa1	<i>Acidithiobacillus ferrooxidans</i>	Y18309	85	56/61 (84)	3e ⁻¹⁷	23
IS652	<i>Bacillus halodurans</i> C-125	NC_002570	404	22/42 (344)	6e ⁻¹⁷	6

^aIST445 contains three open reading frames which is thought to constitute the Tase (Chakraborty et al, 1997). Each polypeptide is designated with an aa extension and a corresponding number to indicate the ORF encoding it.

^b number in brackets indicates the number of amino acids over which the percentage identity/similarity was obtained.

To illustrate the phylogenetic relationships between ISAtcI and the current ISL3 family members a cladogram based on the multiple alignments of the amino acid sequences of the Tases of all the members of the ISL3 family was constructed, as shown in Fig 3.2. The tree was constructed via the European Bioinformatics Institute (EBI) hosted at the EMBL server (European Molecular Biology Laboratory) by using the ClustalW programme. The PHYLIP output data format was selected and by using PHYLIDRAW v0.82 this data was used to construct the cladogram. As reported before, the elements fall into three deeply branching groups (Chandler and Mahillon, 2002). The cladogram demonstrates that the Tases of ISAtcI, ISAFE1, ISAE1, IS1096, IS489 and ISPst2 are more closely related to each other than to the rest of the other Tases.

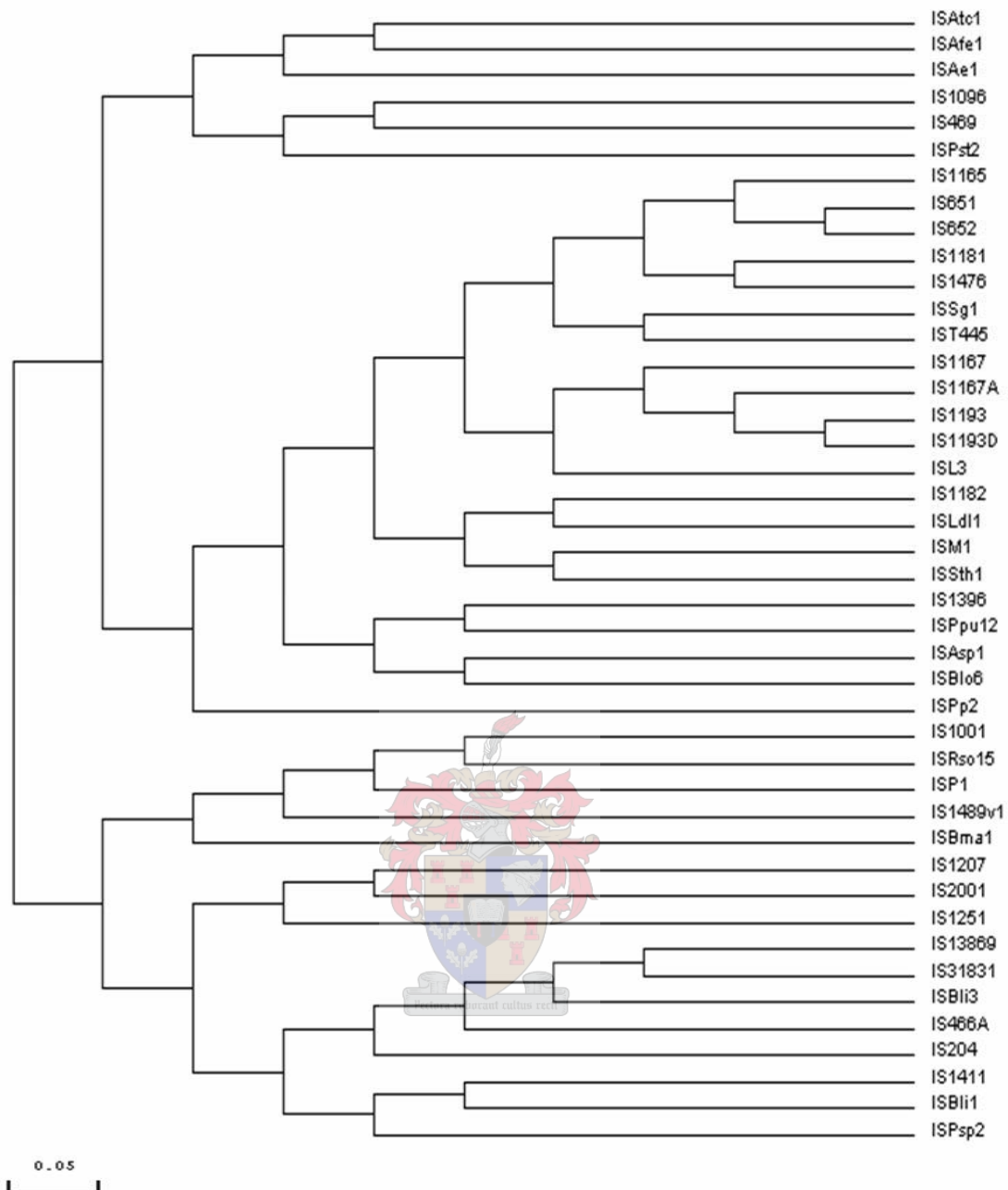


Figure 3.2: Cladogram illustrating the phylogenetic relationship between *ISAtc1* and the ISL3 family of bacterial IS elements. The construction of the cladogram is based on the alignment of the predicted amino acid sequences of the Tpase ORFs of all the members of the ISL3 family of IS elements. Refer to text (“Homology of *ISAtc1* Tpases to Tpases of other IS elements”) for information on the origin of these elements as well as their EMBL sequence accession numbers or Pubmed references numbers.

3.3.3. *ISAtc1* Tpase structure analysis

The translated sequence of the Tpase of *ISAtc1* encodes a polypeptide of 404 amino acid residues and has a deduced molecular weight 46188 Da. The putative protein has a moderately high content of basic amino acids with arginine (R), lysine (L) and histidine (H) constituting

17.33% of the amino acids in the protein, and has a theoretical isoelectric point (pI) of 8.20. This pI is relatively low in comparison to the pI values observed for other Tpsases (Galas and Chandler, 1989) but the data is consistent with the function of Tpsases in DNA binding. This feature is characteristic of Tpsases from IS elements.

A multiple sequence alignment of the amino acid sequences of the Tpsases of *ISAtcI*, *ISAfeI*, *ISAeI*, *ISBlo6* and *IST445* was done using the ClustalW program, as shown in Fig 3.3. These IS elements showed the highest identity and similarity in their primary amino acid sequences to *ISAtcI* during a BLASTP search of the IS database (Table 3.3) and was subsequently selected for alignment purposes. It should be noted that ISL3 is the prototypical member of the SL3 family of IS elements and is the best characterized element of this family in terms of the conserved motifs in its primary amino acid sequence, and is therefore included in the alignments as a control to assist in the discovery of similar conserved motifs. An alignment of all members of the ISL3 family was done, but the data set is too large to present here. The alignment revealed that several amino acids residues were strongly conserved in corresponding positions along the entire protein with most of the conservation being observed in the C-terminal region of these Tpsases, from positions 165 onward. This shows that all the elements examined are closely related to one another and supports the assumption that *ISAtcI* is a member of the ISL3 family. Despite the prevalence of the overall basic amino acids in Tpsases, three critical acidic amino acid residues (two aspartate and one glutamate residue) have been shown to form a catalytic triad and have been shown through sequence analysis to be present in the Tpsases of various bacterial insertion sequences. This triad, known as the DDE motif, has also been identified in retroviral integrases and in the bacteriophage Mu Tpsase. The importance of this triad has been shown through directed mutagenesis and is intimately involved in catalysis where it is devoted to the coordination of the divalent metal ions that assist the various reactions that underlie the cutting and joining events of the transposition process (Chapter1; Chandler and Mahillon, 2002). Based on a multiple alignment of Tpsases from the various families of bacterial insertion sequences, retroviral integrases and bacteriophage Mu (Chandler and Mahillon, 2002), the position of the two aspartate residues and one glutamate residue constituting the DDE motif of ISL3 has been identified. This served as basis in discovering a putative DDE motif for *ISAtcI*. For *ISAtcI* a putative DDE motif is seen at positions 161 (D161), 237 (D237) and 364 (E364). *ISAfeI* also contains a putative DDE motif in identical positions as that seen for *ISAtcI*. Although a number of Tpsases and integrases have a characteristic DD(35)E arrangement, the spacing of the second aspartate and glutamate is larger (ca. 124 amino acids) in the ISL3 family members (data not shown). From the alignment it is

evident that this motif is conserved within the primary sequence of both *ISAeI* and *ISBlo6*. Even though the perfectly conserved D residue at position 181 is the most C-terminal D residue it has been established that the D residue forming part of the DE peptide (position 161 in *ISAtcI*) present in 39 of the 45 members of the ISL3 family is the first D residue of the DDE motif. Various conserved amino acid residues surrounding (upstream and downstream of) the D, D and E residues of the DDE triad are evident from the alignment. This further supports the idea of these proposed amino acid residues constituting the DDE motif in *ISAtcI*, and has previously been documented to be part of three conserved regions of the T_pase protein; i.e. the N2, N3 and C1 regions which encompasses the D, D and E residues respectively. A striking conservation observed on all six T_pase sequences compared in Fig 3.3, is the R-F-H-L(V)-M(V) motif surrounding the D240 which may prove to have an important function during the transposition process. Several highly conserved amino acid residues upstream and downstream of the E364 (position in *ISAtcI*) are also apparent. This includes the perfectly conserved L363 (leucine) upstream of E364. In several T_pases a conserved K (lysine) or R (arginine) residue approximately seven residues downstream from the E residue has been reported and has been shown to interact with the terminal base pairs of the terminal inverted repeats of the several IS elements (Chandler and Mahillon, 2002). This conservation is not evident from the alignment shown in Fig 3.3. However, ISL3 contains a K (K391) residue seven amino acids downstream of its E residue (E384). It has been reported by Berger and Haas (2001) that this K residue is present in only a few IS elements but suggested that other basic or polar residues at this position might play the same role as the K residue. In the corresponding positions in *ISAtcI*, *ISafeI*, *ISAeI* and *ISBlo6* the K residue is absent but a polar glutamine (Q) residue is present. A highly conserved K-A-R-G-Y-R downstream of the E364 could be seen as well a G-K-L-D-L-R(K) motif located at the end of the T_pases. The N2(D) and N3(D) regions of T_pases from other IS elements have been shown to each contain a β -sheet and the C1(E) region to contain an α -helix. This could not be identified for *ISAtcI* using the PSIPRED protein secondary structure prediction server. No obvious DDE motif could be detected for *IST445*. Although the DDE motif has directly been implicated in the joining and cutting reactions, not all IS elements exhibit a well defined DDE motif and appear to have a different mechanism of transposition. Whether this is the case for *IST445* is not known.

Unlike the C-terminal region, containing the highly conserved DDE motif and surrounding amino acids, the N-terminal region is not as strongly conserved. The C-terminal contains the DNA-binding function of the T_pase which plays a role in the binding of both the terminal IRs and possibly the target DNA during the transposition. Different IS elements carry IRs which

differ in their nucleotide sequences and this might require different arrangements of amino acids in the DNA binding domain. A comparison of the IRs of the IS elements aligned in Fig 3.3 is shown in Fig 3.4 and a number of similarities can be found. The right terminal IRs seem to be the best conserved in their nucleotide sequences and in particular the heptanucleotide sequence, AAGAGCC, at the end of the IRRs is conspicuous as is the terminal heptanucleotide, CGCTCTT, in the IRL. The terminal two base pairs of the IRL and IRR have been shown to be important in catalysis. This conservation of the terminal base pairs of the IRs might possibly point to a potentially similar catalytic step in the transposition process. The presence of the DNA-binding domain located in the N-terminal region is a common characteristic amongst Tpsases and the presence of a Helix-turn-Helix (HTH) motif and/or Zinc Finger (ZF) motif and/or Leucine Zipper motifs have been reported to be present in various IS elements.

Analysis of the N-terminal region of *ISAtcI* for its potential to form HTH motifs has been performed using the NPSA prediction service (<http://npsa-pbil.ibcp.fr/>) and predicted that *ISAtcI* contains a HTH motif at position 117-139. However, the probability of *ISAtcI* carrying a HTH motif at this position is given as very low. *ISAfeI*, *ISAeI*, *ISBlo6*, *IST445* and *ISL3* were also analyzed for the presence of a HTH motif and the results are shown in Fig 3.5. The probability for *ISAeI* and *ISL3* carrying a HTH motif is 25 and 100% respectively as determined by the NPSA server. The probability of *ISBlo6* and *IST445* carrying a HTH motif is very low but is included here for the purpose of being complete. Inspection of the alignment reveals some degree of amino acid conservation. Most of the conservation is seen in the first helix (helix 1) in particular the perfectly conserved A (alanine) residue. Helix 1 is thought to ensure the maintenance of the structure of the DNA-binding motif (Rousseau *et al.*, 2004). Helix 2 on the other hand is involved in nucleotide sequence discrimination and binding and this might explain the weak conservation in this region which is possibly dictated by the difference in the nucleotide sequence of the terminal IRs of these elements. Helix 2 appears to be more basic than helix 1 which is consistent with helix 2 being concerned with DNA binding. Interestingly the perfectly conserved A residue seen in helix 1 is highly conserved in 39 of the 45 members of the *ISL3* family and suggests that this residue performs an important function in the stabilization of the DNA-binding motif. No putative ZF or LZ motifs could be detected for *ISAtcI*.

```

ISAtc1 : -MVPEELFSLALGLVPPWLVD-HVTFET---VEEKRLDLHINFPKGSRFACSVCGEEC---PVHDTDRDHTWRHMDFEQHEAYLHARVPRVKCQEHGVHQISVPWAREGSRF : 102
ISAFE1 : -MVPEDLFLSLALGLVPPWLVD-DVTFEQ---VEEKRLDLHINFPKGSRFACPVCGEEC---PVHDTREHTWRHMDFEQHEAYLHARVPRVMCPEHGVHQIPVPWAREGSRF : 102
ISAE1 : --MHSKLFEEALGLISDPWFVR-EVDEN---AQTKTLLTIQIDFVAGSRFSHEEVAGGH---PVHDTVTKRYRHLNFFEHDCYLEVRTPRVKLPDGRVVLVEPDWAGKLSGF : 101
ISB106 : -MDTSYLFATAALQLQSPWKVE-SVDFRDAGDGRQELRIAIGFEAGSRFCCPEPGCDETVCVHDTRETRTWRHPDFERYKAFIHAGVPRVSCPVHGVMPVPAPWARPGSGF : 108
IST445 : -MVPEDLFLSLALGLVPPVVGGRNVTFEQ---VEEKRLDLHINFPKGSRFACPVVRSV---RYMHPGTIPAAHG-LFQHEVLSHAADPG----- : 80
ISL3 : MNPTANYIENLLNVKDFCLNFVDCSQT---TYKGRKVIIMCTAIAELDTCPNCKSAD---YYTYNGFKVVIPIYISADESNPVIKVKKHGIFCKNCELYSYPSTQIVDKY : 104

```

```

ISAtc1 : TLLFEALIMTLVREMPVLTAAARMVGETDKLLWRVIDHYVPEARAADVMANVHAVG---VDETSSRHGHDIITLFDVLDNARRLLLFATPGKDAKTFEKFSGDLQAHGGSAAEA : 209
ISAFE1 : TLLFEALIMTLVREMPVLTARLVGETDTLLWRVIDHYVPEARTRVDMANVHAVG---VDETSSRRGHDIITLFDVLDNARRLLLFATPGKDAKTFEKFSEDLQAHGGSAAEA : 209
ISAE1 : TLLFEAMVVALAQMPFSAVARTVGSWHRVYAIICERYVDLVAELDLAAGVTAATA---VDETSYRRGHNYLTLVADADARKVVFVTEGKDAATVGKFAAHLRENNAAPEQ : 208
ISB106 : TLLFEAWAVEMARHLPVSTLAEQLDETTRLWRSVAHYVDEARRLEDYMGVEAIG---IDETS-RRGRNCITVVADLVEHDVINVTIPGKDPATVERFSRDFMDHNGVPEY : 214
IST445 : -----RRMPVLTVARWS-ASRHAPVRVIDHYVPEARTRVDMACPCPVRRRIENQAVGAVMTTITLFDVLDNARRLLFRIPGKDAKTFEKFSEDLQAHGGSAAEG : 177
ISL3 : CHISNAVKRKIIIVGLTKDHSMTSIAEENGVSVTTVQRYLDNCSAQFTPSFDSLPEHLAFDFEFR-GVGRKLHFICQDGEKHTIVAILENRFKNTIIKYFLQFP--EIVRKT : 211

```

```

ISAtc1 : IT-DVSMDLSPAFQKGA- EHLPNAEITFDRFHLMKLVNE-----AVDDVRKGEVLT-QPNL-KKTRWLWLKNDCLMKKQKEKLOELLKD--QNLKTAQAYQFRLTFQD : 308
ISAFE1 : IT-DVSMDLSPAFQKGA- EHLPNAEITFDRFHLMKLVNE-----AVDAVRKGEVLT-QPNL-KKTRWLWLKNDCLKARQKEKLOELLKD--QNLKTAQAYQFRLTFQD : 308
ISAE1 : IG-VVSIDMSPAFIKGVS- EHLPNARITFDKFKHVVAHASA-----AVDKTRRIEQKT-DPSL-KGLRWTLKDRDGLPAAQRADLDALIANV-TTKRTARAWLYREQLRE : 308
ISB106 : VR-LVACDMSLGFIRKIR- EHLPHARRIVDRFHVARHANE-----AVDKAGKTEGRS-NPLL-KRTKYLWLRNEESTELQVETKRNLTR---QRLKTGRACRMREVLQD : 312
IST445 : INRIVSMALSPAFQKACRSTCEMLRSPSNRFHLMKLVQR-----GRGCRAQGGSPS-PSQISKRPAAGSGSRMIVTPQQAERKAAGIIKR--PEPENGTGLPVSILTFQA : 279
ISL3 : VR-TVSMDLNCYIGDIVR-QIFPNAEIIVDRFHVMQMVNRSFIGFRVQVMKQLDKKSREYKLLKRYWKLKMYKDYEGSKQFYDRCLKVPYTPAQIVDEGLKCNETLKN : 319

```

RFHLM

```

ISAtc1 : IFTIK--NRHQGATLLKAWLENARTSDLPPIVAVAYTL-MNHWDG--VLRWFES-QVSNGLIEGFNSLLQSAKAKARGVTRTHKNFIAMAYLILGKLDLRLPT----- : 404
ISAFE1 : IFTVK--NRHQGATLLKAWMENAKDSGMPMPVKVAYTI-MNHWDG--VLRWFES-QITNGILEGFNSLIQSAKAKARGVTRTHKNFINMAYLILGKLDLRLPT----- : 404
ISAE1 : ILERK--QINVVSAMLEQWCTNVMSKVEPMKEVARMIRKHFEDG--IVAWTQTRQTNGLFLEALNGLFQAARKARGVVSFKTMRTVIFLIAGKLDLFSAINPHAA---- : 408
ISB106 : IYTDSP-TPSEAVRHLRCSWMMHSRLEPMKDFARMIRRHFAE--IVAYFEH-PYTNVAVLEGADSVIQNVKRRASGERNMDYFATMIYLTTCGKLDLKAIVTTT----- : 411
IST445 : FFTVK--NRTRAPTLLKAWMENAKDSGMPMPVKVAYTIPLNNWDGR-CFRWFEKPDYQWNSWKASTASINPPRPRP-GVTDHKNFINMAYLILGKLDLRLPT----- : 377
ISL3 : TYDFMQDFVYALADKDTKKINDLLDSNIGQYICERLKTITRTRKNNRAVINGAKMSYSNGCLEGVNRKIKQIERTAYGVSNFTHLLKRIRLEQNIICFKKALNRRVLFSA : 429

```

E

KARGYR

GKLDLR

Figure 3.3: Multiple sequence alignment of the Tpsase amino acid sequences of ISAtc1, ISAFE1, ISAE1, ISB106, IST445 and ISL3. For information of the respective accession numbers and original host refer to Table 3.3. The D, D and E residues constituting the DDE motif is labeled and boxed. Other conserved motifs are labeled.

Additional well conserved motifs in the N-terminal region include the highly conserved G-S-R-F and the highly conserved P-V-H-D-T motifs located at positions 44 and 57 respectively. Other highly conserved residues are also evident from the alignment in Fig 3.3.



Figure 3.4: Sequence alignment of the IRs of ISAtc1, ISAfe1, ISAe1, ISBlo6, IST445 and ISL3. For information of the respective accession numbers and original host refer to Table 3.3.



Figure 3.5: Sequence alignment of the predicted HTH motifs of ISAtc1, ISAfe1, ISAe1, ISBlo6, IST445 and ISL3. For information of the respective accession numbers and original host refer to Table 3.3

3.3.4. Genomic distribution of ISAtc1 in six *At. caldus* strains

In order to ascertain whether ISAtc1 elements are as widespread amongst *At. caldus* strains as ISAfe1 is amongst *At. ferrooxidans* strains (Holmes *et al.*, 2001), we used the Southern hybridization technique. Genomic DNA was prepared from six *At caldus* strains, two of which originated from Europe, one from Australia and three from South Africa (Table 3.1). BamHI does not cut inside ISAtc1, thus Southern Hybridization was carried out in which BamHI digested genomic DNA was hybridized to a labeled probe prepared from ISAtc1. The result is shown in Fig 3.6. The three South African isolates of *At. caldus* gave positive signals for ISAtc1, but not the three foreign strains. Bands similar in size are apparent but clear differences

in banding patterns are also visible in the three South African isolates. Approximately 11-15 bands were obtained for each of the South African isolates, which suggest that multiple copies of ISAtcI-like sequences are present on the chromosome of these isolates. Because ISAtcI harbor no internal *Bam*HI restriction endonuclease sites, we estimate that there are about 11-16 copies of ISAtcI-like sequences present on the chromosome of these local isolates. However, this is not an accurate estimation because bands showing higher intensity hybridization signals may contain two (or more) copies of ISAtcI-like sequences on the same fragment. On the other hand, bands showing low intensity hybridization signals may contain truncated copies of ISAtcI-like sequences and thus may be an overestimation of ISAtcI-like sequence copy number. The three local strains are partly related as they have some common hybridization fragments but others that differ. This implies that these strains have a degree of restriction site conservation for *Bam*HI. Bands not shared between these isolates may result from combinations of different transposition events, deletions of IS elements and recombination and/or of IS element containing pieces. Alternatively mutations in chromosomal restriction sites or the difference in the conservation of *Bam*HI sites may explain which why the position of ISAtcI is not completely conserved amongst these three strains. Although IS elements have been viewed as parasitic they can contribute to the evolution and adaptation of a bacterial species through the various effects they have on the chromosome of bacteria. The absence of ISAtcI from the three foreign strains may suggest that they have not been in close proximity to this element or ISAtcI plays a specific role in the three local strains and therefore are not a necessary constituent of the genetic material of the three foreign strains. However, these foreign strains might contain IS elements un-related to ISAtcI.

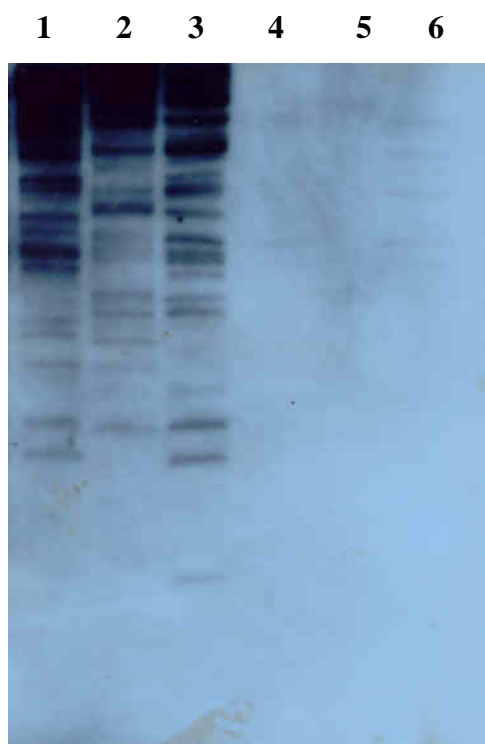


Figure 3.6: DNA hybridization experiment (using ISAtcI as probe) showing the genomic distribution of ISAtcI on the chromosome of *At. caldus* strains. Lane 1, “f”; lane 2, #6; lane 3, MNG; lane 4, KU; lane 5, C-SH12 and lane 6, BC13.

3.3.5. Mobility of ISAtcI in *At. caldus*

Through Southern hybridization experiments we wished to determine whether ISAtcI are mobile within the *At. caldus* f chromosome. Genomic DNA was isolated from *At. caldus* f after various generations of propagation in the laboratory. *At. caldus* have been estimated to have a double time of approximately 9 hours and it was decided to grow *At. caldus* cells for a period of five days after which an inoculum taken from this sample was introduced into fresh *At. caldus* growth media and grown for a further five days. After each growth period genomic DNA was extracted and this procedure was repeated 5 times. The genomic DNA extracted was used in a Southern hybridization experiment using the labeled probe as mentioned above (section 3.3.4). The result is illustrated in Fig 3.7. The frequency of internal transposition of ISAtcI appears not be very high. From the 6 samples of genomic DNA isolated at various generations, only one sample showed a difference in banding pattern. Only in lane 3 could a clear difference be noted (indicated by the arrows in Fig 3.7). It appears that the transposition ISAtcI is strongly controlled within in *At. caldus* f. It is however not known whether the difference in banding pattern is a result of a transposition process or whether it is mediated by chromosomal or plasmid (pTC-F14) mediated rearrangements.

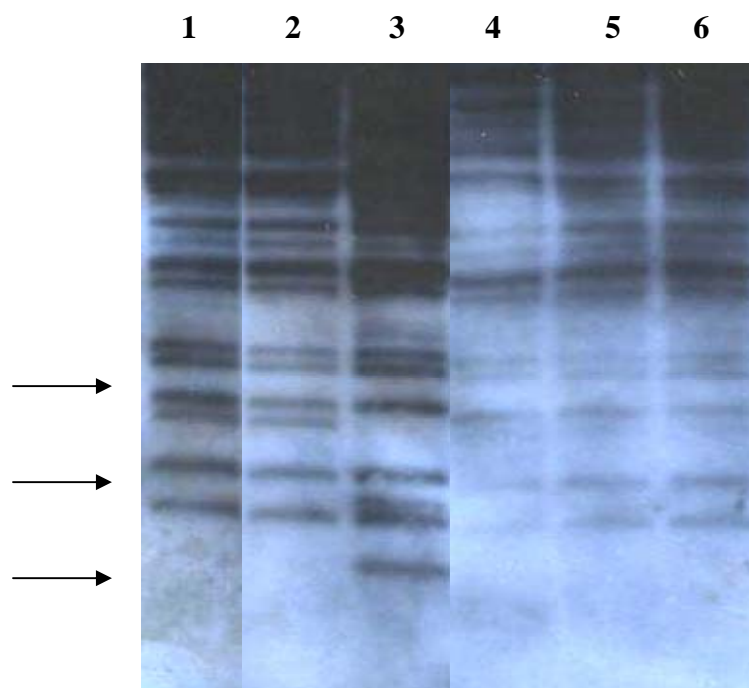


Figure 3.7: Demonstration of the mobility of ISAtcI through Southern hybridization analysis of the genomic DNA prepared from *At. caldus* f at 5 day intervals (lanes 1-6). ISAtcI was used as probe. The mobility of ISAtcI is not extensive and only minor differences in hybridization signals (banding patterns) are evident in lane 3. Arrows indicate difference in banding sizes observed in lane 3 in comparison to the other lanes.

3.3.6. Transposition of ISAtcI in *E. coli*

To determine the ability of ISAtcI to transpose in an *Escherichia coli* S17.1 background we monitored its ability to generate replicon fusions. We constructed pUC19-ISAtcI which contains the 2.8kb *Bam*HI/*Eco*RI fragment of the accessory gene region of pTC-F14 (contains one intact copy of ISAtcI, Fig 3.1) cloned into the non-mobilizable plasmid pUC19. This ampicillin resistant (Amp^r) pUC19-ISAtcI construct served as the donor replicon. The chloramphenicol resistant (Cm^r) mobilizable plasmid pSUP106, which is devoid of any transposable elements served as the IS element transposition “catcher” replicon. These plasmids were transformed into a streptomycin resistant (Str^r) *E. coli* S17.1 donor strain that has an RP4 plasmid derivative integrated into the chromosome to provide the conjugative functions required for plasmid mobilization. Several transformants (Amp^r, Str^r, Cm^r) from the transformation experiments in *E. coli* S17.1 were selected as potential donor strains for independent mating experiments with the nalidixic resistant (Nal^r) *E. coli* CSH56 recipient strain. As negative controls *E. coli* S17.1 was transformed with pUC19 and pSUP106 and spurious transfer of the Amp marker was selected for on LA medium supplemented with Amp and Cm.

To determine whether ISAtcI could transpose from pUC19-ISAtc1 to pSUP106 we evaluated the ability of ISAtcI to form cointegrates either during transposition or following transposition. These cointegrates would be mated from an *E. coli* S17.1 donor strain to an *E. coli* CSH56 recipient strain. Resolution of the cointegrate could occur in the recipient and the pSUP106 “catcher” plasmid could be digested with restriction enzymes to show that it had gained a copy of ISAtcI. A proposed schematic representation of the transposition experiment is given in Fig 3.8. The Str^r *E. coli* S17.1 donor strains harboring pSUP106 and pUC19-ISAtc1 were conjugated with the Nal^r *E. coli* CSH56 strains and transconjugants were selected for on LA media containing Nal, Amp and Cm. Selection was also done on Nal, Str, Amp and Cm LA plates to determine whether the S17.1 donor strains present in the plated out mating mixture gave false positive results by becoming Nal resistant overnight during incubation. The *E. coli* S17.1 strain containing pSUP106 and pUC19 which served as the donor strain in the negative control experiment, was conjugated with *E. coli* CSH56 and transconjugants selected for on Nal, Amp and Cm containing LA media as well as on Nal, Amp LA plates and Nal, Cm LA plates.

Transconjugants were obtained with pUC19-ISAtc1 but not with pUC19. In the negative control no colonies could be recovered either on LA plates containing Nal, Amp and Cm or on the LA Amp plates. This is consistent with pUC19 being a non-mobilizable plasmid. This also indicates that these two replicons have no homology which can serve as substrates for homologous recombination which can subsequently lead to cointegrate formation. Large amounts of colonies were detected when selection was performed on Nal Cm plates as expected since pSUP106 is a mobilizable plasmid which was mobilized by the RP4 plasmid of the *E. coli* S17.1 chromosome. Transconjugants (Amp^r Cm^r Nal^r) were obtained with pUC19-ISAtc1 which implies that the presence of ISAtcI plays a role in the formation of the observed transconjugants in the conjugation experiments with pUC19-ISAtc1 and pSUP106. Since pUC19-ISAtc1 is a non-mobilizable plasmid, the only way the Amp resistance in the transconjugants could be explained is if a replicon fusion (cointegrate formation) occurred between pUC19-ISAtc1 and pSUP106. pSUP106 is mobilizable and when fused to pUC19-ISAtc1 followed by a subsequent conjugation event with a recipient strain would lead to the movement of both these plasmids to the recipient strain. Both the Cm and Amp resistance markers would then be observed in the recipient strain. There is no homology between the donor replicon (pUC19-ISAtc1) and the recipient replicon (pSUP106) which was confirmed by the negative control. One likely way that recombination, which leads to replicon fusion, could take place is after a transposition event in which ISAtcI present on pUC19-ISAtc1 is

transposed to pSUP106 (Fig 3.8 ABD). These two copies, one on each plasmid, would serve as substrates for homologous recombination resulting in to the formation of the fused replicons (cointegrates). It is important to note here that in this option the replicon fusions are formed after a transposition event which implies a nonreplicative process, i.e a cut-and-paste mechanism (Fig 3.8 ABD). In this case the cointegrates are formed between a donor replicon containing an intact copy of *ISAtcI* and a recipient replicon to which an *ISAtcI* element transposed to from another donor replicon. However, a second pathway for the formation of cointegrate-like structures can occur during a replicative transposition event. Here cointegrates are observed as an intermediate transposition product (Fig 3.8 ACD). In the former case cointegrates are a product of homologous recombination and in the latter is a product of replicative transposition, but the presence of fused replicons of two non-homologous replicons points to a transposition event.

Three single colonies from three independent transformations were used in three independent mating experiments. This was done to ensure that each successful experiment was the result of a unique transposition event. The $\text{Nal}^r \text{Amp}^r \text{Cm}^r$ transconjugants obtained after the conjugation experiments were structurally analysed through restriction endonuclease cleavage and Southern hybridization experiments. Transconjugants were restricted with *EcoRV* and compared with the *EcoRV* restriction patterns of pUC19-*ISAtcI* and pSUP106. The result is shown in Fig 3.9A and a schematic illustration of the *EcoRV* restriction map of pUC19-*ISAtcI* and pSUP106 as well as the possible cointegrate intermediates formed, are shown in Fig 3.10. *ISAtcI* is 1,303 bps and contains only one internal *EcoRV* site. The construct pUC19-*ISAtcI* is 5,243 bp and contains two *EcoRV* sites, one located within *ISAtcI*. *EcoRV* cleavage of pUC19-*ISAtcI* yields two bands of more or less 300 bp and 5,000 bp (Fig 3.9A lane 2), which is consistent with the sequencing data. It should be noted that a third faint band can be seen from the gel photo in Fig 3.9A (lane 2) and which is also picked up in the Southern hybridization analysis. This band might be ascribed to being unrestricted nicked pUC19-*ISAtcI* DNA since this band is not seen in all experiments. pSUP106 is 10,900 bp and yields three bands with *EcoRV* restriction with approximate sizes of 3,600; 4,500 and 2,800 bps (Fig 3.9 A, lane 3). In theory, if *ISAtcI* transposed to pSUP106 with the subsequent formation of a cointegrate, pSUP106 should gain two extra *EcoRV* sites, so that on digestion with *EcoRV*, the sizes of two of the bands should stay the same (see doublet at 4.3-4.5 kb in Fig 3.9A lane 3 and Fig 3.9B lanes 2-5) and three different sized fragments would be produced. This is provided that *ISAtcI* transposition does not corrupt one of the *EcoRV* cleavage sites of pSUP106. Thus, a cointegrate formed between pSUP106 and pUC19-*ISAtcI*

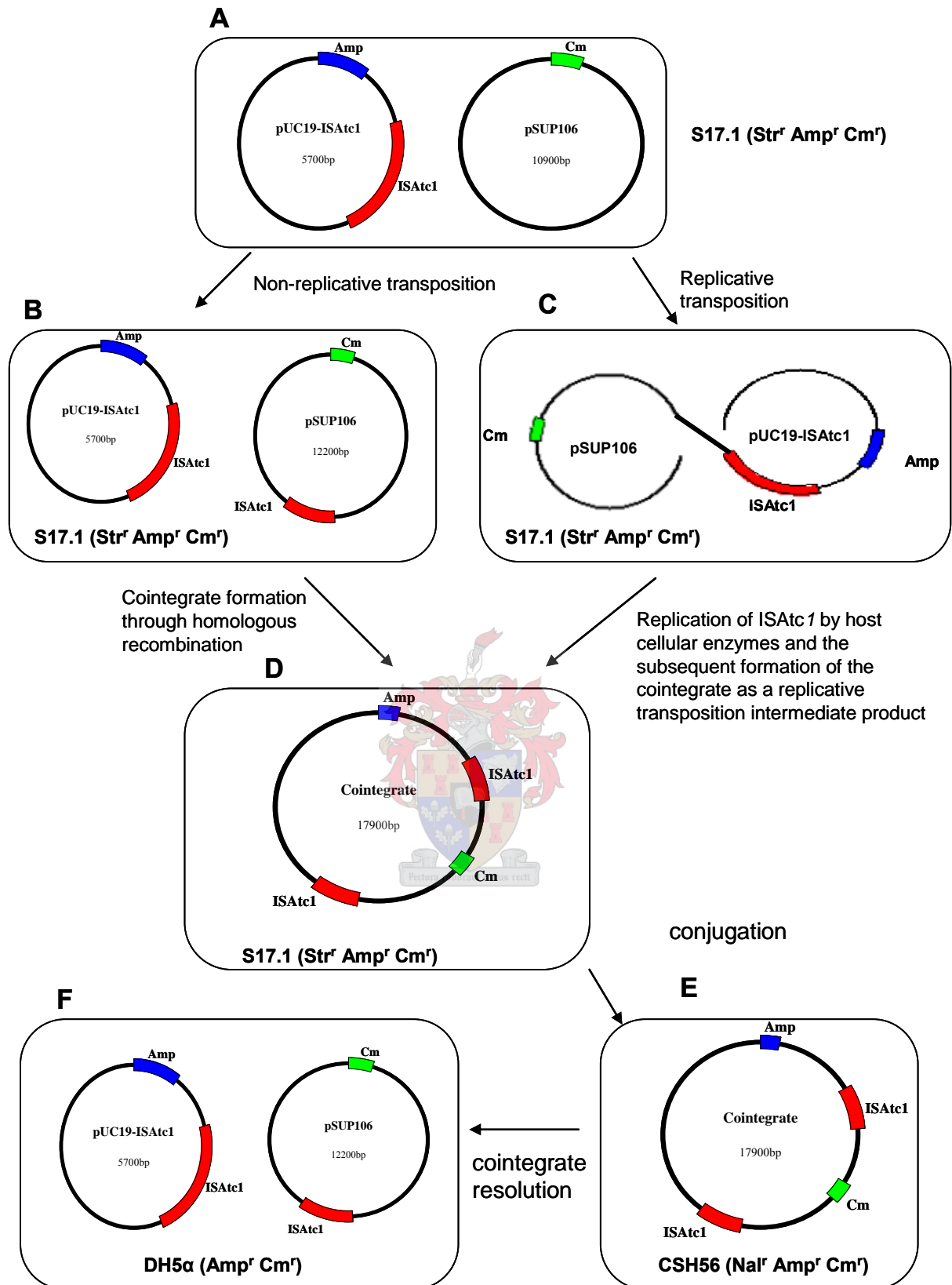


Figure 3.8: Schematic illustration of the proposed transposition of *ISAtc1* from pUC19-*ISAtc1* to pSUP106. Two pathways for cointegrate formation are shown, i.e. through homologous recombination after a nonreplicative transposition event (ABD) and during a replicative transposition event (ACD) as a by product of transposition. This is followed by the conjugal transfer of the cointegrate from the *E. coli* S17.1 donor strain to the *E. coli* CSH56 recipient strain. The cointegrate is subsequently resolved in the recipient strain through a homologous recombination process. Cointegrates were resolved in *E. coli* DH5α.

should exhibit 6 bands on an electrophoresis gel. See Fig 3.10 for an illustration of the possible cointegrate intermediates and their resolved intermediate structures. Our result does not meet this criteria for cointegrate formation between these two replicons. Instead, five bands of different sizes are seen on the gel. It is clear from the gel that the transconjugant plasmids exhibit two bands of similar size to that in pSUP106 and two extra bands are formed. The presence of the 300 bp band, which is not clearly visible on the electrophoresis gel but picked up by the Southern Hybridization (Fig 3.9C), indicates that pUC19-ISA t c1 is present in the transconjugants which is also confirmed by the Amp resistance displayed by the transconjugants. Further analysis of banding pattern of the transconjugants plasmids revealed that they were approximately 1,300 bp shorter than expected. ISA t c1 is 1,303 bp long and it is thought that the transconjugant contains only one copy of ISA t c1. This was confirmed by comparing *B*g/III restrictions of pUC19-ISA t c1, pSUP106 and the transconjugant plasmids. ISA t c1 has a unique *B*g/III site in contrast with pSUP106 and pUC19 which harbours no *B*g/III sites. *B*g/III restriction linearized the transconjugants which confirms the presence of one copy of ISA t c1.

We performed an experiment as to determine if the transconjugant plasmids or cointegrate-like structures (Fig 3.10 C and D) have the ability to resolve in *E. coli* DH5 α . This entailed the transformation of the transconjugant plasmids into *E. coli* DH5 α and then selecting first for Cm and then Amp resistant colonies. The Cm^r colonies that have no Amp resistance are evidence for the plasmid being resolved into the two individual plasmids, pUC19-ISA t c1 and pSUP106, and hence the Amp marker being lost. However, our results were not as expected, as only Amp sensitive colonies could be detected but no Cm sensitive colonies. *E*coRV restriction analyses of the Amp sensitive colonies did not display a banding pattern more or less similar to that for pSUP106 (results not shown). We expected that the Amp sensitive colonies should display at least two bands similar to that for pSUP106 with two different bands appearing if transposition occurred from pUC19-ISA t c1 to pSUP106.

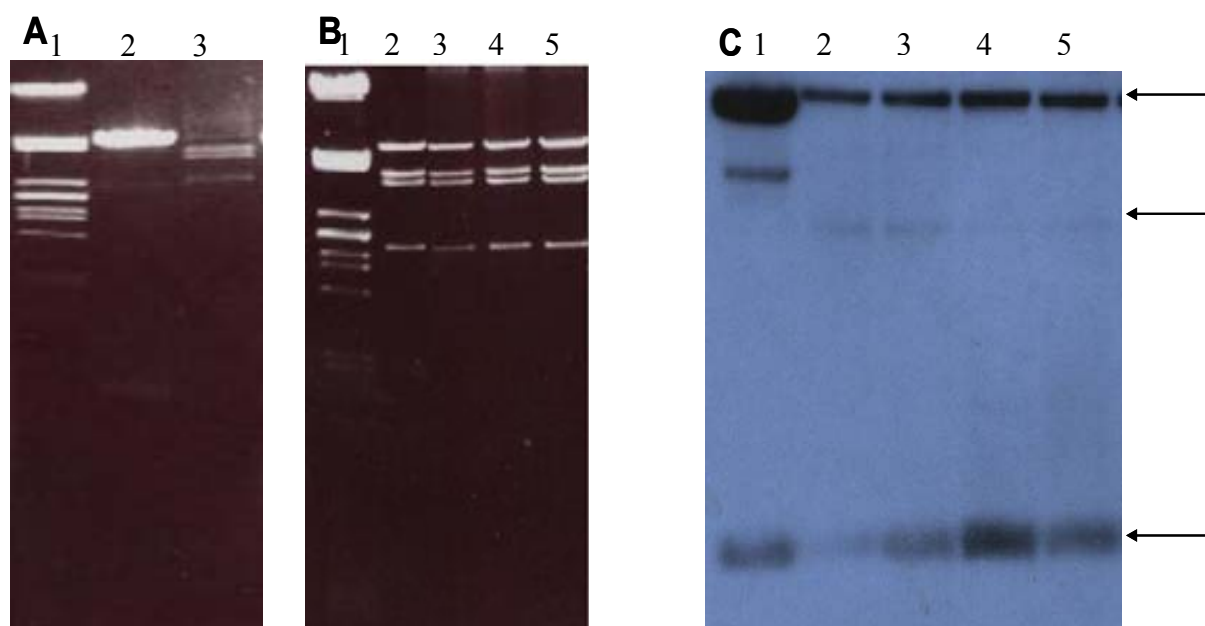


Figure 3.9: Results of the transposition of ISAtc1 in *E. coli*. **A:** Gel electrophoresis photo showing the *EcoRV* restriction banding patterns of pUC19-ISAtc1 (lane 2) and pSUP106 (lane 3). *Pst*I digested lambda DNA was used as marker DNA. **B:** Gel photo illustrating the banding patterns of *EcoRV* digested transconjugants obtained after the conjugation experiments (lanes 2-5). **C:** Southern hybridization analysis of the transconjugants. Transconjugants are shown in lanes 2-4 and pUC19-ISAtc1 included as a positive control is shown in lane 1. The arrows indicate the banding sizes of the transconjugants yielding positive signals.

An experiment similar to that carried out in *At. caldus* f (section 3.3.5), was done in *E. coli* DH5 α to determine whether ISAtc1 has the ability to transpose into the chromosome of *E. coli*. A strain of *E. coli* DH5 α harbouring pTC-F14, which contains ISAtc1, was passaged through several cycles of plating. Genomic DNA was extracted after various cycles of platings and was used in Southern Hybridization experiments using the labelled ISAtc1 as a probe. No signals consistent with the transposition of ISAtc1 into the chromosome of *E. coli* DH5 α were detected. The higher order organization of chromosomal DNA to that of plasmid DNA might explain why ISAtc1 does not transpose readily into the *E. coli* DH5 α chromosome.

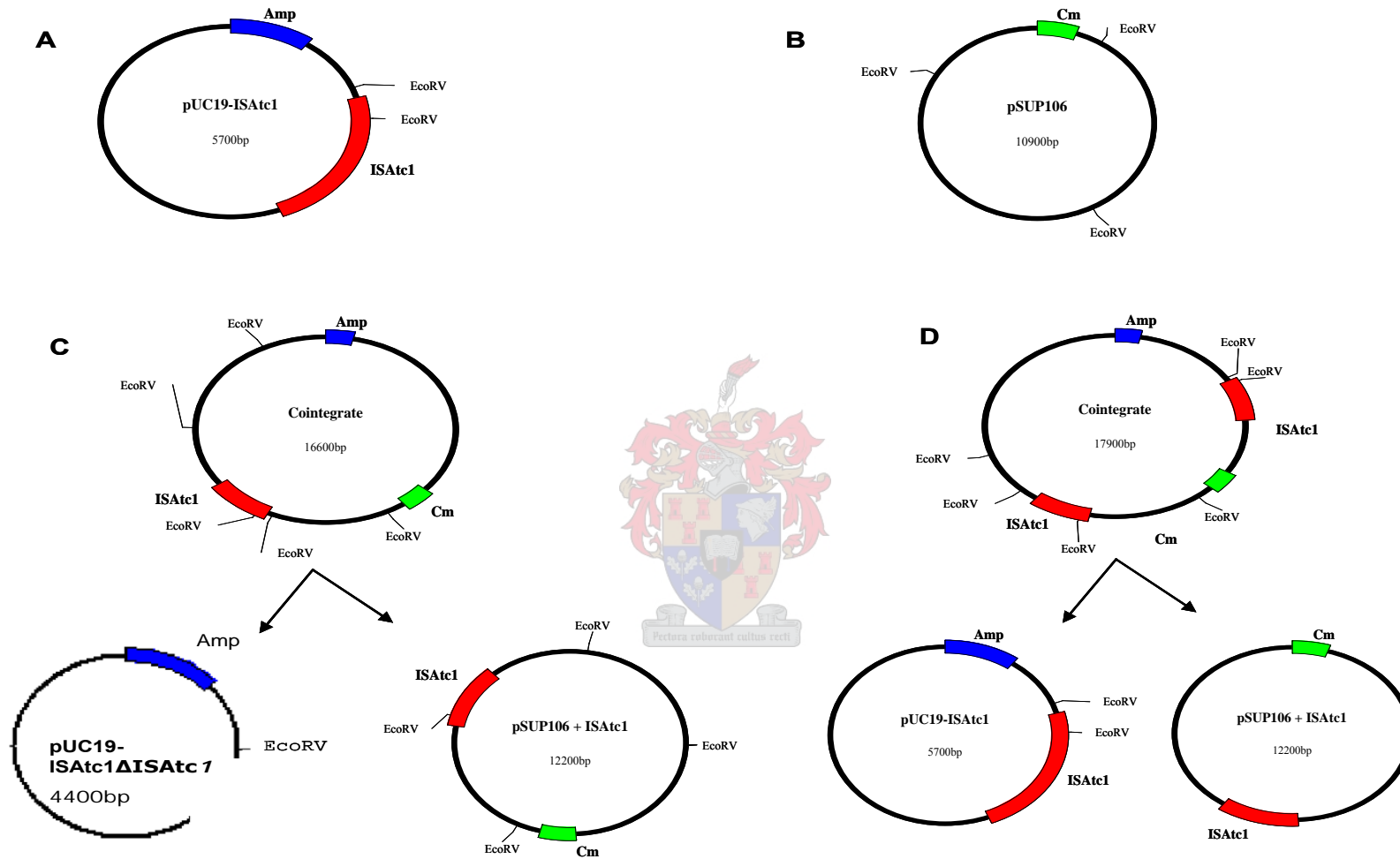


Figure 3.10: Schematic illustration of the donor and recipient replicons as well as the possible cointegrate intermediates and their resolved structures. **A.** pUC19-ISA_tc1. **B.** pSUP106. **C.** Cointegrate like structure containing only one copy of ISA_tc1. **D.** Cointegrate-like structure containing two copies of ISA_tc1.

3.4. Discussion

In this study an insertion element, *ISAtcI*, was detected in the pTC-F14 plasmid of *At. caldus*. A BLASTP analysis of both the GenBank-EMBL and the general IS Database has revealed that *ISAtcI* is a member of the ISL3 family and that it displays typical features of members of the family. It has a length of 1,303bp which is in the range of sizes described for other members of this family (1,300-1,550 bp) (Chandler and Mahillon, 2002). It has imperfectly conserved 26 bp terminal inverted repeats which share clear similarities with IRs of other members of the ISL3 family. The IRs of *ISAtcI* ends with the GG dinucleotide which is a characteristic of this family. No obvious direct target repeats could be detected upstream and downstream of *ISAtcI*. The absence of direct target repeats (DR) could be the result of inter- or intra-molecular recombination events between two *ISAtcI* elements, each with a different DR sequence (Chapter 1). Putative promoter regions were identified through visual inspection and by using web-based promoter identification software trained on *E. coli* σ^{70} promoter sites. A putative outwardly directed -35 promoter sites was identified within the right terminal inverted repeat. These outwardly directed promoters have been shown to be involved in the control of the expression of genes downstream to the insertion point, provided that an appropriately located -10 promoter site is present on the target DNA allowing for a hybrid promoter to be formed. *ISAtcI* has a G+C content of 56% compared to the 63% G+C content of *At. caldus*. This implies that the acquisition of *ISAtcI* is a recent event. The high extent of identity of *ISAtcI* to *ISAfeI* from *At. ferrooxidans*, and other IS elements found in a range of prokaryotes observed through sequence comparisons, indicate that these organisms may have plasmids or other horizontal gene transfer elements (onto which these IS elements can hitchhike) and can that move between their ecological niches. Through these processes these IS elements can become established in a wide variety of hosts and co-evolve with its respective host.

Analysis of the deduced amino acid sequence of the T_pase ORF of *ISAtcI* revealed that it is a basic protein which is consistent with its ability to bind DNA specifically during the transposition process. The multiple alignment of the deduced amino acid sequence of *ISAtcI* T_pase as well as its IRs with T_pases and IRs of *ISAfeI*, *ISAeI*, *ISBlo6*, *IST445* and *ISL3* revealed homologies that may possibly indicate a conserved function or a common transposition mechanism and a possible common evolutionary origin. Our data are also consistent with the T_pase being composed of an N-terminal domain containing the sequence-specific DNA binding determinants and the C-terminal region which contains most of the catalytic residues involved in catalysis. Through sequence analysis we detected a putative HTH

motif in the N-terminal region of *ISAtcI* which shows relatively weak homology to HTH motifs of other ISL3 family elements. No leucine zipper or zinc finger motifs could be detected. The HTH motif provides sequence-specific binding to the inverted repeats of these elements and the divergence in the nucleotide sequence of the IRs of these elements may explain the weak similarity in amino acid sequence of the HTH motifs of the IS elements compared. We have shown that the three acidic amino acids D161, D237 and E364 constitute the DDE motif of *ISAtcI* and are seen in all Tpsases compared except for *IST445*. This motif has been previously documented to be conserved in various integrases and recombinases. Based on this it has been proposed that these Tpsases, together with RNaseH and RuvC, forms part of a superfamily of phosphoryltransferases (reviewed in Chandler and Mahillon, 2002). Several additional highly conserved residues surrounding the DDE motif present on the majority of ISL3 family members are also present on *ISAtcI* which is consistent with observations made with recombinases, integrases and various IS elements.

The mobile character of IS elements allows it to be transferred from the chromosome of bacteria to resident plasmids of that bacterium. Transmissible plasmids in turn can transport the IS elements to bacteria of the same or different species through a horizontal transfer process which can then transpose to the chromosome of that bacterium. The association of IS elements with bacteriophages can also contribute to its distribution in a wide range of bacteria. Therefore, bacteria of different species may contain the same IS element or bacteria of the same species may contain variable numbers of copies of a particular IS element.

IS elements provide potential sites for the plasmids to integrate into the chromosome and increases the level of plasticity of the chromosome. Since pTC-F14 is a broad host range plasmid with a copy of *ISAtcI* this allows the potential for the integration of plasmid pTC-F14 into the chromosome of organisms harboring copies of *ISAtcI* or *ISAtcI*-like sequences. Accordingly, a number of integrated plasmids may excise from the chromosome carrying additional chromosomal DNA (prime plasmid formation) which can subsequently be transferred to other bacterial cells through conjugation (Merlin *et al*, 2001). Through this process chromosomal genes can become available in the horizontal gene pool and IS elements have been associated with the acquisition of accessory genes. This plasticity could be provided by other elements in the non-South African strains. These sequences may be purely parasitical or may have been become an integral part of the chromosome of these strains by performing a specific function. It has been previously reported that transposable elements play an intimate role in genome evolution (Kidwell and Lisch, 2001). It is therefore possible that the

transposition of IS*AtcI* and other IS elements present in the *At. caldus* strains may provide a driving force behind genome plasticity and variation. It would be interesting to discover whether the behavior of IS elements, e.g. IS*AtcI*, have contributed to the adaptation of *At. caldus* to the extreme low pH, inorganic biomining environments. The foreign strains that do not contain IS*AtcI*-likes sequences could possibly not have been in contact with IS*AtcI*-like elements since they are from geographically different regions. This does not imply that there are no IS elements present on the chromosomes of these strains.

We have shown that IS*AtcI* is functional within its native *At. caldus* host. It is however, not known what mechanism of transposition is employed by IS*AtcI* or whether the difference in banding patterns observed in the Southern hybridization experiment is caused by homologous recombination. The activity of IS*AtcI* within the chromosome of the host is not as extensive as seen for its close relative IS*AfeI* in *At. ferrooxidans*. This extent of the transposition is possibly controlled or induced by different factors within these bacteria, i.e. these IS elements might play slightly different roles and are controlled according to the physiological state of the bacterium and the surrounding environmental conditions. It has been proposed that the movement of IS*AfeI* is responsible for the phenomenon of phenotypic switching (phase variation) in *At. ferrooxidans* between wild-type state in which both ferrous iron and reduced sulphur compounds can be oxidized and a mutant state during which the ability to oxidize ferrous iron is lost but the ability to oxidize reduced sulphur is retained (Cabrejos *et al*, 1999). IS*AfeI* has been shown to insert within the *resB* gene which encodes for a putative cytochrome *c*-type biogenesis protein. It is proposed that this insertion results in the loss of activity of this *c*-type cytochrome which is thought to be required for ferrous iron but not sulphur oxidation. It is at present not known whether the mobility of IS*AtcI* in *At. caldus* has a similar role to that of IS*AfeI*, i.e. to regulate the expression of biochemical pathways by moving in and out of specific regions of the chromosome.

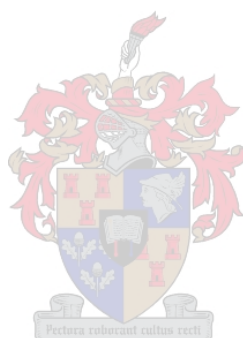
Two major modes of transposition exist for bacterial transposable elements. In the conservative or nonreplicative mode (cut-and-paste) mechanism the T_pase cuts both DNA strands at the terminal ends and the excised elements are transferred to target DNA. During replicative transposition only the 3' ends of the terminal ends are cleaved which is transferred to the target DNA. This leads to cointegrate formation where the donor and target molecules are fused and the element is duplicated as a result of DNA repair and replication by the host cellular enzymes. In this case the cointegrate-like structure is an intermediate structure (product) of the replicative transposition pathway. Subsequent recombination between the two copies of the

elements can resolve the cointegrate with the donor and target molecule each containing a copy of the element. Cointegrate structures can also be formed after a nonreplicative transposition event. Here a cointegrate structure can be formed through homologous recombination between a donor replicon, which did not undergo a cut-and-paste transposition process and thus containing an intact copy of the IS element, and a recipient replicon into which an IS element transposed into. The results of the experiments to determine the transposition ability of ISAtcI presented here is unique in the sense that a cointegrate structure must have been formed but has only one copy of ISAtcI (Fig 3.10C). This has been confirmed through *EcoRV* and *BglIII* restriction endonuclease analyses. The presence of the Amp marker in the transconjugants can only be explained if the nonmobilizable plasmid pUC19-ISAtc1 was fused to the mobilizable plasmid pSUP106 prior to conjugation. In the replicative transposition mode only cleavage of the 3' ends of the IS elements occurs which are transferred to the target molecule. However, at this stage the process can proceed in two ways. Replication and gap repair by the host cellular enzymes results in the formation of a cointegrate-like structure. Additional cleavage of the 5' end of the IS element would result in a cut-and-paste mechanism (Chapter 1). The latter case is a possible explanation for our results. It has been reported that the bacterial transposon Tn7 displays an interesting preference for plasmid DNA undergoing conjugal replication and transposes to plasmids undergoing conjugation (reviewed in Craig, 1997). Plasmid extractions from old cultures of *E. coli* S17.1 harboring both pSUP106 and pUC19-ISAtc1 were done and then re-transformed into *E. coli* DH5 α and selection was done on Amp Cm LA plates. Extracted plasmid DNA was restricted with *EcoRV* and compared to the individual *EcoRV* restriction patterns of these two plasmids as well as with the transconjugants obtained. The results showed that the banding patterns of the re-transformed DNA displayed an identical banding pattern to the individual plasmids, implying that no transposition event has occurred. This lead to the assumption that ISAtcI might display a preference for plasmid DNA undergoing conjugation. Based on the current results, it is difficult to ascertain as to why and how only one copy of ISAtcI is observed in the cointegrates (transconjugants). In this respect the result is unique, in contrast to cointegrate-like structures normally carrying two copies of an IS element and further work is needed to establish whether this is a new transposition mechanism.

Chapter 4: General Discussion and Perspectives

Table of Contents

4.1. General Introduction.....	122
4.2. Plasmid pTC-F14 in perspective.....	123



4.1. General Introduction

As result of a continuously changing environment caused by environmental or man-made stresses, the individuals of any bacterial population are forced to adapt to these changes to ensure their own survivability and proliferation. For any living organism a high fidelity of DNA replication is indispensable but point mutations generated during this process contribute to the adaptation of microbes to changing environmental conditions. It is widely documented that plasmids and other mobile elements such as transposons, insertion sequences and bacteriophages play a crucial role in the adaptation of microbes to hostile environments. Plasmids are tremendously diverse, self-replicating extrachromosomal mobile genetic elements. Plasmids are mosaic in structure encoding essential genes involved in plasmid replication, stable maintenance and transfer (conjugation) functions. These functions constitute what is known as the plasmid backbone. In addition to these essential genes plasmids also carry with them accessory genes (non essential functions) that confer a selective advantage to the host under certain conditions that would otherwise not be needed. These phenotypic traits conferred by plasmids are selected by the host's immediate environment and thus provides the host with a means to survive certain conditions of environmental stress. Plasmids are dependent upon hosts for their survival. Plasmids enhance host competitiveness and survivability in cases of selective pressure but may be dispensable in the absence of these environmental stress conditions. Plasmids impose an energetic drain on its hosts in terms of protein synthesis but are seen as desirable elements for host cells in situations of unfavorable environmental conditions. In contrast, plasmids have also been viewed as purely parasitic or selfish DNA molecules.

The types of phenotypic traits discovered on plasmids have been mentioned in Chapter 2. The mobile character of transmissible plasmids accounts for the plasmid-mediated dissemination of genetic material (phenotypic traits) between the same or different species. It is therefore that plasmids have been considered as the key to the evolution and adaptation of bacterial populations by being the major role players in the horizontal transfer of genetic material. As plasmids move between the same or different species new genetic traits are gathered from the gene pool of the greater bacterial population and are spread between members of this population. The accessory gene regions of plasmids are often associated with transposons and insertion sequences. These mobile elements contribute to moving DNA between bacteria often integrating them into the chromosome, or play a role in the movement of DNA from the chromosome to plasmid. Bacteriophages also play an important role in this respect. Much of the knowledge on plasmids comes from the intensively studied plasmids isolated from a clinical environment. Accessory genes found on these plasmids have been shown to code for

antibiotic resistance and pathogenicity determinants. Although plasmids from environmental samples have been shown to code for resistance to heavy metals, antibiotics and codes for degradative genes and bacteriocins, the knowledge on the diversity of plasmids from non-clinical environment remains limited (Smalla *et al*, 2000). This might be due to the fact few efficient cultivating techniques are available to culture these newly discovered species from other ecological niches. We study bacteria in the laboratory environments which are very unnatural environments, e.g. pure cultures and stable growth conditions in comparison to its natural continuously changing environment. These conditions may place environmental stresses on these microbes who respond by forming viable but nonculturable cells (Roszak and Colwell, 1987). The study of the accessory functions are of great importance to elucidate the understanding of plasmid spread and the ecological role played by plasmids. Our study focused the accessory gene region of a plasmid, pTC-F14, isolated from a relatively unexplored environment in terms of its biology and genetics.

4.2. Plasmid pTC-F14 in perspective

As mentioned before, pTC-F14 is a 14.2-kb broad host range mobilizable IncQ-like plasmid isolated from the acidophilic bacterium, *Acidithiobacillus caldus* f. *At. caldus* f is a moderately thermophilic, sulphur oxidizing bacterium which inhabits a highly specialized ecological niche, the low pH, largely inorganic mineral-rich environments. The sequencing of pTC-F14 has recently been completed and has been shown to consist of four distinct regions (Fig 4.1): a replicon region; a region between the *repB* and *repA* genes containing the *pasAB* genes which constitute the plasmid addiction stability (*pas*) system (post-segregational killing system); a mobilization region and an accessory gene region. The replication region, the *pas* system and mobilization regions has been studied in detail by researchers from this laboratory (Gardner *et al*, 2001; Deane and Rawlings, 2004 and van Zyl *et al*, 2003). The focus of this study was on the accessory gene region but the other regions are mentioned as to provide a more global view of the biological nature of pTC-F14.

Plasmid pTC-F14 was shown to have an IncQ-like replicon that was closely related to, but compatible with the 12.2-kb broad-host range plasmid, pTC-FC2 (Fig 4.1). This plasmid has been isolated from a different but related biomining bacterium, *Acidithiobacillus ferrooxidans*.

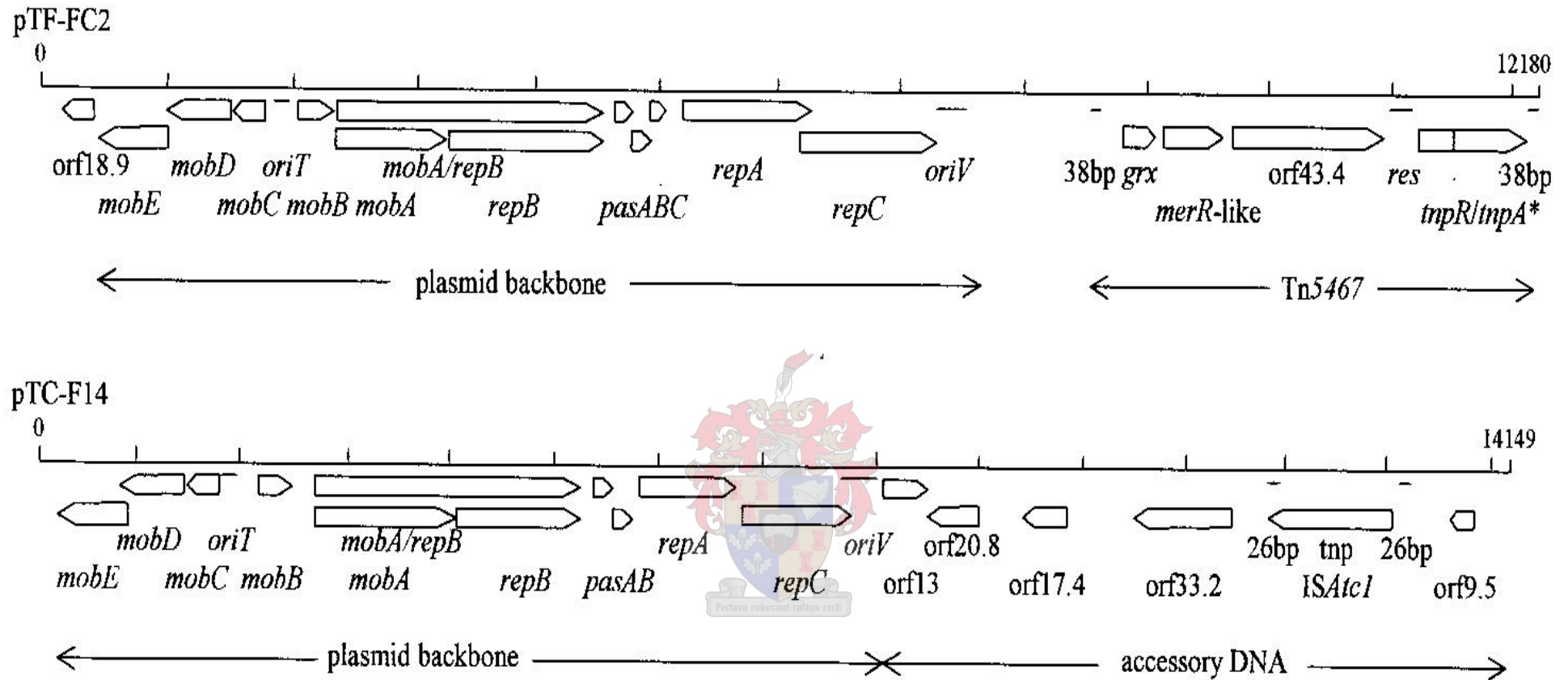


Figure 4.1: Genetic maps of pTC-FC2 and pTC-F14 comparing the replicon regions, the mobilization regions and the accessory gene regions.

Plasmids forming part of the *Escherichia coli* incompatibility group Q have a very broad host range and are capable of replication in a wide variety of both gram-negative and gram-positive bacteria (reviewed and referenced in Rawlings and Tietze, 2001). These plasmids are not self-transmissible but are they are highly mobilizable by IncP α and IncP β plasmids. Their mobile character as well as their ability to replicate within a wide range of hosts makes them highly promiscuous plasmids and thus have been acquired by microbes inhabiting very different environments. Examples of other well studied IncQ plasmids isolated from different environments include RSF1010, R1162 and R300B (Rawlings and Tietze, 2001).

The *mob* genes constituting the mobilization region of pTC-F14, like pTC-FC2, have been shown to be of the IncP type. pTC-FC2 and pTC-F14 inhabit the same environment and because of their promiscuous nature it is likely that these two plasmids will come in contact with each other. In spite of the close relationship between these plasmids, it was surprising that the *mob* genes of pTC-FC2 were more closely related to the *mob* genes of plasmid pRAS3.1 isolated from the neutrophilic Norwegian salmon pathogen, *Aeromonas salmonicida* (van Zyl *et al*, 2003). This finding emphasizes the promiscuous nature of the IncQ plasmid family.

Not much detail will be given on the *pas* system of pTC-F14. It consists of two genes, *pasA* (anti-toxin) and *pasB* (toxin) which enhance the plasmid stability through the post-segregational killing (PSK) of plasmid-free daughter cells. Protein-type PSK systems usually consist of two genes, one encoding a highly expressed but short-lived anti-toxin (*pasA*) and poorly expressed toxin with a longer half-life (Deane and Rawlings, 2004). Because of the nature of the toxin and anti-toxin proteins, daughter cells that fail to inherit the plasmid will be killed as amounts of the short-lived anti-toxin will be insufficient to neutralize the more stable toxin present in daughter cells. This will be returned to later. Plasmid pTC-FC2 has been shown to have a PSK stability system that consists of three genes, *pasA* (anti-toxin), *pasB* (toxin) and *pasC* (regulatory protein).

The aim of this study was to discover what accessory genes are carried on pTC-F14. As mentioned before our findings were rather disappointing in the sense that no genes have been detected that confer a phenotype that is known to provide a selective advantage to the host (Chapter 2). Biochemical, genetic and structural studies are usually done on individual proteins as to identify the possible functions of these proteins. Sequence similarity searches of the freely accessible nucleotide and protein sequences using BLAST programs has been applied in this study to extend our knowledge on the open reading frames (ORFs) detected in the accessory

gene region. Except for the resolvase gene and the insertion sequence, *ISA_{tcI}*, detected, no possible functions could be assigned to the other ORFs. It should be noted however that these ORFs, ORF13, ORF33.2 and ORF17.4, may encode functions that are advantageous to the host, but are at present unknown. There is currently a burst in the output of nucleotide sequences as a result of an increase in the number of genomic sequencing projects of various microbes that are being pursued. Biomining has been well established as an industrial fermentation process and in view of improving the technology of the biomining process its biotechnology will need attention. It is therefore likely that the homologues of the protein products of ORFs 13, 33.2 and 17.4 will be discovered in the near future. Based on the results of the similarity searches of ORFs 20.8, 17.4 and ORF33.2 it is evident that horizontal transfer events occurred between *At. caldus* and other soil bacteria. This view is strengthened by the observation of the high homology of *ISA_{tcI}* to insertion sequences isolated from other soil bacteria (Chapter 3). The interesting question is whether these very different ecological habitats are somehow microbiologically connected? Do the biomining bacteria share a gene pool with bacteria from other environments? Not enough studies have been done on the genetics of microbes from this environment to address this question. Because of the promiscuous nature of pTC-F14 and possible discovery of other IncQ-like plasmids from these biomining environments, it would not be surprising if these biomining microbes share a common gene pool or a subset of genes with other soil bacteria. The average G+C content of the accessory gene region of pTC-F14 is 50% in comparison to the reported 59.3% G+C content of the plasmid backbone (Rawlings and Tietze, 2001), which implies that the addition of the accessory gene region to the pTC-F14 is a recent event.

At present pTC-F14 remains a cryptic plasmid and the question of whether it is purely parasitic in nature or whether it does contain genes coding for local adaptations, remains unanswered. A plasmid may accumulate a set of genes that are beneficial under specific circumstances, and it may be that some of the plasmids that appear to be cryptic carry genes pertaining to a specific set of environmental conditions (Lilly *et al*, 2000). It is possible that one of these uncharacterized ORFs may encode for genes that relate to a certain set of environmental conditions, for which the phenotypic effect may be difficult to observe since we ultimately study bacteria and their genes in an unnatural laboratory environment. On the one hand, it has been argued that cryptic plasmids cannot be maintained unless they code for a selectively advantageous phenotype (Levin and Stewart, 1980; Caugant *et al*, 1981). On the other, whether cryptic or not, plasmid pTC-F14 contains a post-segregational killing system which is a tenacious parasitic trait. This ensures that the host cell contains a copy of the plasmid after cell

division. It could be this property rather than any beneficial effects that ensures pTC-F14 survival.

Future work will be to elucidate the function of the ORFs and possibly also the mechanism of the transposition of *ISAtc1*. One approach to finding the possible functions of these ORFs, in addition to regular database similarity searches or bioinformatic analyses, is to make individual mutations or knockouts of these ORFs and then selecting for a possible phenotype that may be coded for by the specific ORF. However, this task is difficult since these ORFs could contribute to or play a role in one of the hundreds of biochemical pathways present within microbes. For this purpose a well characterized host background is needed where the effect of the inactivation of the specific ORF, or the lack thereof, can be studied. In many cases proteins are expressed in a foreign host but this does not necessarily mean that the protein would be functional within the foreign host. This further complicates the aim of elucidating the possible phenotypic function of a putative protein whose sequence has just become available in the laboratory. In conclusion, this is a great task and requires much time and effort spend on research and it is therefore that so many hypothetical proteins are found in the current GenBank-EMBL database.



Appendix One: Annotated sequence of the accessory gene region

Shown below is the double stranded sequence of the accessory gene region of plasmid pTC-F14 isolated from *Acidithiobacillus caldus* strain f. pTC-F14#1 contained an insert of 5.772-kb and was thought to carry the accessory genes of the pTC-F14. After sequence analysis (DNA and Protein) of this region, it was revealed that an open reading starts 287-bp upstream from the starting point of the initial accessory gene region worked with. It is therefore that the sequence below is joined to 347-bp upstream from the *SphI* site (initial starting point of the accessory gene area in pTC-F14#1; indicated here as position 347). The sequence of this 5.772-kb region as part of the complete sequence of pTC-F14 was deposited in the NCBI database: accession number NC_004734. The protein translations of the six open reading frames encoded on this region are shown below the DNA sequence and the start and end positions of the ORFs are indicated Restriction endonuclease site are labeled and the corresponding sequence are shaded. Putative Ribosomal Binding Sites are indicated in bold italics and start and stop codons of the ORFs are indicated in bold. No further ORFs were detected when the BamHI site was joined to adjacent section of pTC-F14 and no meaningful ORFs can be detected when ISAtc1 is deleted.



1 TTAGACAAATAGATGTAAATTTACTATATTACATTTACAATGTAATTACAGGAGAGTGCCTCA
AACTGTTTATCTACATTTAAATGATATAATGTAATGTTACATTAATGTCTCTCACGGT

NcoI

61 **Start of ORF13,124aa; strong RBS, no meaningful homology**
TGGAAGCAGGACCGATCACACAACAAGACGTTGCCGCCGCCCTGACCAGCTTGGCGGCC
ACCTTCGTCCTGGCTAGTGTGTTGTTCTGCAACGGCGGGCGACTGGTCGAACCGCCGG
M E A G P I T Q Q D V A A A A D Q L G G

121 CCGCCAACACCAATGCCCGGAAAATCCGGGAGATTCTGGGCCGGGGGAGTCTCGCCACCA
GGCGGTTGTGGTTACGGCGCTTTTAGGCCCTCTAAGACCCGGCCCCCTCAGAGCGGTGGT
P A N T N A A K I R E I L G R G S L A T

181 TTCAACGCCATTTGCAGGCCCTTAGAGACGCTCAGAGGGCTCCAGAGCTACCAGAGGCCG
AAGTTGCGGTAAACGTCCGGGAATCTCTGCGAGTCTCCCGAGGTCTCGATGGTCTCCGGC
I Q R H L Q A L R D A Q R A P E L P E A

241 TCCAGACGATTCCAACGCCTCCTGAGGCCGTTTCTGAGGCTTCCAGAGGCATTTGGGCGG
AGGTCTGCTAAGGTTGCGGAGGACTCCGGCAAAGACTCCGAAGGTCTCCGTAACCCGCC
V Q T I P T P P E A V S E A S R G I W A

SphI

301 CGGCCTGGGCAATGGCGGAGCAACGCCACGGGAAAGTCTGGCGCATGCCCTGCTGGG
GCCGGACCCGTTACCGCCTCGTTGCGGTGCGCCTTTCAGACCCGCTACGGGGGACGACCC
A A W A M A E Q R H A E S L A H A P C W

361 TTATAACAGGTATTACTTTTTGCGTACTCTGGCAACGGATGCCAGTCTCTATCTCCCAATA
AATATTGTCCATAATGAAAACGCATGAGACCGTTGCTACGGTCAGAGATAGAGGGTTAT
V I T G I T F A Y S G N G C Q S L S P N

Stop of ORF13

Stop of invertase-like ORF,ORF20.8; 189aa

421 CAACCTGCTTTAGGGCGAAAATCCATGCTATGAACCATTCCC GCCACGCTTACGCTGCA
GTTGGGACGAAATCCCGCTTTTAGGTACGATACTTGGTAAGGGCGGTGCGAATGCGACGT
T T L L * * S G N G G R K R Q L

481 ACTCGGATAGAGGACGGTTTTACTGATCTTGAGGCGGGTGGCGGCCTCGGGACGCTGA
TGAGCGCTATCTCCTGCCAAAATGACTAGAACTCCGCCACCGCCGGAGCGCTGCGACT
E R Y L V T K S I K L R T A A E R V S L

541 GGCCCTGGGCGATCAACTGCTGGGCGCGTTGCTGCTTCTCTGGGGTCATGACGGGTTTGC
CCGGGACCCGCTAGTTGACGACCCGCGCAACGACGAAGAGACCCCAGTACTGCCCAAACG
G Q A I L Q Q A R Q Q K E P T M V P K R

PstI

StuI

601 GTCCCCCTTTGCGGCCACGGGCGACTGCAGCGGGCGAGGCCTGCCAAGGTGCGCTCCCGGA
CAGGGGAAACGCGGTGCCCGCTGACGTCGCGCTCCCGGACGGTTCCACGCGAGGGCCT
G G K R G R A V A A A L G A L T R E R I

661 TAATGTCACGCTCGAACTGTCCGAGCGCCCCAAAGAGATGGAAGGTGAGTCTCCCGGCTG
ATTACAGTGCAGCTTGACAGGCTCGCGGGTTTTCTTACCTTCCAGTCAGAGGGCCGAC
I D R E F Q G L A G F L H F T L R G A P

721 GGGTGGTGGTGTCAATGGATTGCGGTGAGGGACCGAAAACCGCACCCGCGTTGTTCCAGGT
CCCACCACAGTTACCTAAGCCAGTCCCTGGCTTTTGGCGTGGGCGCAACAAGTCCA
T T T D I S E T L S R F G C G R Q L E D

781 CGTTCAGGACCTCGATGAGGTGGGGGAGAGAACGGGCCAGGCGATCCAGTCGCCAGACGG
GCAAGTCTGGAGCTACTCCACCCCTCTCTTGCCCGGTCCGCTAGGTGAGCGGTCTGCC
N L V E I L H P L S R A L R D L R W V T

NcoI

841 TGAGCACGTCGCCAGGGCGGAGGAAGGCCATGGCCTGCACCAATCCGGTGGCTTCCGCCCT
ACTCGTGCAGCGGTCCCGCCTCCTTCCGGTACCGGACGTGGTTAGGCCACGCAAGGCGGA
L V D G P R L F A M A Q V L G T R E A K

901 TGGCCCCGCTGGCGACATCCTCAAAGGTGTGTTTCGCACCCGGCATGGTTCAGGGCATCCC
ACCGGGGCGACCGCTGTAGGAGTTTCCACACAAGCGTGGGCCGTACCAAGTCCCGTAGGG
A G S A V D E F T H E C G A H N L A D R

Start of invertase-like ORF,ORF20.8;189aa

EcoRV

NcoI

961 GCTGTAGGTCGAGGTTCTGGTCTCCGGTGGATATCCGCGCATAACCGATCTGTGCCATGG
CGACATCCAGCTCCAAGACCAGAGGCCACCTATAGGCGCGTATTGGCTAGACACGGTACC
Q L D L N Q D G T S I R A Y G I Q A M

1021 GTGTGGCCTGTGTGCGTCAAGTCGTCCACAATCATAGCGTCCGTCTATCCGTTGAAACA
CACAGCGGACACACGAGTTCAGCAGGGTGTAGTATCGCAGGCAGATAGGCAACTTTGT

1081 GGGGGTTGTGGGACGCTTTTGGGGGTGTCCGTCAGAGGATCGGGATCACCCCTTCGTCTGC
CCCCAACACCCTGCGAAAACCCCCACAGGCAGTCTCCTAGCCCTAGTGGGAAGCAGACG

1141 CAGAGCCGGAYTTTTCAGCCATCCGTCCCCATAGCGCCAGACCCGCCGATAGGTTTCGCGC
GTCTCGGCCTRAAAGTCGGTAGGCAGGGGTATCGCGGTCTGGGCGGCTATCCAAAGCGCG

1201 TCCCAGTCATCCATCCGGTTGTGGGCTTGACGGCGCAGCGCCAGACCGTGAGATCCAGG
AGGGTCAGTAGGTAGGCCAACACCCGAAGTCCCGGTTCGCGGGTCTGGCACTCTAGGTCC

1261 **Stop of ORF17.4**
CGAAAGGGCGGTTCGCGGGCTGAGGCGGAAGTAGTATCCCAT**CTA**TCCGGTCCAGATCCGGC
GCTTTCCCGCCAGCGCCCGACTCCGCTTCATCATAGGGTAGATAGCCAGGTCTAGGCCG
* R D L D P

1321 CGATAACTGCCGGCGGGGTACGGAAGGCAATGGCTAGGCGGTTCCAGGCGTTGATGGCC
GCTATTGACGGCCGCCCCCATGCCTTCCGTTACCGATCCGCCAAGGTCCGCAACTACCGG
R Y S G A P T R F A I A L R N W A N I A

1381 ACGACGGCCAAGGTCAGATCCACCAGTTCCTTCTCCTCGAAATAACGGCGGACGTTTTC
TGCTGCCGGTTCAGTCTAGGTGGTCAAGGAAGAGGAGCTTTATTGCCGCTGCAAAAGT
V V A L T L D V L E K E E F Y R R V N E

1441 TAGAGCGCATCCGAGACGCCTTCATCGATGCGGGTGATAGCATCGGTCCACGCCAGCGCC
ATCTCGCGTAGGCTCTGCGGAAGTAGCTACGCCACTATCGTAGCCAGGTGCGGTTCGCGG
Y L A D S V G E D I R T I A D T W A L A

1501 GCGCTTTCCCGCTCGCTGAAAAATGGCGCTTCCCGCCAAGCGGCGAGGGCATAAAGACGC
CGCGAAAGGGCGAGCGACTTTTTTACCGCGAAGGGCGGTTTCGCCGCTCCCGTATTTCTGCG
A S E R E S F F P A E R W A A L A Y L R

SmaI

1561 TGCTCCATCTCGCCCCGGGCGCGGGCTTCCCTTGCTGTGCATGTTCGATGCAAAAGGCGCAG
ACGAGGTAGAGCGGGCCCGCGCCCGAAGGAACGACACGTACAGCTACGTTTCCGCGTC
Q E M E G R A R A E K S H M D I C F A C

1621 CCGTTGAGGATGGATGCGCGCAGCTTACCAGTTCACGAGCGCTGGTTCCAAGTTGCTT
GGCAACTCCTACCTACGCGGTGCAAGTGGTCAAGGTTCGTCGCGACCAAGGTTCAACGAA
G N L I S A R L K V L E L L A P E L N S

NcoI

1681 CCATGGAGATAAGTCTCCAGATCCCGCAGGCTGCGGATGCCGTTGGGGGATACCAAGCCA
GGTACCTCTATTAGAGGTCTAGGGCGTCCGACGCCTACGGCAACCCCTATGGTTCCGT
G H L Y T E L D R L S R I G N P S V L G

SalI **Start ORF 17.4; 153aa**

1741 TAGTCGACGCGTTTTTCCATC**ATGGACTCCT**GTGGGAACCTTCTTGGAGCGGAAGGGGAT
ATCAGCTGCGCAAAAAGGTAGTACCT**GAGGA**CACCCTTGAAGAACCTCGCCTTCCCCTA
Y D V R K E M M

1801 CTGGGGCCGGATAACCAGATCCCGCCATCGGGACGCCGCCCGCCCAAGGGCTCAAACGAT
GACCCCGCCTATGGTCTAGGGCGGTAGCCCTGCGGCGGCGGGTTCCTCGAGTTTGCTA

1861 GCCTCTGGTCATGCCCTGGCGTCAGCAGGTGATGCCACGGTAAAAACCTGGGCGGTAAGG
CGGAGACCAGTACGGGACCGCAGTTCCTACTACGGTGCATTTTGGGACCCGCCATTC

1921 ACGTCAGCAATGAGAAAGCGATGACAGCGCCACGGGAGGGTGCTCGGCGCACAGGAGGAC
TGCAGTCGTTACTCTTTTCGCTACTGTGCGGGTGCCCTCCCACGAGCCGCGTGTCTCTCTG

StuI

1981 CGTCCTGCGCCTACGCGACCAAGGTGCTCCAAGGCCTCCAATCCTTTCTGGAAAGGCGCC
GCAGGACGCGGATGCGCTGGTTCCACGAGGTTCCGGAGGTTAGGAAAGACCTTTCCGCGG

2041 GTCTGCGCAAGCGACCATATCGGCAGATACTGCCAGGATAGAGCCAGAAGACTGATCTAC
CAGACGCGTTTCGCTGGTATAGCCGTCTATGACGGTCTATCTCGGTCTTCTGACTAGATG

2101 CGGCACAGCGCGACCATTCTAATCACGGATGTTTCCAGAGGTTCCAAAGGAATATCATGAGC
GCCGTGTGCGCTGGTAAGATTAGTGCCTACAAGTCTCCAAGGTTTCCTTATAGTACTCG

2161 GCCACATATCACGATCCACGATACGCTAGGCAGGATGGCTAGACAGGCCCGCTCCGCAGA
CGGTGTATAGTGTAGGTGCTATGCGATCCGTCTACCGATCTGTCCGGCGGAGGCGTCT

HindIII

2221 CGGTATCTGATCCTGAGCGAGCGTTGCCCATAGTAAGCTTGCGGGTATCCTCTAACTGA
GCCATAGACTAGGACTCGCTCGCAACGGGTATCATTTCGAAACGCCCCATAGGAGATTGACT

2281 AGAACGACGGCTTGAGACTGGAAGCAGACGTTTCGGTTGGTGCGCCAAGAAAGGCGCATCG
TCTTGCTGCCGAACCTGTGACCTTCGTCTGCAAGCCAACCACGCGGTTCTTTCCGCGTAGC

2341 AATCTCCTGCCAGTCCAAGGTTAGTCAGCATTGGGCATGATCTGTAAGCTCGCTTGCTAT
TTAGAGGACGGTCAGGTTCCAATCAGTCGTAACCCGTAAGTACTAGACATTCGAGCGAACGATA

2401 CACACCTCCTAATAATCAACTATGATTGACGCTGGCGGTACAATCCGATGAGGAATGGCA
GTGTGGAGGATTATTAGTTGATACTAACTGCGACCCCATGTTAGGCTACTCCTTACCGT

EcoRI

Stop of aminotransferase-like ORF33.2; 286aa

2461 ACGTGAACAAGCAAGAATTCTATCCTGCTTCGTTAAGGGCCTCTAAACATAGCTTCTCA
TGCACCTTGTTCGTTCTTAAGGATAGGACGAAGCAATTCCCGGAGATTTGTATCGAAGAGT

* G A E N L A E L C L K E

2521 TATCCTAGTGGTGTCAATTCAGGTTCTTCTATAAACTGTGTAATTTCAACCTTATGTTTT
ATAGGATCACCACAGTTAAGTCCAAGAAGATATTTGACACATTAAGTTGGAATACAAAA
Y G L P T L E P E E I F Q T I E V K H K

2581 TTTATTTTCATCTGCCAATACATTTTCAAGCATATCGTACAGCATATTAAGACTTAGTTGT
AAATAAAGTAGACGGTTATGTAAGGTTTCGTATAGCATGTCGTATAATTTCTGAATCAACA
K I E D A L V N E L M D Y L M N L S L Q

2641 TTAATCCCGAATAGAACGATAACCGAAATGAATTTTGAAAATATCTCATTTTCAGCAAGA
AATTAGGGTCTTATCTTGCTATGGCTTTACTTAAAACCTTTTATAGAGTAAAAGTCGTTCT
K I G S Y F S V S I F K S F I E N E A L

2701 AGCACTTTTATCTTATCCACACAATCATCTTTTATACTATTTATGTTATCTTTGCTGTCT
TCGTGAAAATAGAATAGGTGTGTTAGTAGAAAATATGATAAAATACAATAGAAAACGACAGA
L V K I K D V C D D K I S N I N D K S D

2761 GGAAATGCTCCCTTAGCATATTCCTTAAAATCCAATGAGCTTGTATTTAGGAACAACGAT
CCTTTACGAGGGAATCGTATAAGGAATTTTAGGTTACTCGAACATAAATCCTTGTGCTA
P F A G K A Y E K F D L S S T N L F L S

2821 TTTATTTCTTTTATTGTGTTTTCTATATCCTCTCCGAATTTATTGTATATCACATTTTCA
AAATAAAGAAAATAACACAAAAGATATAGGAGAGGCTTAAATAACATATAGTGTAAAAGT
K I E K I T N E I D E G F K N Y I V N E

2881 AAAAAATATTCCTTCTCTTTTAAACCATTCTGACCGTCAAGAACATCATATTTGACAACC
TTTTTTATAAGGAAGAGAAAATTGGTAAGGACTGGCAGTTCTTGTAGTATAAACTGTTGG
F F Y E K E K L W E Q G D L V D Y K V V

2941 CTTACCCCTCGAAAAATCTTTCTGACAGCTCGCTGTAATGTGTCTTGTATGATATCTGAA
GAATGGGGGAGCTTTTGTAGAAAGACTGTGCGAGCATTACACAGAACTACTATAGACTT
R G V E F F R E S L E S Y H T K I I D S

EcoRV

3001 ATTTTTTGTGTTGACTCGTTATTGTTTATGGTGTATTCTTTAGTCAGTTTTGCAACAGAA
TAAAAACAAACATGAGCAATAACAAATACCACATAAGAAAATCAGTCAAAACGTTGTCTT
I K Q K Y E N N N I T Y E K T L K A V S

3061 AACGCCATAGATGTGATTATACTCTTTATTTGTTCTTGTCTCAAAAATAAGCATTTGAATCA
TTGCGGTATCTACTAATATGAGAAAATAACAAGAACGAGTTTTATTTCGTAAACTTAGT
F A M S T I I S K I Q E Q E F Y A N S D

3121 TTATCTTCATCCTCTATATCAACATCGCCATAGTCAGTTATGTCATCCTCGTCATCACCG
AATAGAAGTAGGAGATATAGTTGTAGCGGTATCAGTCAATACAGTAGGAGCAGTAGTGGC
N D E D E I D V D G Y D T I D D E D D G

DraI

3181 TCTTCAAAAAAGAACCTTTAAATCGCCATCCTTTAGTTGTTGAAGGGCTATTATTATATCG
AGAAGTTTTTCTTGAATTTAGCGGTAGGAAATCAACAACCTTCCCGATAATAATATAGC
D E F F F K L D G D K L Q Q L A I I I D

3241 TCAGATGAGCATGTCGTGCTTTCCTTGGCGTTATCATACTGCGTCAAGTCTTTCTCTAGT
AGTCTACTCGTACAGCACGAAAGGAACCGCAATAGTATGACGCAGTTCAGAAAGAGATCA
D S S C T T S E K A N D Y Q T L D K E L

Start of aminotransferase-like ORF33.2;286aa

3301 GTTGGAGTAATACGTGTTGCCTTTTTTTTCCAGAAAAACATATAAAAAGGCTCCTAGATTT
CAACCTCATTATGCACAACGGAAAAAAGGTCTTTTTTGTATATTTTCCGAGGATCTAAA
T P T I R T A K K K W F F M

3361 TTTATTGAGAGCGATAACCAGACAAAATGCATCCTTTGGTCCGGTTCGGCCGGGGACCGG
AAATAACTCTCGCTATGGTCTGTTTTACGTAGGAAACCAGGCCAAGGCCGCCCTTGCC

3421 AGCAGAACCCATTGATTCCATGAGAGCCGAGAGCCGAGAGCCGAGAAAGGCCGGCAAACATCTGTCC
TCGTCTTGGGTAACCTAAGGTACTCTCGGCTCTCGGCGTCTTCCGGCCGTTTGTAGACAGG

EcoRI

3481 TTACGGTGAACATGGTCATTTCCACACCGAATCCCCGAATTCACCGTATAGAAGACCTGA
AATGCCACTTGTACCAGTAAAGGTGTGGCTTAGGGGCTTAAGTGGCATATCTTCTGGACT

3541 TGGCACCTAGGATATGAGAGCTATGTTAGTGCCTTAACGAAGCAGGATGGATTCTGAAC
ACCGTGGGATCCTATACTCTCGATAACAATCACGGAATTGCTTCGTCTACCTAAGACTTG

3601 TAACAAAATCGGACCGGCGACCAAGGTGCTGATATAATTGCAAAAAAGAATGATAATTCT
ATTGTTTTAGCCTGGCCGCTGGTTCCACGACTATATTAACGTTTTTCTTACTATTAAGA

3661 TTGGTGGTGCAGTGCAAATATTAATCCGGTGGGCGTTGAACCGAAAAGCAGTCCAGGAAGT
AACCACCACGTACGTTTATAATGAGGCCACCCGCAACTTGGCTTTCGTTCAGGTCCCTTCA

3721 TATCGCTGCGCGTACATATTATTGTGCTGATTATGCCACCGTTGTATCAAACACATACTT
ATAGCGACGCGCATGTATAATAACACGACTAATACGGTGGCAACATAGTTTGTGTATGAA

3781 CACACCTGGCGCGAAAGATTTAGCAAAAAGAGGCGATGTTTTACTATTGGACGAAAACAA
GTGTGGACCGCGCTTTCTAAATCGTTTTTCTCCGCTACAAAATGATAACCTGCTTTTTGTT

DraI

Direct repeat of IS element

3841 GTTGAAGAATATTGACTTTTTTTTAAAAAACAGTTAGGCTCTTCGACGTTTCATGTGGGT
CAACTTCTTATAACTGAAAAAAATTTTTTGTCAATCCGAGAAGCTGCAAAGTACACCCA
* T P

Stop of Irase ORF of ISAtc1

NcoI

3901 AGCCTCAGATCCAGCTTGCCCAGGATGAGGTAAGCCATGGCGATGAAGTTCTTGTGGGTG
TCGGAGTCTAGGTCGAACGGGTCTACTCCATTCCGGTACCGCTACTTCAAGAACACCCAC
L R L D L K G L I L Y A M A I F N K H T

3961 CGGTAGCCACGGGCTTTTCGCTTGGCGGACTGGAGGAGGCTGTTGAAGCCCTCCAGGATG
GCCATCGGTGCCCAGAAAGCGGAACCGCCTGACCTCCTCCGACAACTTCGGGAGGTCTTAC
R Y G R A K A K A S Q L L S N F G E L L

4021 CCATTGCTGACTTGGCTCTCGAACAGCGGAGGACCCCGTCCCAATGGTTCATGAGGGTG
GGTAACGACTGAACCGAGAGCTTGGTCGCCTCCTGGGGCAGGGTTACCAAGTACTCCAC
G N S V Q S E F W R L V G D W H N M L D

4081 TAGGCGACCCTGACGATAGGCGGCAGATCGCTGGTTCTGGCGTTTTCCAACCAGGCTTTC
ATCCGCTGGGACTGCTATCCGCCGTCTAGCGACCAAGACCGCAAAAAGGTTGGTCCGAAAAG
Y A V R V I P P L D S T R A N E L W A K

4141 AAGAGGGTAGCCCCCTGGTGGCGATTCTTGATCGTGAAGATGTCCTGAAAGGTCAGGCGG
TTCTCCCATCGGGGACCACCGCTAAGAAGTACACTTCTACAGGACTTTCAGTCCGCC
L L D A G Q H R N K I T F I D Q F T L R

StuI
PstI

4201 AACTGGTAGGCCTGCGCCGTCTTGAGGTTCTGGTCTTTGAGCAATTCCTGCAGCTTTTCT
 TTGACCATCCGGACGCGGCAGAACTCCAAGACCAGAAACTCGTTAAGGACGTCGAAAAAGA
 F Q Y A Q A T K L N Q D K L L E Q L K E

4261 TTCTGCTTCATTTTGGAGGTTGCAATCGTTCTTGAGCCAAAGCCAGCGGGTCTTTTGGAGA
 AAGACGAAGTAAAACCTCCAACGTTAGCAAGAAGCTCGGTTTCGGTCGCCAGAAAACTCT
 K Q K M K L N C D N K L W L W R T K K L

4321 TTTGGCTGGGTGAGGACTTCCCCCTTGCGCACGTCGTCTACGGCCTCGTTGACGAGCTTC
 AAACCGACCCACTCCTGAAGGGGGAACGCGTGCAGCAGATGCCGGAGCAACTGCTCGAAG
 N P Q T L V E G K R V D D V A E N V L K

4381 ATGAGGTGAAAACGATCGAAAAGTGATCTCCGCATTGGGCAGGTGCTCGGCAGCCCCTTTC
 TACTCCACTTTTGGTAGCTTTTCACTAGAGGCGTAACCCGTCCACGAGCCGTCGGGGAAAAG
 M L H F R D F T I E A N P L H E A A G K

NcoI

4441 TGGAAGGCCGGCGAGAGGTCCATGCTCACATCGGTGATCGCTTCCGCGCTACCAACCATGG
 ACCTTCCGGCCGCTCTCCAGGTACGAGTGTAGCCACTAGCGAAGGCGCGATGGTGGTACC
 Q F A P S L D N S V D T I A E A S G G H

BglIII

4501 GCCTGAAGATCTCCGGAGAATTTCTCAAAGGTCTTGGCATCCTTGCCGGGAGTAGCGAAC
 CGGACTTCTAGAGGCCTCTTAAAGAGTTTCCAGAACCGTAGGAACGGCCCTCATCGCTTG
 A Q L D G S F K E F T K A D K G P T A F

4561 AAGAGTCGCCGGGCATTCAGATCCACGAAGAGCGTGATGTAGTCATGTCCATGCCGACTG
 TTCTCAGCGGCCGTAAGTCTAGGTGCTTCTCGCACTACATCAGTACAGGTACGGCTGAC
 L L R R A N L D V F L T I Y D H G H R S

4621 CTGGTTTTATCGACGCCGACGGCATGGACATTGGCCATATCCACCGCAGCAGAGCTTCG
 GACCAAAGTAGCTGCGGCTGCCGTACCTGTAACCGGTATAGGTGGCGTCGTGCTCGAAGC
 S T E D V G V A H V N A M D V A A R A E

4681 GGCACATAATGGTCAATCACTCGCCACAGGAGCTTGTTCGGTCTCACCGACCATGCGAGCT
 CCGTGTATTACCAGTTAGTGAGCGGTGTCTCGAACAGCCAGAGTGGCTGGTACGCTCGA
 P V Y H D I V R W L L K D T E G V M R A

4741 GCCGTCAATACCGGCATCTCCCGCACAGGGTCATGATCAGCGCTTCAAAGAGCAGGGTG
 CGGCAGTTATGGCCGTAGAGGGCGTGGTCCCAGTACTAGTCGCGAAGTTTCTCGTCCCAC
 A T L V P M E R V L T M I L A E F L L T

EcoRV

4801 AAACGCGAGCCTTCCCGCGCCCAGGGAACAGATATCTGATGCACTCCATGCTCCTGGCAC
 TTTGCGCTCGGAAGGGCGCGGGTCCCTTGTCTATAGACTACGTGAGGTACGAGGACCGTG
 F R S G E R A W P V S I Q H V G H E Q C

4861 TTCACACGAGGTACACGGGCATGGAGATAGGCTTTCATGCTGGAAGAAATCCATGTGCCGC
 AAGTGTGCTCCATGTGCCCGTACCTCTATCCGAAGTACGACCTTCTTTAGGTACACGGCC
 K V R P V R A H L Y A E H Q F F D M H R

4921 CAGGTATGGTCACGGGTATCATGTACCGGACACTCCTCACCACAGACGGAGCAAGCAAAG
GTCCATACCAGTGCCCATAGTACATGGCCTGTGAGGAGTGGTGTCTGCCTCGTTTCGTTTC
W T H D R T D H V P C E E G C V S C A F

4981 CGACTACCTTTGGGAAAGTTGATGTGCAGATCCAGGCGTTTCTCCTCCACCGTGAAAGTC
GCTGATGGAAACCCTTTCAACTACACGTCTAGGTCCGCAAAGAGGAGGTGGCACTTTTCAG
R S G K P F N I H L D L R K E E V T F T

5041 ACATGATCCACCAACCACGGCGGAACCAATCCTAACGCGAGAGAAAAACAGCTCTTCAGGG
TGTACTAGGTGGTTGGTGCCGCCTTGGTTAGGATTGCGCTCTCTTTTGTGAGAAAGTCCC
V H D V L W P P V L G L A L S F L E E P

Start of Tphase Orf of ISAtc1

Direct Repeat of ISAtc1

5101 ACCATCAGCTAACTCCTATTTCAGGGCACGACCGCTCCACAGCATTCTACCCCTACCCAC
TGGTAGTCGATTGAGGATAAGTCCCCTGCTGGCGAGGTGTCGTAAGATGGGGGATGGGTG
V M

EcoRV

5161 TCAATCTGACGAAGAGCCATAAGGAAGGATCTCGATATCATAAGTTTCCAGTGATGTGAA
AGTTAGACTGCTTCTCGGTATTCTTCTAGAGCTATAGTATTCAAAGGTCCTACTACACTT

5221 TCGTTCTTTGGTGCCATCCAACCTCCCATAGAATCTGGCTGTCTCTAAGCGCACGTAATCC
AGCAAGAAACCACGGTAGGTTGAGGGTATCTTAGACCGACAGAGATTTCGCGTGCATTAGG

SphI

5281 ATACAGCCAGAAATCACCGGGCATGTGCTCGGGTCTAACCTAAACTGCATGCTCCCGGC
TATGTCGGTCTTTAGTGGCCCGTACAGCGAGCCAGATTGGATTTGACGTACGAGGGCCG

5341 ATGGATGCTGAAGACAATCTGATGCCTTTCCCATCTAGTGGATATGGTTTACTATFACT
TACCTACGACTTCTGTTAGACTACGGAAAGGGGTAGATCACCTATAACCAATGATAATGA

5401 TCTGTGCTCATCGAAACCATCTTTCTTTTTGGACCAGTATAAGGTTGTGTTTGGCTCAA
AGACACGAGTAGCTTTGGTAGAAAGAAAACCTGGTCATATTCCAACACAAACGCGAGTT

5461 TGTCCTATTATTTGTAATCACTGCGGGAGCAGTGATTAGGCTGGCCTTTTCTGCTGTAC
ACAGGATAATAAACATTAGTGACGCCCTCGTCACTAATCCGACCGGAAAAGGACGACATG

Stop of 86aa ORF

HindIII

5521 CGCCTCCAATATAAACTGGTAGACGGACCCATCTTCAGAACAGAAGATGTTCTGGACTAA
GCGGAGGTTATATTTGACCATCTGCCTGGGTAGAAGTCTTGTCTTCTACAAGACCTGATT
* F L L H E P S F

5581 AGCTTCTGGACAAGGTAAAAAAGCGAAATGTGCGTCTGATCCAAGCAAGTCCTTTAGGG
TCGAAGACCTGTTCCATTTTTTCGCTTTACACGCAGACTAGGTTTCGTTTCAGGAAATCCC
S R S L T F F A F H A D S G L L D K L P

5641 GTCGGGTAGCGGAAAGTTCTCCATCCAAATAGTTCAAGGATGACGCCGATATGCCCAACA
CAGCCCATCGCCTTTCAAGAGGTAGGTTTATCAAGTTCTACTGCGGCTATACGGGTTGT
R T A S L E G D L Y N L S S A S I G L M

5701 TTAGGGATGAAATCAAGAATCTACCTTCAGCTATTAGTAGGTCTTGCAGGAAGCTGCAAC
AATCCCTACTTTAGTTCTTAGATGGAAGTCGATAATCATCCAGAACGTCCTTCGACGTTG
L S S I L F R G E A I L L D Q L F S C C

Start of 86aa ORF, no homology

5761 AATCTTCCCGGTTCCCGAAGGTGCTCCAGAATGTCGGCGCAGACGATATAGTCATAACGC
TTAGAAGGGCCAAGGGCTTCCACGAGGTCTTACAGCCGCGTCTGCTATATCAGTATTGCG
D E R N G F T S W F T P A S S I T M

5821 AGCTTCTGAAGATTGCGCGGAAGATGCTCCCGCTCCAGATCCCACTCCATAAGTGACGA
TCGAAGACTTCTAACGCGCCTTCTACGAGGGCGAGGTCTAGGGTGAGGGTATTCAGTCTGCT

5881 TAGAAGGGTCGTGCCGCTTCTAGTTCAATGTGATTGGCATCGATAACCATCCACCACGCAG
ATCTTCCCAGCACGGCGAAGATCAAGTTACACTAACCCTAGCTATGGTAGGTGGTGCCGTC

5941 TGCAGACTTTCACTCAGGTACTTCCCCAACGTGCCTACACCACAACCAACGTCCAGGACA
ACGTCTGAAAGTGAGTCCATGAAGGGGTTGCACGGATGTGGTGTGGTTGCAGGTCTCTGT

6001 AGGCTGCCGGGCTGAATGCGTCGGGCAATCTTTGCAAGACTATCTTGGCCCTGAGGATCT
TCCGACGGCCCCGACTTACGCAGCCCCTTAGAAACGTTCTGATAGAACCGGGACTCCTAGA

6061 ATCGGGGACGATTATAAACATGCGGGGTATGTGGGACTGATTCATACGGGATCC
TAGCCCCCTGCTAATATTTGTACGCCCCATACACCCTGACTAAGTATGCCCTAGG

*Bam*HI



Literature cited

1. **Adzuma K., and K. Mizuuchi.** 1988. Target immunity of Mu transposition reflects a differential distribution of Mu B protein. *Cell* **53**:257-66.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Meyers, and D. J. Lipman.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
3. **Altshul, S. F., T. L. Madden, A. A. Schaffer, J., Z. Zhang, W. Milller, and D. J. Lipman.** 1997. Gapped BLAST and PSI BLAST: a new generation or protein database search programs. *Nucleic. Acids. Res.* **25**:3389-3402.
4. **Andrake, M. D., and A. M. Skalka.** 1996. Retroviral integrase, putting the pieces together. *J. Biol. Chem.* **271**:19633-19636.
5. **Arciszewska, L. K., D. Drake, and N. L. Craig.** 1989. *cis*-Acting sequences in transposition and transposition immunity. *J. Mol. Biol.* **207**:35-52.
6. **Bainton, R. J., K. M. Kubo, J-N. Feng, and N. L. Craig.** 1993. Tn7 transposition: target DNA recognition is mediated by multiple Tn7-encoded proteins in a purified in vitro system. *Cell* **72**:931-43.
7. **Bairoch, A., P. Bucher, and K. Hofmann.** 1991. PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res.* **25**:217-221.
8. **Bairoch, A., P. Bucher, and K. Hofmann.** 1997. The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**:217-221.
9. **Baker, T. A., and K. Mizuuchi.** 1992. DNA-promoted assembly of the active tetramer of the Mu transposase. *Genes Dev.* **6**:221-231.
10. **Beall, E. L., and D. C. Rio.** 1997. *Drosophila* P element transposase is a novel site-specific endonuclease. *Genes. Dev.* **11**:2137-2151.
11. **Beck, C. F., H. Moyed, and J. L. Ingraham.** 1980. The tetracycline resistance transposon Tn10 inhibits translocation of Tn10. *Mol. Gen. Genet.* **179**:453-455.
12. **Bender, J., and N. Kleckner.** 1992. IS10 transposome mutations that specifically alter target site recognition. *EMBO J.* **11**:741-750.
13. **Bennet, J. C., and H. Tributsch.** 1978. Bacterial leaching patterns on pyrite crystal surfaces. *J. Bacteriol.* **134**:310-317.
14. **Bennett, P. M.** 2004. Genome plasticity: insertion sequence elements, transposons and integrons, and DNA rearrangments. *Methods. Mol. Biol.* **266**:71-114.
15. **Berg, D. E.** 1983. Structural requirement for IS50-mediated gene transcription. *Proc. Natl. Acad. Sci. USA* **80**:792-796.

16. **Berger, B., and D. Haas.** 2001. Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS21. *Cell. Mol. Life Sci.* **58**:403-419.
17. **Berns, K., and R. Linden.** 1995. The cryptic life style of adeno-associated virus. *Bioessays* **17**:237-245.
18. **Bhasin, A., I. Y. Goryshin, and W. S. Reznikoff.** 1999. Hairpin formation in Tn5 transposition. *J. Biol. Chem.* **274**:37021-37029.
19. **Blake, R. C. H, G. T. Howard, and S. McGinness.** 1994. Enhanced yields of iron-oxidizing bacteria by in situ electrochemical reduction of soluble iron in the growth medium. *Appl. Environ. Microbiol.* **60**:2704-2710.
20. **Boeke, J. D.** 2002. Putting mobile DNA to work: the toolbox, p. 24-37. *In* N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz (ed.), *Mobile DNA II*. American Society for Microbiology, Washington D.C.
21. **Bolland, S., and N. Kleckner.** 1995. The two single-strand cleavages at each end of Tn10 occur in a specific order during transposition. *Proc. Natl. Acad. Sci. USA.* **92**:7814-7818.
22. **Bolland, S., and N. Kleckner.** 1996. The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site. *Cell* **84**:223-233.
23. **Braam, L. A, and W. S. Reznikoff.** 1998. Functional characterization of the Tn5 transposase by limited proteolysis. *J. Biol. Chem.* **273**:10908-10913.
24. **Brandl, H.** 2001. Microbial Leaching of Metals, p. 191-224. *In*: H. J. Rehm and G. Reed (ed.), *Biotechnology*. Wiley-VHC, Weinheim.
25. **Breed, A. W., C. J. N. Dempers, G. E. Searby, M. N. Gardner, D. E. Rawlings and G. S. Hansford.** 1999. The effect of temperature on the ferrous iron oxidation kinetics of *Leptospirillum ferrooxidans* in continuous culture. *Biotech. Bioeng.* **65**:44-53.
26. **Brierley, C. L.** 1978. Bacterial leaching. *Crit. Rev. Microbiol.* **6**:207-261.
27. **Brierley, C. L.** 1982. Microbiological mining. *Sci. Am.* **247(2)**:42-51.
28. **Brierley, C. L.** 1997. Mining biotechnology: research into commercial development and beyond, p.3-17. *In* D. E. Rawlings (ed.), *Biomining: theory, microbes and industrial processes*. Springer-Verlag, Berlin, Germany.
29. **Brierley, J. A., and C. L. Brierley.** 1999. Present and future of commercial applications of biohydrometallurgy, p. 81-89. *In* R. Amils and A. Ballester (ed.), *Biohydrometallurgy and the environment towards the 21st Century, Part A*. Elsevier, Amsterdam.
30. **Brown, P. O.** 1997. Integration, p161-203. *In* J. M. Coffin, S. H. Hughes, H. E. Varmus, *Retroviruses*. Cold Spring Harbor Laboratory Press, Plainview, N. Y.

31. **Bryner, L. C. and Jameson, A. K.** 1958. Microorganisms in leaching sulfide minerals. *Appl. Microbiol.* **6**:281-287.
32. **Bryner, L. C., and R. Anderson.** 1957. Microorganisms in leaching sulfide minerals. *Ind Eng. Chem.* **49**:1721-1724.
33. **Bushman, F.** 2002. Lateral DNA transfer: mechanisms and consequences. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
34. **Cabrejos, M. E., H. L. Zhao, M. Guacocano, S. Bueno, G. Levican, E. Garcia, E. Jedlicki, and D. S. Holmes.** 1999. *IST1* insertional inactivation of the *resB* gene: implications for phenotypic switching in *Thiobacillus ferrooxidans*. *FEMS Microbiol. Lett.* **175**: 223-229.
35. **Capy, P., G. Gasperi, C. Biermont, and C. Bazin.** 2000. Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**:101-6.
36. **Caugant, D. A., B. R. Levin, and R. K. Selander.** 1981. Genetic diversity and temporal variation in the *E. coli* populations of a human host. *Genetics* **98**:467-490.
37. **Chaconas, G., B. D. Lavoie, and M. A. Watson.** 1996. DNA transposition: jumping gene machine, some assembly required. *Curr Biol.* **6**:817-820.
38. **Chakraborty, R., C. Deb, A. Lohia, and P. Roy.** 1997. Cloning and characterization of a high-copy-number novel insertion sequence from chemolithotrophic *Thiobacillus ferrooxidans*. *Plasmid* **38**:129-134.
39. **Chalmers, R. M., and N. Kleckner.** 1994. *Tn10/IS10* transposase purification, activation and in vitro reaction. *J. Biol. Chem.* **269**:8029-8035.
40. **Chandler, M., and J. Mahillon.** 2002. Insertion sequence revisited, p.305-366. *In* N. L. Craig , R. Craigie, M. Gellert and A. Lambowitz (ed.), *Mobile DNA II*. ASM Press, Washington, D. C.
41. **Chandler, M., and O. Fayet.** 1993. Translational frameshifting in the control of transposition in bacteria. *Mol. Microbiol.* **7**:497-503.
42. **Coen, E. S., T. P. Robbins, J. Almeida, A. Hudson, and R. Carpenter.** 1989. Consequences and mechanisms of transposition in *Antirrhinum majus*, p. 413-436. *In* D. E. Berg and M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D. C.
43. **Colmer, A. R., and M. E. Hinkle.** 1947. The role of microorganisms in acid mine drainage: a preliminary report. *Science* **106**:253-256.
44. **Coplin, D. L.** 1989. Plasmids and their role in the evolution of plant pathogenic bacteria. *Annu Rev. Phytopathol.* **27**:187-212.
45. **Craig, N. L.** 1995. Unity in transposition reactions. *Science* **270**:253-254.

46. **Craig, N. L.** 1996a. Transposition, p. 2339-2362. *In* F. C. Neidhart (ed), *Escherichia coli* and *Salmonella*, cellular and molecular biology, 2nd edition. American Society for Microbiology, Washington, D. C.
47. **Craig, N. L.** 1996b. Transposon Tn7. *Curr. Top. Microbiol. Immunol.* **204**:27-48.
48. **Craig, N. L.** 1997. Target site selection in transposition. *Annu. Rev. Biochem.* **66**:437-474.
49. **Craig, N. L.** 2002. Mobile DNA: an introduction, p. 3-11. *In* N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz (ed.), *Mobile DNA II*. American Society for Microbiology, Washington D.C.
50. **Craigie, R.** 1996. Quality control in Mu DNA transposition. *Cell* **85**:137-140.
51. **Craigie, R., and K. Mizuuchi.** 1985. Mechanism of transposition of bacteriophage Mu: structure of a transposition intermediate. *Cell* **41**:867-876.
52. **Darzins, A., N. E. Kent, M. S. Buckwalter, and M. J. Casadaban.** 1988. Bacteriophage Mu sites required for transposition immunity. *Proc. Natl. Acad. Sci. USA* **85**:6826-6830.
53. **Davies, D. R., I. Y. Goryshin, W. S. Reznikoff, and I. Rayment.** 2000. Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* **289**:77-85.
54. **Davies, M. A., R. W. Simons, and N. Kleckner.** 1985. Tn10 protects itself at two levels from fortuitous activation by external promoters. *Cell* **43**:379-387.
55. **Deane, S. M., and D. E. Rawlings.** 2004. Plasmid evolution and interaction between the plasmid addiction stability of the two related broad-host-range IncQ-like plasmids. *J. Bacteriol.* **186**:2123-2133.
56. **DeBoy, R., and N. L. Craig.** 1996. Tn7 transposition as a probe of *cis* interactions between widely separated (190 kilobases apart) DNA sites in the *Escherichia coli* chromosome. *J. Bacteriol.* **178**:6184-6191.
57. **Debyshire, K. M., and N. D. Grindley.** 1996. *cis* preference of the IS903 transposase instability and efficient translation. *Mol. Microbiol.* **21**:1261-1272.
58. **Debyshire, K. M., M. Kramer, and N. D. Grindley.** 1990. Role of instability in the *cis* action of the insertion sequence of IS903 transposase. *Proc. Natl. Acad. Sci. USA* **87**:4048-4052.
59. **Diaz-Aroca, E., M. V. Mendiola, J.C. Zabala, and F. de la Cruz.** 1987. Transposition of IS91 does not generate a target duplication. *J. Bacteriol.* **169**:442-443.

60. **Doak, T. G., F. P. Doerder, C. L. Jahn and G. Herrick.** 1994. A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common 'D35E' motif. *Proc. Natl. Acad. Sci USA* **91**:942-946
61. **Doetsch, R. N., T. M. Cook, and Z. Vaituzis.** 1967. On the uniqueness of the flagellum of *Thiobacillus thiooxidans*. *Antonie Leeuwenhoek J. Microbiol. Serol.* **33**:196-202.
62. **Doolittle, W. F., T. B. Kirkwood, and M. A. Dempster.** 1984. Selfish DNAs with self-restraint. *Nature* **307**:501-502.
63. **Dorman, C. J., and P. Deighan.** 2003. Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Gen. Dev.* **13**(2):79-184.
64. **Dyda, F., A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, and D. R. Davis.** 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other plynucleotidyl transferases. *Science* **266**:1981-1986.
65. **Eberhard, W. G.** 1989. Why do bacterial plasmids carry some genes and not others? *Plasmid* **21**:167-174.
66. **Echols, H.** 1990. Nucleoprotein structures initiating DNA replication, transcription, and site-specific recombination. *J. Biol. Chem.* **265**:14697-14700.
67. **Edwards, K. J., T. M. Gihring, and J. F. Banfield.** 1999. Seasonal variations in microbial populations and environmental conditions in an extreme acid mine drainage environment. *Appl. Environ. Microbiol.* **65**:3627-3632.
68. **Ehrlich, H. L.** 1997. Microbes and metals. *Appl. Microbiol. Biotechnol.* **48**:687-692.
69. **Ehrlich, H. L., and C. L. Brierley.** 1990. *Microbial mineral recovery.* McGraw-Hill, New York, N. Y.
70. **Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston, and J. Sved.** 1990. High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* **62**:515-25.
71. **Escoubas, J. M., M. F. Prere, O. Fayet, I. Salvignol, D. Galas, D. Zerbib, and M. Chandler.** 1991. Translational control of transposition activity of the bacterial insertion sequence *IS1*. *EMBO J.* **10**:705-712.
72. **Espejo, R. T., and J. Romero.** 1997. Bacterial community in copper sulfide ores inoculated and leached with solution from a commercial-scale copper leaching plant. *Appl. Environ. Microbiol.* **63**:1344-1348.
73. **Farabaugh, P. J.** 1997. *Programmed alternative reading of the genetic code.* R. G. Landes Company, Austin, Texas.

74. **Fayet, O., P. Raymond, P. Polard, M. F. Prere, and M. Chandler.** 1990. Functional similarities between retroviruses and the IS3 family of bacterial insertion sequences? *Mol. Microbiol.* **4**:1771-1777.
75. **Federoff, N.** 2002. Control of mobile DNA, p. 997-1007. *In* N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz(ed.), *Mobile DNA II*. American Society for Microbiology, Washington, D. C.
76. **Fowler, T. A., P. R. Holmes, and F. K. Crundwell.** 1999. Mechanism of pyrite dissolution in the presence of *Thiobacillus ferrooxidans*. *Appl Environ. Microbiol.* **65**:2987-93.
77. **Galas, D. J., and M. Chandler.** 1981. On the molecular mechanisms of transposition. *Proc. Natl. Acad. Sci. USA* **78**:5848-5862.
78. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109-162. *In* D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D. C.
79. **Garcillan-Barcia, M. P., I. Bernalez, M. V. Mendiola, and F. de la Cruz.** 2002. IS91 rolling circle transposition, p. 891-904. *In* N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz (ed.), *Mobile DNA II*. American Society for Microbiology, Washington, D. C.
80. **Gardner, M. N., and D. E. Rawlings.** 2000. Production of rhodanese by bacteria present in bio-oxidation plants used to recover gold from arsenopyrite ores. *J. Appl Microbiol.* **89**:185-190.
81. **Gardner, M. N., S. M. Deane, and D. E. Rawlings.** 2001. Isolation of a new broad-host range IncQ-like plasmid pTC-F14, from Acidiphilic bacterium *Acidithiobacillus caldus* and analysis of the plasmid replicon. *J. Bacteriol.* **183**:3303-3309.
82. **Gehrke, T., J. Telegdi, D. Thierry, and W. Sand.** 1998. Importance of extracellular polymeric substances from *Thiobacillus ferrooxidans* for bioleaching. *Appl. Environ. Microbiol.* **64**:2743-2747.
83. **Geourjon, C., and G. Deleage.** 1995. SOPMA: Significant improvement in protein secondary prediction by consensus prediction from multiple alignments. *Comput. Appl Biosci.* **11**:681-684.
84. **Gesteland, R. F., and J. F. Atkins.** 1996. Recoding: dynamic reprogramming of translation. *Annu. Rev. Biochem.* **65**:741-768.
85. **Goldschmidt, G. K., M. N. Gardner, L. J. van Zyl, S. M. Deane, and D. E. Rawlings.** 2003. A promiscuous broad-host range, IncQ-like plasmid isolated from an industrial strain of *Acidithiobacillus caldus*, its accessory DNA and potential to

- replicate in the horizontal gene pool of biomining and other bacteria. International Biohydrometallurgy Symposium, IBS2003, Athens, Greece.
86. **Goosen, N., and P. van de Putte.** 1986. Role of Ner protein in bacteriophage Mu transposition. *J. Bacteriol.* **167**:503-507.
 87. **Goryshin, I. Y., and W. S. Reznikoff.** 1998. Tn5 *in vitro* transposition. *J. Biol. Chem.* **273**:7367-7374.
 88. **Goryshin, I. Y., J. A. Miller, Y. V. Kil, A. Lanzov, and W. S. Reznikoff.** 1998. Tn5/IS50 target recognition. *Proc. Natl. Acad. Sci. USA* **95**:10716-10721.
 89. **Grindley, N. D. F., and R. R. Reed.** 1985. Transpositional recombination in prokaryotes. *Annu. Rev. Biochem.* **54**:863-896.
 90. **Hagemann, A. T., and N. L. Craig.** 1993. Tn7 transposition creates a hotspot for homologous recombination at the transposon donor site. *Genetics* **133**:9-16.
 91. **Hallberg, K. B., and E. B. Lindström.** 1994. Characterization of of *Thiobacillus caldus*, sp. nov., a moderately thermophilic acidophile. *Microbiol.* **140**:3451-3456.
 92. **Hallet, B., R. Rezsöházy, J. Mahillon, and J. Delcour.** 1994. IS231A insertions specificity: consensus sequence and DNA binding at the target site. *Mol Microbiol.* **14**:131-139.
 93. **Halling, S. M., and N. Kleckner.** 1982. A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. *Cell* **28**:155-63.
 94. **Haniford, D. B., and G. Chaconas.** 1992. Mechanistic aspects of DNA transposition. *Curr. Opin. Genet. Dev.* **2**:698-704.
 95. **Haniford, D. B., H. W. Benjamin, and N. Kleckner.** 1991. Kinetic and structural analysis of a cleaved donor intermediate and a strand transfer intermediate in Tn10 transposition. *Cell* **64**:171-179.
 96. **Haren, L., B. Ton-Hoang, and M. Chandler.** 1999. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53**:245-81.
 97. **Haren, L., M. Betermier, P. Polard, and M. Chandler.** 1997. IS911-mediated intramolecular transposition is naturally temperature sensitive. *Mol. Microbiol.* **25**:531-540.
 98. **Haren, L., P. Polard, B. Ton-Hoang, and M. Chandler.** 1999. Multiple oligomerization domains in the IS911 transposase: a leucine zipper motif is essential for activity. *J. Mol. Biol.* **283**:29-41.
 99. **Harrison, A. P.** 1984. The acidophilic *Thiobacilli* and other acidophilic bacteria that share their habit. *Annu. Rev. Microbiol.* **38**:265-292.

100. **Hauer, B., and J. A. Shapiro.** 1984. Control of Tn7 transposition. *Mol. Gen. Genet.* **194**:149-158.
101. **Heuer, T. S., P. O. Brown.** 1998. Photo-cross linking studies suggests a model for the architecture of an active human immunodeficiency virus type I integrase-DNA complex. *Biochemistry* **37**:6667-6678.
102. **Hickman, A. B., Y. Li, S. V. Mathew, E. W. May, N. L. Craig, and F. Dyda.** 2000. Unexpected structural diversity in DNA recombination: the restriction endonuclease reaction. *Mol. Cell. Biol.* **5**:1025-1034.
103. **Hirschel, B. J., D. J. Galas, and M. Chandler.** 1982. Cointegrate formation by Tn5, but not transposition, is dependent on *recA*. *Proc. Natl. Acad. Sci. USA* **79**:4530-4534.
104. **Holmes, D. S., H-L. Zhao, G. Levican, J. Ratouchniak, V. Bonnefoy, P. Varela, and E. Jedlicki.** 2001. IS*AfeI*, an ISL3 family insertion sequence from *Acidithiobacillus ferrooxidans* ATC 19859. *J. Bacteriol.* **183**:4323-4329.
105. **Holmes, I.** 2002. Transcendent elements: whole genome transposon screens and open evolutionary questions. *Genome Research* **12**:1152-1155.
106. **Iida, S., J. Meyer, and W. Arber.** 1983. Prokaryotic IS elements, p.159-222. *In* J. Shapiro (ed.), *Mobile genetic elements*. Academic Press Incorporated.
107. **Ingledeu, W. J.** 1982. *Thiobacillus ferrooxidans*: the bioenergetics of an acidophilic chemolithotroph. *Biochem. Biophys. Acta.* **683**:89-117.
108. **Jain, C., and N. Kleckner.** 1993. Preferential *cis* action of IS*I0* transposase depends upon its mode of synthesis. *Mol. Microbiol.* **9**:249-260.
109. **Jenkins, T. M., D. Esposito, A. Engelman, R. Craigie.** 1997. Critical contacts between HIV-1 integrase and viral DNA identification by structure-based analysis and photo-cross linking. *EMBO J.* **16**:6849-6859.
110. **Johnson, R. C.** 2002. Bacterial site-specific DNA inversion systems, p.230-271. *In* N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz (ed.), *Mobile DNA II*. American Society for Microbiology, Washington D.C.
111. **Jordan, E., H. Saedler, and P. Starlinger.** 1968. O^o and strong polar mutations in the gal operon are insertions. *Mol. Gen. Genet.* **102**:680-681.
112. **Kallastu, A., R. Horak, and M. Kivisaar.** 1998. Identification and characterization of IS*I411*, a new insertion sequence which causes transcriptional activation of the phenol degradation genes in *Pseudomonas putida*. *J. Bacteriol.* **180**:5306-5312.
113. **Katzman, M., J. P. Mack, A. M. Skalka, and J. Leis.** 1991. A covalent complex between retroviral integrase and nicked substrate DNA. *Proc. Natl. Acad. Sci. USA* **88**:4695-4699.

114. **Kaufman, P. D., and D. C. Rio.** 1992. P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. *Cell* **69**:27-39.
115. **Kelly, D. P.** 1988. Evolution of the understanding of the microbiology and biochemistry of the mineral leaching habitat, p. 3-14. *In* P. R. Norris and D. P. Kelly (ed.), *Biohydrometallurgy -87: science and technology letters*. Kew, Surrey, UK.
116. **Kelly, D. P., and A. P. Harrison.** 1989. Genus *Thiobacillus* Beijerinck 1904b, 597, p. 1842-1858. *In* J. T. Staley, M. P. Bryant, N. Pfennig and J. G. Holt (ed.), *Bergey's manual of systematic bacteriology*, 1st ed, vol. 3. Williams and Wilkins, Baltimore.
117. **Kelly, D. P., and Wood A. P.** 2000. Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen. nov., *Halothiobacillus* gen. nov., and *Thermothiobacillus* gen. nov. *Int. J. Syst. Evol. Microbiol.* **50**:511-516.
118. **Kelly, D. P., P. R. Norris, and C. L. Brierley.** 1979. Microbiological methods for the extraction and recovery of metals, p. 263-308. *In* A. T. Bull, D. G. Ellwood and C. Ratledge (ed.), *Microbial Technology: Current State and Future Prospects*. Cambridge Univ. Press, Cambridge.
119. **Kennedy, A. K., A. Guhathakurta, N. Kleckner, and D. B. Haniford.** 1998. *Tn10* transposition via a DNA hairpin intermediate. *Cell* **95**:125-134.
120. **Khan, E., J. P. Mack, R. A. Katz, J. Kulkosky and A. M. Skalka.** 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**:851-60.
121. **Kidwell, M. G., and D. R. Lisch.** 2001. Perspective: transposable elements, parasitic DNA and genome evolution. *Evolution Int. J. Org. Evolution.* **55**:1-24.
122. **Kleckner, N., R. M. Chalmers, D. Kwon, J. Sakai, and S. Bolland.** 1996. *Tn10* and *IS10* transposition and chromosome rearrangements: mechanisms and regulation in vivo and in vitro, p. 49-82. *In* H. Saedler and A. Gierl (ed.), *Transposable elements*. Springer-Verlag KG, Heidelberg, Germany.
123. **Krebs, M. P., and W. S. Reznikoff.** 1986. Transcriptional and translational initiation sites of IS50. Control of transposase and inhibitor expression. *J. Mol. Biol.* **192**:781-791.
124. **Kretschner, P. J., and S. N. Cohen.** 1979. Effect of temperature on translocation frequency of the Tn3 element. *J. Bacteriol.* **139**:515-519.
125. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell. Biol.* **12**:2331-2328.

126. **Labrador, M., and V. C. Corces.** 2002. Interactions between transposable elements and the host genome, p. 1008-1023. *In* N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz, Mobile DNA II. American Society for Microbiology, Washington, D. C.
127. **Lacey, D. T., and F. Lawson.** 1970. Kinetics of the liquid-phase oxidation of acid ferrous sulfate by the bacterium *Thiobacillus ferrooxidans*. *Biotechnol. Bioeng.* **12**:29-50.
128. **Lavoie, B. D., and G. Chaconas.** 1996. Transposition of phage Mu DNA. *Curr. Top. Microbiol. Immunol.* **204**:83-102.
129. **Lavoie, B. D., B. S. Chan, R. G. Allison, and G. Chaconas.** 1991. Structural aspects of a higher order nucleoprotein complex-induction of a altered DNA structure at the Mu-host junction of the Mu type I transposome. *EMBO J.* **10**:3051-3059.
130. **Lawson, E. N.** 1997. The composition of mixed populations of leaching bacteria active in gold and nickel recovery from sulphide ores, p. QP4.1-4.10. *In* International Biohydrometallurgy Symposium, IBS97. Glenside, Australian Mineral Foundation, South Australia.
131. **Lederberg, E. M.** 1981. Plasmid reference center registry of transposon (Tn) allocations through July 1981. *Gene* **16**:59-61.
132. **Lee, C-H., A. Bhagwat, and F. Heffron.** 1983. Identification of a transposon Tn3 sequence required for transposition immunity. *Proc. Natl. Acad. Sci. USA* **80**:6765-6769.
133. **Lee, S., D. Butler, and N. Kleckner.** 1987. Efficient Tn10 transposition into a DNA insertion hot spot *in vivo* requires the 5-methyl groups of symmetrically disposed thymines within the hot-spot consensus sequence. *Proc. Natl. Acad. Sci. USA* **84**:7876-7880.
134. **Lenich, A. G., and A. C. Glasgow.** 1994. Amino acid sequence homology between Piv, an essential protein in site-specific DNA inversion in *Moraxella lacunata*, and transposases of an unusual family of insertion elements. *J. Bacteriol.* **176**:4160-4164.
135. **Levin, B. R., F. M. Stewart.** 1980. The population biology of bacterial plasmids: a priori conditions for the existence of mobilizable nonconjugative factors. *Genetics* **94**:425-443.
136. **Lichens-Park, A., and M. Syvanen.** 1988. Cointegrate formation by IS50 requires multiple donor molecules. *Mol. Gen. Genet.* **211**:244-251.
137. **Lilly, A., P. Young, M. Bailey.** 2000. Bacterial population genetics: do plasmids maintain bacterial diversity and adaptation?, p. 287-297. *In* C. M. Thomas (ed.), The

- horizontal gene pool: bacterial plasmids and gene spread. Harwood Academic Publishers.
138. **Little, J. W., and D. W. Mount.** 1982. The SOS regulatory system of *Escherichia coli*. *Cell* **29**:11-22.
 139. **Livesey-Goldblatt, E., P. Norman, and D. R. Livesey-Goldblatt.** 1983. Gold recovery from arsenopyrite/pyrite ore by bacterial leaching and cyanidation, p. 627-641. *In* G. Rossi and A. E. Torma (ed.), Recent progress in biohydrometallurgy. Iglesias, Assoc. Mineraria Sarda, Italy.
 140. **Lundgren, D. G., and M. Silver.** 1980. Ore leaching by bacteria. *Annu. Rev. Microbiol.* **34**:263-283.
 141. **Luthi, K., M. Moser, J. Ryser, and H. Weber.** 1990. Evidence for a role of translational frameshifting in the expression of transposition activity of the bacterial insertion element *IS1*. *Gene* **88**:15-20.
 142. **Machida, C., and Y. Machida.** 1989. Regulation of *IS1* transposition by the *insA* gene product. *J. Mol. Biol.* **208**:567-574.
 143. **Maekawa, T., J. Amemura-Maekawa, and E. Ohtsubo.** 1993. DNA binding domains in Tn3 transposase. *Mol. Gen. Genet.* **236**:267-274.
 144. **Mahillon, J. and M. Chandler.** 1998. Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**:725-774.
 145. **Malamy, M. H.** 1970. Some properties of insertion mutations in the lac operon, p. 359-373. *In* J. R. Beckwith and D. Zipser (ed.), The lactose operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
 146. **Markosyan, G. E.** 1972. A New iron-oxidizing bacterium *Leptospirillum ferrooxidans* nov. gen. nov. sp. *Biol. J. Armenia* **25**:26-29
 147. **Mayard-Smith, J.** 1982. Overview-unsolved evolutionary problems, p.375-382. *In* G. Dover and R. Flavel (ed.), Genome evolution. London, Academic Press.
 148. **Mayr, E.** 1982. The growth of biological thought. Harvard University Press.
 149. **Mendiola, M. V. , and F. de la Cruz.** 1989. Specificity of insertion of *IS91*, and insertion sequence present in the alpha-haemolysin plasmids of *Escherichia coli*. *Mol. Microbiol.* **3**:979-984.
 150. **Mendiola, M. V., I. Bernales, and F. de la Cruz.** 1994. Differential roles of the transposon termini in *IS91* transposition. *Proc. Natl. Acad. Sci. USA* **91**:1922-1926.
 151. **Merlin, C., J. Mahillon, J. Nesvera, and A. Toussaint.** 2000. Gene recruiters and transporters: the modular structure of bacterial mobile elements, p 363-409. *In* C. M.

- Thomas (ed.), The horizontal gene pool: bacterial plasmids and spread. Harwood Academic, Amsterdam.
152. **Miller, W. J., and P. Capy.** 2004. Mobile genetic elements as natural tools for genome evolution. *Methods Mol. Biol.* **260**:1-20.
 153. **Mizuuchi, K.** 1992a. Transpositional recombination: mechanistic insights from studies of Mu and other elements. *Annu. Rev. Biochem.* **61**:1011-1051.
 154. **Mizuuchi, K.** 1992b. Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.* **267**:21273-21276.
 155. **Mizuuchi, K., and T. A. Baker.** 2002. Chemical mechanisms for mobilizing DNA, p. 12-23. *In* N. L. Craig, R. Craigie, M. Gellert and A. Lambowitz, *Mobile DNA II*. American Society for Microbiology, Washington, D. C.
 156. **Mizuuchi, M., and K. Mizuuchi.** 1993. Target site selection in transposition of phage Mu. *Cold Spring Harbor Symp. Quant. Biol.* **5884**:515-523.
 157. **Mizuuchi, M., T. A. Baker, and K. Mizuuchi.** 1992. Assembly of the active form of the transposase-Mu DNA complex: a critical control point in Mu transposition. *Cell* **70**:303-311.
 158. **Morisato, D., and N. Kleckner.** 1987. Tn10 transposition and circle formation *in vitro*. *Cell* **51**:101-111.
 159. **Nicholaidis, A. A.** 1987. Microbial mineral processing: the opportunities for genetic manipulation. *J. Chem. Tech. Biotechnol.* **38**:167-185.
 160. **Norris, P. R.** 1990. Acidophilic bacteria and their activity in mineral sulfide oxidation, p. 3-27. *In* H. L. Ehrlich and C. L. Brierley (ed.), *Microbial Mineral Recovery*. McGraw Hill, New York.
 161. **Norris, P. R., and D. B. Johnson.** 1998. Acidophilic microorganisms, p. 133-154. *In* K. Horikoshi and W. Grant (ed.), *Extremophiles: Microbial Life in Extreme Environments*. Wiley, New York, NY.
 162. **Ohtsubo, E., and Y. Sekine.** 1996. Bacterial insertion sequences. *Curr. Top. Microbiol. Immunol.* **204**:1-26.
 163. **Ohtsubo, H., and E. Ohtsubo.** 1978. Nucleotide sequence of an insertion element, IS1. *Proc. Natl. Acad. Sci. USA* **75**:615-619.
 164. **Orgel, L. E., and F. H. C. Crick.** 1980. Selfish DNA: the ultimate parasite. *Nature* **284**:604-607.
 165. **Orgel, L. E., F. H. C. Crick, and C. Sapienza.** 1980. Selfish DNA. *Nature* **288**:645-646.

166. **Osborn, M., S. Bron, N. Firth, S. Holsappel, A. Huddleston, R. Kiewiet, W. Meijier, J. Seegers, R. Skurray, P. Terpstra, C. M. Thomas, P. Thorsted, E. Tietze, and S. L. Turner.** 2000. Evolution of bacterial plasmids, p. 301-361. *In* C. M. Thomas (ed.), The horizontal gene pool: bacterial plasmids and gene spread. Harwood Academic Publishers.
167. **Pizarro, J., E. Jedlicki, O Orellana, J. Romero, and R. T. Espejo.** 1996. Bacterial populations in samples of bioleached copper ore as revealed by analyses of DNA obtained before and after cultivation. *Appl. Environ. Microbiol.* **62**:1323-28.
168. **Polard, P., and M. Chandler.** 1995. Bacterial transposases and retroviral integrases. *Mol. Microbiol.* **15**(11):13-23.
169. **Polard, P., B. Ton-Hoang, L. Haren, M. Betermier, R. Walczak, and M. Chandler.** 1996. IS911-mediated transpositional recombination *in vitro*. *J. Mol. Biol.* **264**:68-81.
170. **Polard, P., M. F. Prere, O. Fayet, and M. Chandler.** 1992. Transposase-induced excision and circularization of the bacterial insertion sequence IS911. *EMBO. J.* **11**:5079-5090.
171. **Priefer, U. B., R. Simon, A. Puhler.** 1985. Extension of the host range of *Escherichia coli* vectors by incorporation of RSF1010 replication and mobilization functions. *J. Bacteriol.* **163**:324-330.
172. **Pronk, J. T., K. Liem, P. Bos, and J. G. Kuenen.** 1991. Energy translation by anaerobic ferric iron respiration in *Thiobacillus ferrooxidans*. *Appl Environ. Microbiol.* **57**:2063-2068.
173. **Rawlings, D. E.** 1995. Restriction enzyme analysis of 16S rRNA genes for the rapid identification of *Thiobacillus ferrooxidans*, *Thiobacillus thiooxidans* and *Leptospirillum ferrooxidans* strains in the leaching environments, p 9-17. *In* C. A Jerez, T. Vargas, H. Toledo and J. V. Wiertz, Biohydrometallurgy Processing, vol. 2. University of Chile, Santiago.
174. **Rawlings, D. E.** 1997. Mesophilic, autotrophic bioleaching bacteria: description, physiology and role, p. 229-246. *In* D. E. Rawlings (ed.), Biomining: theory, microbes and industrial processes. Springer-Verlag, Berlin, Germany.
175. **Rawlings, D. E.** 2002. Heavy metal mining using microbes. *Annu. Rev. Microbiol.* **56**:65-91.
176. **Rawlings, D. E.** 2003. The microbially assisted dissolution of minerals and its use in the mining industry. SEANAC/IUPAC conference, Gaborone, Botswana.
177. **Rawlings, D. E., and E. Tietze.** 2001. Comparative biology of IncQ and IncQ-like plasmids. *Microbiol. Mol. Biol. Rev.* **65**(4): 481-496.

178. **Rawlings, D. E., and S. Silver.** 1995. Mining with microbes. *Bio. Technol.* **13**:773-778.
179. **Rawlings, D. E., H. Tributsch, and G. S. Hansford.** 1999a. Reasons why '*Leptospirillum*'-like species rather than *Thiobacillus ferrooxidans* are the dominant iron-oxidizing bacteria in many commercial processes for the biooxidation of pyrite and related ores. *Microbiology* **145**:5-13.
180. **Rawlings, D. E., N. J. Coram, M. N. Gardner, and S. M. Deane.** 1999b. *Thiobacillus caldus* and *Leptospirillum ferrooxidans* are widely distributed in continuous flow biooxidation tanks used to treat a variety of metal containing ores and concentrates, p. 777-786. In R. Amils and A. Ballister (ed.), *Biohydrometallurgy and the environment towards the mining of the 21st century*, Part A. Elsevier, Amsterdam, The Netherlands.
181. **Rehm, B. H. A.** 2001. Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. *Appl. Microbiol. Biotechnol.* **57**:579-592.
182. **Reyes, I., A. Beyon, C. Mignotte-Vieux, and F. Richard.** 1987. Mini-Mu transduction: *cis*-inhibition of the insertion of Mu transposons. *Plasmid* **18**:183-192.
183. **Reznikoff, W. S.** 1993. The Tn5 transposition. *Annu. Rev. Microbiol.* **47**:945-963.
184. **Rezsohazy, R., B. Hallet, J. Delcour, and J. Mahillon.** 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* **9**:1283-1295.
185. **Rice, L. B.** 2002. Association of different mobile elements to generate novel integrative elements. *Cell Mol. Life Sci.* **59**:2023-2032.
186. **Rice, P., and K. Mizuuchi.** 1995. Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* **82**:209-220.
187. **Rice, P., R. Craigie, and D. R. Davies.** 1996. Retroviral integrases and their cousins. *Curr. Opin. Struct. Biol.* **6**:76-83.
188. **Roberts, D., B. C. Hoopes, W. R. McClure, and N. Kleckner.** 1985. IS10 transposition is regulated by DNA methylation. *Cell* **43**:117-130.
189. **Rojas, J., M. Giersig, and H. Tributsch.** 1995. Sulfur colloids as temporary energy reservoirs for *Thiobacillus ferrooxidans* during pyrite oxidation. *Arch. Microbiol.* **163**:256-352.
190. **Rojas-Chapana, J. A., and H. Tributsch.** 2001. Biochemistry of sulfur extraction in bio-corrosion of pyrite by *Thiobacillus ferrooxidans*. *Hydrometallurgy* **59**:291-300.

191. **Rojas-Chapana, J. A., C. C. Bärtels, L. Pohlman, and H. Tributsch.** 1998. Cooperative leaching and chemotaxis of thiobacilli studied with spherical sulfur/sulfide substrates. *Process Biochem.* **33(3):**239-248.
192. **Rossi, G.** 1990. *Biohydrometallurgy.* McGraw-Hill, Hamburg, Germany.
193. **Rost, B., R. Schneider, and C. Sander.** 1997. Protein fold recognition by prediction based threading. *J. Mol. Biol.* **270:**471-480.
194. **Roszak, D. B., and R. R. Colwell.** 1987. Metabolic activity of bacterial cells enumerated by direct viable count. *Appl. Environ. Microbiol.* **53:**2889-2893.
195. **Rousseau P., E. Gueguen, G. Duval-Valentin, and M. Chandler.** 2004. The helix-turn-helix motif of bacterial insertion sequences IS911 transposase is required for DNA binding. *Nucleic. Acids. Res.* **32:**1335-1344.
196. **Sakai, J., and N. Kleckner.** 1997. The *Tn10* synaptic complex can capture a target DNA only after transposon excision. *Cell* **89:**205-211.
197. **Sakai, J., R. M. Chalmers, and N. Kleckner.** 1995. Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in *Tn10* transposition. *EMBO J.* **43:**74-83.
198. **Sambrook, J. E., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
199. **Sand, W., T. Gehrke, P-G. Jozsa, and A. Schippers.** 2001. (Bio)chemistry of bacterial leaching-direct vs. indirect. *Hydrometallurgy* **59:**159-175.
200. **Sand, W., T. Gherke, T. Hallman, and A. Schippers.** 1995. Sulphur chemistry, biofilm and the (in)direct attack mechanism-a critical evaluation of bacterial leaching. *Appl. Microbiol.* **43:** 961-966.
201. **Sapienza, C., and F. W. Doolittle.** 1981. Genes and things you have whether you want them or not. *Cold Spring Harbor Symp. Quant. Biol.* **45:**177-182.
202. **Sarnovsky, R. J., E. W. May, and N. L. Craig.** 1996. The *Tn7* transposase is a heteromeric complex in which DNA breakage and joining activities are distributed between different gene products. *EMBO J.* **15:**6348:6361.
203. **Sasakawa, C., Y. Uno, and M. Yoshikawa.** 1981. The requirement for both DNA polymerase end 5' to 3' exonuclease activities of DNA polymerase I during *Tn5* transposition. *Mol. Gen. Genet.* **182:**19-24.
204. **Schippers A., P-G. Jozsa, and W. Sand.** 1996. Sulfur chemistry in bacterial leaching of pyrite. *Appl. Environ. Microbiol.* **62:**3424-3431.

205. **Schippers, A., and W. Sand.** 1999. Bacterial leaching of metal sulfides proceeds by two indirect mechanisms via thiosulfate or via polysulfides and sulfur. *Appl Environ. Microbiol.* **65**:319-321.
206. **Schippers, A., P-G. Jozsa, and W. Sand.** 1996. Sulfur chemistry in bacterial leaching of pyrite. *Appl. Environ. Microbiol.* **62**:3424-3431.
207. **Schnell, H. A.** 1997. Bioleaching of copper, p. 21-43. *In* D. E. Rawlings (ed.), *Biomining: theory, microbes and industrial processes*. Berlin: Springer-Verlag.
208. **Sekine, Y., and E. Ohtsubo.** 1989. Frameshifting is required for expression of IS1 transposase. *Proc. Natl. Acad. Sci. USA* **86**:4609-4613.
209. **Sekine, Y., N. Eisaki, and E. Ohtsubo.** 1994. Translational control in production of transposase and in transposition of insertion sequence IS3. *J. Mol. Biol.* **235**:1406-1420.
210. **Sengstang, C., S. Iida, R. Hiestand-Nauer, and W. Arber.** 1986. Terminal inverted repeats of prokaryotic transposable elements IS186 which can generate duplications of variable length at an identical target sequence. *Gene* **49**:153-156.
211. **Shapiro, J. A.** 1969. Mutations caused by the insertions of genetic material into the galactose operon of *Escherichia coli*. *J. Mol. Biol.* **40**:93-105.
212. **Shapiro, J. A.** 1979. Molecular model for transposition and replication of bacteriophage Mu and other transposable elements. *Proc. Natl. Acad. Sci. USA* **76**:1933-1937.
213. **Shapiro, J. A.** 1999. Transposable elements as the key to a 21st century view of evolution. *Genetica* **107**:171-179.
214. **Sheratt, D.** 1989. Tn3 and related transposable elements: site specific recombination and transposition, p. 163-184. *In* D. E. Berg and M. M Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D. C.
215. **Shiga, Y., Y. Sekine, Y. Kano, and E. Ohtsubo.** 2001. Involvement of H-NS in transpositional recombination mediated by IS1. *J. Bacteriol.* **183** (8):2476-2484.
216. **Simon R., U. Priefer, and A. Puhler.** 1983. A broad host range mobilization system from in vivo genetic engineering: transposon mutagenesis in gram-negative bacteria. *Bio/Technol.* **1**:784-790.
217. **Simons, R. W., and N Kleckner.** 1983. Translational control of IS10 transposition. *Cell* **34**:683-691.
218. **Smalla, K., A. M. Osborn, and E. M. H. Wellington.** 2000. Isolation and characterization of plasmids from bacteria, p. 207-248. *In* C. M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood Academic Publishers.

219. **Smith, M. W., and R. F. Doolittle.** 1992. A comparison of evolutionary rates of the two major kinds of superoxide dismutase. *J. Mol. Evolu.* **34**:175-184.
220. **Southam, G., and T. J. Beveridge.** 1993. Examination of lipopolysaccharide (O-antigen) populations of *Thiobacillus ferrooxidans* from two mine tailings. *Appl. Environ. Microbiol.* **59**:1283-1288.
221. **Stalder, R., P. Caspers, F. Olasy, and W. Arber.** 1990. The N-terminal domain of the insertion sequence 30 transposase interacts specifically with the terminal inverted repeats of the element. *J. Biol. Chem.* **265**:3757-3762.
222. **Sternglanz, R., S. DiNardo, K. A. Voelkel, Y. Nishimura, Y. Hirota, K. Becherer, L. Zumstein, and J. C. Wang.** 1981. Mutations in the gene coding for *Escherichia coli* DNA topoisomerase I affect transcription and transposition. *Proc. Natl. Acad. Sci. USA* **78**:2747-2751.
223. **Surette, M. G., S. J. Buch, and G. Chaconas.** 1987. Transposomes: stable protein complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell* **49**:253-262.
224. **Surette, M. G., T. Harkness, and G. Chaconas.** 1991. Stimulation of the MuA protein-mediated strand cleavage reaction by the MuB protein, and the requirement of DNA nicking for a stable type I transposome formation. *In vitro* transposition characterization of mini-Mu plasmids carrying terminal base pair mutations. *J. Biol. Chem.* **266**:3118-3124.
225. **Syvanen, M.** 1988. Bacterial insertion sequences, p. 331-356. *In* R. Kucherlapati and G. R. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D. C.
226. **Syvanen, M.** 1994. Horizontal gene transfer: evidence and possible consequences. *Annu. Rev. Gen.* **28**:237-261.
227. **Syvanen, M., J. D. Hopkins, and M. Clements.** 1982. A new class of mutants in DNA polymerase I that affects gene transposition. *J. Mol. Biol.* **158**:203-289.
228. **Tavakoli, N. P., J. de Vos, and K. M. Derbyshire.** 1997. Defining functional regions of the IS903 transposase. *J. Mol. Biol.* **274**:491-504.
229. **Ton-Hoang, B., P. Polard, and M. Chandler.** 1998. Efficient transposition of IS911 circles in vitro. *EMBO J.* **17**:1169-1181.
230. **Top, E. M., Y. Moënné-Loccoz, T. Pembroke, and C. M. Thomas.** 2000. Phenotypic traits conferred by plasmids, p. 249-286. *In* C. M. Thomas (ed.), *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood Academic Publishers.

231. **Toussaint, A., and C. Merlin.** 2002. Mobile elements as a combination of functional modules. *Plasmid* **47**:26-35.
232. **Tributsch, H.** 2001. Direct versus indirect bioleaching. *Hydrometallurgy* **59(2-3)**:177-185.
233. **Tributsch, H., and J. C. Bennet.** 1981a. Semiconductor-electrochemical aspects of bacterial leaching: I. Oxidation of metal. *J. Chem. Technol. Biotechnol.* **31**:565-577.
234. **Tributsch, H., and J. C. Bennet.** 1981b. Semiconductor-electrochemical aspects of bacterial leaching: II. Survey of rate controlling metal sulphide properties. *J. Chem. Technol. Biotechnol.* **31**:627-635.
235. **Turlan, C., and M. Chandler.** 2000. Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.* **8**:268-274.
236. **van Zyl, L. J., S. M. Deane and D. E. Rawlings.** 2003. Analysis of the mobilization region of the broad-host range IncQ like plasmid pTC-F14 and its ability to interact with a related plasmid, pTF-FC2. *J. Bacteriol.* **185**:6104-6111.
237. **Vasquez M., and R. T. Espejo.** 1997. Chemolithotrophic bacteria in copper ores leached at high sulfuric acid concentration. *Appl. Environm. Microbiol.* **63**:332-334.
238. **Waksman, S. A., and J. S. Joffe.** 1922. Microorganisms concerned in the oxidation of sulfur in the soil. II *Thiobacillus thiooxidans*, a new sulfur-oxidizing organism isolated from the soil. *J. Bacteriol.* **7**:239-256.
239. **Weinrich, M. D., L. Mahnke-Braam, and W. S. Reznikoff.** 1994. A functional analysis of the Tn5 transposase. Identification of the domains required for DNA binding and multimerization. *J. Mol. Biol.* **241**:166-177.
240. **Wiater, L. A., and N. D. Grindley.** 1988. Gamma delta transposase and integration host factor bind cooperatively at both ends of gamma delta. *EMBO J.* **7**:1907-1911.
241. **Wiater, L. A., and N. D. Grindley.** 1990. Integration host factor increases the transpositional immunity conferred by gamma delta ends. *J. Bacteriol.* **172**:4951-58.
242. **Wiegand, T. W., and W. S. Reznikoff.** 1994. Interaction of the Tn5 transposase with the transposon termini. *J. Mol. Biol.* **235**:486-495.
243. **Wu, Z., and G. Chaconas.** 1995. A novel DNA binding and nuclease activity in domain III of Mu transposase: evidence for a catalytic region involved in donor cleavage. *EMBO J.* **14**:3835-3843.
244. **Yang, W., and T. A. Steitz.** 1995. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **3**:131-134.

245. **Yanisch-Perron, C., J. Viera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains:nucleotide sequence of M13mp18 and pUC19 vectors. *Gene* **33**:103-119.
246. **Yin, J. C. P., and W. S. Reznikoff.** 1987. DnaA essential host gene, and Tn5 transposition. *J. Bacteriol.* **169**:3714-3725.
247. **Zerbib, D., P. Gamas, M. Chandler, P. Prentki, S. Bass, and D. Galas.** 1985. Specificity of insertion of *IS1*. *J Mol. Biol.* **185**:517-524.
248. **Zheng, R., T. M. Jenkins, and R, Craigie.** 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization and enhances catalytic activity. *Proc. Natl. Acad. Sci. USA* **93**:13659-13664.

