Studies on a Novel Insertion Sequence, IS*Rl1*, Isolated from *Rhizobium leguminosarum* biovar *viciae*.

John O'Brien

Ph. D. 1993

Studies on a Novel Insertion Sequence, ISR*l1*, Isolated from *Rhizobium leguminosarum* biovar *viciae*.

Thesis
Presented for the Degree of
DOCTOR OF PHILOSOPHY
by John O'Brien, B.Sc. (Biotechnology)
under the Supervision of
Dr. M. O'Connell.

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of a doctorate degree is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged with the text of my work.

Signed: John O'Brien

School of Biological Sciences,
Dublin City University.
July 1993.

This thesis is dedicated to my parents,

in gratitude for their support over many years

and their steadfast belief in me.

Table of Contents.

List of Figures.

List of Tables.

ABSTRACT

An insertion sequence, IS*Rl1*, isolated from *Rhizobium leguminosarum* biovar *viciae* was studied in detail. Both strands of the entire element were sequenced. The inverted repeats (13bp) and target duplications (8bp) of the element had features in common with three other IS elements, IS*Rm2*, IS*66* and IS*866* (from *Rhizobium meliloti* and *Agrobacterium tumefaciens*).

The GENBANK and PIR sequence databases were searched for similarities to IS*Rl1*. The homologies found were to other insertion sequences, unidentified open reading frames and to several genes. All these sequences were from species of *Rhizobiaceae* and all mapped to four small regions of the element.

Two approaches were taken to identifying coding regions within IS*Rl1*. Gene search by signal identified twenty two open reading frames. The sequence of all these ORFs was translated to protein sequences and compared to the sequence databases. Four ORFs showed significant homology to five sequences from *Rhizobiaceae* none of which were insertion sequences. The second approach, gene search by content, was a statistical method based on an assumed codon bias in the element and was unsuccessful.

Five frameshifting motifs were found in IS*Rl1*, four of which were in frame with the ORFs in which they were located. Four binding sites for DnaA protein, one for integration host factor (IHF) and a potential promoter (which may be associated with one of the ORFs) were also found.

The distribution of IS*Rl1* throughout the *Rhizobiaceae* was examined by Southern hybridization. The element was found to be widespread but not ubiquitous in these species. The banding patterns observed were not sufficiently different for IS*Rl1* to be used as a DNA fingerprinting probe on its own. No homologous sequences were detected outside the *Rhizobiaceae*.

Using the element's inverted repeats as primers for polymerase chain reaction experiments, a family of related insertion sequences was discovered in the *Rhizobiaceae*. The elements ranged in size from c. 5 - 0·7kb and were present in some strains which showed no homology to IS*Rl1* in the Southern blots and absent from some strains which did.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Michael O'Connell, for his help, advice and encouragement throughout this project.

I would also like to acknowledge the help of everyone at the NCBI with the sequence analysis utilities there.

Sincerest thanks to the many people who are due it for their friendship and help at various stages of this long, strange, letter-collecting trip, especially the cast of thousands in the Good Time and Rowing Clubs.

Particular thanks are due to all my good friends, new and old, in the John Barry labs for their unparalleled company through both the mundane and the sublime days of my postgrad. They were always quick to help with technical difficulties and made every occasion memorable. I am glad to have known them and hope to see them again often.

I am deeply grateful to Dolores for her unreserved support and help on many occasions and in many ways.

Chapter 1. Transposable Elements in Prokaryotes.

Introduction.

Transposable elements have a number of remarkable properties. These properties are the subject of this review, but since many of them are interconnected, it is useful to have a brief overview before considering them in detail.

Briefly, the main features are as follows. Firstly, their characteristic property is that they can move from one locus to another which has no discernible homology to the first; a phenomenon known as transposition. Associated with transposition is a small duplication of target DNA sequence, the length of which is characteristic to each element. Secondly, they can induce various rearrangements of genetic material, including deletion and inversion of adjoining sequences and fusion of independent replicons. Their third noteworthy characteristic is that they can have strong genetic effects, in some cases switching genes on and in others switching off a gene into which they have inserted and also, in some instances, genes in which they have not.

Structurally, they consist simply of a linear stretch of double stranded DNA, the terminal sequences of which are inverted repeats of each other. One class, the compound transposons, are bound by insertion sequences (themselves bound by inverted repeats). At least some of the functions required for transposition are thought to be carried on the transposable elements and some classes carry additional functions unrelated to transposition.

Research History.

Transposable elements were discovered through the elegant genetic and cytogenetic studies of Barbara McClintock in maize (*Zea mays*) more than forty years ago.[250]

While studying the inheritance of colour and the distribution of pigmentation in plants that had undergone repeated cycles of chromosome breakage, McClintock found that the activities of particular genes were being turned

on or off at abnormal times. Because some of these genes were associated with the development of pigments in kernels, as well as in the plant itself, certain kernels were mottled, showing patches of pigmentation against an otherwise colourless background. The patterns of this variegation were reproduced in successive generations and could be analyzed like any other heritable trait. McClintock concluded that the variegation observed was the result of the action of distinct genetic units, which she called controlling elements. Direct microscopic examination of maize chromosomes containing controlling elements showed that these elements served as specific sites for the breakage and rejoining of DNA. Their physical characterization in *Z. mays* and in other eukaryotes has awaited the analytical techniques of the present.[251]

In prokaryotes the study of transposable elements dates from the discovery in the late 1960's by several researchers (most notably Malamy;[242] Saedler and Starlinger;[329,330] Brachet *et al.*[41], Adhya and Shapiro[1,348,351] and Jordan *et al.*;[198]) of a new type of spontaneous mutation in the galactose and lactose operons of a laboratory strain of *Escherichia coli* and in the early genes of bacteriophage λ. The mutations were unusual in that their effects were detectable beyond the borders of the mutated genes themselves (polar mutations). Many of these mutations were soon shown by hybridization and heteroduplex analysis to be insertions of the same few segments of DNA in different positions and orientations.[111,160] These fragments were named insertion sequences or IS elements. Their similarity to the genetic elements discovered by McClintock became clear when it was recognized that insertion sequences are natural residents of the *E. coli* genome and that the observed insertion mutations were examples of their movement to new genetic locations. Such DNA elements were shown to mediate recombination processes leading to gross DNA rearrangements such as deletion, inversion and replicon fusion.[176]

At about the same time that IS elements were discovered, other researchers observed that genes known to code for antibiotic resistance could be transferred from one DNA molecule to another. In 1974, Hedges and Jacob[151] found that the transfer between plasmids of a gene conferring

resistance to antibiotics such as penicillin and ampicillin was always accompanied by an increase in size of the recipient plasmid and that the recipient could thereafter donate this resistance to other plasmids which thereupon showed a similar increase in size. They postulated that the gene for ampicillin resistance was carried on a DNA element that could move from one molecule to another and called this element a transposon.

Prior to the discovery of transposable elements the only other type of genetic recombination known was generalized recombination (mediated by the *recA* gene), a process that relies on genetic crossover between a pair of sequences with a high degree of homology but which was not believed to be capable of producing sudden or dramatic changes in a genome. McClintock's discovery that genes were not the stable entities they had been thought, was "a discovery comparable to observing spontaneous atomic decay".[350] The discovery of bacterial insertion sequences and their capacity for modifying gene expression, sequestering genes and promoting genome rearrangements has made an important contribution to destroying the concept of the genome as a fundamentally stable entity. The concept of chromosomes as invariable entities of genetic information transmitted unchanged from generation to generation has now been abandoned in favour of a more dynamic view.

Variety, Types.

The two principal types of transposable elements are insertion sequences and transposons. Insertion sequences (IS elements) are defined as mobile genetic elements containing no detectable genes unrelated to insertion functions. Transposons (abbreviated as Tn) then are transposable elements that do contain additional, detectable genes unrelated to transposition functions. IS elements are distinguished from transposons solely on this basis, although some researchers have used the terms "transposon" and "transposable element" interchangeably.[138,214]

There is no one satisfactory method for the classification of all mobile elements. However, although it is an oversimplification, many reviews divide bacterial transpos-

able elements into three groups which do share many properties, even though there is an increasingly large number of elements which clearly do not fall into any of these groups (Tn7 is an important example). The groups are (a) the insertion sequences together with those compound transposons (see below) which have IS constituents, (b) the Tn*3* family of transposons, and (c) the transposing bacteriophage Mu and related phages. Although this review is not organized according to this scheme, the common features of these groups are described in this section.

Noting the structural and functional similarities among linear DNA plasmids, adenoviruses, bacteriophage ϕ29 and the various transposable elements found in eukaryotes and prokaryotes, Sakaguchi[332] proposed that all these elements be considered as a coherent grouping called invertrons.

IS Elements.

Because the whole range of possible IS elements in various prokaryotic organisms has not yet been detected, their definition and classification remain rather arbitrary. For the time being, any IS element in the range between 0·7 and 1·8kb is considered as regular and other IS elements are classified as exceptions. The termini carry perfect or nearly perfect inverted repeats of about 10-40bp and there are usually several open reading frames. Each of these characteristics is discussed in greater detail in later sections.

Several plasmid-associated insertion sequences have not yet been demonstrated to have chromosomal equivalents. Many (but not all) of these elements form part of compound (or composite) transposons. These are elements composed of two flanking insertion sequences in direct or inverse orientation and a large central, non-self-transposable DNA segment. For example, the inverted repeats of Tn*9* are in fact copies of IS*1*.[237] A large variety of interstitial genes have been observed as part of compound transposons, although the majority of known transposons of this type carry antibiotic resistance genes. This bias may not be representative of natural populations, but may, rather, reflect the widespread use of antibiotics in medicine and animal husbandry over the past thirty years and also the ease of

manipulation of this type of element in the laboratory.

Although they do not cite any examples, Galas and Chandler[124] say that several IS elements have only been observed to occur alone (as part of the bacterial chromosome or plasmids), others are found both alone and in compound transposons, while others, despite subsequent demonstrations that they can transpose autonomously, have only been detected as part of compound transposons.

Although many ISs are known, only a few have been subjected to mechanistic studies. Mainly IS*1*, IS*10*, IS*50* and IS*903*, which are constituents of the composite transposons Tn*9*, Tn*10*, Tn*5* and Tn*903* respectively. Although these ISs are very similar in their recombinational behaviour and genetic organization, there is no indication that they are related to one another; generalizations, therefore, have to be treated with caution.

The Tn*3* Family.

An entire group of transposons has been described whose similarity in structure and function suggest a common evolutionary origin. Tn*3* is the best characterized member of this group and Sherratt in his review of the group,[354] used the term Tn*3* to refer to both Tn*3* itself and its close Ap^r (ampicillin resistance) relatives such as Tn*1* and Tn*2*, as he considered that some of these elements have been confused with each other during distribution and subsequent use.

All the elements of this class generate 5bp repeats of target DNA during insertion; most have 38 to 40bp repeats at their ends, and those repeats are very similar among different elements. They are all more than 5kb long and most encode accessory determinants in addition to transposition functions. They carry two nonoverlapping transposition genes (*tnpA* and *tnpR*) and an important site called *res*, which is involved in the transposition process. These functions are described in more detail below.

Three members of the Tn*3* family are worth mentioning at this point, IS*101*, Tn*1000* and Tn*1721*. IS*101* is the shortest known insertion sequence (209bp), it consists only of two terminal repeats and a *res* site and depends on other elements for transposition functions.[256] Fewer than 10 nucleotides in this element are nonfunctional for transpos-

ition. Tn*1000* is another cryptic transposon. It was classified as an insertion sequence and called $\gamma\delta$[144] but has fundamental differences to regular IS elements. It is 5·7kb long and is more complex than regular IS elements, but still lacks detectable genes unrelated to insertion functions. It shows a close structural and functional relation to the ampicillin resistance transposon Tn*3* and for these reasons was reclassified as a transposon. Tn*1721* is a composite transposon with a copy of Tn*3* at either end.[341]

Transposons of the Tn*3* family can be divided into two subgroups based on their genetic organization and on the ability of recombination functions to complement one another.[354] One group is represented by Tn*3* itself and Tn*1000*, the other subgroup is represented by Tn*501* and Tn*21*.[143]

Bacteriophage Mu.

In 1963, Austin Taylor reported the discovery of a peculiar type of bacteriophage which could integrate into the chromosome at multiple sites,[379] thereby causing many different types of mutations in the host bacterium. Because of this property Taylor called his phage Mu, for "mutator". Further studies have shown that Mu is actually a transposable element that can also exist as an infectious virus. Unlike λ, phage Mu's reproduction depends entirely on its ability to move through the bacterial genome and combine with many regions of the host DNA.

The genome of phage Mu (and related phages) is integrated into host sequences at all stages in its life cycle, including when packaged as mature virions. The phage DNA is a linear molecule of about 38kb of Mu DNA with about 1·5kb of host DNA at one end and 50-150bp at the other. These attached host sequences reflect the life style of the phage - to replicate, it undergoes repeated cycles of transposition in the cell (about one hundred events per cell during the lytic cycle), and is finally packaged along with adjacent host sequences into virus particles.[381] As with other transposons, Mu integration results in duplication of a small target sequence (5bp). However, in contrast to other transposons, the ends of Mu are not inverted repeats,[201] even though they are clearly the sites of recombination and are both bound by purified Mu transposase (the A protein) *in vitro*.[78]

A related phage, D108, shares 95% sequence homology, similar functional organization, and partial functional cross reaction with Mu, although the two are heteroimmune.[285] Mutator phages that may be analagous to Mu and D108 have also been identified in *Vibrio cholera*[194] and *Pseudomonas*.[3]

Bacteriophages such as λ are often included in reviews of transposable elements[212,350] and in fact Campbell *et al.*[52] included λ in their definition of transposable elements. The similarities are obvious but there are also important differences. For the purpose of this review, λ will not be considered as a transposable element but will be referred to for the purposes of comparison or contrast. Also, even though it is a well studied element and could fittingly be the subject of several reviews on its own account, phage Mu is a very specialized element and so conclusions drawn from studies on Mu are not always applicable to the other mobile elements and for that reason this review does not examine phage Mu in detail but focuses, rather, on the IS elements and transposons.

Nomenclature.

Campbell *et al.*[52] drafted a set of rules for the nomenclature of transposable elements based on previously accepted conventions for the Mu prophage,[167] for bacteriophage λ,[157] for bacterial plasmids[92] and for bacterial genetics.[272] Some of their proposals, which have since been adopted, are summarized here.

The designation for an insertion sequence is the letters IS and an italicized number, and for transposons the designation is the letters Tn and an italicized number. Genes carried by a transposon can be indicated after the transposon name, e.g. Tn*9* *cam*$^+$ (Tn*9* carries a gene that confers chloramphenicol resistance). Mutations in inserted elements are listed after the symbol for the element e.g. Tn*9* *cam4* is a mutant of Tn*9* that no longer confers chloramphenicol resistance.

Whereas all insertion sequences that appear indistinguishable by hybridization, heteroduplex or restriction analysis (but not nucleotide sequencing) are referred to by

one name, only those transposons that have been shown to be identical through sequence analysis are grouped under the same name; i.e. IS designations are generic while Tn designations are specific.

The designation for insertion of an IS element into a genome or a particular gene or region is a double colon, for example,

λ::IS2 (IS2 has inserted into phage λ);

F8 42kb-7::Tn2 (mutation 7 at 42kb on F8 plasmid is a Tn2 insertion).

Orientation can be designated with respect to direction along the genome using a plus or minus sign after the symbol of the element (the plus orientation is arbitrary and the minus is opposite to it); or with respect to the polarity of the operon using Roman numerals I and II (again orientation I is arbitrary).

Distribution and Occurrence.

Although transposable elements have been most thoroughly studied in bacteria, they have been found in most organisms that have been examined and in a variety of contexts. In bacteria, in addition to their presence on the chromosome, insertion sequences were found on bacteriophages such as λ[41] and P1[175] and on plasmids such as the fertility factor, F,[168-170] and the R factors.[169,297] They are particularly frequent as components of natural plasmids. In the case of the F plasmid, IS elements have been implicated in the formation of Hfr strains and F' plasmids.[124] IS elements often appear to occur more frequently (per unit length of DNA) in plasmids than in bacterial genomes.[124]

The life cycle of retroviruses involves events very similar to transposition, they are bounded by repeat sequences, generate target repeats and when in the provirus stage are capable of movement throughout the genome and of promoting genetic rearrangements. Evolutionary relationships have been established between retroviruses and transposable elements[386] and it is tempting to think that the evolution of the former may have in some way have involved the latter as is generally accepted to be the case with

phage Mu.[285] As a class, these elements are called retrotransposons (sometimes retroposons) which comprise elements ranging from the retroviruses to elements with features of retroviruses but without the ability to code for transposition proteins.

Prokaryotes.

IS elements are normal constituents of bacteria such as *E. coli*. and, although this review is biased towards those transposable elements found in enteric bacteria, because of the preponderance of published work on these organisms, insertion sequences have been found in all bacterial species examined. In fact, they can be found in the genomes of many different bacteria at multiplicities of between a few and a few hundred per genome, (based on the copy number of the known IS elements, it has been calculated that at least 1-2% of the *E. coli* K12 chromosome consists of the various IS elements).[176]

There are documented examples in many of the other enterics, such as *Shigella*,[278] *Klebsiella*,[273] *Serratia*,[273] *Salmonella*,[227] *Yersinia*,[73] *Citrobacter*[269] and *Enterobacter*.[269] Less information is available concerning the occurrence and distribution of insertion sequences in nonenteric gram-negative bacteria. Several insertion sequences and sequences having structural features clearly related to those of ISs have been described in the *Rhizobiaceae*, these are described in some detail in the last section of this review. At least twelve different insertion sequences have been detected in the pseudomonads,[124] another important group of gram-negative bacteria.

Transposable elements are also common among the gram-positive bacteria. Many of these elements have counterparts among the gram-negative elements, but there are many novel mobile elements found in the gram-positive bacteria. These include an element called the 2·6kb minicircle, which is found in *Streptomyces coelicolor* as free circles and is capable of chromosomal integration at different sites.[236] Another example is Tn*1545* from *Streptococcus pneumoniae* which encodes a conjugal transfer system.[74] (This property of Tn*1545*, and its relative Tn*916*, are described on page 34.) The elements found in *Staphylococcus aureus* are particularly important in that this organism is a major cause

of nocosomial infection and death in hospitals world-wide and many examples of transposon-encoded drug-resistance are known in *S. aureus*.[263]

Several insertion sequences have been found in *Halobacterium halobium*, a species of archaebacteria. One of these, ISH*1*, was shown to be present in other species of halobacteria also.[63]

Transposable elements often exist in multiple copies per cell. The copy number of IS*1*, one of the best documented insertion sequences, can vary from strain to strain. Most laboratory strains carry between six to ten copies, while *E. coli* C (from two sources) has three, and certain strains of *Shigella* spp. have been estimated to carry more than fifty copies, perhaps as many as two hundred.[273,278] It is interesting that the high copy number of ISs in *Shigella* spp. is not restricted to IS*1*, relatively large copy numbers of many previously known and novel ISs can be detected in a single strain.[248] Like most *Salmonella* strains, *E. coli* W is apparently devoid of IS*1*. Similar variations in copy number have been observed in natural isolates of *E. coli*.

In general, the copy number of other insertion sequences is lower than for IS*1*. Although certain individual natural isolates of *E. coli* have been shown to carry up to seventeen copies of IS*2*, fourteen copies of IS*4*, and twenty one copies of IS*5*, the average copy number in a collection of seventy one independent isolates was determined to be 6·4 for IS*1*, 2·7 for IS*2*, 1·7 for IS*3*, 2 for IS*4*, 1·3 for IS*5* and 0·9 for IS*30*.[336] The average copy number of IS*5* was found to be significantly higher in a collection of *Enterobacteriaceae* isolated in the "preantibiotic era" (between 1917 and 1954 - this is the Murray collection),[83] an observation which may reflect accumulation of copies during storage of the strains.[133] Extensive data concerning the distribution and genomic copy number are available for IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, IS*30*, and IS*200*.[124]

Measurement of copy number by the techniques normally used (hybridization) does not imply that the insertion sequences are identical or even active. Sequence divergence has been observed in the case of copies of IS*1*-like sequences analyzed in *Shigella dysenteriae*, where three major classes of IS*1*-like sequences have been shown to occur

having approximately 99, 90 and 60 percent homology with the original IS*1*.[275,278]

Eukaryotes.

Although it is outside the scope of this review, it is worth pointing out that mobile elements are also widespread in eukaryotes. In the book "Mobile DNA"[19] there are several chapters dealing with mobile elements in organisms such as yeast, maize, snapdragon, fruitflies, nematodes, cellular slime molds and frogs. Together these organisms form a representative cross-section of the eukaryotic spectrum.

Transposons (Ty elements) are found in both nuclear and mitochondrial DNA in *Saccharomyces cerevisiae*, though no single transposon has been found in both. The Ty elements have terminal direct repeats, generate target duplications and exist in multiple copies throughout the genome. Most yeast transposons discovered so far are retrotransposons.

Transposable elements in the fruitfly *Drosophila melanogaster* (P elements) are used for germ line transformation by the injection of DNA into embryos. These P elements share structural similarities with mobile elements found in *Zea mays* (maize), *Antirrhinum majus* (snapdragon) and *Xenopus laevis* (frog).

As with prokaryotic species, it is quite probable that mobile elements of one form or another will be found in any eukaryotic species examined.

Host-Specificity.

An organism in which a particular IS element was discovered originally need not necessarily be the original host. The same kind of IS element can be found on plasmids, phage genomes and the chromosomes of different bacteria. Sometimes the isolates from different species represent variants (see, for example, ref. 264).

Several IS elements can function efficiently in bacteria that do not contain the same IS element in their chromosome, for example IS*10* is active in both *E. coli* K12 and *Salmonella typhimurium* although their chromosomes contain no IS*10* sequence.[315] Other IS elements, such as IS*R1*, are active in one microorganism but not in others.[294]

IS*200* appears to be restricted to *Salmonella* sp. strains, although material that shows some low-level cross

hybridization has been observed in two *Shigella* sp. strains.[228,248] Other IS elements, (i.e. IS*1*, IS*2*, IS*3*, IS*4*, IS*5*, IS*30*), appear in members of the *Enterobacteriaceae* such as *E. coli* and *Shigella* spp., but not in most *Salmonella* spp.[124] In view of the relatedness of *S. typhimurium* and *E. coli* and the documented transfer of genetic information between these species, this apparent species specificity is somewhat unexpected.

Structure.

As mentioned earlier most IS elements studied range in size from 0·7 to 1·8kb, they terminate in inverted repeats of about 10-40bp and they usually possess several open reading frames. The smallest known of these "regular" insertion sequence is IS*1*, which is 768bp in length, IS*200* is also about this size.[124] Ohtsubo and Ohtsubo[280] have sequenced IS*1* in its entirety.

Transposons are distinguished from insertion sequences because transposons carry detectable genes other than those required for transposition functions, but it is clear that there is a close relationship between IS and Tn elements.[50]

Inverted repeats.

Transposable elements carry inverted repeat sequences at their ends. The repeats vary from 8bp to 40bp and may show perfect homology (38/38 for Tn*3*) or imperfect homology (12/18 for IS*50*).[138]

Transposons often terminate in long (800-1500bp) inverted or direct repeats and these repeats are IS or IS-like elements. Many transposons thus represent a segment of DNA that is mobile as a result of being flanked by IS units, such transposons are called compound or composite transposons. Since the IS elements themselves terminate in short inverted repeats, a transposon that is flanked by two IS elements, even in direct repeat as in Tn*9*, will also be flanked by inverted repeats. Virtually all the transposable elements analyzed at the sequence level have a terminal inverted repeat.[50] However, in contrast to other transposons, the ends of Mu are not inverted repeats, homologies do occur, but are not symmetrically arranged.[201] Other exceptions are Tn*554* from *S. aureus*[264] and IS*902* from *Mycobact-*

erium avium.[261]

The ends of transposable elements must be intact for transposition to occur. Deletion of one or both ends produces a noncomplementable transposition defect.[114,153,183] However, Machida *et al.*[238] reported the generation of new transposons flanked by a copy of IS*102* at one end, whereas the other end was formed by a very short sequence partially homologous to the end of IS*102* and carried in inverse orientation. Apparently, the IS*102* transposition enzyme can recognize not only the inverted repeat of IS*102*, but also a sequence with homology to the inverted repeat and that is carried on the donor replicon. This finding may imply that IS-mediated DNA rearrangements can sometimes initiate at sequences only partially homologous to the inverted repeats of the IS element, thereby generating a wider variety of gene rearrangements.

Comparison of the available sequences of inverted repeats reveals several patterns of homology. When grouped according to the size of the repeat generated upon integration there are regions of homology shared by many members of each group and this has been proposed as a criterion for classification.[50]

Compound transposons.

Compound transposons carrying closely related antibiotic genes but flanked by different insertion sequences are not infrequent. The aminoglycoside phosphotransferase gene, which specifies resistance to aminoglycosides such as kanamycin, has been found flanked by inverted or directly repeated copies of IS*903* (Tn*903* and Tn*602*, respectively),[271,372] by direct repeats of IS*1* (Tn*2350*),[65] by direct repeats of IS*6* elements (Tn*6*)[21] or by direct repeats of IS*15* (Tn*1525*).[226] In all these cases significant sequence divergence is observed both upstream and downstream from the aminoglycoside phosphotransferase gene and in one case, Tn*2680*, an IS*903*-like sequence is also located between the flanking IS*26* elements.[257]

Insertion sequences are also versatile, or promiscuous, in that a given IS is often found flanking different interstitial genes. Perhaps one of the most promiscuous is IS*1*, it forms part of the chloramphenicol resistance transposon Tn*9*,[237] the kanamycin resistance transposon Tn*2350*,[65]

the heat-stable enterotoxin transposon Tn*1681*;[362] it is part of compound transposons encoding iron uptake[70,253] and of a large transposon of unknown function apparently resident in the chromosome of *E. coli* HB101.[314]

Tn*5*, Tn*903* and Tn*1681* carry almost no extraneous sequences which are neither part of the IS elements nor the accessory element.[11,136,362] Tn*10* contains about 4kb whose function is not known, but the repeated isolation of tetracycline-resistance transposons having the same size and structure as Tn*10* argues that this extra information does play a role in the natural environment.[200]

In most composite transposons, it has been directly confirmed that all of the information pertinent to transposition is encoded in the IS modules themselves and that one or both IS modules are still capable of independent transposition as individual units.[214] Studies with Tn*5*[20] and with Tn*10*[50] indicate that these transposons encode trans-acting transposition functions in their long IS-like inverted repeats. The evidence is that one of the inverted repeats units differs slightly from the other which is dominant in supplying transposition functions. It is highly probable that the long inverted repeats of these transposons are, or once were, independent insertion sequences.

Functional Properties.

Duplication of target sites.

Another characteristic of the majority of known transposable elements is that they generate small, directly repeated duplications of the target DNA at the point of insertion. This was shown first for the integration of IS*1* into the *lac* operon[49] and the *gal* operon.[134] The length of this duplication is a characteristic proper to each element and ranges from 2 to 13bp.[124] Transposition of IS*1*,[134] Tn*5*,[338] Tn*9*,[196] Tn*10*[213] and Tn*903*[281] results in a 9bp sequence of the target sequence being duplicated on either side of the inserted element; IS*2*,[128] Tn*3*,[378] Tn*501*,[206] Tn*551*,[50] Tn*1000*[304] and Mu[4] result in a 5bp repeated sequence.

When a transposable element located at a single site integrates into other regions of the DNA, numerous integration points are found. The analysis of their distribution

and the different repeated sequences[124] indicates that the repeat sequence is generated during the process of integration rather than resulting from a recombination event between a sequence in the transposable element and a homologous sequence in the target DNA, as is the case for λ integration.[229] Thus, when presented with a target containing the same 9bp which are present on both sides of the donor element, Tn*9* does not preferentially integrate into this 9bp sequence.[196]

The direct repeats do not appear to play any structural role in transposition, since experimentally-constructed elements not flanked by direct repeats still transpose and mediate cointegrate formation and deletion.[204,276] Calos and Miller[51] showed that Tn*9* derivatives, flanked by differing nine base pair sequences, transpose normally. Similarly Tn*3*,[279] Tn*10*[213] and Tn*1000*[304] have been shown to transpose without being flanked by a repeated sequence. Precise excision, however, is thought to require this small duplication.[116]

It is simplest to propose that the origin of the target duplication is that a staggered cut is made in the target DNA and the transposable element attaches to the protruding end of each strand. Filling in the staggered cuts by repair synthesis would result in the repeated sequence.[50] The element IS*91* from *E. coli* produces no target duplication,[97] but there is no theoretical difficulty in assuming that this element produces blunt ended cuts in the target sequence.

All of the several models proposed for the mechanism of transposition of mobile elements account for the flanking duplications of a short target sequence in this way. Because the size of the duplicated segment is characteristic of the transposon, it seems likely that the transposase itself either cleaves, or directs cleavage of, the target DNA.[138]

Although many elements appear to induce a duplication of a fixed number of base pairs, four exceptions have been reported: IS*1*, IS4, IS*21* and IS*186*. IS*1* will serve as an example, the majority of sequenced IS*1* insertions (52/68) generate a duplication of 9bp. However, duplications of 7bp (one example), 8bp (thirteen examples), 10bp (one example)

and 14bp (one example) have been observed, and an example of an 11bp duplication has also been reported.[124]

Two proposals have been advanced to explain these observed variations.[174] One possibility is that cleavage of the target occurs with a 9bp spacing, but that in some situations, bases are removed from one or the other of the cleaved target ends by nuclease action. An alternative possibility is that duplication of 7, 8, 10, 11 and 14bp reflects a variation in the spacing of the initial cleavages, which has been attributed to small sequence variations in the different IS*1* elements used. Factors such as DNA helix geometry (known to be influenced on a local scale by nucleotide sequence)[98] may also be involved. Local variations in parameters such as helix twist or flexibility may well play a subtle role in target cleavage, although no systematic study of this aspect has been reported. Since variation from the 9bp IS*1*-induced target duplication is not uncommon, similar variations may also occur with other insertion sequences for which few sequenced insertions are available.[124]

DNA Rearrangements.

The characteristic property of mobile genetic elements is essentially that they are capable of generating a new join between nonhomologous DNA segments. This enables them to insert themselves into unrelated segments of DNA independently of the host's recombination systems, but they are also observed to mediate other rearrangements leading to the formation of novel genomic configurations. So that, as well as inserting into and excising from non-homologous DNA, they can promote the deletion of genetic material, the fusion of two separate and independently replicating replicons, the transposition of segments of DNA from one replicon to another (or within the same replicon) and the inversion of segments of DNA. As will be seen in the following sections, some transposable elements can participate in specific rearrangements of DNA more frequently than they do in transposition events. Because IS elements are not of themselves phenotypically detectable, their activities are often experimentally monitored as cointegration or deletion rather than as transposition itself.[176]

Combinations of these DNA rearrangements would lead to

more complex DNA restructuring and transposable elements also contribute to DNA rearrangements by providing homologous DNA for general recombination systems. However, when carefully examined, not all elements are found to exhibit all these capabilities. The relative frequencies of occurrence of each vary widely among mobile elements.

Most prokaryotic transposons promote transposition and rearrangements at frequencies of 10^{-4} to 10^{-7} per generation. These low frequencies are attributable to stringent regulation, inefficient translation/transcription and/or inefficiencies intrinsic to the transposition process itself.[214]

It is important to keep in mind that the pathways by which each end product arises may be multiple and include quite distinct processes. It is thought that in some cases the rearrangements observed may be the result of defective transposition pathways and in others they may represent additional properties of the element in question. Also in many transposition models, deletions and duplicative inversions arise as the intramolecular consequence of the same steps that yield cointegrates by intermolecular transposition.

All the above rearrangements are known to arise in the absence of *recA* function and must in general be accounted for as direct transposition products. However, suitable combinations of transposition and homologous recombination can generate products indistinguishable from those generated by transposition alone and can interconvert several of those products.[214]

(i) Insertion.

Each insertion of a particular element contains exactly the same set of nonpermuted transposable sequences. Thus, each insertion event involves the joining of exactly the same nucleotides at the ends of the mobile element to a broken target DNA molecule. The same nucleotides at the ends of the element are also joined to target DNA sequences during other transposon-promoted events.*

Most of the translocatable elements can insert at readily detectable frequencies into many different sites. The absolute frequencies of translocation of different

* References 2,50,124,138,176,214,368

elements vary over several orders of magnitude. Integration of phages λ and Mu are extremely efficient, consistent with the important role of these events in the life cycles of these phages.[75] By contrast, translocation of IS elements into the *gal* and *lac* operons of *E. coli* can be found in about one per 10^7 cells in a culture;[242,369] this is the sum of frequencies of insertion for all IS elements in the cell and does not reflect heterogeneities among different elements in their translocation frequencies or in their specificities for different targets. This frequency is comparable to the spontaneous mutation rate in *E. coli*, and IS insertions do appear to constitute a significant proportion of spontaneous mutations.[370] Early estimates of the frequency of insertion mutations were made for the *galK* region,[199,348] the *lac* region,[243] phage P2[231] and the *cII* region of phage λ.[32] In each case, it was estimated that between five and fifteen percent of spontaneous mutations are IS insertions. It is important to realize that specificity of insertion can vary greatly from gene to gene and from element to element.

The transposition frequencies of individual IS elements vary between 10^{-3} or, perhaps more usually, 10^{-5} and 10^{-9} per cell division.[176] Many factors can influence transposition frequencies. Factors such as the location of the DNA target (chromosomal or extrachromosomal), the specific affinity of the target for a given insertion sequence (insertion specificity), the variety and number of insertion sequences carried by the bacterium in question, the growth conditions of the culture and the selection or screening procedure are all expected to influence frequency estimation.[124]

Different antibiotic resistance elements also translocate at different frequencies. Translocation of Tn*5*, Tn*9* and Tn*10* from the genome of phage λ to the *E. coli* chromosome and translocation of Tn*5* and Tn*10* from phage P22 to the Salmonella chromosome have been tested under comparable conditions. The frequencies of translocation per infective phage are approximately 10^{-3} (Tn*5*), 10^{-5} (Tn*9*) and 10^{-6} (Tn*10*) under conditions where all infected cells survive infection.[212]

Actual transposition frequencies may be greater

than observed if a fraction of cells does not survive the transposition process, as a consequence of insertion into (or deletion of) essential genes.

It has not been possible to observe directly successive cycles of translocation of any particular IS element, because multiple copies of each type of insertion sequence are present in *E. coli*, and there are as yet no useful genetic markers on any one sequence. Tn*10* (a tetracycline resistance element) has been shown to translocate from its original position in an extrachromosomal plasmid (R factor) to the genome of *Salmonella* phage P22, from P22 to known structural genes in the *Salmonella* chromosome, from the *Salmonella leu* operon into coliphage λ, from λ into the *E. coli trp* operon, and from there back to λ again.*

(ii) Excision.

Insertion of a transposable element into a gene inactivates the function of that gene. For most of the transposable elements revertants can be detected at rates varying between 10^{-9} and 10^{-6} in the population. Reversion is associated with loss of the inserted elements and it is assumed that in most cases the duplication of target nucleotides is also eliminated, this is what is meant by precise excision.[50] A different type of excision is observed that relieves the polar effects but does not restore the function of the particular gene that carried the insertion. This is called *nearly precise* or, more commonly, *imprecise excision*.[317] For elements other than λ, imprecise excision occurs much more frequently than precise excision. Partial excisions are either not observed or are attributable to other processes.†

The frequency of reversion is generally quite low, varies from element to element and depends also on the position of the insertion in an operon; but not all mutations due to IS insertion revert at measurable rates.[124] For IS*1*, reversion may be as high as 10^{-3}/cell plated.[198] IS*4* inserted in *galT* reverts at 10^{-9}/cell plated.[368] The reversion of Mu integrated into *galT* is detected at a frequency below 10^{-10}/cell plated.[199]

* References 40,115,212,216,353

† References 16,40,45,216

For Tn*10*, reversion frequencies for different mutations can vary as much as one hundred fold even among closely linked insertions in a single gene where the element is in the same orientation in all insertions.[217] Mutants generated by Tn*5* show similar variability.[16] These results suggest that the junctions between the elements and the adjoining chromosome, which will differ from one insertion to another, directly affect precise excision.[212]

The failure to detect transposition events associated with reversion led to the idea that the two processes are independent.[16,116] This has been verified for Tn*5*[104] and Tn*10*,[215] by experiments with derivatives carrying inactivated transposition functions. When integrated these altered transposons are still eliminated by reversion events that restore gene function. Therefore precise excision of a transposable element from a gene, being detected as a reversion of the latter, may not be an intermediate step in transposition of the element, but rather a different event either unrelated to transposition, or deviating from its normal pathway.[368]

Excision is also largely independent of host *recA* functions.[104,111] Host functions must be involved in some way because several chromosomal mutations affecting precise excision have been identified.[165] The discovery that many spontaneous deletions in *E. coli* occur between short repeated sequences[109] has led to the suggestion that the same host pathway involved in spontaneous deletion formation may be responsible for eliminating inserted transposable elements,[196] since the latter are found between short repeated sequences. This does not easily explain the wide range of reversion frequencies observed with Tn*5* or Tn*10* and leaves wide open the question of whether the long inverted repeats of such elements play a passive role in the precise excision of an inserted element.[50]

It is thought that excision occurs following intrastrand pairing between inverted repeat sequences, leading to the extrusion of the intervening DNA segment and its deletion by replication slippage across the small directly repeated target sequences (generated by the insertion of the element). In this model, excision should be influenced by factors affecting the efficiency of formation

and repair of extruded DNA. The necessary intrastrand pairing might be expected to be facilitated within single-stranded regions of DNA and indeed, excision of elements carried by the F plasmid appears to occur at a much higher frequency than a corresponding chromosomal insertion[104,166] and is stimulated by conjugation,[23,377] a process involving the transfer of single-stranded DNA. Tn*5* insertions in single-stranded bacteriophage have been reported to be unstable.[23,155] Factors affecting the stability of the extruded structure would also be expected to influence excision. The excision frequency has been observed to decrease with decreasing length of the inverted sequences.* This length dependence could explain the differences in frequency of precise excision observed from element to element.[166,233] In addition, extrusion of palindromic sequences in a model plasmid-based system has been shown to depend both on supercoiling and on the local sequence environment.[375] Such context effects may explain the observed dependence of excision on the location (nucleotide sequence environment) of the insertion element.[116,166] Many host mutations which increase excision are in genes implicated in DNA repair. Although their precise role is not known, such (wild-type) functions could intervene by repairing intermediate structures generated by replication slippage.[124]

In no case is the fate of the excised DNA known, it is rarely, if ever, reinserted elsewhere in the genome.[2,16,284] For Tn*5* and Tn*10*, precise reconstruction of the "donor site" does not seem to be integrally linked to translocation: precise excision of these elements occurs ten to ten thousand fold less frequently than translocation, and very few precise revertants carry the element reinserted at a new location (10^{-2} for Tn*5*, 10^{-3} for Tn*10*).[16,40]

In conclusion, precise excision of transposons, resulting in the reconstruction of an uninterrupted target gene to its wild-type sequence, occurs at much lower frequencies than transposition, is not transposon-promoted (except in Mu, where excision is linked to the lytic cycle) and is not usually accompanied by reinsertion of the trans-

* References 23,69,104,116

poson at a new position. These observations do not rule out the possibility of suicidal excision from the donor chromosome during transposition (see the section on models for transposition below).[214] A better knowledge of the relation of precise excision to transposition will also be necessary if we are to understand why the number of IS elements within the *E. coli* K12 chromosome is relatively constant, as has been found for IS*1* and IS*2*,[327] IS*3*[94] and IS*4*.[211]

(iii) Cointegrate formation.

Cointegrate formation, the fusion of two replicons, was initially observed in the case of members of the Tn*3* family of transposons[354] and bacteriophage Mu[285] and is perhaps the best-studied type of genetic rearrangement induced by transposable elements. The resulting cointegrate structure is composed of donor and recipient replicons separated at each junction by a single, directly repeated copy of the transposable element. Cointegrates that carry directly repeated copies of an IS element at the junctions have been isolated even if only one of the two parental replicons contained this element.[124] The fused replicon can resolve to yield two replicons, each carrying one copy of the IS element that had mediated the process. Duplication of the target DNA sequence is also observed in cointegrates.[176]

Cointegrates generated by composite transposons may contain, as the duplicated element, the entire transposon (a Tn cointegrate) or either one of the individual IS modules (IS cointegrates). The three types of cointegrate structures occur with different relative frequencies for different elements; for Tn*5*,[24] Tn*5* > IS*50*-R > IS*50*-L; for Tn*9*,[214] IS*1*A > IS*1*B >Tn*9* and for Tn*903*Δ1,[137] IS*903* > Tn*903* > IS*903*Δ1. Variables that might account for these differences include: dependence of transposition or cointegration on the length of the element; preferential *cis*-action of functions on the IS module from which they were transcribed/translated; effects of nontransposon sequences adjacent to the ends of IS elements or intrinsic differences in relevant sites on different transposons.[214]

Note that replicon fusions can also be generated by inverse transposition, in which the DNA segment located between repeated insertion sequences carries an entire copy

of the plasmid. Such structures can be generated by dimerization of the donor plasmid followed by transposition of one of the plasmid copies flanked by two insertion sequences,[18,34] or, more simply, if a plasmid contains a compound transposon as the following example illustrates.

An unusual example of transposition-generated replicon fusion is the case of IS*21*, which exhibits significant activity only when present as a directly repeated tandem dimer.[310,399] For convenience the duplicated sequence is referred to as IS*21.2*. The sequence IS*21.2*, isolated from the plasmid R68.45, is able to promote replicon fusion at frequencies approaching 10^{-3}. The original duplicated copies of IS*21* in IS*21.2* are separated by 3bp. Cointegrates generated by this element have only one copy of the IS*21* element at each junction[309] and have lost the intervening three base pair sequence present in the IS*21.2* element. Replicon fusion by IS*21.2* thus appears to occur by a simple transposition of the entire plasmid using two neighbouring ends, one from each copy of IS*21*.[124]

The products of IS*6* transposition appear to be exclusively cointegrates,[44,383] whereas cointegrates appear to occur, if at all, as only infrequent products of the transposition of IS*10*, IS*50*,[161,162] IS*91*[96] and IS*3411*.[185]

Simple insertions of a transposable element may be related to cointegrate structures in two possible ways. Stable, covalently closed cointegrate structures could be normal intermediates in transposition, and such cointegrate molecules could then be reduced by reciprocal recombination to give back an unaltered donor genome and a simple insertion product (i.e. they are only susceptible to efficient resolution during the actual transposition event itself when appropriate structures and/or functions may transiently exist). Alternatively, cointegrates and simple transpositions can arise as alternative products of a single starting structure or transposition intermediate. Some elements clearly use the first pathway, others are likely to use the second.[214]

An example of the former case is the intermolecular transposition of Tn*3* and related transposons. Transposition of these elements depends on the two genes mentioned previously and proceeds in two distinct stages.[135]

First, through the action of the *tnpA* gene product, donor and target replicons fuse to form a cointegrate. This cointegrate is a transient intermediate and, by site-specific recombination mediated by the resolvase (*tnpR*), is rapidly broken down at the two *res* sites. This yields the final transposition product (a simple insertion of the transposon in the target replicon) and regenerates the donor replicon.[354]

Ordinary insertion sequences normally generate both simple transpositions and cointegrate structures at reasonable frequencies, although the proportions of the two products vary from element to element. In contrast to the Tn*3* family of elements, fused replicons generated by most insertion sequences are relatively stable in the absence of the host homologous recombination system (and even in *recA*$^+$ hosts).[22,137] These properties are consistent with either of the two possible relationships between simple insertions and cointegrates described above. The observation that IS*1* (Tn*9*)-mediated cointegrates are stable, while transposition of Tn*9* (without accompanying replicon fusion) is a relatively frequent event, led to the proposal that cointegrates cannot be obligatory intermediates in the transposition of IS*1*.[123]

(iv) Deletion.

The capacity of certain transposable elements to generate deletions of neighbouring DNA was first noted for an IS*1* located in the *gal* operon. Reif and Saedler[305] showed that IS*1*-mediated deletions in the *gal* operon could form on either side of the element (at frequencies 30- to 2,000-fold higher than the spontaneous rate), seemed to retain the IS*1* and had endpoints in the neighbouring chromosome that were variable but not random. At the nucleotide sequence level, Ohtsubo and Ohtsubo[280] showed by DNA sequence analysis that the endpoint near IS*1* is located precisely between the last nucleotide of the element and the first nucleotide of adjacent DNA (removing one of the flanking repeats[51,280]). Subsequent studies with other transposable elements showed that they could also mediate deletion formation and that the element seemed to be left intact.[305]

The large deletions generated by IS*4* seem to differ in structure from the usual IS-generated deletion.

These deletions remove the IS element and adjacent DNA at both termini to varying extents. They are found at a frequency of more than 10^{-3}. Their origin is presently unexplained.[208]

In the case of IS*903* and IS*102*, it has been found that deletion results in the appearance of an additional copy of the element. Both products of the deletion event (the plasmid carrying the deletion and the deleted fragment) have been isolated and both carry copies of the insertion element.[27,29,392]

When selection for loss of antibiotic resistance[217] or relief of polarity are employed,[16,113] elimination of parts of compound transposons can be detected. Many of these events can now be understood as resulting from deletions or inversions mediated by the internal ends of the IS-like inverted repeat segments[316] or from events resembling spontaneous deletions mediated by host pathways.[317]

Different elements promote deletion formation at different frequencies. For the insertion sequences and translocatable drug-resistance elements, the frequency of deletion formation is often as high or higher than the frequency of translocation and is often well above spontaneous levels. IS*1*, for example, can increase deletion frequencies in *gal* as much as two thousand fold (to 10^{-4} per cell).[305,306] In general, the frequency of deletions adjacent to IS*1* varies with the position of the element. For this element, deletion formation is temperature-dependent,[267] at 30°C they can occur at 10^{-3}/cell plated, while at 37°C no adjacent deletions are detected above the background of deletions formed in the absence of IS*1*.

In considering the mechanism for deletion formation by a transposable element, the sequence results, site specificity and the requirement for both ends of the transposable element[106] are consistent with the event being closely related to transposition. That this was the case for IS*1* was suggested by the observation that IS*1* (in Tn*9*)-mediated deletions into the *lacI* gene terminate preferentially within regions that are preferred for Tn*9* insertion[316] and by the observation that impinging, external transcription reduces IS*1*-mediated transposition, cointeg-

rate formation and deletion formation.[124] However Nevers and Saedler[268] have described a mutant in which the frequency of deletion formation adjacent to IS*1* is decreased one hundred-fold; transposition is not affected in this mutant.

(v) Inversion.

Another property often considered as characteristic of insertion sequences is their ability to invert a neighbouring segment of DNA. Inversion of a DNA segment, accompanied by the appearance of a second copy of the insertion sequence flanking the segment in inverted orientation to the first, has been reported for several elements.

In IS-mediated inversions, the small repeats should be inverted with the sequence between the IS elements, although this expectation has not yet found experimental verification.[176]

Care must be taken to distinguish a single-step inversion event from more complex, multistep events. Simple insertion of a second copy of the element in inverse orientation followed by homologous recombination between the two copies, for example, would also lead to the type of end product observed.[124] In at least two of the cases cited in the literature,[73,316] the products were isolated in recombination-proficient hosts. It is therefore possible that the observed inversions occurred via this postulated pathway.

(vi) Homologous recombination.

In addition to their ability to induce the formation of new DNA joints without the aid of significant homology, insertion sequences can also be used as effective substrates for the homologous recombination system of the host by acting as portable segments of homology and thus serve as sites for duplication, inversion, deletion or replicon fusion via homologous recombination.[124]

Among the fully sequenced IS elements, IS*136* alone carries a Chi sequence, which is the activation site for *recBCD* endonuclease. Although this Chi site has not been shown to be functional, its presence on IS*136* gives this insertion sequence an additional capacity in recombination as a portable enhancer of recombination.[124]

Intramolecular recombination results in either the inversion of the DNA between the elements or deletion of this DNA plus one copy of the IS element depending on

the orientation in which the two interacting IS elements are carried. Intermolecular recombination results in replicon fusion. The efficiency of these recombinational processes depends on the length of the IS element, the presence of particular sequences within or near the IS elements and probably on the spatial proximity of the two interacting elements.[176]

Processes of apparently homologous recombination also occur at low rates in *rec*⁻ mutant bacteria, suggesting that more than one mechanism may be involved, at least some of these might represent site-specific rather than general recombination.[176] For example, *recA*-independent, RNA polymerase-dependent, homologous recombination has been reported.[172,173]

The compound transposon, Tn*2571* (and deletion variants) can transpose by a homology-dependent "directed transposition" process. Homologous recombination between the direct repeats (IS*1*) produces a circle containing a single IS*1* element; this segment then integrates by homologous recombination into another IS*1* element which is naturally present in the target genome, phage P1.[58,87]

Investigation of the behaviour of plasmids such as the sex factor F (IncFI) and multiple-antibiotic-resistance plasmids of the IncFII class in various bacterial hosts provided some of the earliest examples of the importance of insertion sequences in (homologous) recombination. Homologous recombination between insertion sequences IS*2*, IS*3* and Tn*1000* located on the F plasmid and at various positions on the *E. coli* chromosome appears to be responsible for F integration and the formation of many Hfr strains (F plasmid-chromosome cointegrates). Insertion sequences are also involved in the excision of the integrated F plasmid, together with flanking DNA, to form F′ plasmids. This often involves (but need not) homologous recombination.[85] The homology-dependent pathways are simply more efficient than those promoted directly by the transposons themselves.[124] There is a strong correlation between the position of IS*2*, IS*3* and Tn*1000* elements on the chromosome and the point of origin of several Hfr strains,[93] and Hfr formation occurs at significantly reduced frequencies in *recA* mutant strains.[79] In addition it has been shown that

a chromosomal copy of Tn*1000* acts as a sex factor affinity site (*sfa*) promoting polarized mobilization of chromosomal genes by a resident F^+ plasmid.[145] The process, presumably due to transient integration, depends on the presence of Tn*1000* both on the plasmid and in the chromosome and is RecA dependent.[124]

The presence of two directly repeated copies of IS*1* flanking the drug resistance genes of plasmids such as R100.1[169,297] can provide an important adaptive function. The resistance genes of this type of plasmid (the r-determinant or *r-det*) undergo spontaneous amplification by tandem duplication in certain species, with a concomitant increase in the level of antibiotic resistance.[58] Such behaviour is observed in *Proteus mirabilis* when plasmid-carrying cells are grown in the presence of high levels of chloramphenicol;[150,321] not only do tandem multimers occur, but circular *r-det* molecules can be observed. Other species, however, show quite different behaviour. In *S. typhimurium*, a rapid loss of resistance occurs,[390] presumably because high levels of recombination between flanking copies of IS*1* lead to a rapid loss of the interstitial resistance genes. In *E. coli*, the plasmid is stable, but the *r-det* does not undergo amplification. Amplification can be observed, however, using the IS*1*-based transposon Tn*9* (2·6kb), which is structurally similar to the *r-det* but is much smaller[254] (the *r-det* is approximately 20kb). Amplification in this case is RecA dependent and probably occurs by unequal crossing over between the flanking IS elements during replication of the host plasmid.[56] It is perhaps worth noting that plasmid-coded functions can also influence this type of recombination. Amplification of DNA flanked by IS*1* elements by tandem duplication in a ColE1-based plasmid has also been observed.[60]

Functional sites.

Several detailed studies of different insertion sequences have demonstrated that the ends are essential components of the transposition apparatus. This has been directly demonstrated for IS*1*,[126] IS*10*,[259] IS*50*,[287] and IS*903*[95] by the analysis of deletions and internal point mutations. In addition to sites concerned with transposase recognition and binding, the ends of insertion sequences carry an

assortment of binding sites for other proteins.[124]

Given the range of genetic effects that insertion sequences have and the fact that they are presumed to encode their own transposases, there must also be specific sites inside the elements that are involved in gene expression or other functions. These sites have, in many instances been found and are described in greater detail in the section on the genetic effects of transposable elements.

Several elements carry candidate integration host factor, IHF, recognition sequences (AAnnnnTTGAT) near their ends. These include IS*1*, IS*903*, IS*5*, IS*10* and IS*50*. There are also several occurrences of adjacent sites, one in each orientation, these occur in IS*6*, ISSHO*1*, IS*136*, IS*10* and IS*186*. Dam methylation sites (GATC) are also found near the ends of several elements; for example, IS*3* has sites at either end, and IS*4* and IS*50* have clusters of sites at one end. DnaA sites (TTATCCACA) occur near the ends of only two insertion sequences, IS*5* and IS*50*. Another interesting site, due to its role in homologous recombination, is Chi (GCTGGTGG).[361] Only a single site was found, in IS*136*, among eighteen sequenced insertion sequences. It should also be noted that IS*50* carries a sequence that can generate a Chi site by a single-base-pair change.

Another protein identified for its affect on site specific inversion, Fis (factor for inversion stimulation),[189] has at least two binding sites in ISs. The binding of Fis to two sites in IS*5* has been detected by footprinting experiments. There has been no general experimental survey of known ISs for binding sites and the consensus sequence for Fis binding is too ill-defined to allow the scanning of known sequences for sites. The function of Fis in IS*5* is not known.[124]

Tn*3* contains three sites required *in cis* for normal transposition.[153] The two ends of the transposon are necessary, presumably because they are the recognition sites for the transposition proteins, and a site localized to within a 546bp region near the centre of the transposon (the *res* site) is required. If this third site is deleted, fused replicons result, in which the donor and recipient plasmids are fused with a copy of Tn*3* in direct repeat at each juncture point,[10] these cointegrate structures are intermediates

in transposition.

The *res* sites of Tn*3* and Tn*1000* have been extensively mapped and characterized. They consist of three separate resolvase sites, spanning a total of about 120bp.[354] Site I, the most *tnpR*-distal site, contains the recombinational crossover point. In the case of Tn*1000* the entire 120bp segment has been shown to be both necessary and sufficient for efficient cointegrate resolution *in vivo* and *in vitro*.[393] The *res* sites of the Tn*501* subgroup have the same organization.[354]

Coding properties.

The definition, isolation and functional characterization of IS-encoded proteins is receiving much attention in various laboratories. These studies are not only a prerequisite to understanding transposition processes at a molecular level, but should provide an important contribution to the understanding of DNA-protein and protein-protein interactions in general.[124]

Although the entire DNA sequences of some IS elements are now known, it is possible that the nucleotide sequences (particularly of regular IS elements) still hide unknown functional features. If these sequences were used as much as is potentially possible, the existence of overlapping genes and multifunctional sequences might be expected.

(i) Transposition functions.

The first evidence that transposable elements coded for highly specific enzymatic mechanisms capable of recognizing their own inverted repeats and cleaving DNA at precisely these locations was provided by Heffron *et al*.[152,153] by mutagenizing Tn*3*. This has since been shown to be true for other transposable elements also.[67]

Many of the sequenced IS elements (with the notable exception of IS*1*) contain a single long coding region that begins just inside the inverted repeat at one end of the element and continues through to the other end, stopping just short of, or a few basepairs within, the second inverted repeat.[128,209,280] They often carry one or more smaller open reading frames (ORF) within the large reading frame on the opposite strand.[176] One may assume that the large open reading frame encodes one component of a *transposase*, that is, an enzyme involved in transposition, and

apparently this is the case for IS*10*, IS*50*, IS*903* and Tn*1000*.[124]

However, many elements have a set of smaller ORFs, of various lengths and positions, making it difficult to ascertain from sequence data alone which of them actually code for proteins. Only in a few cases has it been shown which ORFs are required for transposition. These include IS*1*, IS*10*, IS*50*, IS*903* and Tn*1000*. The frequency versus size distribution of the smallest of these ORFs is consistent with that expected for random occurrence in DNA, but for longer ORFs (in the range of 90 - 140 codons, for example), it is significantly higher than expected.[124] This suggests that several of the ORFs in this size range are probably real. It should be noted here perhaps that the demonstration that a given ORF is essential for transposition does not necessarily mean that it encodes its own protein. It is possible that in some cases, where one ORF overlaps or closely follows a second on the same strand, that some level of control is exercised by low-frequency frameshifting between the two to generate a fusion protein.

IS*1* is different in several respects from most of the other elements. It does not have a single large ORF, but has a surprisingly large number of ORFs (eight) for its length. Evidence that the two ORFs, *insA* and *insB*, which together span most of the length of IS*1*, are required for transposition was provided by Ohtsubo and co-workers.* Using site directed mutagenesis, it has now been demonstrated that *insA* and *insB* are the only ORFs essential for IS*1* transposition.[179] Two additional ORFS, *InsC*[42] and *InsE*,[187] may also be involved but are not essential. Using gel retardation techniques, it has been shown that the ends of the element are specifically recognized and bound by the InsA protein.[410] There is no evidence, beyond the demonstration that it is required in transposition, that suggests what role is played by the InsB protein.[124] Experiments involving IS*1* are complicated by the fact that in the *E. coli* K12 chromosome several additional copies of IS*1* are present and may provide missing functions.[327]

Some information has been obtained on how the InsA protein of IS*1* interacts with the end of the

* References 239,277,278,280

element.[411] The minimum essential sequence required for transposition activity of an end of IS*1* is the final 23bp of either end (they are almost perfect inverted repeats). This sequence, from either end, is sufficient to bind InsA specifically. It was determined that there are two distinct functional domains in the ends of IS*1*: one that contains the specificity determinants for recognition by InsA, and one (at the very ends) that determines the rate of some step in the transposition process other than InsA binding. Since we know that InsA is essential for IS*1* transposition, it seems likely that it is part of the protein-DNA transposition complex which forms on the ends of IS*1*. Its precise role, however, is unclear. It is possible that in addition to recognizing the ends, InsA itself carries out the cleavage and ligation reactions necessary for transposition. Alternatively, InsA may provide an anchor via protein-protein interaction for the binding of InsB or possibly other proteins. The specific sequence at the very ends is clearly essential for transposition, although it is probably not involved in InsA recognition. It must, therefore, be involved in another sequence-specific process. Possible candidates for this process are addition of other proteins to the complex, strand cleavage or some conformational transition of the complex.[124]

Mutagenesis studies on Tn*3* identified three proteins whose synthesis is directed by Tn*3*. In addition to β-lactamase, which confers ampicillin resistance, the transposon codes for a transposase (TnpA) and a resolvase (TnpR) which regulates itself and the transposase. The site of cointegrate resolution, *res*, lies between the two genes. The transposase consists of approximately 1015 amino acids and the resolvase is about 185 residues in length. All three of these proteins have been synthesized in minicells and isolated on gels. The transposase and the resolvase are organized in a bidirectional, self-regulated transcription unit.[354]

TnpA has been purified to homogeneity; it has a tendency to aggregate under some conditions, and has been shown to recognize and bind to the ends of Tn*3*.[354] Grindley and Wiater[140] have found, using footprinting techniques, that the Tn*1000* transposase, which is similar to that of

Tn*3*, can recognize and bind to the 38bp inverted repeats at the ends of its element.

Based on their genetic organization and on the ability of recombination functions to complement one another, transposons of the Tn*3* family can be divided into two subgroups. In one group, represented by Tn*3* itself and Tn*1000*, the *tnpA* and *tnpR* genes are divergently transcribed from a shared regulatory region.[354] The *tnpR* functions, but not the *tnpA*, are interchangeable.[207] In the Tn*501* subgroup the region that contains *res* and *tnpR* is inverted with respect to its orientation in Tn*3*. The best characterized members of this subgroup are Tn*21*, Tn*501* and Tn*1721*.[100,142,313]

The presence of regions with conserved open reading frames common to all studied derivatives of the related elements IS*102*[28] and IS*903*[136] suggests that these sequences are coding for active gene products.

All of the three largest ORFs of IS*5* have been shown to encode proteins.[235,299,300] The arrangement of the promoters internal to IS*5*[342] further suggests that the proteins are important components for IS*5* function. Unfortunately, however, no genetic or biochemical evidence attributes a function to any of these genes.[124]

Studies on IS*21* have detected two proteins of 43kD and 29kD but, as in the case of IS*5*, it is not clear whether either or both of these proteins are essential. Similarly, a 42kD protein product corresponding to the long, open reading frame (presumably the transposase) of IS*30* has been detected.[124]

A notable property of the characterized transposase proteins is that many are quite basic. InsA and InsB, for example, are extremely basic, with estimated pI values of 10·9 and 11·0, respectively. Among the known sequence-specific DNA-binding proteins, only IHF of *E. coli* is as basic. This property of the transposases may play some role in the unusual behaviour exhibited by all of those mentioned above. They act efficiently only on IS ends that are present on the DNA molecule encoding the transposase, that is they act largely *in cis*. Complementation of a transposase-defective element is very inefficient unless the transposase gene is *in cis*, and, in several instances, close

linkage is apparently preferred. This *cis* effect is large, in some cases more than one hundred-fold, and it has been demonstrated for IS*1*,[239] IS*10*,[260] IS*50*[183] and IS*903*.[136] The only plausible explanation for this effect appears to be that the proteins bind to the DNA very close to the DNA segment from which they were made, presumably because of the strong non-specific binding affinity arising from their positive charges. The proteins may tend to remain associated with the DNA and diffuse along it to specific binding sites.[401]

In Mu, the products of genes *A* and *B* are involved in transposition.[285] Mutations in gene *B* result in the lowering of the transposition frequency, while transposition is completely abolished in mutants lacking gene *A* function.[285] The polypeptides specified by these genes have been identified on gels.[131,241]

(ii) Other genes.

In addition to transposons encoding antibiotic resistance genes, transposons encoding genes for heat-stable enterotoxins,[362,363] for lactose utilization[72] and for chromosomal conjugal transfer[117] have been identified in nature; transpositions of large chromosomal or plasmid segments encoding genes for arginine[405] or histidine[402] biosynthesis or for raffinose utilization,[247] have been observed in the laboratory. Other known examples include transposons encoding traits such as resistance to toxic mercury compounds and synthesis of bacterial toxins.[67] Elements of even higher order complexity have also arisen in nature by insertion of one drug resistance element into another.[219,220]

The conjugative transposable elements, typified by Tn*916* and Tn*1545*, are ubiquitous among streptococci. They encode resistance to one or more antibiotics and promote conjugation by a mechanism independent of the normal endogenous or plasmid-specified conjugation systems of the host. The mechanism of conjugation and transposition are not well understood, but mutagenesis has shown that over half of Tn*916* specifies transfer functions.[346]

Both Tn*916* and Tn*1545*, are capable of transfer among many different species, those in common include all the streptococci, *Staphylococcus aureus* and *Bacillus sub-*

tilis, while Tn*916* (originally found in *Enterococcus faecalis*) has been transferred to *Acholeplasma laidlawii* and *Mycoplasma pulmonis* and Tn*1545* can be transferred to *Bacillus thuringiensis* and *Listeria monocytogenes*.[263]

Interactions between transposable elements.

Compound transposons contain two flanking copies of the same IS element. For two of these transposons, Tn*5*[318] and Tn*10*,[114] it has been clearly shown that the two copies of the flanking IS elements, IS*50* and IS*10* respectively, are structurally and functionally different from each other. One of the copies is more active and transposition is therefore driven by the more active element. This indicates that an active IS element can complement an inactive one, which may mean that proteins produced from the active element can efficiently recognize its interaction sites on both the active and inactive IS element.[176]

Deletions of one end of an IS*1* completely abolished cointegration mediated by this mutated IS*1*.[239] Deletions or a short insertion internal to IS*1* reduced, but did not completely abolish, its ability to cointegrate. This reduction was partially but not fully restored by the presence of an additional intact IS*1* on the same plasmid. From these observations, it was concluded that IS*1*-encoded proteins can function *in trans*, but that they act preferentially on the ends of the IS*1* from which they are produced. The same conclusion was drawn from experiments with IS*903*.[137]

Negative interaction (interference) between IS elements was found in studies on Tn*5* transposition. Transposition of Tn*5* is most frequent immediately after entry of Tn*5* into a cell and it later falls under negative control by a factor encoded within the element itself. Studies with a series of deletions seem to suggest that the repressor or inhibitor in Tn*5* is encoded by IS*50*. Identification and characterization of these repressor or inhibitor proteins are still lacking.[32]

Mechanism and Regulation of Transposition.

Although bacteria carry a wide variety of insertion sequences, apparently unrelated in sequence and in organization, one essential common feature distinguishes the rear-

rangements they induce from generalized recombination: transposition is entirely independent of sequence homology and leads to the formation of novel DNA joints between the ends of the IS element and the target DNA. There are, of course, several biochemical events that must form part of this process (a minimum number of strand cleavages and rejoinings) and others, such as replication, that may or may not be involved. Only relatively recently have experimental results from a limited number of systems been obtained that can allow the critical distinction among some of the possibilities for a particular system, but the diversity of transposable elements leads to serious doubts about the general applicability of any particular model.

In the attempt to determine what pathway actually accounts for the transposition of a particular element, most of the evidence is derived from examining the structures of transposition products from well defined donor and target molecules and comparing these with the predictions of the models. Great caution, however, must be taken in interpreting the observation of certain products as evidence for a particular mechanism. The assumption usually inherent in such interpretations is that all possible mechanisms have been foreseen and that the investigator must simply decide among them on the basis of the observed products. This is seldom the case.

Host Factors.

As relatively simple genetic units, ISs are expected to require some components, for transposition, other than those encoded by the element itself. Several *E. coli* chromosomal mutants affecting IS element activities have been isolated and characterized by using selective markers carried on compound transposons.* The role of host factors in the transposition process is in most cases poorly understood. Only in the case of bacteriophage Mu, for which a defined *in vitro* transposition system has been developed,[77] are these factors understood to some degree.[124]

Mutations in *top*, the structural gene for topoisomerase I, were reported to affect the transposition of some mobile elements but not others.[374] Since *top*$^-$ mutations affect transcription of several bacterial operons, they may

* References 64,165,180,181,335,360,374

affect transposition at the mechanistic and/or the regulatory level. The effect could be explained either by a direct participation of topoisomerase I in the cleavage/ligation step or by the requirement of a particular DNA topology for the transposition process.

Syvanen and colleagues[64,376] have found that mutations in *polA*, the gene encoding polymerase I, affecting the 5′→3′ exonuclease activity of the enzyme lead to reduced frequency of transposition of both Tn*5* (IS*50*) and Tn*10* (IS*10*). Other *polA* alleles, however, were found to exhibit only slightly reduced transposition frequencies. Similar results have been found by Sasakawa *et al.*[335] The role of *polA* in the transposition process remains a matter for speculation but these results indicate that the *polA* gene is somehow involved with the replication processes associated with transposition.

Many of the well-studied, site-specific recombination reactions have proven to be sensitive to the degree of supercoiling of donor or recipient DNA. It has been proposed that the energy of supercoiling is used to drive one or more steps of the multistep recombination reactions.[266] Studies on the transposition of Tn*5*[184] have provided convincing evidence for the involvement of GyrA and GyrB in its transposition and suggests that the level of supercoiling of the target rather than that of the donor molecule is important in this case.

Binding sites for the histone-like protein IHF have been found at both ends of IS*1* within the inverted repeats.[125] They occur between the -35 and -10 regions of promoters that direct transcription across the element. IHF induces a strong bend on binding to these sites *in vitro*,[291] protects the ends of IS*1* against DNase I digestion and increases the accessibility of the terminal phosphate bonds to DNase I cleavage.[125] Addition of IHF prior to, or at the same time as, RNA polymerase (RNAP) results in preferential binding of IHF to the left end, although prebound RNAP cannot be displaced by subsequent addition of IHF.[125] Moreover, preliminary competition experiments between IHF and the IS*1*-specified protein InsA suggest that both proteins might bind the ends simultaneously. In spite of these suggestive observations, Galas and Chandler[124] have been un-

able to show a convincing effect of IHF *in vivo* using a number of different assays for transposition activity. Thus a role for this protein in transposition remains to be demonstrated.

An interesting speculation is that interaction between RNA polymerase and the ends of transposons could be important for transposition. Inceptors are proposed to be sites at which transcription terminates in such a way as to yield a primer for subsequent DNA synthesis.[163] For example, when coupled to a suitable promoter in "out-to-in" orientation, the ends of IS*5* can direct DNA replication of a tester plasmid.[235]

Many insertion sequences have been shown to carry Rho-dependent transcription termination signals that are presumably involved in the control of expression of element-encoded genes. The effects of several *rho* mutations on the transposition of IS*1* (or Tn*9*), IS*5* and of Tn*5* have been investigated.[84] Reductions in frequency of ten- to thirty-fold were observed. Such an effect on IS*1* may be a reflection of the direct effect of Rho on the *insB* gene, which lies downstream from the resident IS*1* terminator.

Dam methylation sites are found at the ends of IS*10*, IS*50* and IS*903*, three elements whose transposition is thought to be largely non-replicative (see below). Dam sites are also found in several other elements. Dam methylation at the ends of IS*10* has been shown not only to influence the expression of IS*10*-encoded transposition functions, but also to determine the intrinsic activity of the ends in transposition.[311] The fact that the ends are more active when these sites are unmethylated or hemimethylated than when fully methylated provides an elegant mechanism by which the transposition of this element could be coupled to the passage of a host replication fork by generating transiently hemimethylated (newly replicated) DNA. A similar picture has emerged from studies of IS*50*, where Dam methylation controls the relative levels of transposase and repressor and the activity of the end(s).[20] Involvement of Dam in the control of expression of IS*903*-encoded genes has also been reported.[311]

IS*50* carries a site at one end, resembling the consensus sequence for DnaA protein-binding sites.[190] This

essential protein, involved in the initiation of chromosome replication[221] has been shown to recognize and bind the IS*50* site *in vitro*[120] and influences the activity of the end *in vivo*.[404]

It should be pointed out that some of these host mutations may affect IS functions indirectly (e.g. altering the transcriptional characteristics) and that it is very likely that different classes of IS elements require different host functions.[176]

Insertion specificity.

Although some of the events mediated by translocatable elements involve highly specific sites on the element itself, there is often little or no specificity for which sites on the participating target chromosome molecule are involved.[212] In order to fully understand the mutagenic effects and evolutionary impact of transposable elements it will be necessary to first understand the rules which govern the choice of target sites. Since the choice of site must reflect the way in which the transposition complex interacts with the target DNA, these rules may also provide insights into the mechanisms of transposition.

A wide range of insertion specificities have been noted for different elements. At one extreme are elements which insert at many points in different targets with little recognizable specificity. A well studied example in this category is phage Mu. At the other end of the spectrum are elements (such as IS*4*) which are found in only one or a few sites. The majority of elements show specificities between these two extremes. For example Tn*10* shows a marked preference for certain hotspots but integrates at other points as well.[50]

Several factors may, individually or collectively, contribute to the degree of specificity observed with any one element. Among these may be recognition of a specific target sequence or of a structural (e.g. a bend) or functional (e.g. active transcription) feature of the DNA. Sequence specificity will be determined by the size of the site recognized (which decides the expected frequency of the site in the bacterial genome) and the degree of degeneracy of the site tolerated by the IS (which will increase the expected frequency).

Based on the observed DNA sequences of target DNA and of the junctions of IS*1*-induced deletions, Saedler *et al.*[326] suggested that small sequence homologies between the target area and the ends of IS*1* play a role in the choice of target sequences for this element. The presence of nearby regions of limited homology with the ends of ISs has often been suggested to play a role in site selection*, but no convincing evidence of either a direct or statistical nature has been presented.[124]

Certain elements, such as Tn*7*, insert preferentially at a single site in the bacterial genome and, at significantly lower frequencies, at various secondary sites.[76] This type of behaviour is reminiscent of the integration, by site-specific or specialized recombination, of temperate bacteriophage such as λ, in which the principal "target" sequence (*attB*) is relatively long and in which variant sites with single-base-pair differences (the secondary sites) are much less attractive.

Although a statistically significant sample of insertions is not available for many ISs, it is clear that specific consensus target sequences of various lengths and degrees of degeneracy are involved in the transposition of several, among them IS*4*,[249] IS*10*[146] and IS*30*.[54] Recent results of Rak and co-workers on IS*150* suggest that this element may also be highly specific in its target selection.[342,343]

The insertion hot spots for Tn*10* (a composite transposon based on IS*10*) tend to conform to a symmetrical, consensus sequence, GCTNAGC, whose degree of match with a target sequence is directly related to the relative insertion frequency at that sequence. It has been suggested that specific protein contacts with the consensus base pairs are involved, since the thymine methyl groups are necessary for high-frequency insertion.[230]

A simple consensus target sequence was determined for IS*5*, this preferential sequence coincides with the 4bp duplication sequence.[105] Whether the presence of the consensus sequence alone is enough to serve as a target for IS*5* integration has yet to be determined.[176]

A comparison of the DNA sequences around the only

* References 39,121,234,257,331

three known insertion sites of IS*4* reveals common structural features including a hyphenated palindromic consensus sequence outside the duplicated target sequence.[209]

Other insertion sequences exhibit little specificity for a given target sequence. On the other hand, some of these elements exhibit pronounced "regional" specificity for segments of DNA of the order of one hundred base pairs in length and insert at many sites within these regions. Among those ISs for which such preferences are observed are IS*2*,[347] IS*186*[222] and, to a lesser extent, IS*50*.[25]

IS*2* (and IS*1*) integrates preferentially into the leader sequence of the *gal* operon.[325] The leader sequence comprises less than one percent of the *gal* operon,[265] while mutations caused by the integration of IS*1* and IS*2* into this region constitute approximately twenty percent of all insertions into this operon.[368]

Independent insertions of Tn*10* within selected small targets are, in each case, not randomly distributed. For example, 27 out of 55 Tn*10* insertions in the Salmonella *his* operon (about 10kb long) map at a single site in *hisG*; the rest are scattered among 10 to 15 other sites.[40,217] Similar results have been found for Tn*10* insertions in *lacZ*.[212] Genetic analysis of the insertions in *hisG* has shown that insertions at the single "site" are not identical. They differ in orientation of the inserted element and reversion frequencies of the insertion mutants to *his*$^+$, but there is no correlation between orientation and reversion frequency.[217]

Tn*2*, Tn*3*, Tn*4* and Tn*5* have all been shown to insert at many different positions within small targets.[21,154,220] For example, nineteen independent insertions of Tn*5* in *lacZ* all map at different positions.[16] The specificity with which these elements are inserted at particular positions is much less than for Tn*10* or IS elements.[212]

The determinants of regional specificity are more difficult to define and interpret. IS*1*[409] and IS*186*[222] (the latter based on limited number of insertions) show pronounced preferences for regions rich in A+T and G+C, respectively, while IS*2* insertions are found repeatedly in a small region of the phage λ[331] and P1[347] genomes, that appears to have no obvious sequence features that could

account for this property.[124]

In the case of IS*1*, the initial observations that IS*1* hot spots in the *lac* operon and in bacteriophage P1 were particularly rich in A·T base pairs[121,255] were subsequently extended to insertions in the plasmid pBR322.[409] This plasmid exhibits two major regions of insertion which have similar (high) A+T densities. However, one region attracts over ninety percent of all IS*1* insertions, suggesting that A+T density is not the only factor influencing IS*1* site selection. A 60bp poly(dA)·poly(dT) tract (100% A+T), while attracting some insertions, does not exhibit a strength compared to the major site. A striking feature of both the pBR322 hot spots is their aberrant migration on polyacrylamide gels,[373] suggesting that these regions are bent. It has been proposed that such structural anomalies may provide the necessary signal.[124]

A peculiar preference has been detected in the case of Tn*501*.[141] Here the sites of integration into plasmid RP1 are determined by the presence or absence of another unrelated transposon, Tn*801*, even though the sites of integration of the former are physically distant from the latter.[368]

Transposon-induced deletions are thought to arise by pathways similar to transposition. Thus, we might expect that the selection of deletion endpoints would follow the same preferences as for insertion specificity. Surprisingly few studies have compared these specificities directly.[51,270] The data that do exist confirm that the preferences of target sites for both deletions and intramolecular transposition are the same.

Regulation.

Transposition at very high frequency would certainly be detrimental to the host cell, and indeed all known ISs exhibit relatively low frequencies of transposition. On the other hand, transposition is essential to the propagation and survival of IS elements in a bacterial population. It would be surprising, therefore, if transposition activity was regulated not only by the IS element itself, but also by host factors, perhaps those involved in signalling the physiological state of the cell. The few cases in which control of transposition activity has been studied in some detail have provided examples of some novel regulation

mechanisms.

Several observations are at least consistent with the interesting possibility that transposable elements might respond to signals from their environment in such a way that transposition could preferentially occur under conditions of stress to the host. For example, there are a number of indications that transpositions may be recovered more frequently in bacterial inocula which have been stored for long periods of time.[8] Cornelis[71] has reported that transposition of Tn*951* is increased by incubation of storage inocula at 4° for days or months. Read and Jaskunas[303] reported that an unexpectedly high proportion of *E. coli* strains which have acquired one new IS insertion have in fact also acquired a second insertion elsewhere; this could be an indication that transposition occurs in bursts, perhaps in response to some signal.[271]

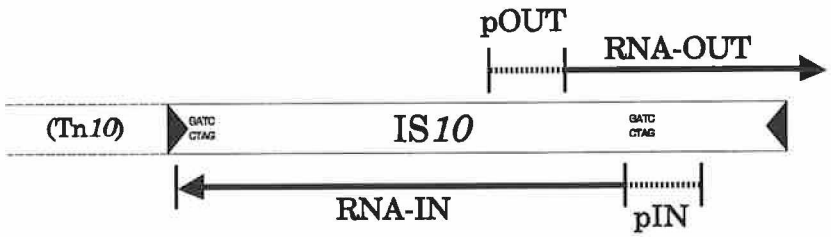

Figure 1. Structure of IS*10*.

IS*10* is subject to regulation by several systems (see figure 1). It encodes a transcript (RNA-OUT) which acts as an anti-sense RNA repressor of translation.[215,359] This RNA is apparently responsible for multicopy inhibition of transposition[359] by limiting the translation of the transposase mRNA. A second system makes use of the only two sites in IS*10* for adenine methylation by the host Dam methylase (GATC). Both of these sites are implicated in the regulation of transposition by the host cell. One site, in the transposase promoter (pIN), reduces the transcription of the transposase gene when methylated.[359] This may have the double effect of lowering the transposition activity of the end and of timing bursts of transposase synthesis with the passing of a replication fork, which produces transiently hemimethylated DNA. Another site, near the inside of the element, sharply reduces the activity of this end when methylated.[359] This has the effect of increasing the coher-

ence of the compound transposon in situations where methylation is efficient.[124]

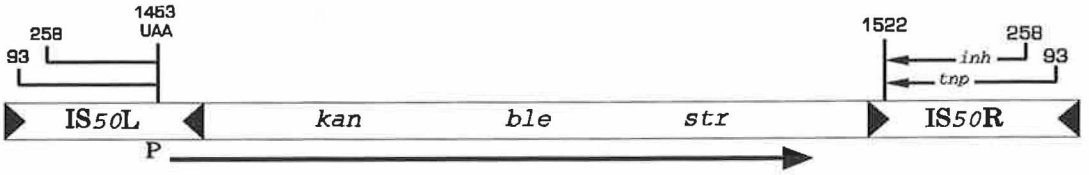

Figure 2. Structure of Tn*5*.

IS*50* transposition is also subject to several levels of control.[20] Two proteins are encoded by IS*50*. The first is the transposase - *tnp*, and the second, the inhibitor - *inh*, is the product of a later translation start site within the same ORF.[182,192] IS*50* was found by Biek and Roth[32] to become increasingly inhibited for transposition after it is first introduced into the cell. This negative control acts *in trans*, and some evidence indicates that the second, shorter protein is responsible.[193,232] McCommas and Syvanen[252] have recently shown that the two different but overlapping transcripts from which the two proteins are expressed have different stabilities. Transposon Tn*5* (see figure 2) is a composite transposon based on IS*50*. The two copies of IS*50* (IS*50*L and IS*50*R) differ functionally, IS*50*L contains an ochre mutation that results in truncated versions of the same two polypeptides specified by IS*50*R.[319] Tn*5* specifies a diffusible negative regulator which suppresses the activity of a Tn*5* element resident in a cell and reduces the activity of any additional Tn*5* element which is newly introduced. The net effect of this regulation is that Tn*5* transposition is more frequent when an element enters a new host than when it has established permanent residence. Both IS modules of Tn*5* are required for this regulatory activity.[32] When a Tn*5* is introduced into a naïve cell,[191,252] the initial mRNA ratio of 1:2 drops within three hours to 1:80.[124]

As for IS*10*, regulation involving host factors also occurs at the ends of IS*50*. The outside end has a binding site for the DnaA protein, which, while not essential for transposition, stimulates the activity of this end.[404] Both ends carry Dam methylation sites (the inside end carries a cluster of these sites). When methylated, the activity of the inside end is reduced by about ten-fold.[403]

Tn*3* and related transposons (such as Tn*1*, Tn*501*, and

Tn*1717*) exhibit a phenomenon known as transposition immunity.[312,389] Plasmids carrying two copies of one of these elements (in inverted orientation) are stable; and two copies of such an element can insert into a target simultaneously. However, when a copy of the element is already stably integrated in a plasmid, a second copy of the same element cannot efficiently insert itself. This inhibition of secondary insertion is also seen for intramolecular transpositions, but to a lesser degree. The inhibition is *cis*-acting. The presence in a cell of a plasmid containing one copy of the element inhibits secondary insertion into that plasmid but not into another plasmid which does not already contain a copy of the element.[354,389] This unusual phenomenon is unique to the Tn*3* family

The insertion of an IS element into a transcription unit might be expected *a priori* to result in the gratuitous expression of IS-encoded genes driven by external transcription. Since the activation of transposition has never been reported, mechanisms must exist to "protect" most ISs from external activation. This might be accomplished simply if transposition were dependent on two (or more) IS proteins encoded on opposite strands of the element. As pointed out in the preceding section, many IS elements have ORFs arranged in this fashion, although in most cases their relevance to transposition remains to be demonstrated.[124]

For many elements (e.g. IS*10*, IS*50*, IS*903*, Tn*1000*), only one reading frame seems to be essential, or if more than one is involved (e.g. IS*1*), these are encoded on the same strand. The presence of transcription termination signals within an IS element could protect downstream genes from over-expression by external transcription. Such terminators have been detected in several ISs (for example, IS*1*, IS*2*, IS*3*, IS*5*, and IS*30*). External transcription into the transposase gene of IS*10* generates an mRNA with a secondary structure which sequesters the translation initiation signal(s) and prevents translation. The transcript from the resident IS*10* promoter, pIN, does not include 5′ sequences necessary for the formation of this secondary structure, and hence the initiation signals are available for translation.[382] Schwartz *et al.*[343] have suggested that a similar arrangement in IS*150* may protect it from impinging

transcription.

For IS*1*, external transcription has been shown greatly to increase the expression of the *insA* protein.[410] Moreover, such transcription has a large inhibitory effect on transposition activity. Transcription impinging on either end of the element has a comparable effect, indicating that inhibition is not due to the overproduction of InsA itself.[33,61] IS*50* also exhibits a reduced activity when subject to impinging transcription.[334] As in the case of IS*1*, this effect cannot be ascribed to an influence on the expression of IS-encoded genes. It is possible that in both cases transcription disrupts the structure of the DNA at the ends of the element, inhibiting the formation of a transposition complex.[62]

The Mu *A* and *B* genes are located, adjacent to one another, close to one end of the phage genome. They are separated from that end by a small (less than 1kb) regulatory region which encodes a classical phage repressor and an immunity determinant, both of which negatively regulate expression of the *A* and *B* genes.[167]

Mechanisms.

A model, at the level of biochemical detail considered here, consists in knowing which strands of the element are cleaved, which pairs of ends are joined, how replication proceeds, if it does and in what order each of these events take place. Some care must be taken about simple classification of models and, more importantly, in the interpretation of tests of models which rely on the examination of transposition products, as there may be several different pathways leading to the same end result.

It is possible to classify most transposition models in several distinct ways. One could focus attention, for example, on whether replication of the donor occurs and call them "replicative" or "conservative", or focus on whether the ends of the element are treated identically throughout the process and call them "symmetric" or "asymmetric".[124]

At present, we do not know whether all regular IS elements transpose by the same pathway. We cannot even rule out the possibility that the same IS element transposes by different processes depending on conditions.[176] However any

proposed model will have to take into account properties of mobile elements such as the various genetic rearrangements they mediate and structural characteristics, such as terminal inverted repeats, which are common to all such elements.

The transposition process as envisioned from indirect genetic and physical evidence, can be separated conceptually into two parts: (a) precise breakage of transposon sequences and the target DNA molecule which are then joined in a very particular way at the insertion site; and (b) specific replication of the transposon segment *without* replication of adjacent chromosomal sequences. It is likely (though not experimentally established) that all transposons move by a combination of break/join and recombination processes. However, different classes of elements appear to differ significantly in the way they resolve transposition intermediates to give final separation of donor and target DNA molecules.[214]

So far, most of the investigations that have been carried out are structural analyses of DNA molecules before and after transposition and DNA rearrangements. Detailed elucidation of transposition reactions must await reconstitution of IS-mediated processes in the test tube. Such *in vitro* systems could allow us to examine the proteins, factors and substrates required for the transposition processes. They would also be expected to yield information regarding intermediate steps, the mode of DNA replication, the possible role of superhelicity of DNA or other structural particularities and finally, to reveal the mechanisms regulating the transposition reactions. Until then, however, it is important to bear in mind the potential for multiple mechanisms of illegitimate recombination. Transposition phenomena that are analogous genetically may not be similar biochemically.

Replication of the transposable elements may play a major role in their transposition. Displacement synthesis by DNA polymerase I has been described to occur in the replication of plasmid colE1.[371] This synthesis proceeds for approximately 600bp, at which point the formation of a replication fork occurs with pol III taking over. It is not known presently whether the segment replicated by pol I in

colE1 represents an upper limit for displacement synthesis. If it were, this might provide an explanation for the observation that IS elements very often are approximately 1·5kb long. If this is approximately twice the length of maximum displacement synthesis it would provide an upper limit to transposition by the mechanism suggested above. Larger elements, like the transposons or bacteriophage Mu, might then need a site, at which a different mode of replication might take over. Polymerase I cannot be the only enzyme involved in this process, because transposition occurs in *polA*$^-$ strains.[368]

The simplest way to imagine insertion of an excised, unreplicated transposon is a cut-and-paste process.[16] This is the simple insertion pathway, in which both ends of a single copy of the transposon are joined to the target sequences. The simple insertion pathway, however, could be either a replicative process in which a copy of the transposon is retained at the donor site, or a conservative process in which the transposon is excised from the donor site without replication.[138]

It is widely accepted that IS elements are duplicated during the transposition processes since in strains where a particular element is known to have inserted at a new position, an intact copy of the element still remains integrated at the original donor site.* However, such results are consistent with "suicide" models in which transposon sequences are excised from one copy of the chromosome and inserted into another, with concomitant destruction and non-recovery of the donor molecule. Therefore it is not clear whether the IS element on the donor molecule is *always* recovered in the transposition process, or whether there exists a strict relationship between duplication and recombination during transposition of IS elements.[214]

In principle, excision of an element either could destroy the donor replicon or could be accompanied by repair of the donor. The latter possibility is ruled out by observations made with several transposons.[214] Firstly, "precise excision" (as measured by restored function of an inactivated gene) does not require transposase function, is not accompanied by reinsertion of the transposon, and oc-

* References 15,110,210,302

curs at a lower frequency than transposition. Secondly, in strains in which a transposition event has occurred, the transposon is retained at the donor site; this is compatible either with replicative transposition or with a destructive transposition process (from a second copy of the replicon within the same cell).[138] If destruction of the donor replicon were a necessary step in the transposition process this would mean that successful transposition is never truly intramolecular.

It seems, for the elements IS*1*, IS*10*, IS*50* and IS*903*, at least, that transposition proceeds most commonly by the simple insertion pathway with little or no replication of the element. Much less frequently (about one to five percent for IS*1* and IS*903*, virtually undetectable for IS*10* and IS*50*) a truly replicative process occurs that preserves both donor junctions while creating two new target junctions.[138] The need to account for a new copy of a transposable element during transposition has thus led to models involving specific replication of the transposable element.[50]

In the cointegrate pathway two copies of the element are formed and each copy retains a parental joint at one end while the other end is attached to new target sequences. For intermolecular transposition, this results in the two replicons forming a cointegrate structure and hence the name of the pathway. The cointegrate pathway must be a replicative pathway since two copies of the transposon are produced. All of the models proposed account automatically for transposition-promoted deletions and for duplicative inversions as cointegration events in which the target site occurs on the same molecule as the transposon itself. The two different structures correspond to the two possible orientations of a target site with respect to the element.[108,349]

None of the explanations proposed thus far to account for simple insertion and cointegrates as alternative products of a common intermediate are particularly simple. The correct explanation may remain to be suggested.

Replicative transfer of genetic information is best achieved by transferring a single strand of the DNA segment; complementary strands are then synthesized in donor

and recipient. The conjugational transfer of bacterial DNA occurs by such a mechanism[400] and, together with the rolling circle mechanism of replication,[221] had provided a basis for models of replicative transposition (e.g. Bennet *et al.*[15]). Although conjugational transfer and rolling circle replication are asymmetric processes, the symmetry of most transposons (i.e. the usually identical terminal inverted repeats) suggests that transposition might be a symmetric process and therefore basically different from the integration mechanism of plasmids or temperate bacteriophage.[138]

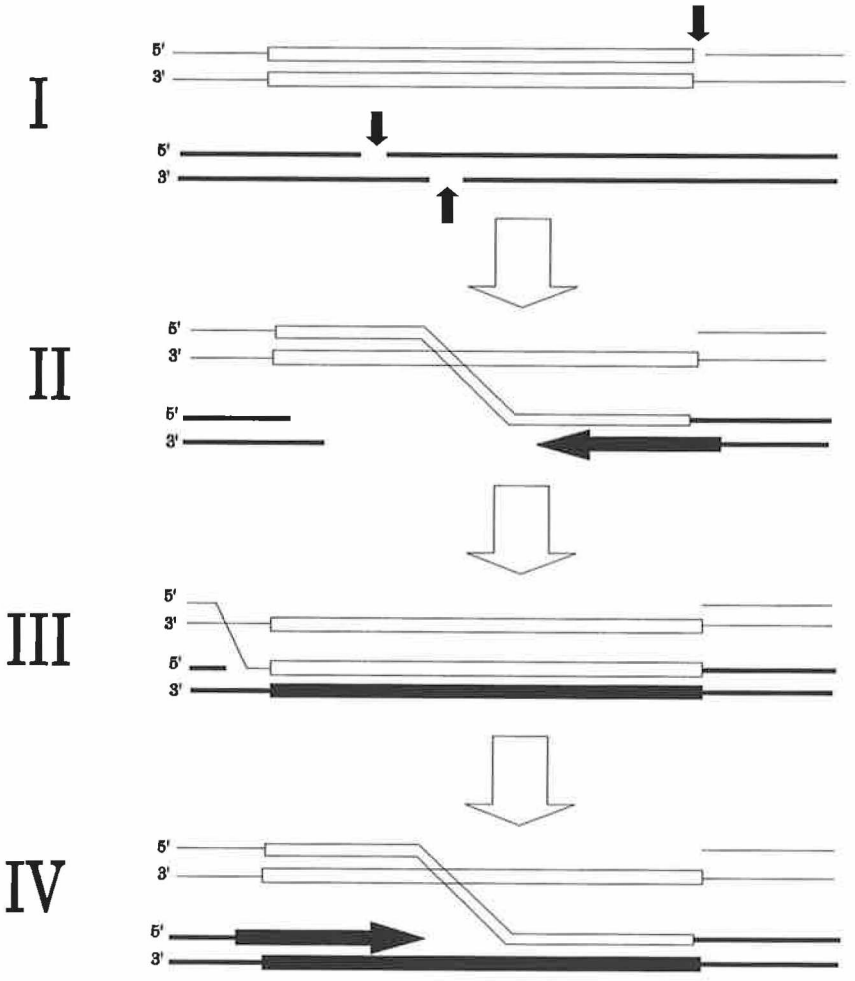

Figure 3. Grindley/Sherratt Model.

The first model proposed, (Grindley and Sherratt,[139] see figure 3), an asymmetric model, starts from the notion that staggered nicks are introduced in the recipient DNA. They assume that the protruding single strand is the 5′ terminus and suggest that one 3′ terminus of the IS element is to be linked to this single-stranded protrusion (I). DNA synthesis, in the form of displacement synthesis, could then start at the 3′ terminus of the complementary strand and continue beyond the recipient replicon's protruding end

into the transposable element (II). When this growing strand meets and is rejoined to the 5′ donor strand (III), displacement synthesis can then take place on the other strand (IV). The manner in which the free ends in part IV are rejoined determines whether the resulting structure is a simple insertion or a cointegrate.

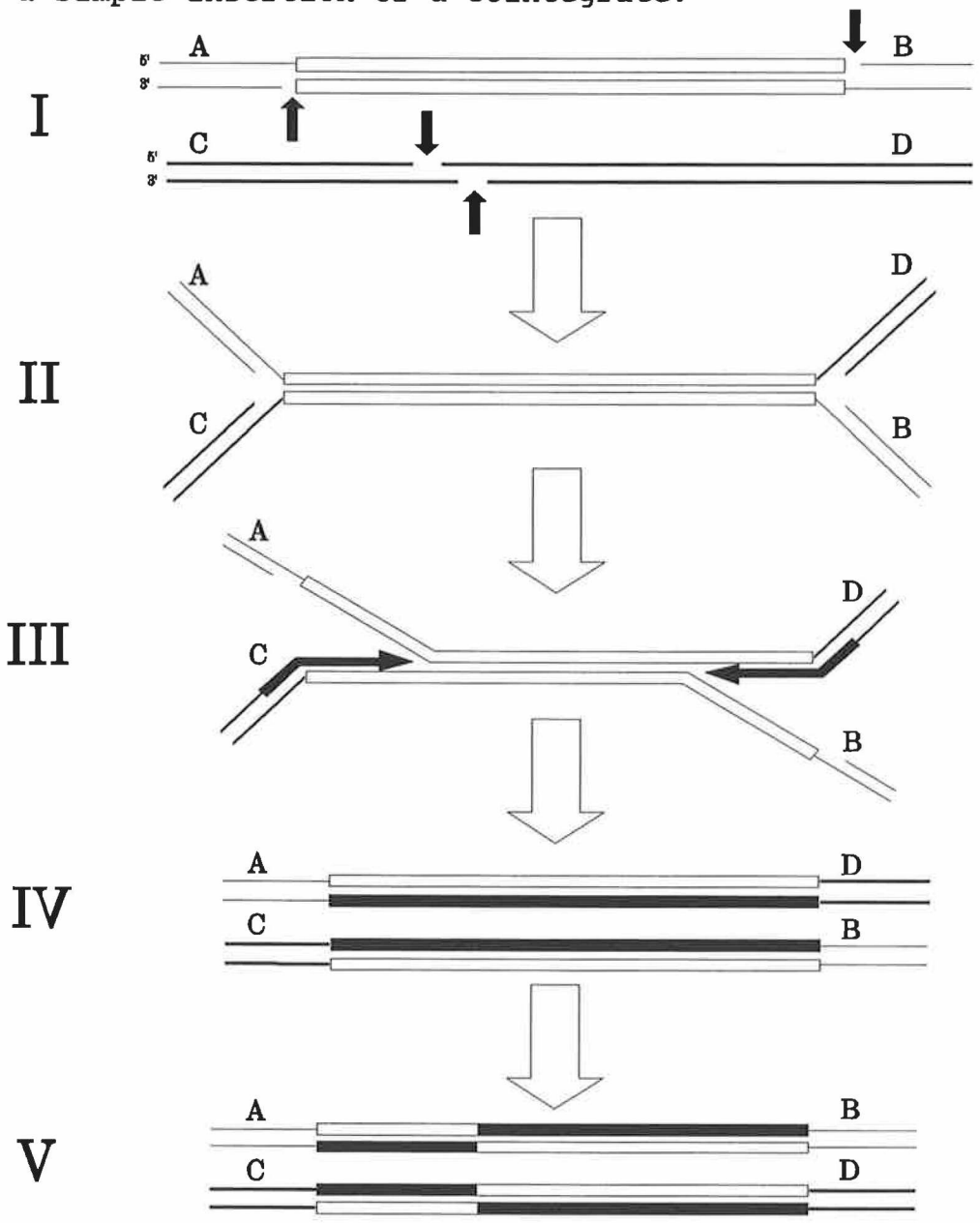

Figure 4. Shapiro Model.

Shapiro[349] has simplified this model by suggesting two changes (see figure 4). Firstly, he proposes that IS element DNA is cleaved at both of its 3′ termini (I), and is ligated to both of the 5′ termini of the staggered nicks of the cleaved recipient DNA (II). (This model is symmetric). Secondly, he proposes that replication forks are formed at these ligation points, which move towards each other (III).

This has two consequences. When the two forks meet, the problems of normal termination of DNA replication of circular replicons arise, and Shapiro suggests that the machinery responsible for this process is used in the resolution of the replicating IS elements. The second consequence of this proposal is that the first product of this process is not a transposed IS element and an otherwise unaltered recipient replicon (IV). Rather the result of the replication is a fused replicon. If the element has transposed from one replicon to another, the fused replicons must be dissociated by recombination of the two copies carried on the cointegrate (V). Since transposition is independent of *recA* in all cases tested, Shapiro suggests that transposons possess a site-specific recombination function of their own. The internal deletions of Tn*3* which, in the experiments by Gill *et al.*[129] and by Arthur and Sherratt,[10] caused replicon fusion rather than transposition were, in some cases, shown to have removed part of the *tnpR* gene which is required for resolution of cointegrate structures (see above).

According to the Shapiro model if transposition occurs into the same replicon and the model is followed through with elimination of the last step, a deletion results. If the orientation of the event is reversed, an inversion will result instead. Inversions in Mu[107] and IS*1*[50] are of the form predicted by the model but in Tn*10* they are more complex, since both IS*10* copies participate simultaneously, using their inside ends.[217,316] An insertion sequence that generated a nearby inversion would result in the intervening material being flanked by ISs in inverted orientation which might have been the origin of transposons like Tn*5* and Tn*10*.[349]

Starlinger[368] argued against Shapiro's model on the grounds that the products of the proposed single-strand cleavages and ligations essentially formed a nick, and that nicks are not known to serve as origins of replication in bacteria. He found the original proposal of Grindley and Sherratt of a displacement synthesis starting at the free 3′ ends to be more appealing. If this displacement synthesis started simultaneously at both of the termini of an IS element, it would have to proceed for about half the length

of the element. At which point, the polymerases pass each other and afterwards the synthesis would continue as the filling of a gap along a single strand. After continuing through the rest of the IS element the polymerase would encounter a free 5′ end of the DNA adjacent to the IS element in its old position. Depending on the polymerase, synthesis could proceed for a short distance in the form of nick translation, but this would not lead to a net duplication of adjacent DNA and would soon come to a halt.[368]

Some of the proposed models contain features not illustrated by the prototypes. The roll-in replication model of Harshey and Bukhari[148] specifies protein-mediated association between the ends of the element and the target, a feature for which there is physical evidence in the case of Mu.[285] Read *et al.*[302] suggest that donor-primed replication of one strand generates a free single-strand intermediate which then inserts and is replicated at the new site. Franke and Clewell[117] have identified a transposon (Tn*916*) which encodes genes for conjugal transfer; genetic evidence suggests that transfer and transposition could be mechanistically linked.[66,117,118]

Genetic effects.

Insertion of an element within a structural gene interrupts the continuity of the gene and causes a mutation. Deletion or inversion events promoted by an inserted element can fuse two previously unconnected genes or operons to produce new functional units. In addition, there can be particular determinants within the translocatable element itself which can influence gene expression.

Polar effects.

In addition to their role in linking nonhomologous DNAs, transposable elements can affect expression of neighbouring genes. Insertion of one or more elements within an operon not only abolishes expression of the gene interrupted by the insertion, but can also greatly reduce the expression of genes located promoter-distal to (or downstream of) the position of the insertion. This type of mutation is called a polar mutation and is the property of insertion sequences which led to their discovery.

Polar mutations were first described by Franklin and Luria[119] and Jacob and Monod[186] and were subsequently shown to be of the nonsense type.[246] These mutations lead to a moderate decrease in the expression of the distal genes (only rarely is the residual synthesis less than a few percent of the wild type level), and the residual synthesis in the mutants is inducible, or repressible to the same extent as in the wild type.[370]

In 1966, a new class of polar mutations was described, which differed from known polar mutations.* The degree of polarity was much more pronounced in these mutants, to the extent that residual synthesis of the products of distal genes was either completely absent or, at most, barely detectable. Most of these mutants reverted spontaneously to wild-type, but the reversion rate was not enhanced by mutagens. In addition, fine structure genetic mapping did not reveal a gradient of polarity associated with map position, a phenomenon often observed with polar mutations of the nonsense type. These mutations were caused by the insertion of IS elements and transposons and are also associated with insertions of phage Mu.[81] It has been directly shown that insertions of IS*1*, Tn*10* and Mu are polar when they are inserted in either of the two possible orientations with respect to transcription of the operon.[46,160] Tn*5* is almost certainly the same, since all insertions are polar.[16] It seems probable, however, that IS*2* is polar in only one orientation as is the case for Tn*2*[328] and Tn*3*.[153]

The simplest explanation for this behaviour is that the inserted elements carry internal transcription terminators which, depending on their location, may interfere with the regulation of expression of adjacent genes. The DNA sequences of IS*1*[280] and IS*2*[128] display nonsense codons in all three reading frames which would be responsible for some of the observed polarity. Transcription termination sites have been detected in several other insertion sequences, including IS*5*[224] and IS*30*.[80] The role of these sites in polarity remains to be determined, but it might be expected that transcription entering the element from flanking DNA is modulated by such signals.[124]

Insertion of IS*1*, Tn*9* and Tn*10* within a ribosomal RNA

* References 198,242,329,330

operon, the transcripts of which are not translated, show incomplete polarity, whereas their insertions into other operons are usually polar.[43,258] In operons which are not translated, termination of transcription may occur at Rho-dependent transcription terminators in the downstream flanking DNA.[288]

Early studies demonstrated that IS*1* exerts a polar effect when inserted into a number of operons in either orientation,[111] but that the polarity can be partially suppressed by mutations in the *rho* gene or by placing transcription under the control of the λ p_L promoter in the presence of the λ antitermination protein N.[307] It has been shown more recently that transcription from the *E. coli* rRNA promoter (*rrnX*) is also resistant to IS*1*- (and Tn*9*-) induced termination.[43,355] Although the absence of a detectable Rho-sensitive site *in vitro*[88] has been taken to indicate that IS*1* exerts its polar effect by introducing nonsense mutations,[82] it now seems clear that a transcription terminator responsible for polarity is located in a region between two IS*1* genes and that partially Rho-dependent termination can occur in either orientation.[171]

Polarity effects, partially suppressible by mutations in *rho*, have also been noted for IS*2* in orientation I[30,82] and for IS*4*.[30] IS*2* has been shown to contain a site which is sensitive to Rho *in vitro*.[88,128] While the polarity effects of IS*1* and IS*2* can be suppressed by the phage λ protein N system,[307] this does not appear to be the case for IS*4*.[127]

It is still possible that IS-mediated polarity has some additional unanticipated features. For example, the suggestion that polarity may sometimes result from the presence within the element of a strong promoter directing transcription which collides with read-through transcription has not been ruled out.[369] Also, the interesting possibility remains that IS element polarity is mediated by unique determinants which occur inside the element as a consequence of their role in the transposition process. Particularly attractive are speculations that transposition could involve unique interactions between RNA polymerase and the ends of the element; such interactions could have the result of terminating transcription entering the element from outside as a by-product of their normal transpos-

ition role.[214]

Promoters.

Some of the genetic phenomena associated with IS-induced mutations were at first difficult to understand. For example, most insertions were highly polar, although a few did not reduce the expression of genes downstream of the insertion points and in some cases even turned on unexpressed genes. The idea gained currency that insertion sequences could turn genes either off or on. It was proposed that at least some insertion sequences carried portable promoters that could activate the transcription of flanking genes.[124]

The earliest example of an IS element turning gene expression on was in the case of IS*2* insertions in the *gal* operon,[328] where it was concluded that IS*2* carries a strong promoter at the end of the element proximal to the *gal* operon, which is able to direct transcription. Although it seems likely that, in that case, IS*2* sequence rearrangements[324] were selected and that these were responsible for *gal* activation,[158,176] subsequent investigations have provided many examples of the activation of gene expression by the spontaneous insertion of IS*2*.[124]

Studies on IS*2* activation of the divergent *argECBH*[130] operon indicated that IS*2* in orientation II was not sufficient for gene activation, implying that IS*2* does not carry an active outward promoter. Glansdorff *et al.*[130] suggested that a partial promoter sequence may be present within the end of IS*2* and that for gene activation, this sequence must be correctly placed with respect to resident signals. This hypothesis was confirmed in a study of IS*2* activation of the *E. coli ampC* gene,[188] where it was shown that IS*2* can provide a -35 promoter region which when correctly spaced from the resident *ampC* -10 region, directs initiation from the normal start point of transcription. In the absence of contradictory data, it is reasonable to assume that activation by the spontaneous insertion of IS*2* in the other cases cited above also occurs by this mechanism.[124]

Similar results have been found in studies with IS*1*.[130] Insertion in either orientation in the promoter region of the pBR322 beta-lactamase (*bla*) gene can result in the IS*1*-directed transcription using a promoter composed of the

resident *bla* -10 region together with a -35 region located in each end of IS*1*.[292] Runoff transcription experiments showed that IS*1*-directed transcription initiates at the normal start point of *bla* transcription.[124]

IS*50* is the flanking element of Tn*5*.[11] IS*50*L differs from IS*50*R in only one base pair.[320] This alteration produces a nonsense codon that was shown to affect an IS*50* gene product involved in transposition. At the same time the mutation generates a better promoter for the expression of the neighbouring kanamycin resistance gene of Tn*5*. (See diagram p44). Similar functional differences are reported for the two IS*10* elements carried on Tn*10*.[114]

For IS*10*, transcription of the transposase gene is driven by a promoter, pIN, another promoter, pOUT, is involved in the control of IS*10* transposition activity and is located at one end of the element.[215] In the case of IS*10* inserted into the *hisG* gene of *S. typhimurium*, in which transcription from pOUT terminates at a Rho-dependent site upstream of the next gene in the operon, *hisD*, this polarity can be suppressed by mutation in *rho*, resulting in IS*10*-driven transcription of *hisD*.[269]

Recent experiments with IS*21* have indicated the presence of an outward facing -35 region located close to one end. The tandem duplication of IS*21*, IS*21.2*, which leads to increased transposition, appears to generate a functional promoter composed of the outward facing -35 region of one IS*21* and a -10 region located in the end of the adjacent element. It has been proposed that this hybrid promoter directs transcription across the second element and drives the expression of transposition functions.[124]

Potential outward-facing -35 regions can be seen within the ends of many insertion sequences.[124] However, since direct evidence concerning the position of transcription initiation is not always available, it remains possible that in several of these cases transcription is initiated from promoters located within the element. This phenomenon has been reported with the transposons Tn*3* and Tn*1000*, where the transcription of the transposase gene can continue into flanking sequences.[354]

From the above observations one can assume that several different mechanisms operate to activate genes adjacent

to a newly integrated IS element: an active promoter can be carried on the IS element; the integration can generate a new promoter at the junction with the IS insertion; a pre-existing promoter can be activated and sequences necessary for negative control can be inactivated. Thus the degree and mode of IS-mediated activation of an adjacent gene depend on the sequence formed by IS integration and on the gene in question.

Other Genetic Effects.

In addition to activation resulting from functional or partial promoter sequences provided by IS elements, IS-induced gene activation can clearly occur by other mechanisms. One example, still poorly understood, involves the *bgl* operon of *E. coli*, which is normally cryptic but can be activated by insertion of either IS*1* or IS*5*.[308] Activation does not involve transcription driven by internal or hybrid promoters furnished by these elements. Activating insertions in both orientations have been observed to occur over a 47bp region (between -124 and -77bp). Since transcription initiates at the same sites in these cases, proximal to the beginning of the first gene, stimulation appears to occur at a distance.[308] One possible explanation for this enhancer-like phenomenon is that the presence of the element enhances binding of the cyclic AMP-binding protein (CAP), which is also involved in *bgl* expression. Indeed, *bgl* activation can also result from point mutations that increase CAP binding.[308] Since mutations in either gyrase subunit are known to activate the *bgl* operon,[99,296] conformational changes introduced directly by the insertion or indirectly, by proteins which bind to the element, are implicated (an alternative explanation is that the operon contains an operator site that is disrupted by the insertion sequence).[308] The effect of insertion sequences on local superhelicity and DNA conformation has not been extensively investigated.[124]

Tandem duplication and further amplification to partially oligomeric states of genomes has been observed in many systems.[7] This can happen as a result of recombination between two small repeats of about 10bp[103] and has also been observed between repeats of the same IS element.* The

* References 57,178,254,341,352

IS elements involved need not actively transpose in this process, but they can provide homology for the *rec* pathways or a site-specific recombination system, such as that of Tn*1000*. Therefore, a DNA segment flanked by directly repeated IS elements is ready to amplify under appropriate conditions. In addition, a DNA segment adjacent to the IS element could also be amplified after the formation of flanking direct repeats by the transposition of the element. Interestingly, amplified structures in the form of tandem repeats of IS*1*-flanked Cm^r transposons are sometimes relatively stable in *E. coli* K12 rec^+ cells even in the absence of the drug.[254]

Recombination between two inverted repeats causes the inversion of the DNA segment between them. Such an inversion can lead to altered expression of genes located on the inverted segment or immediately outside. Inversions may be mediated by a site specific recombination system, as was discovered to regulate the synthesis of alternative flagellar antigens in *Salmonella*[179] and to determine two alternative host ranges in Mu[384] and probably phage P1.[177] Recombination between inversely repeated IS elements can also act as a genetic switch for gene expression.[17,218]

IS-mediated deletion followed by excision of the element could result in operon fusion. Among revertants of Gal^+ strains originally carrying IS*1* in either orientation in the regulatory region of the *gal* operon,[364] new promoters were found to be fused to the operon. Similarly, deletions removing one IS*1* plus neighbouring sequences and fusing Tn*10* genes with *lacZ* were found by Beck.[14] The resulting chimeric proteins had β-galactosidase activity and their synthesis was under the control of a Tc repressor. Fusion of the *his* operon to Tn*5* operons has also been reported.[31]

Transposable Elements in the Rhizobiaceae.

The family *Rhizobiaceae* is a group of gram-negative aerobes which are characterized by their close association with plants. Bergey's Manual of Systematic Bacteriology divides this family into four genera - *Rhizobium*, *Bradyrhizobium*, *Agrobacterium* and *Phyllobacterium*.[197]

The genera *Rhizobium* and *Bradyrhizobium* both stimulate

nodule formation on the roots of legumes, in which the bacteria form a symbiotic relationship with the plants by making atmospheric dinitrogen available to the plant as ammonia in exchange for carbon sources. The two genera were classified as one (*Rhizobium*) and share many characteristics but have been reclassified separately on the basis of important differences such as growth rate, host specificity and metabolic differences.

The genus *Rhizobium* is comprised of three species, *Rhizobium leguminosarum*, *Rhizobium meliloti* and *Rhizobium loti* (*R. leguminosarum* is divided among three biovars, bv. *viciae*, bv. *trifolii* and bv. *phaseoli*) as well as several strains (*Rhizobium* sp.) of uncertain affiliation. All these species were originally defined by the plant which they nodulated as this seemed to be a highly specific association. Since then, however, it has become clear that there are many examples of plants which can be nodulated by different rhizobia and rhizobia which can nodulate different plants.[6] The three biovars of *R. leguminosarum* were classified separately on this basis but an overwhelming amount of evidence suggested that they were all closely related and so they were grouped as one species. A fourth species, *Rhizobium fredii*, is not described in Bergey's Manual. This species nodulates soybean and has been called *Rhizobium japonicum*. Another species (*Rhizobium lupini*) which was included in earlier editions, has been eliminated.[55]

All rhizobial species which nodulated soybeans were classified as a single species, *Rhizobium japonicum*. However it became clear that this was a very heterogeneous grouping and the strains were divided into *R. japonicum* (sometimes *R. fredii*) and a new genus, *Bradyrhizobium*, of which there is only one recognized species, *Bradyrhizobium japonicum*, which nodulates siratro as well as soybean, although there are many examples of related strains nodulating other plants. The bradyrhizobia, or 'slow-growers' is still a very heterogeneous group with many strains which have not been assigned any species name. Instead they are often referred to as *Bradyrhizobium* sp. followed by the name of the appropriate host plant in brackets. This has been the fate of the species *R. lupini* the members of which are now designated *Bradyrhizobium* sp. (*Lupinus*).

The plant diseases, crown gall and hairy root, are caused by members of the genus *Agrobacterium*. Although, as with the rhizobia, there remain uncertainties with regard to the taxonomy of this group,[282] there are five recognized species, the best studied of which is *Agrobacterium tumefaciens*. DNA homology studies and ribosomal RNA analysis suggests that *Agrobacterium* and *Rhizobium* are much more closely related to each other than either is to *Bradyrhizobium*.[6]

Phyllobacterium causes nodule production on the leaves of certain plants and the proposed new genus *Azorhizobium* nodulates the stems of sesbania. Since no insertion sequences have been found in either of these two genera, they will not be mentioned any further in this review.

Flores *et al.*[112] determined that a large part of the genomes of several species of *Rhizobiaceae* consists of repeated sequences. They anticipated that if the entire genome was thoroughly screened, an average of two hundred repeated DNA families would be found per genome. As well as insertion sequences, these families would include rDNA, repeated genes and extended regulatory genes.

R. meliloti.

Three insertion sequences have been characterized in *R. meliloti*.

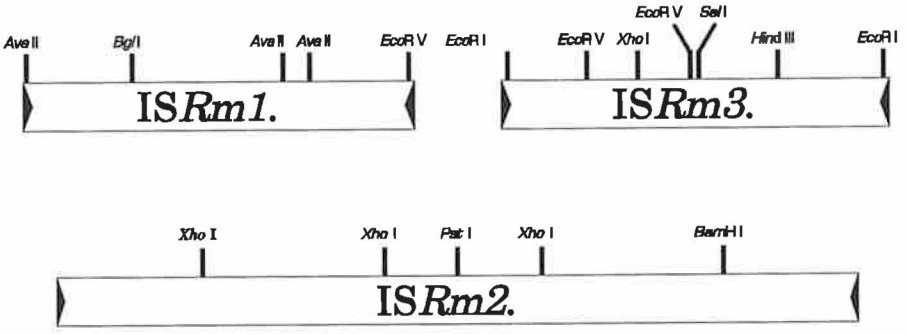

Figure 5. IS Elements from *R. meliloti.*

Ruvkun *et al.*[322] found IS*Rm1* while screening symbiotically defective mutants of *R. meliloti*. These mutants were isolated from several independent Tn*5* mutagenesis experiments and a large number of them were found to be due to the transposition, not of Tn*5*, but of an endogenous insertion sequence, IS*Rm1*. IS*Rm1* is 1·4kb long and was found to transpose preferentially to target regions with a frequency of 10^{-2} to 10^{-3} in two unrelated *R. meliloti* strains.[322,396]

It has been fully sequenced and shows significant homology to IS*2* from *E.coli* and to IS*426* from *A. tumefaciens*.[391]

Despite early indications that it was restricted in its host range,[322,396] IS*Rm1* was shown to be present, throughout the genome, in many different *R. meliloti* strains worldwide, and also to have some degree of relatedness to an element in *R. leguminosarum* bv. *phaseoli*.[397] It was found that the presence and transposition of IS*Rm1* in *R. meliloti* had little impact on nodulating competitiveness, symbiotic effectiveness or frequency of occurrence of strains.[13]

Although the discovery of IS*Rm1* was serendipitous, Dusha *et al.*[101] found a second element, IS*Rm2*, by deliberately using the same strategy. The terminal repeats of IS*Rm2* are 22bp imperfect, inverted repeats. It generates 8bp target duplications and is c. 2·7kb long. Similarities to IS*1* were reported.[101]

Ogawa *et al.*[274] discovered a third insertion sequence in *R. meliloti* during a mutagenesis study of *nod* (nodulation) genes. At the same time, Wheatcroft and Laberge[395] had found an insertion sequence by probing Southern blots of total cellular DNA from different *R. meliloti* strains with plasmid pAT2, an indigenous plasmid from one strain. Wheatcroft and Laberge subsequently showed by hybridization that the elements from the two different studies were the same. This element, IS*Rm3*, was shown to be 1·3kb long, to have perfect, terminal inverted repeats of 30bp and to generate 8 or 9bp duplications of target sequences.[395]

Labes and Simon[225] reported the discovery of five IS elements in *R. meliloti*, one of which was similar to IS*Rm1*. Each of these elements was isolated using a promoterless antibiotic resistance gene carried on pSUP104, indicating that each of the elements contained an outward-facing promoter at the end proximal to the resistance gene. In a later study, Simon *et al.*[357] demonstrated that insertion sequences could be used as means of strain identification, by using hybridization patterns of sixteen IS elements isolated from *R. meliloti*, against total DNA from various *R. meliloti* strains.

B. japonicum.

Kaluza *et al.*[202] found two repeated sequences, RSR*jα*

and RSR$j\beta$, clustered around the *nif* region of *Bradyrhizobium japonicum* and having many of the properties of insertion sequences (terminal repeats, target repeats and a large open reading frame) but without direct evidence of their transposition. Despite their association with the *nif* region, they found no evidence that they had any functional involvement with symbiotic nitrogen fixation.

Burkardt *et al.*[47] reported a series of mutations that were caused by a defined DNA sequence of c. 1·15kb, which they called ISR*1*, in *Rhizobium lupini*. Priefer *et al.*[294,295] confirmed that this segment was indeed an insertion element, with an imperfect, terminal inverted repeat of 42bp and a 4bp target duplication.

A. tumefaciens.

Crown gall disease is a common disease of dicotyledonous plants caused by the transfer of a segment of DNA (T-DNA) from the tumour-inducing plasmid (Ti plasmid) of *Agrobacterium tumefaciens* into the plant genome. This plant-bacteria interaction has recently been reviewed in detail.[406] Several IS elements have been found in this bacterial species.

Insertion elements have been found in both the major types of Ti plasmid i.e. nopaline (IS*426*[385]) and octopine (IS*66*,[35,344] IS*866*[38]). Bonnard *et al.*[38] isolated an insertion sequence (IS*866*) which was found, in many cases, on both the Ti plasmid and the chromosome. IS*66* is found in both *A. tumefaciens* and *R. fredii*.[301]

Two of the elements found in *A. tumefaciens*, IS*426*[385] and IS*427*,[89] have been analyzed in detail. They are very nearly the same length (1·313kb and 1·271kb respectively) and are both bound by imperfect inverted repeats. IS*426* has 32/30bp inverted repeats, contains three overlapping open reading frames (only one of which is preceded by potential promoter sequences) and generates a 9bp target duplication. IS*427* is bounded by 16bp inverted repeats, contains four open reading frames and generates a 2bp target duplication. Comparison of the deduced amino acid sequences of ORF1 from IS*426* and ORF4 from IS*427* revealed a remarkable sequence homology at the carboxyl termini which is more significant at the amino acid level than at the nucleotide level.[91]

T-DNA itself exhibits one of the central properties of

transposons by behaving as a discrete unit of DNA which can integrate into non-homologous DNA sequences.[408] It is not bounded by repeat sequences, although there is an imperfect 22bp repeat outside the T-DNA, which is not transferred to the plant genome.[407] Although the right boundary is precise, with regard to transfer to the plant genome, the left one is less so.[407]

T-DNA consists of three distinct domains, two of which, T_L and T_R, flank the central domain T_C, and are believed to be of eukaryotic origin.[12] De Meirsman *et al.*[90], observing that there was significant homology between segments of the T_C-DNA and the terminal inverted repeats of two insertion elements, speculated that transposition may have played a role in the origin of T_C-DNA. This theory was supported by the work of Machida *et al.*[240] who postulated that the T_R and T_L regions may have originally formed one contiguous eukaryotic sequence which was interrupted by prokaryotic insertion sequences and the DNA rearrangements that often accompany them.

Paulus *et al.*[286] found an IS element, IS*868*, with 82% homology to a region of T_C-DNA, which they described as being IS*51*-like, and 65% homologous to IS*51* itself (an insertion sequence from *Pseudomonas syringae* subsp. *savastanoi*).

A review by Otten *et al.*[282] on the subject of Ti plasmid evolution, detailed the importance of IS elements in the origin of one of two T regions (the TA region) found in biovar III strains of *A. tumefaciens*.

ISR*l1* from *R. leguminosarum* bv. *viciae*.

Angela Ryan (M.Sc. Thesis, DCU. 1987) isolated an insertion sequence, IS*Rl1*, from *R. leguminosarum* bv. *viciae* by introducing the plasmid pSUP104-*rpsL* into the streptomycin resistant strain, 897, and screening for maintenance of the streptomycin resistance phenotype. The ribosomal gene, *rpsL*, confers streptomycin sensitivity and is dominant over streptomycin resistance. However, as the plasmid is maintained by the presence of a tetracycline resistance marker, the cell can recover its streptomycin resistance phenotype by inactivation of the *rpsL* gene, for example by transposition of an IS element into the gene. So any strain growing in the presence of both antibiotics (tetracycline

and streptomycin) and showing an increase in the size of the plasmid pSUP104-*rpsL* is likely to have a transposable element, newly transposed into the *rpsL* gene.

IS*Rl1* was found to be 2·7kb long and to hybridize in Southern blots to several strains of rhizobia and bradyrhizobia. Although Simon *et al.*[357] have since reported the isolation of several elements from *R. leguminosarum* this was the first known example of an insertion sequence from this species. The work presented in this thesis is based on this element.

Materials.

Table 1. Bacterial Strains.

Strain	Phenotype/Genotype	Source
Escherichia coli		
JM83	*ara*, Δ(*lac-pro*), *Sm*^{r}, *thi*, *ϕ80dlacZ*ΔM15.	Vieira and Messing.[387]
S17-1	RP4-2(Tc::Mu)(Km::Tn7), *recA*$^-$, *hsdR*, *pro*, *Tp-Sm*, *res*$^-$, *mod*$^+$.	Simon.[356]
NCIB 9517		DCU Stocks.
Bradyrhizobium japonicum		
U, SD, NJ-2.		T. McLoughlin, Agrigenetics, Wisconsin.
AR6, AR34, 5-7, 27-2, 40-2.	Serogroup 123 wild-type isolates from American soils.	A. Ryan, DCU. " "
USDA123 USDA110		USDA Culture Collection.
Rhizobium meliloti		
65	Wild-type.	M. O'Connell, DCU.
65C	Spontaneously cured of a cryptic plasmid.	"
220-3, 220-5, 220-7, 220-15, 220-16, 220-20.	Wild-type isolates.	A. Pühler, University of Bielefeld.

Table 1. (continued).

Strain	Phenotype/Genotype	Source
Rhizobium leguminosarum biovar *viciae*		
897	Nod$^+$, Fix$^+$, *phe*, *trp*, *str*^{r}.	Johnston and Beringer.[195]
VF39	Wild-type isolate.	University of Bielefeld.
DCU300	Wild-type isolate.	DCU Stocks.
Rhizobium leguminosarum biovar *trifolii*		G. Reigh,
S10 H6, S10 H8,		DCU.
S10 H17, S10 H27.		"
S12-5, S12-11,		"
S12-14, S12-31.		"
S30-2, S30-3,	Wild-type	"
S30-7, S30-11.	isolates.	"
S34-2, S34-16,		"
S34-13, S34-2 H9.		"
SB-H2, SB-H7,		"
SB-2 H1, SB-2 H4.		"
Agrobacterium tumefaciens C58		University of Bielefeld.
Arthrobacter globiformis		DCU Stocks.
Bacillus megaterium		"
Bacillus subtilis		"
Bacillus thuringiensis		"
Lactobacillus amylovorus		"
Lactobacillus plantarum		"
Pediococcus acidilactici		"
Proteus vulgaris		"
Pseudomonas aeruginosa		"
Pseudomonas flourescens		"
Pseudomonas putida		G. Mulcahy, DCU.
Salmonella typhimurium		DCU Stocks.
Serratia marcescens		"
Streptococcus faecalis		"
Streptococcus lactis		"
Vibrio natriegens		"

Table 2. Plasmids.

Plasmid	Relevant Characteristics	Source/Reference
pSVB30	Amp^r	Arnold and Pühler.[9]
pSVB30-*rpsL*::IS*Rl1*	Amp^r	This study.
pSVB30-*rpsL*	Amp^r	This study.
pα1 - pα14	Nested deletions of pSVB30-*rpsL*::IS*Rl1*	This study.
pSUP104-*rpsL*	MobRP4, Tet^r, *rpsL*	R. Simon, Bielefeld.
pSUP104-*rpsL*::IS*Rl1*	MobRP4,Tet^r	Angela Ryan, DCU.

Media.

LB Medium (Maniatis *et al.*[244]).

Used for routine culturing of *Escherichia coli*.

Tryptone	10g
NaCl	10g
Yeast extract	5g
H_2O	1 litre
pH	7·5

XGal (5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside, Sigma) when necessary, was added by spreading 100μl (diluted 1:4) on the surface of the agar and drying briefly.

AG Medium (Cole and Elkan[68]).

Used for routine culturing of *Bradyrhizobium japonicum*.

Arabinose	1g
Sodium gluconate	1g
Yeast extract	1g
NH_4Cl	0·32g
$Na_2SO_4 \cdot H_2O$	0·25g
$MgSO_4 \cdot 7H_2O$	0·18g
Na_2HPO_4	0·125g
$CaCl_2 \cdot H_2O$	0·013g
$FeCl_3 \cdot H_2O$	0·004g
H_2O	1 litre
pH	7·0

TY Medium (Beringer[26]).

Used for routine culturing of rhizobia and of *Agrobacterium tumefaciens*.

Tryptone	5·0g
Yeast extract	3·0g
$CaCl_2 \cdot 2H_2O$	0·7g
H_2O	1 litre
pH	7·2

Media was solidified, where necessary, by the addition of 1·2% Oxoid No. 3 agar. All these media were prepared in distilled water and sterilized by autoclaving at 120°C (1 - 1·5kg/cm^2) for 20 minutes.

Buffers and Solutions.

The following solutions were made up in distilled water and sterilized, where necessary, by autoclaving at 120°C for twenty minutes. They were all stored at room temperature unless stated otherwise. The abbreviations EDTA, SDS, tris and HEPES stand for ethylenediaminetetraacetic acid, sodium dodecyl sulphate, tris[hydroxymethyl]-aminomethane and N-2-hydroxyethylpiperazine-N'-2-ethansulphonic acid respectively. The chemicals and reagents listed here or in the subsequent sections were all obtained either from Sigma Chemical Company or from Riedel de Haen R.G. unless stated otherwise.

		(10:1)	(50:1)
TE Buffer	Tris-HCl	10mM	50mM
	Na_2-EDTA	1mM	1mM
	pH	8·0	8·0
TES Buffer	NaCl	50mM	
	Tris-HCl	10mM	
	Na_2-EDTA	1mM	
	pH	8·0	

(i) For Plasmid Detection and Preparation.

Lysis Solution	Sucrose	8% (w/v)
	Ficoll (type 400)	2% (w/v)
	Tris-HCl (pH 8·0)	50mM
	EDTA	25mM
STET Buffer	Sucrose	8% (w/v)
	Triton X-100	5% (v/v)
	EDTA	50mM
	Tris-HCl (pH 8·0)	50mM
STE Buffer	Sucrose	15% (w/v)
	EDTA	50mM
	Tris-HCl (pH 8·0)	50mM
TTE Buffer	EDTA	50mM
	Tris-HCl (pH 8·0)	50mM
	Triton X-100	0·1% (v/v)
Lytic Mix	EDTA	50mM
	Tris-HCl (pH 8·0)	50mM
	Triton X-100	2% (v/v)

(ii) For Total DNA Purification.

Kirby Mix	Phenol	50ml
	Chloroform	48ml
	Isoamyl alcohol	2ml

Stored under 0·1M Tris (pH 7·5), at 4°C in a dark bottle.

(iii) For Electrophoresis.

50X TAE Buffer	Tris-HCl	242g
	Acetic acid	57·1ml
	0·5M Na_2-EDTA (pH 8·0)	40·0ml
	H_2O	to 1 litre

10X TBE Buffer	Tris-HCl	121·1g
	Boric acid	63·4g
	Na_2-EDTA	7·4g
	H_2O	to 1 litre
	pH	8·3

Stock Acrylamide Solution	10X TBE	100ml
	Urea	234g
	Acrylamide	76g
	N,N'-Methylene-bis-acrylamide	4g

Make up to one litre in water, filter and store in the dark.

Gel Loading Buffer (6X)	Sucrose	40% (w/v)
	Bromophenol blue	0·25% (w/v)

Dissolve in water, filter sterilize and store at 4°C.

(iv) For DNA Fragment Recovery from Agarose.

NEW Wash	Ethanol	50% (v/v)
	Tris-HCl	10mM
	EDTA	1mM
	pH	7·5

(v) For Southern Hybridization.

Denaturing Solution	NaCl	1·5M
	NaOH	0·5M

Neutralization Solution	NaCl	1·5M
	Tris-HCl (pH 8·0)	1·0M

SSC (20X)	NaCl	3·0M
	tri-Sodium citrate	0·3M

Denhardt's Solution (50X)	Ficoll	1% (w/v)
	Polyvinylpyrrolidone	1% (w/v)
	Bovine serum albumin (fraction V)	1% (w/v)

Prehybridization Buffer	6X SSC
	5X Denhardt's solution
	0·5% (w/v) SDS
	100μg/ml denatured salmon sperm DNA

(vi) For Radiolabelling DNA.

Deoxynucleotides	Tris-HCl (pH 7·0)	3·0mM
	dNTP (dCTP, dGTP or dTTP)	0·5mM
	Na_2EDTA	0·2mM

Random Primers Mixture	HEPES	0·67M
	Tris-HCl	0·17M
	2-Mercaptoethanol	33mM
	$MgCl_2$	17mM
	Bovine serum albumin	1·33mg/ml
	Oligonucleotide primers (hexamers)	18 OD_{260} units/ml
	pH	6·8

Klenow Fragment	Klenow fragment	3 units/μl
	Potassium phosphate buffer (pH 7·0)	100mM
	2-Mercaptoethanol	10mM
	Glycerol	50% (v/v)

(vii) For Nested Deletion Formation.

Exonuclease III	Exonuclease III enzyme	90-130u/μl
	KCl	66mM
	Tris-HCl (pH 8·0)	25mM
	Dithiothreitol	1mM
	Glycerol	50% (v/v)

Exo III Buffer	Tris-HCl (pH 8·0)	400mM
	$MgCl_2$	4mM
S1 Nuclease	S1 Nuclease enzyme	40-60u/µl
	NaCl	50mM
	Tris-HCl (pH 8·0)	20mM
	$ZnCl_2$	0·1mM
	Glycerol	50% (v/v)
S1 Buffer	NaCl	1·25M
	Potassium acetate	150mM
	$ZnSO_4$	5mM
	Glycerol	25% (v/v)
	pH	4·6
S1 Stop Solution	Tris base	303mM
	EDTA	50mM
5X Ligase Mix	T4 DNA ligase	0·13u/µl
	Tris-HCl (pH 7·6)	300mM
	$MgCl_2$	50mM
	Dithiothreitol	25mM
	ATP	5mM
	Spermidine	0·5mM
	Glycerol	50% (v/v)
dNTPαS Mix	dATPαS, dCTPαS, dGTPαS and dTTPαS	400mM each
Klenow Fragment	Klenow fragment	5 - 10u/µl
	Potassium phosphate	50mM
	Dithiothreitol	1mM
	Glycerol	50% (v/v)

10X Klenow Buffer	Tris-HCl (pH 7·5)	500mM
	$MgCl_2$	100mM
	Dithiothreitol	1mM

(viii) For Sequencing.

Annealing Buffer	Tris-HCl (pH 7·6)	1M
	Dithiothreitol	160mM
	$MgCl_2$	100mM

'A' Mix-Short	dCTP, dGTP, dTTP	840μM each
	dATP	93·5μM
	ddATP	14μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM

'C' Mix-Short	dATP, dGTP, dTTP	840μM each
	dCTP	93·5μM
	ddCTP	17μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM

'G' Mix-Short	dATP, dCTP, dTTP	840μM each
	dGTP	93·5μM
	ddGTP	14μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM

'T' Mix-Short	dATP, dCTP, dGTP	840μM each
	dTTP	93·5μM
	ddTTP	14μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM

'A' Mix-Long	dCTP, dGTP, dTTP	840μM each
	dATP	93·5μM
	ddATP	2·1μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM
'C' Mix-Long	dATP, dGTP, dTTP	840μM each
	dCTP	93·5μM
	ddCTP	2·8μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM
'G' Mix-Long	dATP, dCTP, dTTP	840μM each
	dGTP	93·5μM
	ddGTP	2·8μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM
'T' Mix-Long	dATP, dCTP, dGTP	840μM each
	dTTP	93·5μM
	ddTTP	2·8μM
	NaCl	50mM
	Tris-HCl (pH 7·6)	40mM
Enzyme Dilution Buffer	Tris-HCl (pH 7·5)	20mM
	Dithiothreitol	5mM
	Bovine serum albumin	100μg/ml
	Glycerol	5%
Labelling Mix	dCTP, dGTP, dTTP	1·375μM
	NaCl	333·5mM

Stop Solution	Bromophenol blue	0·3%
	Xylene cyanol FF	0·3%
	EDTA (pH 7·5)	10mM
	Deionized formamide	97·5%

(ix) For Polymerase Chain Reaction.

PCR Buffer	KCl	0·5M
	Tris-HCl (pH 8·3)	100mM
	$MgCl_2$	15mM
	Gelatin	1%

Enzymes.

Restriction and ligation enzymes and buffers were obtained from BRL and used according to the manufacturers instructions.

Protease type XXV was obtained from Sigma and, before use, was dissolved in TES, allowed to self digest at 37°C for two hours and stored in aliquots at -20°C.

RNase A (Sigma) was dissolved at 10mg/ml in 10mM tris/15mM NaCl (pH 7·5). The solution was boiled for 15 minutes, allowed to cool slowly to room temperature and stored in aliquots at -20°C.

Lysozyme was also obtained from Sigma, prepared immediately prior to use and required no pretreatment.

Antibiotics.

All antibiotics were obtained from Sigma, filter sterilized and stored at -20°C. Ampicillin was dissolved in water to 25mg/ml, tetracycline was dissolved in 50% ethanol to 12·5mg/ml and streptomycin was dissolved in water to 20mg/ml. For use in agar plates, these antibiotic stocks were used at a 1/500 dilution and at 1/1000 for broth cultures.

Methods.

Storage and Culturing of Bacterial Strains.

Working stocks were stored on agar plates at 4°C and all strains were stored in glycerol at -20°C. These were

prepared by growing a culture (in appropriate liquid medium and with antibiotics where necessary) to late log phase. 1ml of this culture was mixed with an equal volume of 80% glycerol and immediately frozen.

Broth cultures were grown from single colony isolates in 10ml of medium. *Bradyrhizobium* cultures required between five and seven days incubation at 30°C, *Rhizobium* and *Agrobacterium* one to two days at the same temperature, while *Escherichia coli* was grown overnight at 37°C. All other species were grown on nutrient agar or broth (*Vibrio natriegens* required 2% NaCl) at 30°C or 37°C.

Transformation.

Competent cells (*E. coli* JM83) were prepared from 100ml of culture grown to an OD_{600} of 0·3-0·4. The culture was chilled on ice for at least 30 minutes. After this interval, the cells (typically 10ml) were harvested by centrifugation at 3000g and 4°C, washed in 5ml of ice-cold 0·1M $MgCl_2$ and resuspended in 5ml of ice-cold 50mM $CaCl_2$. After 30 minutes on ice, the cells were harvested as before and resuspended in 2ml of ice-cold 50mM $CaCl_2$. The cells were then competent and were stored on ice (for no more than 16 hours) until use.

A 20µl aliquot of plasmid DNA was added to 200µl of competent cells and the mixture left on ice for at least 1 hour. The transformation mix was transferred directly from ice to a 42°C water bath for a 2 minute heat-shock and returned immediately to ice. After addition of 0·7ml of LB the cells were left at 37°C for no more than 1 hour and then spread-plated onto selective LB agar.

An aliquot of competent cells (200µl) was put through the above procedure on its own as a negative control.

Modified Eckhardt Procedure for Rapid Plasmid Detection.

The following procedure is an adaptation of the original method of Eckhardt.[102]

A patch of culture was scraped from an overnight plate and emulsified in 10µl of TE with a sterilized toothpick. Separately, 500µl of lysis mix, 10µl of RNase A (10mg/ml) and 1mg of lysozyme were mixed in an microfuge tube. The emulsified culture was mixed with 15µl of this solution and

immediately loaded onto a 0·7% agarose gel, made up in TBE and 0·1% sodium dodecyl sulphate (SDS).

When all the samples were loaded on the gel an empty well was loaded with gel loading buffer to follow the progress of the electrophoresis. The gel was run at 15V for 15 minutes and then at 100V for two hours or more. The gel was stained in an ethidium bromide solution and examined on an ultraviolet transilluminator.

Small Scale Plasmid Purification.

There were two methods used. The first of these, the rapid boiling method, (adapted from Holmes and Quigley[164]) was the quickest and gave the higher yield but occasionally didn't work. The second method, alkaline/SDS lysis, (adapted from Birnboim and Doly[36]) was more reliable but took longer and yielded less DNA.

(i) Rapid Boiling Method.

A patch of growth was scraped from an overnight culture of *E. coli* (grown on appropriate medium and antibiotics), suspended in 300μl of STET buffer in an microfuge tube and mixed with 20μl of lysozyme (10mg/ml in STET). The tube was left at room temperature for 10 minutes, then placed in boiling water for 60 seconds and immediately centrifuged for 10 minutes. The supernatant was removed to a new microfuge tube and mixed with an equal volume of isopropanol. After precipitation at -20°C for 20 minutes, the tube was centrifuged for 5 minutes and the supernatant discarded. The pellet was washed with ether and dried under vacuum. The DNA was resuspended in 50μl of TE. This method also worked with 1·5ml of broth culture.

(ii) Alkaline/SDS Lysis.

An aliquot of 1·5ml of an overnight broth culture was centrifuged for 2 minutes, the supernatant discarded and a further 1·5ml of culture added to the tube and centrifuged as before. When all the supernatant had been carefully removed, the pellet was resuspended in 100μl of 50mM glucose (in 25mM EDTA/50mM tris, pH 8·0) and left on ice for 4 minutes. After this interval, 200μl of 0·2M NaOH/1% SDS was added, mixed and the tube returned to ice for a further 10 minutes, with occasional mixing. Then 150μl of 3M sodium acetate was added and the tube replaced

on ice for a final 10 minutes. After centrifugation, 400µl of supernatant was extracted against an equal volume of Kirby mix and then precipitated with 0·8ml of cold ethanol at -20°C for 15 minutes. Then the pellet was harvested by centrifugation for 15 minutes. The pellet was washed twice in 70% ethanol, briefly dried under vacuum and resuspended in 50µl of TE.

Large Scale Plasmid Purification.

The strain harbouring the plasmid of interest was grown in 100ml of medium (with relevant antibiotic) at 37°C for 24 hours. The cells were harvested by centrifugation (3000g for 5 minutes) and washed in 50ml of 0·2M tris (pH 8·0). After resuspension in 3ml of 25% sucrose (in 0·5M tris, pH 8·0), the cells were removed to a 50ml centrifuge tube and kept on ice for the rest of the procedure. After an interval of 5 minutes, 0·6ml of lysozyme (20mg/ml in 0·2M tris) was added and the tube replaced for another 5 minutes, following which 4·8ml of EDTA (0·25M, pH 8·0) was added and the tube replaced for a final 5 minutes. Following the addition of 4·8ml of lytic mix the tube was left for 20 minutes during which time it was occasionally mixed by inversion. At the end of this time the tube was centrifuged for 40 minutes at 26000g and the supernatant removed to a new tube. To form a caesium chloride gradient, 9·6g of CsCl was dissolved in 9·2ml of the supernatant and the solution transferred to polyallomer ultracentrifuge tubes. Ethidium bromide (0·2ml of a 10mg/ml solution) was added and the tube filled with mineral oil. The tubes were heat-sealed and spun at 170000g (50000rpm in a 70.1 Ti Beckman rotor) for 24-48 hours at 18°C.

The plasmid band was lit with ultraviolet light (although it was quite visible in ordinary light), and removed from the tube with a sterile needle and syringe (the top of the tube must be punctured beforehand). The ethidium bromide was removed by repeated extraction against isoamyl alcohol or against butanol saturated with water and the caesium chloride was removed by dialysis in 5l of distilled water (changed 4 times over 48 hours).

Rapid Large-scale Plasmid Purification (Tetsuro[380]).

An overnight culture (grown in 50ml of appropriate medium) was centrifuged at 4°C and 2000g. The cells were resuspended in 2ml of ice cold TE (50:1) buffer, divided between two microfuge tubes and centrifuged again for 1 minute. The supernatant was discarded and the cells resuspended in 300µl of ice-cold STE buffer. After the addition of 100µl of lysozyme (5mg/ml in STE), the suspension was left on ice, with occasional inverting, for 10 minutes. After this time 370µl of ice-cold TTE was added and the suspension left on ice as before. After 10 minutes 2µl of diethyl pyrocarbonate was added and the tubes placed in a boiling water bath for 45 seconds. After this the tubes were immediately spun at full speed for 15 minutes. The pellets were removed with a sterile toothpick, the supernatants respun if necessary and removed to new microfuge tubes where they were mixed with 1ml of ethanol and immediately centrifuged for 3 minutes. The supernatant was discarded and the pellet dried, for no more than 10 minutes, on the bench.

The pellets were resuspended in water and pooled to give a final volume of 700µl which was incubated first at 37°C for 30 minutes with 5µl of RNase A (10mg/ml) and then at 50°C for 15 minutes with 1µl of pronase (10mg/ml). After this, 35µl of 4M NaCl was added and the solution was extracted once with phenol, twice with Kirby mix and once with chloroform/isoamyl alcohol (24:1). After this, the DNA solution was mixed with 800µl of ethanol (at -20°C) and immediately centrifuged for 5 minutes. The pellet was washed twice with cold 70% ethanol, dried under vacuum and resuspended in 100µl of TE.

Total DNA Purification from *Bradyrhizobium* Strains.

The strain was grown to late log phase in 10ml of AG broth and pelleted by centrifugation at 4000g for 10 minutes. The pellet was washed in 3ml of TES buffer and then in 3ml of 25% sucrose in TE buffer. All the supernatant was carefully removed and the pellet resuspended in 600µl of 25% sucrose in TE buffer. Following the addition of 120µl of lysozyme (5mg/ml in water) and 480µl of 0·1M EDTA (pH 8·0), the suspension was left at room temperature for 10

minutes. After this interval, 1·2ml of 0·2% SDS and 300µl of protease (1mg/ml in TES) were added and the mixture incubated at 37°C for 1 hour. The solution was then extracted vigourously with 2·7ml of phenol for 15 minutes. The phases were separated by centrifugation at 3000g for 10 minutes and the upper phase was reextracted as before. Following a second centrifugation, the upper phase was removed to a corex tube and centrifuged for 15 minutes at 27000g. The supernatant was removed to a new tube and extracted vigourously against an equal volume of Kirby mix and then against ether. The ether was removed with a pipette and by leaving the solution under vacuum for 10 minutes. The DNA was precipitated by first adding 270µl of 3M sodium acetate and then 6ml of cold ethanol. The solution was left at -20°C for an hour and, following pelleting at 6000g for 10 minutes, the DNA was transferred to an microfuge tube, washed twice with 70% ethanol, dried under vacuum and dissolved in 200µl of TE.

Total DNA Purification from Strains Other Than Bradyrhizobia.

One millilitre of an overnight culture was centrifuged in a benchtop microfuge for 10 minutes and the pellet resuspended in 0·5ml of 6·7% sucrose in TE. To this suspension was added 100µl of lysozyme (10mg/ml in TE) and the tube was incubated at 37°C for 5 minutes. Next 50µl of 20% SDS was added and the tube was again incubated at 37°C for a further 5 minutes. After this, 25µl of 0·5M EDTA (pH 8·0) and 0·5 ml of phenol were added and the sample was gently mixed, 100µl of chloroform:isoamyl alcohol (24:1) was added and the mixture centrifuged for 5 minutes. This phenol extraction was repeated as necessary. After the final spin the supernatant was extracted with an equal volume of chloroform:isoamyl alcohol and centrifuged once more. The supernatant was mixed with an equal volume of isopropanol and left at -20°C for 10 minutes. After a 10 minute centrifugation the DNA pellet was washed twice in 70% ethanol, dried under vacuum and dissolved in 90µl of TE.

Agarose Gel Electrophoresis.

DNA fragments were analyzed on 0·7% agarose gels,

which were made up and run in TAE buffer. Gels were run in horizontal apparatus at constant voltage (100V typically, but 40V for large, overnight gels). Gel loading buffer was added to the DNA samples prior to loading, to increase the sample density and to monitor the progress of electrophoresis. To facilitate size determination 1kb DNA ladder was run alongside the samples. These markers ranged in size from 12kb to 75bp with increments of 1kb (except at the lower sizes) and were obtained from Gibco BRL.

The DNA was stained by soaking the gel in ethidium bromide (5μg/ml) for 20 minutes or more, destained by immersion in water for 10 minutes and visualized on an ultraviolet transilluminator. The gels were photographed on black and white, 100asa film (T-max), through an A.003, red Cokin filter.

Estimation of DNA Fragment Lengths from Agarose Gels.

The SEQAID™ computer package (described below, page 90) includes an program to estimate the size of DNA fragments given their mobility and the size and mobility of (at least three) known standards. The program is based on a method developed by Schaffer and Sederoff[337] which is described briefly here as it is important for discussion in a later section.

The basic reciprocal relationship between length of the DNA fragment (L) and its mobility (m) can be expressed as $(m-m_o)(L-L_o) = c$, where c is a constant and the o subscripts refer to offset values. The reasons for including these offset values and their significance are fully explained in the original reference but their use has two important consequences. Firstly, the mobilities may be measured from any convenient point and secondly, the consistency of the lengths and migration distances of the various fragments with respect to each other can be tested.

When the program is run, the user supplies the size and mobility of the standard fragments (the mobility can be measured very accurately by projecting an enlarged image of the gel from a photographic negative onto a wall) and the program returns a value for each indicating the internal consistency of the data. Then the user simply enters the mobility of each fragment (in millimetres) and the computer

returns the estimated size in basepairs.

Purification of DNA Fragments from Agarose Gels (Geneclean).

The DNA fragment of interest was cut out of the gel with a scalpel blade, chopped up and dissolved in 2-3 volumes of sodium iodide at 55°C (usually this takes 5 minutes). One microlitre of a glass bead slurry (silica 325 mesh, Stratech Scientific Ltd.) was added to the solution which was mixed and left to stand at room temperature for 5 minutes. The glass beads were pelleted using a 5 second, high speed spin in a microfuge. The beads were washed three times in 50µl of NEW wash (10mM Tris-HCl (pH 7·5), EDTA 1mM in 50% (v/v) ethanol) and finally resuspended in 10-20µl of TE buffer. The DNA was eluted from the beads by incubation at 55°C for 3 minutes and the beads removed by centrifugation and transfer of the supernatant to a new microfuge tube.

Transfer of DNA to Nitrocellulose Filters.

*Eco*R I restricted DNA was electrophoresed in a large 0·7% agarose gel overnight. Also included on the gel was a 1kb DNA ladder (Sigma) as size markers. The gel was stained and photographed as described above.

The gel was soaked in 0·25M HCl for 10 minutes after which it was rinsed with tap water and then placed in 500ml of denaturing solution for 40 minutes with occasional shaking, after this time the gel was left in 500ml of neutralizing solution for 1 hour.

A sheet of Whatman 3mm chromatography paper was placed on the platform of an empty gel box, such that its ends hung down into the reservoirs at either end. The box was partly filled (to below the platform) with 10X SSC and the paper was soaked with same. The gel was placed face up on this and was itself covered with a sheet of nitrocellulose paper (Schleicher and Schuell) which had been cut to the exact size of the gel and presoaked in 2X SSC. On top of the nitrocellulose two sheets of chromatography paper were laid, similarly cut and soaked, next a single dry sheet cut to the same size and above this c. 10cm of tissue paper all of which was kept compressed with a weight. None of the

layers from the gel upwards was allowed to dip into the 10X SSC in the bottom of the gel box. The blotting was left to proceed overnight.

The nitrocellulose and gel were removed from the apparatus, the position of the wells were marked on the filter which was then soaked in a bath of 6X SCC for 5 minutes, air dried and baked at 80°C between two sheets of chromatography paper for 2 hours. The gel was restained in ethidium bromide to check the transfer of DNA.

Radiolabelling of DNA.

Radiolabelled probe was prepared by the random priming method, using a kit supplied by BRL Life Technologies, Inc. (cat. no. 8187SA). Twenty five micrograms of the probe was linearized by cutting with restriction endonucleases. One microlitre of diluted 1kb DNA ladder (1μg/ml) was added and the DNA denatured by boiling for 10 minutes. After this it was placed on ice and 2μl each of dCTP, dGTP and dTTP were added along with 15μl of primer buffer. Lastly, 4μl of ^{32}P radio-labelled dATP was added and enough distilled water to bring the final volume to 49μl. One microlitre of Klenow fragment (3u/μl) was added, the solution was mixed and placed at 25°C for one hour.

DNA-DNA (Southern^{365}) Hybridization.

The filter was soaked for five minutes in 2X SSC and then placed in a heat-sealable bag, three of the sides of which were then sealed. Prehybridization solution was added (0·2ml/cm^2) and the fourth side was heat-sealed. The bag was put on a shaking-table at 64°C for at least 6 hours.

After this time the radiolabelled probe was boiled for 10 minutes and the prehybridization solution discarded. The probe and fresh prehybridization solution (50μl/cm^2) were placed in the bag with which was then resealed and returned to the incubator and left overnight (16 hours).

The following day, the bag was cut open and the filter washed twice in 2X SSC/0·1% SDS (15 minutes at 64°C) and twice in 0·2X SSC/0·1% SDS (15 minutes at 64°C) and left to dry behind a perspex screen.

When dry the filter was placed in another heat-sealable bag and exposed for 3-4 weeks to Kodak X-ray film

which was subsequently developed and fixed with Kodak materials.

Sepharose Spin-columns.

To change the buffer of a DNA sample, it was passed through a mini-sepharose column which was prepared as follows. A hole was pierced in the bottom of a 0·7ml microfuge tube, a small amount of glass beads (40 mesh, BDH Ltd.) was placed in the tube to cover the hole. Then 0·5ml of 70% Sepharose (CL-6B, Sigma Chemical Co.) was added and the tube was placed into a 1·5ml microfuge tube.

The column was centrifuged at 1·5ml for exactly 2 minutes and the liquid in the bottom of the larger microfuge tube was discarded. The timing and speed of each subsequent spin were kept exactly the same. A 20-50μl aliquot of TE (the same volume as the DNA sample) was added to the top of the column, which was centrifuged as before. This washing procedure was repeated until the volume of eluate equalled the volume of buffer added to the top and then the wash was repeated once more. Then the DNA was added to the column, which was removed to a new, sterile 1·5ml microfuge tube and spun as before. This time the eluate contained the DNA.

Nested Deletions of *rpsL*::IS*Rl1*.

Nested deletions of *rpsL*::IS*Rl1* were made using enzymes and reagents from Pharmacia's double-stranded nested deletion kit (cat. no. 27-1682-01). (To facilitate this, and subsequent sequencing, the fragment was cloned into pSVB30.) The following procedure follows the manufacturer's instructions.

The strategy used to generate nested deletions is shown in figure 6 (page 86). The plasmid was cut with the restriction endonuclease, *Hin*d III, and then passed through a sepharose spun-column. Following this 10μl of the DNA (0·2μg/ml) was incubated for 15 minutes at 37°C with 1μl of Klenow fragment (0·05 units), 1μl of 10X buffer and 1μl of deoxyribonucleotide 5'-α-thio-triphosphates (dNTPαS). The reaction was stopped by heating to 65°C for 20 minutes and the DNA precipitated with 20μl of a NaCl/glycogen solution (0·25M and 0·25μg/μl respectively) and 75μl of ethanol.

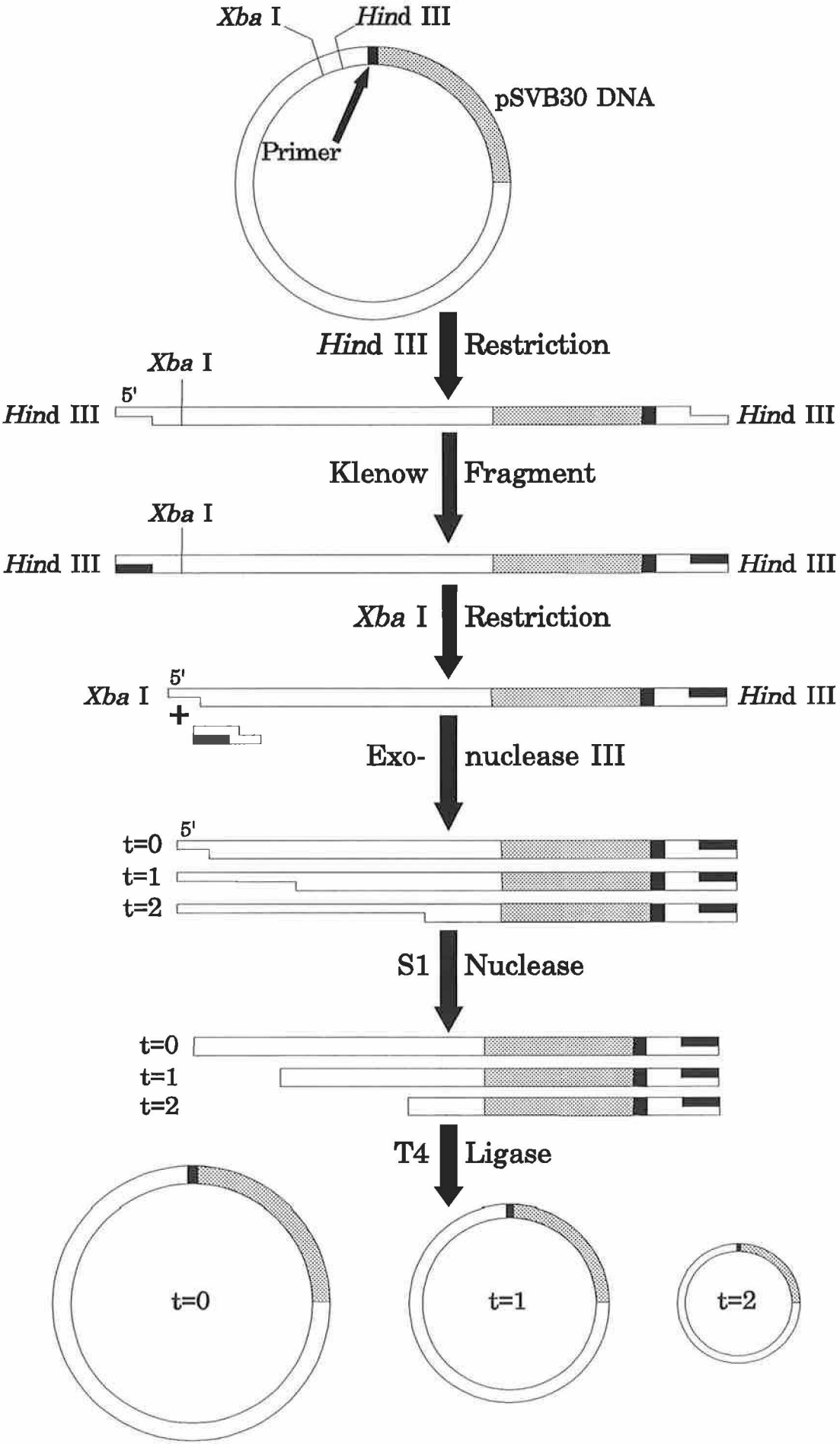

Figure 6. Generation of Nested Deletions.

After centrifugation, washing with 70% ethanol and drying under vacuum the DNA was resuspended in 10μl of water. Finally, the DNA was cut again, this time with *Xba* I in a total volume of 20μl. The end result of these manipulations was that the linear DNA had a blunt end which was exonuclease III-resistant (because of the incorporation of thionucleotides) and the other end was a 5′ overhang which was susceptible to exonuclease III action.

One microlitre of S1 nuclease was mixed with 99μl of 1X S1 buffer and this mixture dispensed in 3μl aliquots to 20 microfuge tubes which were all stored on ice. Twenty microlitres of 2X exonuclease III buffer was prepared from a 6X concentrate and water (no NaCl was added) and mixed with the 20μl of DNA. The tube was left to equilibrate at the reaction temperature (37°C) for 2 minutes. A 2μl aliquot was removed from the tube as a "time = 0" sample and mixed with 3μl of the S1 nuclease solution. A 1μl aliquot of exonuclease III was added to the reaction tube and mixed in gently. After every two minutes a sample was removed from the reaction and mixed with 3μl of S1 nuclease, all of which were left on ice until the last sample was taken. At that time, the tubes were incubated simultaneously at room temperature for 30 minutes. A 1μl aliquot of S1 stop solution was added to each sample and they were all incubated at 65°C for 10 minutes.

At this point the size of the deletions was checked by electrophoresing 3μl of every second sample on an agarose gel following which, all the deletions were recircularized using 17μl of prepared ligation mix and transformed into *E. coli* JM83. Several transformants from each plate (one plate per time sample) were analyzed on an Eckhardt gel alongside cells harbouring the intact plasmid and cells harbouring the vector. Suitable deletants were selected.

DNA Sequencing.

Sequencing was performed using enzymes and solutions from Pharmacia's ^{T7}Sequencing kit (cat. no. 27-1682-01). 16μl of DNA (0·2μg/ml - prepared by either of the large scale methods above) was mixed with 4μl of 2M NaOH and left at room temperature for 20 minutes. After this time the solution was passed through a sepharose spin column and

10μl of the eluate mixed with 2μl of annealing buffer and 2μl of primer and the mixture incubated at 37°C for 20 minutes and then at room temperature for at least 10 minutes.

For each DNA template to be sequenced, a set of four microfuge tubes, each containing 2·5μl of one of the dideoxynucleotides ddATP, ddCTP, ddGTP or ddTTP, was prepared and stored on ice. Enzyme premix was prepared by adding (for each template) 1μl of water, 3μl of labelling mix A, 2μl of T7 polymerase (diluted to 1·5u/μl with dilution buffer) and 1μl of dATPα^{35}S to an microfuge tube. A 6μl aliquot of this enzyme premix was added to the tube containing the annealed template and primer and the labelling reaction left to proceed at room temperature for 5 minutes during which the four microfuge tubes containing the ddNTPs were placed in a 37°C waterbath. Next 4·5μl of the labelling reaction was added to each of the four ddNTPs (using a fresh pipette tip for each transfer), which were incubated at 37°C for a further 5 minutes after which time 5μl of stop solution was added to each tube. The tubes were stored at -20°C until they were loaded on a sequencing gel.

Acrylamide Gel Electrophoresis (Sequencing).

The glass plates were washed in a dilute solution of RBS 25 (AGB Scientific Ltd.), rinsed in distilled water and left to dry. The plates were polished thoroughly with ethanol and then rinsed with Repel Silane (LKB Ltd.), allowed to dry, rinsed in distilled water and polished with ethanol again. Spacers (0·2mm) were used to separate the plates which were taped together along three edges using packing tape and left lying with the open end slightly raised.

Acrylamide was prepared by mixing 60ml of stock acrylamide solution, 60μl of NNN'N'-tetramethylethylenediamine (LKB Ltd.) and 224μl of 10% ammonium persulphate. This was poured, without delay, between the plates using a 10ml syringe. The sharkstooth combs were inserted with the teeth facing outwards and with a clamp at either edge and the gel was left to polymerize (usually less than an hour).

The tape, combs and clamps were removed from the plates, which were then clamped into the gel rig. TBE was poured into the buffer reservoirs and the loading well was

rinsed out before the combs were reinserted with the teeth facing downwards. The gel was prerun at 1500V for one hour. The samples were heated to 75-80°C for 2 minutes prior to loading and during this time the wells to be loaded were rinsed out with TBE in a syringe to remove air bubbles and unpolymerized acrylamide. A 3μl aliquot of each reaction was loaded in the order A, C, G, T. The stop solution contains two dyes (see materials section) and these were used to monitor the progress of the electrophoresis. The gel was typically run until 1 hour after the second band of dye (xylene cyanol) had run off the end of the gel and then another set was loaded and run until the first band of dye (bromophenol blue) had run off.

The gel was taken off the rig and the plates carefully separated. The gel, and the glass plate to which it was stuck, were immersed in a 10% methanol/10% acetic acid aqueous solution for 20 minutes. The plate was carefully removed from the solution, first repositioning the gel if necessary, and then placed on the bench. Two sheets of Whatman No. 1 filter paper (46x57cm) were placed on top of the gel and left for a few minutes. The filter paper was peeled away from the plate, taking the gel with it and placed on a flat surface, gel upwards. The gel was covered with cling film and any excess gel, filter paper and cling film cut away. This 'sandwich' was put, cling film upwards, on a slab gel drier and left to dry at 80°C under vacuum for 2·5 hours.

When the gel was dry, the cling film was removed and the gel exposed to Kodak X-OMAT (35x43cm) X-ray film in a light proof cassette for 36 hours. The autoradiogram was developed with Kodak LX-24 X-ray developer and fixed with Kodak FX-40 X-ray liquid fixer (both diluted 1 in 5).

The sequence was read from the autoradiogram manually with error checking using a specially written computer program (see appendix A).

Computer Analysis.

Sequence entry (from an autoradiogram or, occasionally, from a publication), with error-checking, to the computer was performed using a specially written computer program. This program is written in BASIC under the VAX/VMS

operating system and is listed in appendix A. The program produces a file that holds the data in a format compatible with all the analysis programs i.e. sequence entered in capital letters in lines of less than eighty characters and with the sequence name in the first line preceded by a 'greater than' symbol - >.

Comparison of sequences, restriction site searches and searches for open reading frames (and translation to protein) were carried out using the program SEQAID II™ (Rhoads, D.D. and Roufa, D.J., Molecular Genetics Laboratory, Kansas State University). This program runs on IBM compatible micro-computers and is available from the European Molecular Biology Laboratory file server at IN%"netserv@embl-heidelberg.de".

Searches for most functional sites were performed with the SEQAID II program, searches for translational frame-shifting motifs were performed using FRAME.BAS, a program specially written for this purpose (listed in appendix A), and promoter sequences were sought using SIGNAL SCAN[293] a VAX/VMS program which is available from the EMBL file server at IN%"netserv@embl-heidelberg.de".

Gene seach by content (for a detailed explanation of this method see refs. 366 and 367) was perfomed using a module in the SEQAID package to construct a codon bias table for *R. leguminosarum* from protein coding regions found in sequences from this organism in GENBANK release 76. Information from the GENBANK entries for these sequences are found in appendix B listing the name and accession number of each sequence, a brief description and the position of the coding regions within it. The codon bias table is presented in the results section (page 129). Another module from the SEQAID package was then used to compute values for the expression log[P/(1-P)] for each reading frame in intervals of 5 codons, where P is a measure of the probability of each interval coding for protein.[341a,341b] This expression is then plotted against base pair position for each reading frame for the entire length of the element.

Searches of DNA and protein databases for similar sequences were performed with the BLAST network server (Altschul *et al.*[5]). A program (also called BLAST) was ac-

cessed by remote login to the VAX mainframe computer (vax1-.tcd.ie) in the Genetics Department of Trinity College Dublin. This program creates an text file, containing the query sequence and the parameters for the search, and mails it to the BLAST network server at IN%"blast@ncbi.nlm.nih-.gov". Full documentation on this service is available by sending an e-mail message to that address with the word HELP in the body of the message.

Two databases are available to users of this service. A peptide sequence database which comprises all the sequences in the SWISS-PROT, PIR, GENPEPT, KABATPRO and TFD databases and a nucleotide sequence database which comprises all the sequences in the GENBANK, EMBL, KABATNUC, EPD and DBEST databases (see the help file for further information on these databases). In both cases the database includes daily updates of the components databases, where available. The server returns a file containing information on any sequence which meets the required score, previously specified by the user.

The BLAST service is used to find sequences with close similarities to the query sequence, to find weaker similarities a program called BSEARCH was used. Through this program the latest release (but not daily updates) of the GENBANK nucleotide and PIR peptide databases can be searched. The program returns information on all the best matches to the query sequence (the number of which is specified by the user).

Also at this site (vax1.tcd.ie), the ACNUC utility (Guoy *et al.*[132]) was used to examine the GENBANK and PIR databases, for example, to find all sequenced IS elements from the *Rhizobiaceae*.

Polymerase Chain Reaction.

The inverted repeat at the ends of IS*RI1* was chosen as the primer (for both ends) of a polymerase chain reaction (PCR) to be performed on various DNA samples. This 13mer (GTAAGCGACAAGC) was synthesized by Pharmacia and supplied at a concentration of 438μg/ml (≅1mM).

The PCR was carried out in a reaction volume of 50μl - 3μl Mg_2Cl (25mM), 1μl of each dNTP (10mM in H_2O), 5μl PCR buffer, 0·5μl Taq polymerase (Promega), 1μl primer, 2μl DNA

and 35μl H_2O. The reaction solution was mixed and covered with an overlay of mineral oil (20μl).

The reaction was performed in a programmable temperature cycler (Omnigene™ - Hybaid). The program used was as follows. The first stage was to completely denature the DNA by heating at 94°C for 5 minutes followed by cooling to 42°C and incubation at this temperature for 30 seconds. The following three steps were repeated in a cycle 30 times. The reaction was heated slowly (ramping rate of 2sec/°C) to 72°C and held there for 2·5 minutes. Then it was heated further to 94°C for 30 seconds and lastly, cooled to 42°C for 30 seconds. After 30 cycles the polymerization reaction was run to completion by heating (again with a ramping rate of 2sec/°C) to 72°C for 5 minutes.

CHAPTER 3. SEQUENCING OF ISRL1.

Introduction.

The discovery of IS*Rl1* has already been described (page 64 above). Insertion sequences are interesting both in themselves and in their relationship to their host, also there is considerable interest in the use of these elements as tools in genetic studies and engineering.

IS*Rl1* was found to be 2·7kb long (subsequently shown, in this study, to be 2495bp). This is quite large for an insertion sequence the sizes of which are normally in the range 0·7 - 1·8kb. It is possible therefore, that IS*Rl1* is a transposon and encodes some function unrelated to transposition. To investigate this possibility and to determine its relationship to other IS elements from the *Rhizobiaceae*, it was decided to sequence IS*Rl1*.

The method chosen was dideoxy chain-termination, or Sanger,[333] sequencing. At best, this method yields 400 - 500bp of sequence data from a given template, which is not enough to sequence the whole element. The best way around this problem is to make a set of nested deletions of the fragment *rpsL*::IS*Rl1*, sequence each of them and compile the data.

As a first step the *rpsL*::IS*Rl1* fragment was cloned into pSVB30 (described below, page 94). This plasmid is one of a family of sequencing vectors[9] (all of which are based on pUC8 and like pUC8, carry the universal M13 sequencing primer), differing only in their multiple cloning sites (mcs). The pSVB vectors carry the gene for ampicillin resistance (*amp*) and the *lacZα* gene allowing for the direct identification of inserted DNA in JM83 by white colonies on LBamp plates with XGal[387] (see figure 8, page 95) although in some of these vectors the reading frame of *lacZα* is disrupted.

To generate nested deletions it was necessary to clone the fragment into a vector that has two non-overlapping, restriction endonuclease sites in the multiple cloning site which are not present in the fragment and to facilitate the initial cloning experiment it helps if the reading frame of *lacZα* is conserved. Using these criteria, pSVB30 was chosen

as the best vector candidate.

Cloning of IS*R11* into pSVB30.

The fragment *rpsL*::IS*R11* was cut from the plasmid pSUP104-*rpsL*::IS*R11* with *Eco*R I and purified by the gene-clean technique (the isolation of an IS*R11* insertion into *rpsL* was described by A. Ryan, M.Sc. thesis, DCU 1987). The vector pSVB30 was also cut with *Eco*R I and then incubated at 65°C for 10 minutes to inactivate the enzyme. The *rpsL*::IS*R11* fragment was ligated into pSVB30 and then used to transform *E. coli* JM83, transformants carrying inserts being detected as white colonies on LB agar containing ampicillin and XGal.

The plasmids used in these ligations and the ligation products are illustrated in figure 7, which shows an agarose gel bearing the various fragments used in the cloning experiments, run alongside size markers.

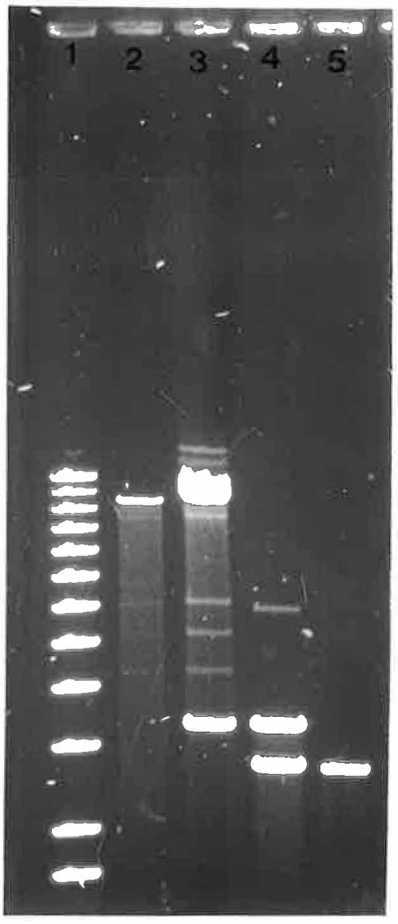

Figure 7. Agarose Gel Showing Cloning of *rpsL*::IS*R11* into pSVB30.

Lane 1. 1kb DNA Ladder.
Lane 2. pSUP104 Cut with *Eco*R I.
Lane 3. pSUP104-*rpsL*::IS*R11* Cut with *Eco*R I.
Lane 4. pSVB30-*rpsL*::IS*R11* Cut with *Eco*R I.
Lane 5. pSVB30 Cut with *Eco*R I.

When IS*R11* was first isolated, a restriction map of the element and *rpsL* was drawn (A. Ryan, M.Sc. thesis, DCU). This map has since been found to be in error in several regards. Firstly, IS*R11* is about 200bp shorter than had been thought and secondly, the positions of two *Pst* I sites in IS*R11* had been reversed. A third, important error (the absence of an *Xba* I site) is discussed below on page 106. Finally, through the sequencing work presented below, the position of IS*R11* in *rpsL* has been precisely determined. The amended map of *rpsL*::IS*R11* (in pSVB30) is presented in figure 8.

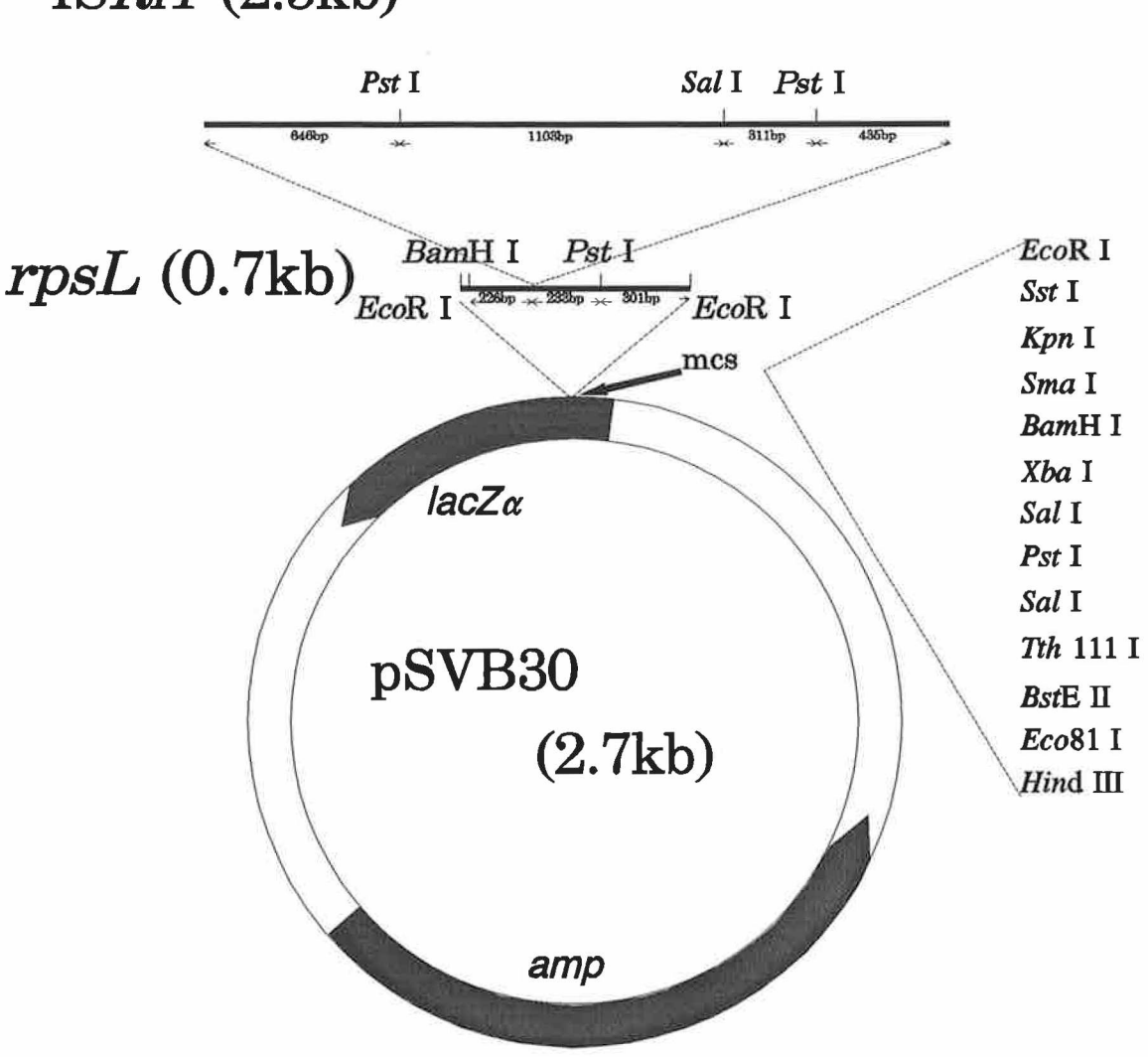

Figure 8. Map of pSVB30-*rpsL*::IS*Rl1*.

Nested Deletions of *rpsL*::IS*Rl1*.

The cloning of *rpsL*::IS*Rl1* into pSVB30 has already been described (page 94, above). It was necessary to choose two restriction sites, which are present in the pSVB30 vector but not in the *rpsL*::IS*Rl1* insert, for use when making the nested deletions. The multiple cloning site of pSVB30 is shown in figure 9 and consists of the following restriction endonuclease sites (listed in order of decreasing distance from the universal primer) *Eco*R I, *Sst* I, *Kpn* I, *Sma* I, *Bam*H I, *Xba* I, *Sal* I, *Pst* I, *Sal* I, *Tth*111 I, *Bst*E II, *Eco*81 I and *Hind* III. The sites *Sal* I, *Pst* I, *Bam*H

I, *Sst* I and *Sma* I were shown by restriction digest to be present in the fragment *rpsL*::IS*Rl1* and are drawn above the line in figure 9, while the sites *Hin*d III, *Xba* I, *Kpn* I and *Eco*R I were shown to be absent (these sites are drawn below the line in figure 9. The enzymes *Eco*81 I, *Bst*E II and *Tth*111 I were not available so their presence/absence remained undetermined (shown in brackets in figure 9).

Of those sites which were absent, *Eco*R I was clearly unsuitable as *rpsL*::IS*Rl1* was cloned into this site, and *Kpn* I was unsuitable as this site cuts to give a 3′ overhang which is resistant to the action of Exo III, the enzyme used to generate nested deletions. Ordinarily this property is highly desirable in an experiment such as this (see page 85 for nested deletion strategy) but in this case there is no suitable restriction site between the *Kpn* I site and the *rpsL*::IS*Rl1* fragment which yields susceptible ends (5′ overhang or blunt) when cut (see figure 9).

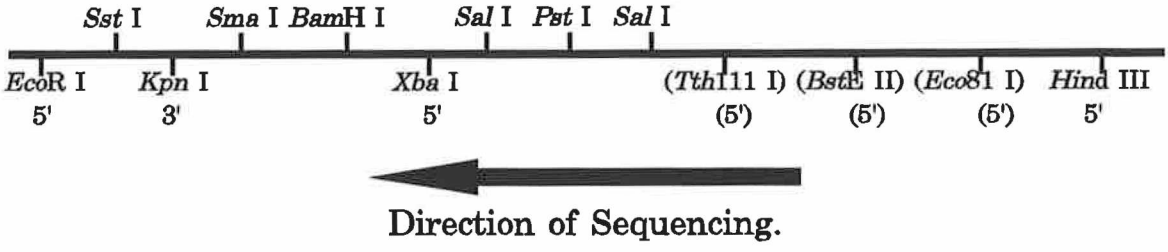

Figure 9.

Multiple Cloning Site of pSVB30.

There was no obstacle to using the two remaining sites, *Xba* I and *Hin*d III. Figure 10 shows the plasmid pSVB30-*rpsL*::IS*Rl1* cut, separately, with both enzymes to show that the sites are unique, and double digested with each enzyme and *Eco*R I to show that these unique sites are in the multiple cloning site.

Figure 10. Agarose Gel Showing pSVB30-*rpsL*::IS*Rl1* Cut with *Xba* I and *Hin*d III.

Lane 1. 1kb DNA Ladder.
Lane 2. pSVB30-*rpsL*::IS*Rl1* cut with *Hin*d III.
Lane 3. pSVB30-*rpsL*::IS*Rl1* cut with *Xba* I.
Lane 4. pSVB30-*rpsL*::IS*Rl1* cut with *Hin*d III/*Eco*R I.
Lane 5. pSVB30-*rpsL*::IS*Rl1* cut with *Xba* I/*Eco*R I.

The nested deletions were generated as described above (page 85). Figure 11 shows all the chosen deletants on an agarose gel between pSVB30-*rpsL*::IS*Rl1* and pSVB30, all cut with *Eco*R I, in decreasing order of size. The deletants were named pα1, pα2, pα3 and so on to pα14 (pα1 being the biggest plasmid and pα14 the smallest) and in many of the figures and tables presented in this chapter pSVB30-*rpsL*::-IS*Rl1* is designated as pα0.

I, *Sst* I and *Sma* I were shown by restriction digest to be present in the fragment *rpsL*::IS*Rl1* and are drawn above the line in figure 9, while the sites *Hind* III, *Xba* I, *Kpn* I and *Eco*R I were shown to be absent (these sites are drawn below the line in figure 9. The enzymes *Eco*81 I, *Bst*E II and *Tth*111 I were not available so their presence/absence remained undetermined (shown in brackets in figure 9).

Of those sites which were absent, *Eco*R I was clearly unsuitable as *rpsL*::IS*Rl1* was cloned into this site, and *Kpn* I was unsuitable as this site cuts to give a 3' overhang which is resistant to the action of Exo III, the enzyme used to generate nested deletions. Ordinarily this property is highly desirable in an experiment such as this (see page 85 for nested deletion strategy) but in this case there is no suitable restriction site between the *Kpn* I site and the *rpsL*::IS*Rl1* fragment which yields susceptible ends (5' overhang or blunt) when cut (see figure 9).

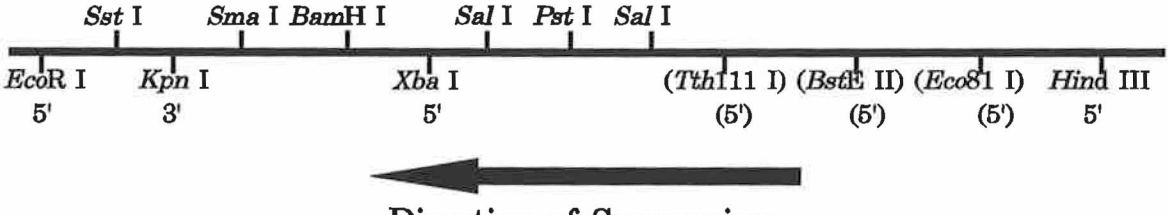

Figure 9.

Multiple Cloning Site of pSVB30.

There was no obstacle to using the two remaining sites, *Xba* I and *Hind* III. Figure 10 shows the plasmid pSVB30-*rpsL*::IS*Rl1* cut, separately, with both enzymes to show that the sites are unique, and double digested with each enzyme and *Eco*R I to show that these unique sites are in the multiple cloning site.

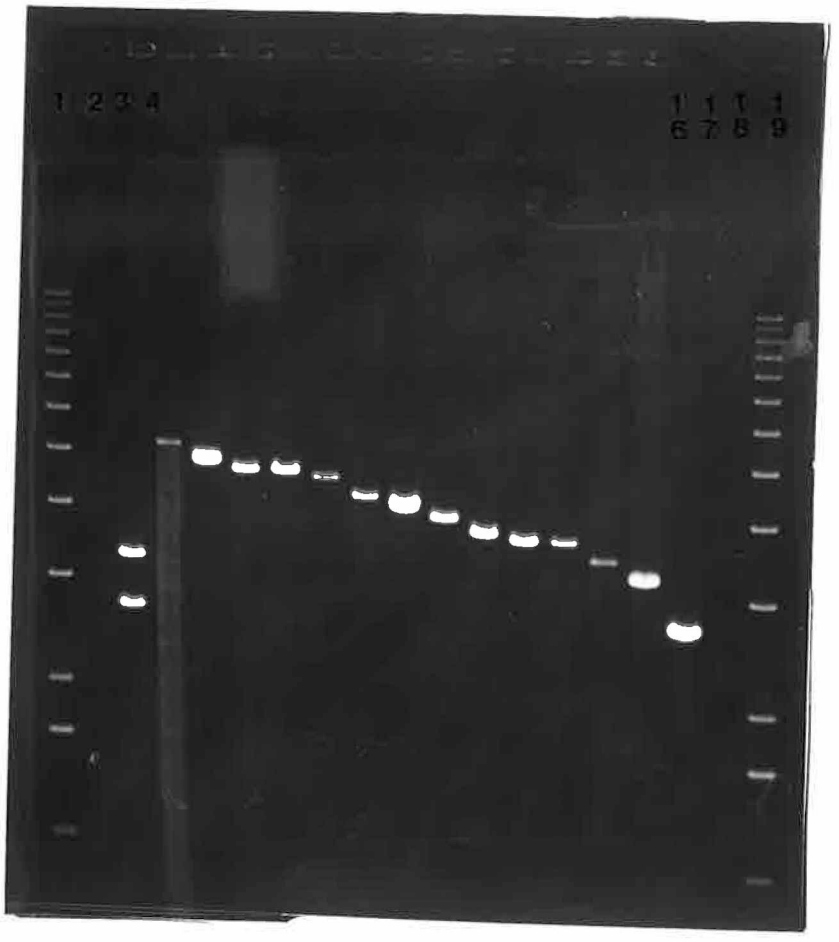

Figure 11. Agarose Gel Showing *Eco*R I-cut Nested Deletions.

Lane 1. 1kb DNA Ladder.
Lane 2. Empty.
Lane 3. pSVB30-*rpsL*::IS*R11* cut with *Eco*R I.
Lanes 4 - 16. Deletants pα1 to pα14 cut with *Eco*R I.
Lane 17. pSVB30 cut with *Eco*R I.
Lane 18. Empty.
Lane 19. 1kb DNA Ladder.

pSVB30-*rpsL*::IS*R11* is cut twice by *Eco*R I while all the deletants are cut only once. This is because the *Eco*R I site is lost almost immediately after the deletion procedure begins, so in figure 11 there are two bands in lane 3, (the biggest being the *rpsL*::IS*R11* fragment and the smaller is the pSVB30 fragment - matching the band in lane 17) and one in lanes 4 to 17.

Two deletants have been omitted from figure 11. On an

agarose gel similar to that shown in figure 11, these deletants gave bands smaller than that from pSVB30 implying that all the *rpsL*::IS*R11* fragment had been deleted and that these deletants were therefore of no interest.

Sequencing of *rpsL*::IS*R11* through Nested Deletions.

Figure 12 shows a map of the deletants relative to IS*R11*. Each box represents a deletant and is left open to indicate that only a part of the deletant is shown. The solid part of each box represents the sequence data obtained from that deletant and is positioned such that it is under the corresponding stretch of the bar representing IS*R11*.

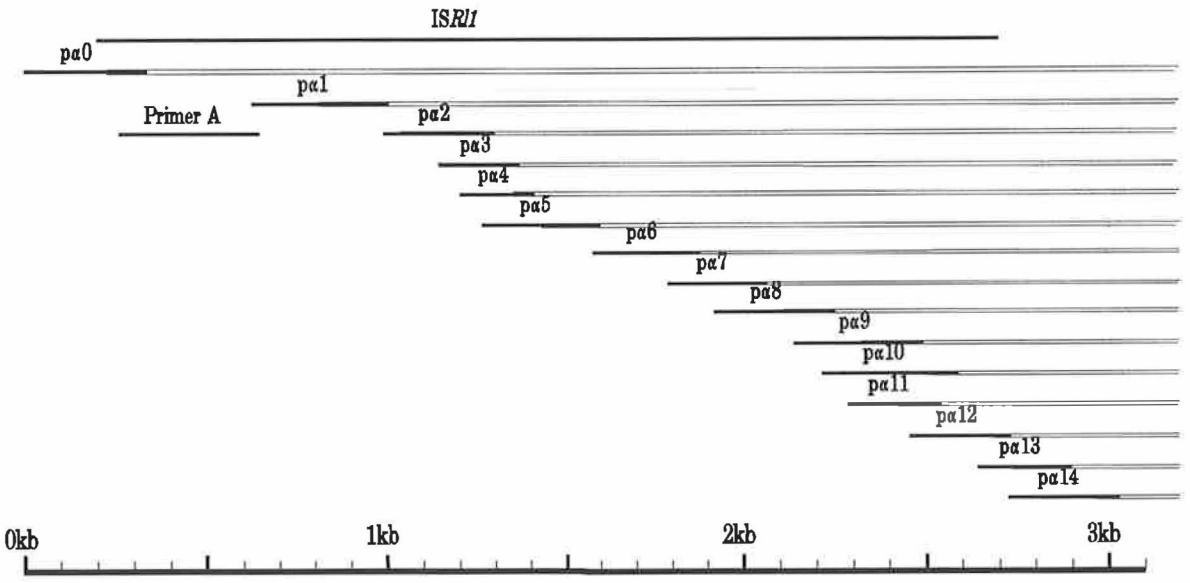

Figure 12. Map of Regions Sequenced with Nested Deletions.

Table 3 shows the beginning and end of the sequence data obtained from each deletant and the amount of overlap with the previous deletant (a minimum of twenty basepairs was required to be certain the overlap was genuine). The positions referred to are based on IS*R11*, where basepair (bp) number one is the first base of the (left hand) inverted repeat.

When all the sequence data was collated, it was found

that there was a gap in the data, approximately 300bp long. This gap was filled by using the data obtained from pα0 (i.e. pSVB30-*rpsL*::IS*R11*) to design a 20mer oligonucleotide. This oligonucleotide was used in sequencing reactions in the place of the universal primer with pSVB30-*rpsL*::IS*R11* as the template and is referred to as primer A. The data obtained by this strategy completed the nucleotide sequence of IS*R11* on one strand and is represented by a solid bar in figure 12 and is included in table 3. The DNA sequence of primer A is given at the bottom of table 3.

Table 3. Regions Sequenced with Nested Deletions.

Deletant.	From (bp No.)	To (bp No.)	Overlap with Previous Deletant (bp).
pα0	-200	135	-
Primer A	58	447	77
pα1	423	802	24
pα2	788	1101	24
pα3	1148	1365	47
pα4	1201	1266	164
pα5	1066	1397	200
pα6	1372	1674	25
pα7	1590	1862	84
pα8	1721	2052	141
pα9	1936	2288	116
pα10	2012	2389	276
pα11	2084	2340	305
pα12	2252	2531	88
pα13	2439	2696	92
pα14	2527	2825	169

Primer A = TTCACGGCGTAGTTCCATGG

The sequence data obtained from these experiments is presented below. The data consists of all of the nucleotide sequence of IS*R11* and most of the sequence of *rpsL*.

It is possible to compile the whole of the sequence of *rpsL* from the literature using information from various published sequences[289,290] and cloning experiments.[86,357] By comparing these two sequences the precise position and

boundaries of *ISR11* were located. In the sequence below and throughout this thesis, position one refers to the first nucleotide of the first inverted repeat (i.e. the left hand repeat which is the first to be encountered in the data below), regardless of which strand is under discussion.

```
-249                    GAATTC CCCGGATCCC CGGTTTGACT GGTCAAATTT
-199  CGAACGACAT CATTCTCATT GACTATACTG ATTTCGTCAG ACTTACGGTT
-149  AAGCACCCCA GCCAGATGGC CTGGTGATGG CGGGATCGTT GTATATTTCT
 -99  TGACACCTTT TCGGCATCGC CTAAAATTCG GCGTCCTCAT ATTGTGTGAG
 -49  GACGTTTTAT TACGTGTTTA CGAAGCAAAA GCTAAAACCA GGAGCTATTT
   1  GTAAGCGACA AGCCGGGGCC GTTCAGGCGG CGTAGTTCCA TGGCATGAGG
  51  GCGTCGATTT CACTGTTGGG CCATCCATTC GCAAGGGCCT CAAGCGTCTG
 101  GGTCAACCAG GCTTGCGGAT CGACGGCATT CATCTTTGCC GTTTGCAGGA
 151  GCGTCGCGAT GGTCGCCCAA GTCCGGCCGC CACCGTCGCT TCCAGCGAAT
 201  AGACTATTTT TTCTCGTGAT CGCTTGGGGT CTGATTGCTC GCTCGACTAT
 251  GTTGGAGTCG AGTTCGATGC GGCCGTCGGT CAGGAAGCGC TCGAAGATGT
 301  CTCGACGCGA GATGGCATAA CGCAGAGCCT CGGCGAGGTT TGGATTTTCC
 351  AGAGACCCGC GGCAGGGTCG CCTGCCAGAG AGTGAAGAGG TCACGGACGA
 401  TCGCTGCGGA GATCTCCTGG CGTGCAGCGA CACGAGCTTC AGGGCTTTGG
 451  GCGCGCACAC GCTCCTCGAT CTCCCATAAC CTCGCCATAC GCTCGACGGT
 501  GTCGGTAGCG ACTTTCGAGC TCTTTGCGAC ATGCAACTCG TAGAACTTTC
 551  TCCGGCTGTG CGACCAACAG CCGGCCAGGA TGGCGCTGTT ATTGCCTCCA
 601  TCTTTCCGGA CGAGCTTGTT ATAAGCGGCA TATCCATCGA CCTGCAGGAT
 651  TCCACGATAG CCGTTCAGGT GTCGTGCGAC ACACTCGGTG GCTCGACTGT
 701  CCTCGAAGCG ATAGGCTACC ATCGGTGGAC CGCTGCCCCC AAAAGTCGGT
 751  CATCCCTGGC ATACGCCCAT AGCCATGCCG TCTTCGCTGA TCCGGAACCA
 801  GGAGCAAGTG TGGGCAAGGT CGTTTCGTCG GCAAAGATCC GTTCGGCCTT
 851  CTTGATCTCA TCGAGGATGT AGTCGGCGAG GATATTGAGC TCGAACCCCA
 901  GCTTGCCCAT CCATTGAGCC ATTAGCTTGC GTCGAGTTCG ACCTTGTCGC
 951  GTGCGTAGAT GGCTTCCTGG CGATAGAGTG GCAGCCCGTC GGCATATTTG
1001  GAAACGGCGA TCTGGGCCAG AAGTGCTTCC GTGGGAATGC CACCTTCGAT
1051  GATGTGTGCC GGAGCCGCCG CCTGGACGAC ACCATCCTCG TTCCTAAAGG
1101  CATACTTCGG CCGGCGGGTG ACGATGACAC GGAACTTCGC TGCCACGACA
1151  TCCAGCCGCT CGGAGACGTC TTCGCCAATC AGGATCTTGG TCTTACCTGC
1201  ATGTTCTGGA AGCTCTTCCG GTTCGATCAC GACCTCGACG CGCTCAAGAT
1251  GAGGCGCAAA GCCCTTGCGT GGCCGCGGAG CGCGTTTGCT GTCCGCGCTA
1301  CCGCGGCCTT TCGTGACCTG AGCCTTGATC GTCGCAATGC CGGTCTCGAT
1351  CTCTTCGAAG ACAAAGGCAT GCTGCTCGTC ATCGACCGGT GGATGATGCA
1401  AGCCTTTCCG AGCGTCGGCC AAATCGGGCG CGATCGAAAG CCTTCAGGAT
1451  CTTGCGTCAG CCGCTCGATG CGCTCATCGG CGTCCGCGTT GCGGCCTCTA
```

1501 GATCGTCAAC CTGTTTCTCC AAAGCCTCGA CGCGGGCTGC TTTTGCACAG
1551 TGGCAAGAAC CATGGCTTTG AGAGCCTCTA CATCATCCGG GAGCTCAATC
1601 TCGGGCCGCG TCATGGGTAT GATCAGAGCA TATTTTGCCG CGCTCCGCCC
1651 ATGCTTTCAG GGACCTGATT CACTTCGCCG CAGCGGTTTT ACCCAACAAT
1701 CTCGGGAGGC TTCACCGGCG CGGATCGAAC TCGCTTCCAG TCCATTCCGT
1751 CGACCAAAGC CAGAAGCTGG GCATGATTGA GCTGGACGCG ATGATGGCCA
1801 TCCGGGGCCA GCAGAACTGG GCTTTTTCCA GCCGCTTCGC ATAGAGGCAA
1851 ACGCCGAGCC ATCCCACCAC ACGATCTTGA CTCGGTCTGC TCTCTTGGCC
1901 CGAAAGACAT AAAGTGCACC GTTGAACGAT CACTGCCAGC ATCCCGTACC
1951 AGCGAAAGCA AGCTGTCGGC CCCTTTCGGA AGTCGATGGG ATGGCTCGCA
2001 AGAAAGACCT TCACGCCGGA CGGGATCATG CCGACCGCAC CGCTCGGATC
2051 ACCCGCTGCA GCTGTGCTTC GCCGATATCT GCGTCGGCAC GGATAATGAC
2101 GTCGCCGATC ACAAGCTCGA TCATCGCACG TGTGCCAATC GTCCTGGTCC
2151 CTGTCTGACC AGCCCGATCT TGCGAGAAGG ATCGTTGTCC GTCACGAGCA
2201 TCACGACGCC AGCCAAATAG CTGCGAGGGA TGTATGCCGA TGCGATGCGC
2251 AATGGCCGAG ACGCTTGCGC CGGGCTCCAT TGCCTCAGCC ACAGCCTGCG
2301 CTTTAAAATC ATCGGACCAG CGGCGCCGAA ATTGCCGTGG GGCGCCCTCA
2351 AGCCGTTCCG GAACAGCCTC AATCATATGG AAGTTTCTAG TTCCAGAGCT
2401 AGGCGCAGAC ATAGAAGCTC ACAGTTTGTG AATCGTCTGT CCGCAATACA
2451 GCGGCCAACG CTACCGACGC CAGATGGGGT CAGCTTGTCG CTTACAGCTA
2501 TTTAATGGCA ACAGTTAACC AGCTGGTACG CAAACCACGT GCTCGCAAAG
2551 TTGCGAAAAG CAACGTGCCT GCGCTGGAAG CATGCCCGCA AAAACGTGGC
2601 GTATGTACTC GTGTATATAC TACCACTCCT AAAAAACCGA ACTCCGCGCT
2651 GCGTAAAGTA TGCCGTGTTC GTCTGACTAA CGGTTTCGAA GTGACTTCCT
2701 ACATCGGTGG TGAAGGTCAC AACCTGCAGG AGCACTCCGT GATCCTGATC
2751 CGTGGCGGTC GTGTTAAAGA CCTCCCGGGT GTTCGTTACC ACACCGTACG
2801 TGGTGCGCTT GACTGCTCCG GCGTTAAAGA CCGTAAGCAG GCTCGTTCCA
2851 AGTATGGCGT GAAGCGTCCT AAGGCTTAAT GGTTCTCCGT TAAGTAAGGC
2901 CAAACGTTTT AACTTAAATG TCAAACTAAA CTCGTAGAGT TTTGGACAAT
2951 CCTGAATTAA CAACGGAGTA TTTCCATGCC ACGTCGTCGC GTCATTGGTC
3001 AGCGTAAAAT TCTGCCGGAT CGGGAATTC 3029

Chapter four (page 114) deals extensively with the analysis of this sequence data but several features are mentioned here first, each of which is underlined in the sequence above.

Insertion sequences are characteristically terminated by short inverted repeats. In the case of IS*R11* these repeats (1 - 13 and 2483 - 2495) are 13bp perfect repeats. The rest of the IS*R11* sequence was searched for the presence of these repeats to check the possibility of IS*R11* being

a compound transposon. No matching sequences were found even when the search was broadened to find matches of down to 75% (this was because it is not uncommon for the terminal repeats of IS elements to contain several mismatches).

The second characteristic of IS elements is the duplication of target sequences. In this case these are 8bp in length (AGCTATTT) and are located at positions -7 - 0 and 2496 - 2593 (double underline in sequence listing above). IS*R11* was searched for the presence of this sequence and its complement (AAATAGCT). The latter was found at position 2215 but, as discussed below (page 109), this is unlikely to be of any significance.

A look at the data from *rpsL* reveals two differences from the published sequence.[289,290] These differences are illustrated in figure 13 (page 105), which shows part of an autoradiogram from the sequencing of pα0. Some of the sequence read from this autorad is presented above the corresponding sequence from the literature. The lanes in this autorad are in the order G,A,T,C, the sequence is read from the bottom up and the positions given refer to the sequence data above (page 102). It can be seen that three bases present at position -122 and one base present at position -84, are clearly not present in the sequence determined in this work.

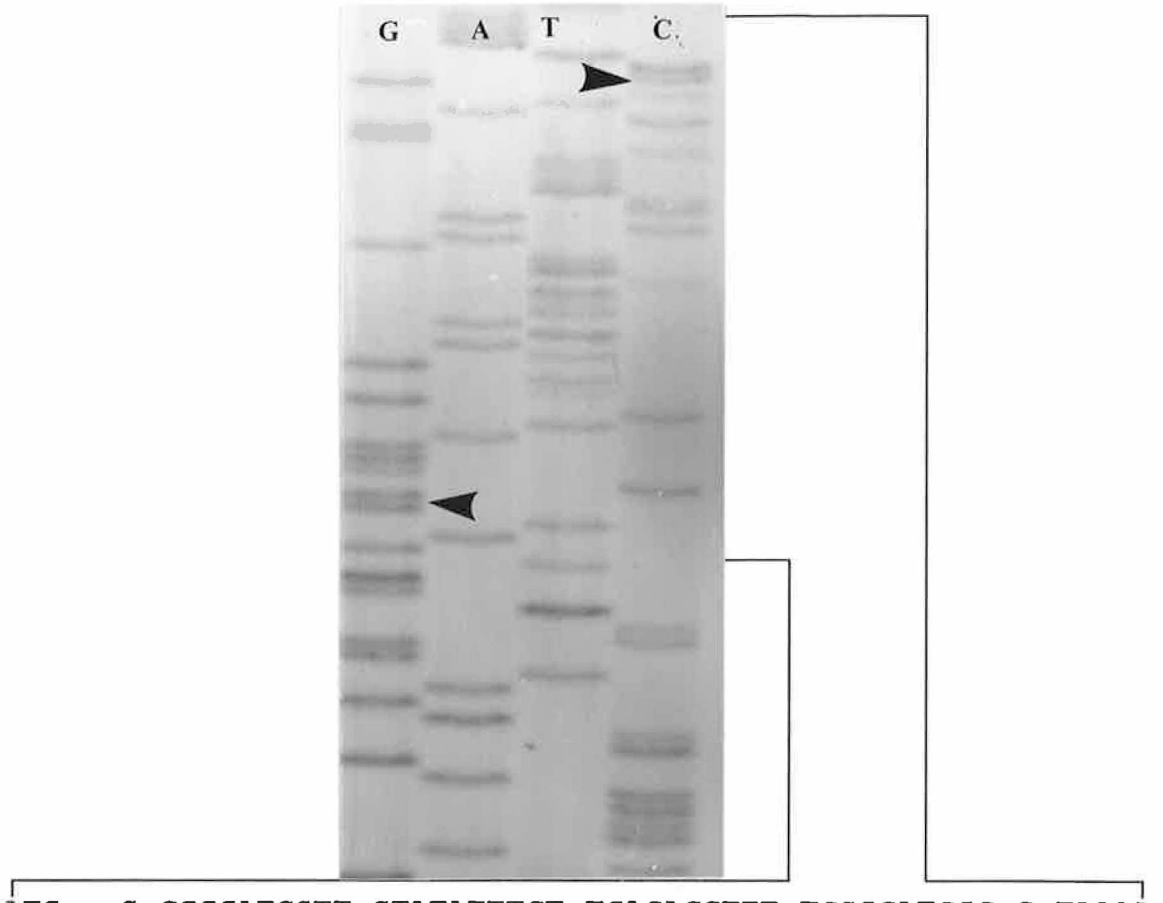

Figure 13. ^{35}S Autoradiogram Showing *rpsL* Sequences.

For complete confidence in the sequence of *ISR11* both strands must be sequenced. For this purpose, another set of nested deletions (this time on *rpsL*::*ISR11* cloned into pSVB30 in the opposite orientation) was required. However, despite repeated attempts, the appropriate deletions were not generated.

One possible reason for this is the presence of an *Xba* I site in the sequence (bases 1497 - 1502). Figure 14 (page 106) shows part of an autoradiogram from the sequencing of pα6, which clearly shows this *Xba* I site. The lanes in the autorad are in the order A,C,G,T and the sequence is read from the bottom up. The surrounding sequence data is written below the figure with the *Xba* I site (TCTAGA) underlined.

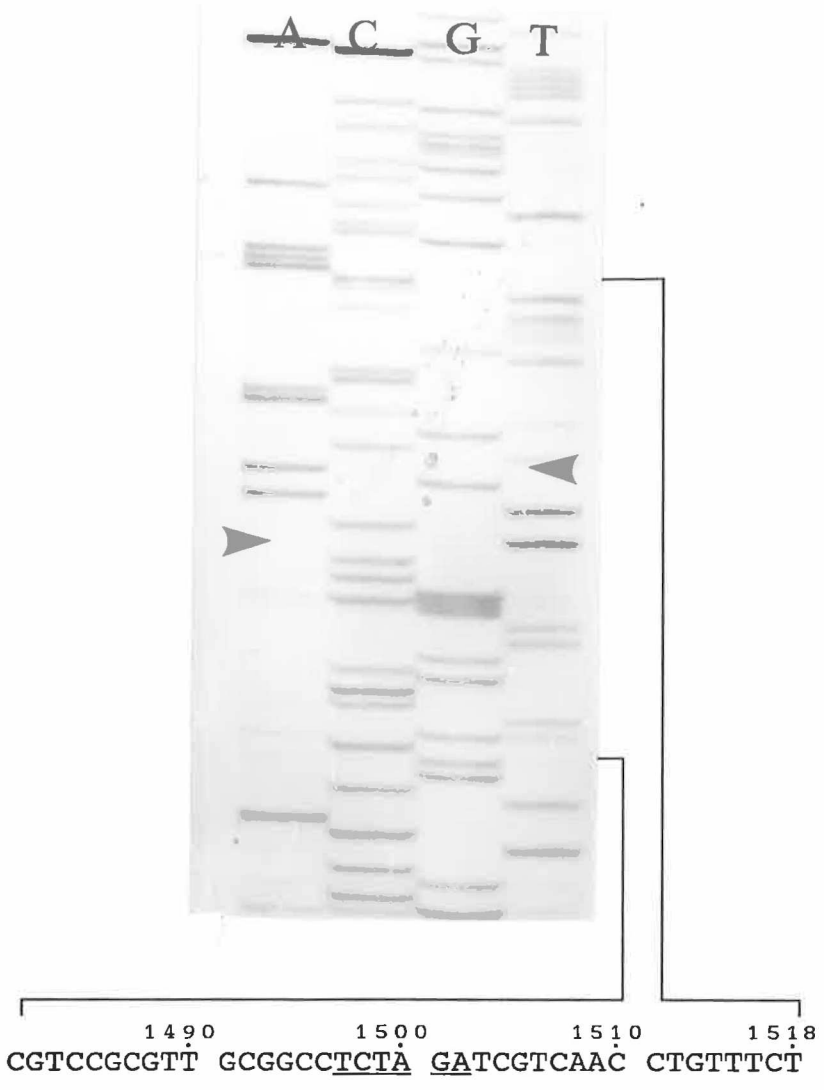

Figure 14. ^{35}S Autoradiogram Showing *Xba* I Restriction Site.

The significance of this site is discussed below (page 111) but as its presence seemed to indicate that the attempt to make more nested deletions was unlikely to ever succeed, it was decided to sequence the other strand by primer walking

Sequencing of IS*R11* by Primer Walking.

Sequencing by this method involves the use of known sequence data to choose a primer for further sequencing experiments from which the next primer is chosen and so on until all the fragment is sequenced.

Table 4 gives the nucleotide sequence of all of the

primers used to sequence the second strand, the position of the first base of the primer and the positions of the first and last bases of the sequence data obtained in each case. To facilitate comparisons with the first strand the same reference system is used for both strands, making it appear that the second strand was sequenced backwards.

Table 4. Primers and Regions Sequenced.

No.	Sequence	First Base	From/To
1	TTGCGAGCACGTGGTTTGCG	2548	2500 - 2272
2	ATGATTTTAAAGCGCAGGCT	2311	2275 - 1895
3	GCAGTGATCGTTCAACGGTG	1936	1907 - 1430
4	ACAGGTTGACGATCTAGAGG	1514	1474 - 1119
5	AAGAGCTTCCAGAACATGCA	1217	1163 - 768
6	TCGCCGACTACATCCTCGAT	879	847 - 496
7	AGTTGCATGTCGCAAAGAGC	538	498 - 88
8	CGAACTCGACTCCAACATAG	266	206 - -93

Figure 15 shows a map of the regions sequenced from each primer, positioned relative to IS*RI1*. The primers are named primer 1 to primer 8 in the order in which they were used and the direction of sequencing is indicated by the arrowhead at the end of each bar.

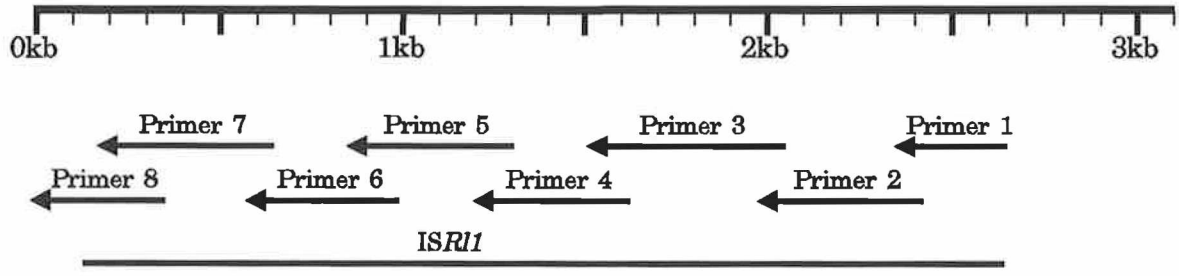

Figure 15. Map of Regions Sequenced by Primer Walking.

When all the sequencing experiments were completed and all inconsistencies resolved, only one discrepancy was discovered between the two strands. At position 2268, one strand reads GC, while the other strand reads CG.

Figure 16 shows the relevant sequences from each strand (pα10 in the case of the nested deletion strand and primer 2 in the case of the primer walking strand). The lanes in both of these autorads are in the order A,C,G,T and both sequences read from the bottom of the autorad upwards. A sequence complementary to the sequence from primer 2 is written under the sequence from pα10 for complete clarity and the anomaly is underlined in all cases.

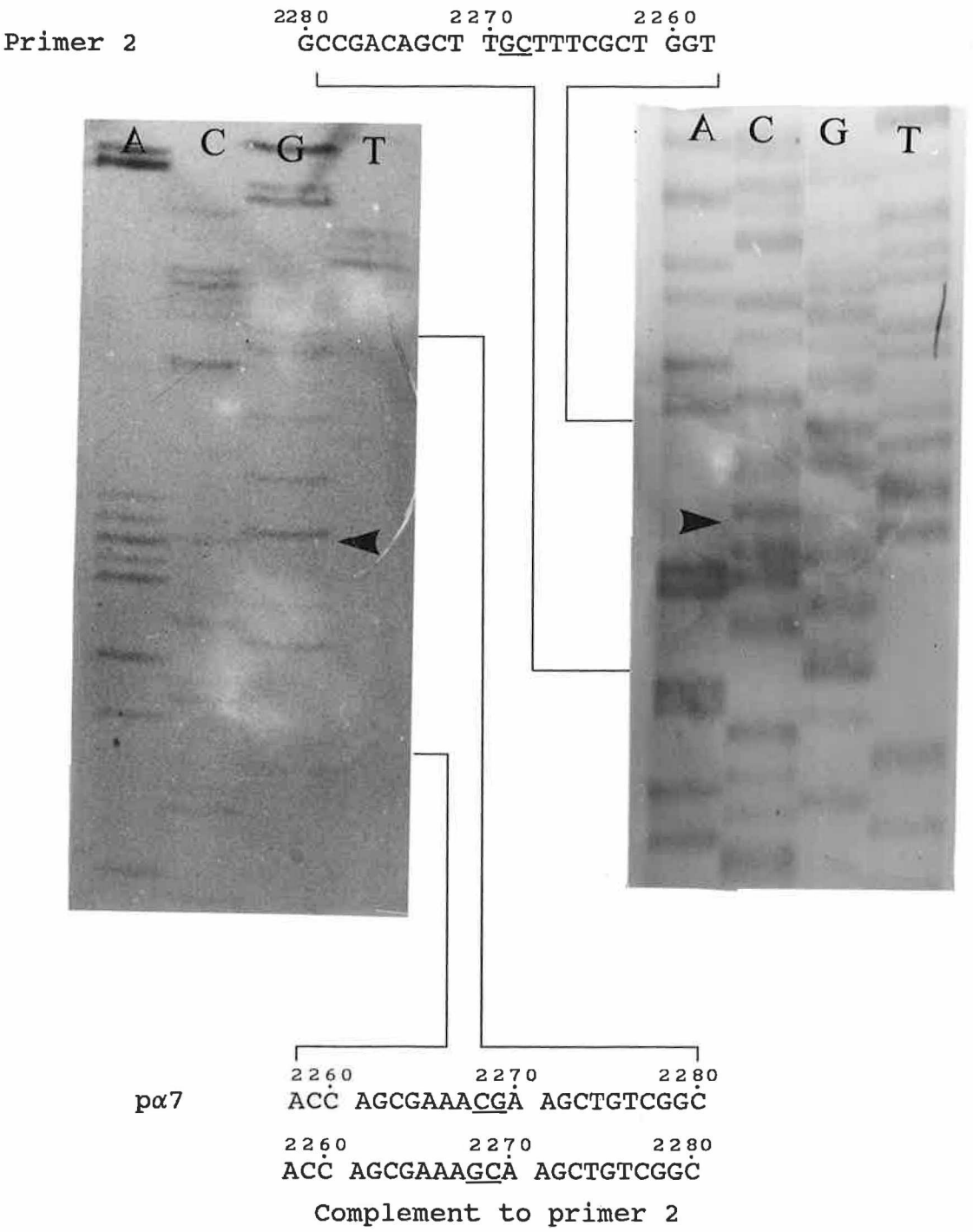

Figure 16. GC/CG Anomaly in Sequenced Strands.

Discussion.

One mismatch was found between the sequences of the two strands of IS*Rl1*. As all the primer walking was done from the same template, intact pSVB30-*rpsL*::IS*Rl1*, which was also the starting material for the nested deletions from which the first strand was sequenced, it is likely that a mutation has been picked up in deletion pα7 and that the sequence actually reads AAAGCAA, as per the sequence read from primer 2 (page 108), this is the version that was submitted to the GENBANK database.

Two differences from the published sequences[289,290] were found in the regions of *rpsL* that were sequenced in this study (figure 13). The *rpsL* fragment was obtained from pSUP104-*rpsL*[357] and was cloned into this plasmid from pNO1523.[86] If either of these mutations removed the function of *rpsL*, neither pNO1523 or pSUP104-*rpsL* would have served the purposes for which they were constructed.[86,357] It seems that the loss of the base C from position -84 would have had just such an affect as it is an important part of the promoter sequence for *rpsL*, but no significance can be attached to the ATG triplet at position -127.[289] So it seems likely that the cytosine was lost from position -84 since the transposition of IS*Rl1* into the gene, but it is impossible to say when the other mutation occurred.

IS*Rl1* is 2496bp long, terminates in 13bp perfect inverted repeats and (at least in this instance) produces a target duplication of eight base pairs.

The sequence of IS*Rl1* was searched for matches to the inverted repeat, and none were found. This means that IS*Rl1* is not a compound transposon but the possibility that it encodes a function unrelated to transposition (or, in other words, that it is a transposon) is not ruled out. Though it is true that IS*Rl1* makes a large IS element, it is equally true that it would make a small transposon as these are usually larger than 5kb.

The sequence duplicated at the target site was also found inside the element in inverse orientation, at position no. 2215. This is of only marginal interest as in no case has homology between an IS element and its target site

been shown to be responsible for transposition.[124]

ISR1 (1260bp), which was isolated from the old species *R. lupini*,[47,294] now designated as *Bradyrhizobium* sp. (*Lupinus*),[197] also has 13bp inverted repeats, but these are part of a longer, imperfect repeat sequence of 41/42 nucleotides.[295] There is no similarity between the inverted repeat of ISRl1 and the 13bp perfect repeat of ISR1 and ISR1 generated target duplications of only 4bp.

Sequences in *B. japonicum* with similarities to IS elements have been reported,[202] but the termini of these putative insertion sequences are poorly defined.

ISRm2, from *R. meliloti*, generates a target duplication of 8bp and terminates in 22bp perfect inverted repeats (although one isolate is known with two single base inhomologies).[101] The first 6bp of the inverted repeats of ISRl1 and ISRm2 are identical. An 8bp homology between the inverted repeats of ISRm2 and IS1 was reported[101] but this sequence does not overlap with the 6bp match to the inverted repeats of ISRl1. As ISRm2 (2·7kb) is a similar size to ISRl1 it would be interesting to compare their nucleotide sequences more fully but ISRm2 remains unsequenced, so comparisons must be limited to restriction digest patterns. Since ISRm2 carries two *Hin*d III sites (which are absent from ISRl1), three *Xho* I sites (ISRl1 has four) and only one *Pst* I site (ISRl1 has two) in can be concluded that these elements have significant differences.

Two other IS elements from *R. meliloti* have been reported, ISRm1[322] and ISRm3.[395] Neither of these elements show similarities to ISRl1 in the sequence or length of their inverted repeats (31/32bp and 30bp, ISRm1 and ISRm3, respectively), the length of the target repeats (5bp and 8 or 9 bp) or the length of the elements themselves (1319bp and 1298bp).

A number of insertion sequences have been isolated and characterized from *A. tumefaciens*. The first of these to be detected, IS66,[35] shows the same kind of similarities to ISRl1 that ISRm2 showed. The inverted repeats (20bp) begin with the same six nucleotides, the target duplication is 8bp and the element is 2548bp long.[240] It seems likely that these three elements are members of the same family of transposable elements.

IS*866*,[38] which shows 70% homology to IS*66* in an 85bp region, is another good candidate for membership of this family. Like the other elements it produces an 8bp target duplication and is in the same size range (2732bp). The terminal sequences are 27bp imperfect inverted repeats, seven of the first eight bases in the inverted repeat match the same bases in IS*R11*. None of the other sequenced IS elements from *A. tumefaciens* shows significant similarity to IS*R11*.

With the discovery of an *Xba* I site in IS*R11* (at position 1497), it becomes clear why it proved impossible to generate a second set of nested deletions. In fact, the difficulty now lies in explaining how it was possible to make the first set rather than why subsequent attempts failed.

Throughout the course both of this and a previous study (A. Ryan, M.Sc. thesis, DCU), *Xba* I repeatedly cut pSVB30-*rpsL*::IS*R11* only once (see figure 10, page 98). The conditions for the deletion reaction were chosen to maximize the rate of Exo III digestion (300nucleotides/min though the enzyme is capable of deleting 500nucleotides/min). So if the *Xba* I enzyme was cutting out 1·8kb of the insert it is clear that the deletions would quickly extend into pSVB30 sequences. It is the experience of other researchers in this laboratory, that the pUC18 plasmid (pSVB30 differs from pUC18 only in its multiple cloning site) will only sustain deletions of less than 500bp without being completely lost (M. Duffy, personal communication).

The reason that the *Xba* I site remained 'hidden' for so long is likely to be due to the fact that *Xba* I is sensitive to methylation in its target site and pSVB30-*rpsL*::IS*R11* was isolated from JM83, a *hsd*$^+$ strain of *E. coli*. The recognition site of the methylation enzyme(s) from this strain is not known, but if it includes or overlaps the *Xba* I site it would be possible for one of the *Xba* I site in pSVB30-*rpsL*::IS*R11* to be methylated or hemimethylated and the other unmethylated.

So why should *Xba* I suddenly cut pSVB30-*rpsL*::IS*R11* at both sites, after consistently cutting at only one? The explanation may lie in the deletion protocol. Prior to the

Xba I restriction step, the plasmid is treated with Klenow fragment, to fill in the sticky ends produced by *Hind* III restriction with thionucleotides. As well as filling in these ends, the Klenow fragment will also proceed rapidly along the plasmid from any nicks in either strand, replacing all the nucleotides it encounters. If the fragment passes over the methylated *Xba* I site it will leave it at least hemimethylated and perhaps unmethylated.

So if this is the case, a successful deletion experiment would depend on a sufficient population of plasmid molecules being unnicked upstream of the second *Xba* I site (on both strands) and/or conditions being right for *Xba* I to recognize a hemimethylated site.

For one reason or another, none of the pUC vectors or the other pSVB vectors could have been used for the deletion experiment except for pSVB26 and pSVB28[9], but the use of either of these vectors would also have entailed using *Xba* I. So prior knowledge of the *Xba* I site would have meant looking further for a suitable vector.

Even without hidden restriction sites to contend with, the generation of nested deletions seems to be a problematic procedure. It was the experience of other researchers from this laboratory that the procedure would only work with the most concentrated and pure plasmid preparations (R. Mac Síomóin, personal communication). Also, once the deletions have been made a separate plasmid prep. is necessary for each 200-300bp to be sequenced and, if the deletion termination points are unevenly spaced, it is possible that the sequence of certain regions may not be accessible (see for example the necessity of using primer A between pα0 and pα1, page 101).

Primer walking seems to be a simpler means of sequencing long regions of DNA. No manipulations of the fragment to be sequenced are necessary, one good plasmid prep will suffice to sequence the entire fragment and sequencing of the other strand can be achieved without having to reorient the fragment in the vector. The drawback to this method is that it is more expensive and, unless the laboratory is equipped with an oligonucleotide synthesizer, slower than other approaches.

The sequence of IS*R11* has been submitted to the GEN-

BANK database and has been given the accession number L19650.

Chapter 4. Sequence Analysis.

Introduction.

There were two main objectives to the analysis of the nucleotide sequence of IS*Rl1*. Firstly, to detect any similarities between IS*Rl1* and any other DNA sequence, particularly transposable elements, and secondly, to extract any information on the function and properties of the element. In light of the similarities between their inverted repeats, special consideration is given to comparisons with IS*Rm2*, IS*66*, and IS*866*.

As described above the data from the two strands sequenced differs in two basepairs. There is a chance that these bases are part of an important site so, to protect against this possibility, all the analyses described in this chapter were performed on both 'versions' of the sequence, but no difference was found in any case.

Sequences Homologous to IS*Rl1*.

Two sources were used to search for sequences with similarities to IS*Rl1*. The BLAST network server permits the rapid identification of such sequences from the most up-to-date database in the world. To use this service the sequence of one strand (arbitrarily called the plus strand) is sent to a computer at the internet site blast@ncbi.nlm.nih-.gov in a particular format and with certain instructions (see above, page 89). At this site the computer transcribes the minus strand and uses both to search for similar sequences in the database it maintains. The computer uses a scoring system to assess which target sequences give the best matches to the query system.[205] This score is partly based on the statistical significance of the matching segments, which is measured by the Poisson probability, P, the lower the value of P, the more significant the match. Information on any sequence with a score equal to or above a cutoff score determined by the user is returned by e-mail.

The results of this search are presented below. Table 5 lists the locus entry used by GENBANK (a notation based on the source of the sequence and its function) to describe

the sequence, of the matching (or target) sequences and the score and lowest Poisson probability for each.

Table 5. Sequences with Close Similarities to IS*R11*.

Target Sequence.	High Score.	Poisson Probability, P.
TIPIS866A	452	2·7e-27
ATTAIAAH	452	2·8e-27
RHMLPSZ	225	2·1e-08
MPSAFTERQ	146	0·078

To find weaker matches a program called BSEARCH (available at vax1.tcd.ie) was used to search the GENBANK database, release 76. This program only considers one strand of the sequence so it is necessary to run the program a second time, with the other strand, to be thorough. As before, the best matches to the query sequence are returned to the user, but in this case the best thirty matches are reported regardless of their score (a different scoring system from that used by the BLAST program is used) and it is left to the user to determine their significance. Another difference is that the database available to this program is less up-to-date than that used by the BLAST program.

Significant matches turned up by this method which went unnoticed by the BLAST program are presented in table 6. The target sequences are listed by their locus name, score and their length.

Table 6. Sequences with Weak Similarities to IS*R11*.

Target Sequence	Score	Length (bp)
ATUIS22A	74	2888
ATACH5	48	24595
TIPISA66	42	2608
RHBGLNA	40	554
RHBGLNB	40	900

All the target sequences reported in tables 5 and 6 are presented below alongside the matching stretch of IS*R11*

sequence. Only the high scoring segments are shown (i.e. those segments that caused the program to report the match) and in the case of matches to the minus strand (the strand read by primer walking) the positions given refer to the plus strand (i.e. the system used on page 102). Poisson probabilities were not available for the weak matches.

Close Similarities.

TIPIS866A
Score = 452, P = 2·7e-27.
Identities = 188/310 (60%), Strand = Minus.

```
Query:  339 ACCTCGCCGAGGCTCTGCGTTATGCCATCTCGCGTCGAGACATCTTC
            |||||| | ||||  ||  ||||    ||   ||||  |||  ||||
Sbjct: 2284 ACCTCGGCAAGGCCATGGCTTATATGCTCAAACGTCAGGACGGCTTC

Query:  292 GAGCGCTTCCTGACCGACGGCCGCATCGAACTCGACTCCAACATAGT
              ||  ||||||  |||||||||||  ||||  ||||||||||| | ||
Sbjct: 2331 CGGCTGTTCCTGGACGACGGCCGCGTCGACATCGACTCCAACCTCGT

Query:  245 CGAGCGAGCAATCAGACCCCAAGCGATCACGAGAAAAAATAGTCTAT
            |||    || ||| |   ||  || || |   |    |||  ||| |
Sbjct: 2378 CGAAAACGCGATCCGTAGCCCGGCCATGAACCGCCGCAATGCTCTCT

Query:  198 TCGCTGGAAGCGACGGTGGCGGCCGGACTTGGGCGACCATCGCGACG
            | || ||    || |  |||||||| | ||||||     |||| |
Sbjct: 2425 TCGCCGGCCATGATGAAGGCGGCCGTAATTGGGCCCGTTTCGCCAGC

Query:  151 CTCCTGCAAACGGCAAAGATGAATGCCGTCGATCCGCAAGCCTGGTT
            ||  |    ||    |||||||| |  || || ||  | ||||   |
Sbjct: 2472 CTGATTGGCACATGCAAGATGAACGGTGTTGAACCATACGCCTACCT

Query:  104 GACCCAGACGCTTGAGGCCCTTGCGAATGGATGGCCCAACAGTGAAA
            |  | |     |       || || |||||    | |  ||  || |
Sbjct: 2519 GCGCGATCTCTTCATCAGTCTCGCAAATGGCCACCTCGCCAAAGACA

Query:   57 TCGACGCCCTCATGCCATGGAACTACGC   30
            ||||||||||||||| |||  ||| ||
Sbjct: 2566 TCGACGCCCTCATGCCCTGGGCCTATGC 2593
```

Sequence TIPIS866A (accession M25805) is the sequence of the IS element IS*866*, isolated from the Ti plasmid of a strain of *A. tumefaciens*.[38] The sequence from position 2284 to position 2593 contains the end of ORF1 (2284 - 2393) and the first five codons of ORF2 (2299 - 2284), on the oppos-

ite strand.

IS*866* was isolated from the TA *iaaH* gene of pTiTm4. The sequence of this gene was reported, including the sequence of IS*866* and is listed in the GENBANK, as ATTAIAAH (accession X56185). This is the next sequence in table 5. As the matching segment from this sequence is exactly the same as that from TIPIS866A, it would be superfluous to repeat the comparison. It will be noted that these two sequences have the same scores in table 5, but that ATTAIAAH has a slightly higher P value, this is because it is a longer sequence than TIPIS866A and so there is a slightly higher chance of the match being coincidental.

RHMLPSZ
Score = 225, P = 2·1e-08.
Identities = 57/72 (79%), Strand = Minus.

```
Query:  281 GACCGACGGCCGCATCGAACTCGACTCCAACATAGTCGAGCGAGCAA
            ||||  | | ||||||||| ||||||||||||||    | |||||||||
Sbjct: 1742 GACCTGCCGTCGCATCGAGCTCGACTCCAACATCCCTGGGCGAGCAA

Query:  234 TCAGACCCCAAGCGATCACGAGAAA  210
            || | ||||| || |||| ||||||
Sbjct: 1789 TCCGGCCCCAGGCCATCAAGAGAAA 1813
```

The sequence RHMLPSZ (accession M59853) is the gene *lpsZ* from *R. meliloti*, a lipopolysaccharide gene involved in symbiosis,[48] and surrounding sequences. The segment which matches IS*R11* (1742 - 1813) is downstream of the *lpsZ* gene and there are no in-frame start codons after the end of *lpsZ*.

MPSAFTERQ
Score = 146, P = 0·078.
Identities = 70/121 (57%), Strand = Minus.

```
Query: 1180 GATTGGCGAAGACGTCTCCGAGCGGCTGGATGTCGTGGCAGCGAAGT
            | | ||||||||||||| |||||  |||| | |   | |      ||
Sbjct:  769 GCTGGGCGAAGACGTCACCGAGACGCTGAAAGAGATTCCGCGTCGGT
```

```
Query: 1133 TCCGTGTCATCGTCACCCGCCGGCCGAAGTATGCCTTTAGGAACGAG
             ||   || ||||  ||    |||  ||| | | | |   | ||
Sbjct:  816 TCAAAGTGATCGAGACGGTGCGGGAGAAATTTACTTGCCGTGACTGC

Query: 1086 GATGGTGTCGTCCAGGCGGCGGCTCCG 1060
             || |   ||  |||| || |||| |||
Sbjct:  863 GAGGCGATCAGCCAGCCGCCGGCGCCG  889
```

MPSAFTERQ is the sequence of a short open reading frame downstream of *R. meliloti nodQ1* (Schwedock, J. and Long, S.R., unpublished - acc. L08667). This ORF shows nucleotide sequence homology to IS*A66*, an insertion sequence from *A. tumefaciens*.[240] The segment of MPSAFTERQ which matches IS*Rl1* (769 - 889) is in the centre of this ORF.

Weak Similarities.

ATUIS22A

Score = 74, Identities = 63/95 (66%), Strand = Minus.

```
Query:  294 CCTGACCGACGGCCGCATCGAACTCGACTCCAACATAGTCGAGCGAG
            | |   ||| |||||  |||||  ||||   ||||||||  |||||||
Sbjct: 2200 CATCGACGATGGCCGTGTCGAAATCGATAACAACATAGCTGAGCGAG

Query:  247 CAATCAGACCCCAAGCGATCACGAGAAAAAATAGTCTATTCGCTGG
            | ||  | ||| |  |   ||   ||||||||  |  |||||||  ||
Sbjct: 2247 CGATGCGTCCGCTGGGCCTCGGCAGAAAAAACTGGTTATTCGCAGG
```

ATUIS22A (accession M82888) is the sequence of an insertion element, IS*1131*, isolated from the Ti plasmid of *A. tumefaciens*.[388] The sequence between nucleotides 2200-2247 is not part of any of the ORFs found in this element.

ATACH5

Score = 48, Identities = 40/59 (67%), Strand = Plus.

```
Query:  107 CCAGGCTTGCGGATCGACGGCATTCATCTTTGCCGTTTGCAGGAGCG
            ||||||  |||||||||| |  ||| | ||| ||  |    |  || |
Sbjct:14673 CCAGGCCTGCGGATCGATGTTATTGAGCTTGGCGCTCATGATCAGTG

Query:  156 TCGCGATGGTCGCC   167
            ||||  |||  ||||
Sbjct:14720 TCGCCATGAACGCC 14733
```

The sequence ATACH5 (accession X00493, J05108, X00282) is the nucleotide sequence of the T-DNA region from the

Agrobacterium tumefaciens octopine Ti plasmid pTi15955.[12] The segment 14673 - 14733 is part of an open reading frame (ORF14).

```
TIPISA66
 Score = 42, Identities 36/56 (64%), Strand = Minus.

Query: 1178 TTGGCGAAGACGTCTCCGAGCGGCTGGATGTCGTGGCAGCGAAGTTC
            |||| ||||||||  |||||   |||||  |||  |   |     |||
Sbjct: 1352 TTGGTGAAGACGTTACCGAGACCCTGGAGGTCATCCCGCGCCAGTGG

Query: 1131 CGTGTCATC 1123
              ||||||
Sbjct: 1399 AAGGTCATC 1147
```

TIPISA66 (accession M10204) is the nucleotide sequence of IS*66* an insertion sequence from the *A. tumefaciens* plasmid pTiA66.[240] The segment 1352 - 1407 is part of an open reading frame, ORF2. This is the element to which the sequence MPSAFTERQ (see above, page 117) showed homology, and the scoring segment from the MPSAFTERQ-IS*R11* match contains the scoring segment from the TIPISA66-IS*R11* match.

```
RHBGLNA
 Score = 40, Identities = 31/43 (72%), Strand = Minus.
Query: 2143 GACGATTGGCACACGTGCGATGATCGAGCTTGTGATCGGCGACG
            |||     || || | | || ||||||| | ||||||||||||
Sbjct:    4 GACTTCCTGCCCAAG-GTGAAGATCGAGATCGTGATCGGCGACG
```

```
RHBGLNB
 Score = 40, Identities = 24/28 (85%), Strand = Minus.
Query: 2126 GCGATGATCGAGCTTGTGATCGGCGACG
            | || ||||||| | ||||||||||||
Sbjct:  512 GTGAAGATCGAGATCGTGATCGGCGACG
```

RHBGLNA (accession M10926) and RHBGLNB (accession M26753) are overlapping sequences from *B. japonicum* which contain the genes *glnA*[53] and *glnB*[245] respectively. The sequence RHBGLNA contains sequences upstream of *glnA* that contain part of the *glnB* sequence and it is to the same sequence of this gene that both the matches with IS*R11* are made.

Functional Sites within IS*R11*.

Several types of functional sites are of interest. Sites such as recognition sites for Dam methylase, binding sites for DnaA protein and IHF (integration host factor) and Chi sites have been implicated in the transposition of IS elements.[124] The presence of translational and transcriptional control sites can reveal much about the functioning and properties of IS elements. Lastly, the presence of unique restriction endonuclease recognition sites provides a simple means of identifying similar sequences and can help further investigations into the element.

Searches for sites in the first category were performed using the SEQAID II package (see above page 89). The criteria used by Galas and Chandler[124] were used in searching for IHF sites (target = AAnnnnTTGAT, one mismatch tolerated), DnaA sites (target = TTATCCACA, two mismatches tolerated), Dam sites (target = GATC, no mismatches tolerated) and Chi sites (target = GCTGGTGG, no mismatches tolerated).

Sites not considered by Galas and Chandler are the translational frameshift motifs. Any sequence of the form XXX YYY Z is sufficient to cause -1 frameshifting, but the sequence AAAAAAG has been found to be particularly effective.[59] A specific search for this site was conducted using the SEQAID II program and a more general search was conducted with the program FRAME.BAS which was written expressly for that purpose (see page 89 and appendix A).

The results from the search with SEQAID are presented in table 7, the GATC site is symmetrical and so is present on both strands in each case but for the DnaA protein and IHF binding sites this information is indicated by a + (plus strand) or a - symbol (minus strand).

Table 7. Results from SEQAID Search.

Functional Site	Position
Dam Methylase (GATC)	118, 218, 399, 411, 468, 789, 836, 854, 1009, 1183, 1225, 1327, 1348, 1432, 1448, 1501, 1621, 1723, 1873, 1928, 2024, 2047, 2107, 2119, 2166, 2180.
DnaA Protein. (TTATCCACA)	345 (TTATCCACA) +
	817 (TTATCCACA) -
	1515 (TTCTCCAAA) +
	1541 (TTTTGCACA) +
IHF (AAnnnnTTGAT)	1188 (AAGATCCTGAT) -

Table 8 lists the results from the FRAME.BAS search. The nature of the frameshift motif means that if it is present at all, it is present on both strands, but as it is not completely symmetrical the sites from each strand will rarely be identical and so both sites are reported in table 8 and the positions of the first bases of each are given in the format plus strand/minus strand. The sequence of the site is given for each strand with the reading frame of the site in brackets.

Table 8. Results from FRAME.BAS Search.

Position	+ Strand	- Strand
207/212	TTTTTTC (3)	AAAAAAT (1)
446/451	TTTGGGC (2)	CCCAAAG (2)
738/743	CCCAAAA (3)	TTTGGGG (1)
1971/1976	CCCTTTC (3)	AAAGGGG (1)
2302/2307	TTTAAAA (1)	TTTAAAG (3)

The program SIGNAL consists essentially of a list of published consensus sequences for promoters and other regulatory sites. When run, it scans a sequence for the presence of sequences matching any of these consensus sequences. Two control sequences were found in *ISR11*. Three incidents of MalT binding sites (GGAKGA) were located at

positions 911 (minus strand), 1393 (plus strand) and 1744 (minus strand). MalT is a protein, involved in regulating the *malPQ* operon in *E. coli*.[298] In this operon there are three MalT binding sites upstream between positions -78 and -33. As the three sites found in IS*R11* are more than 300bp apart this motif was not considered further.

Potentially more interesting is the sequence TTTTGCA, found at position 1543 on the plus strand. This heptamer is a consensus sequence from the -15 region of promoters from many genes involved in nitrogen assimilation, including *nif* genes and glutamine synthetase genes (*gln*).[283] Since IS*R11* was found to have homology with *B. japonicum glnB* sequences the nucleotide sequence around position 1543 was examined for promoter structure.

Taking the first base of the sequence TTTTGCA as position -15 of the putative promoter, the region around -35 reads TTCTCC and that around -10 reads CACAGT, where underlined bases indicate homology to the consensus sequences. While it is true that this amounts to, at best, a very weak promoter it must be remembered that this is exactly what would be expected in transposable elements, where too efficient a transposition mechanism would ultimately be detrimental to the element in much the same way that a too virulent disease organism quickly runs out of hosts to infect.

However, if this sequence is a promoter, it is unconnected to the *glnB* homology as it is on the opposite strand and is 500bp away from this region.

The last sequence type to be considered was restriction endonuclease sites. Table 9 lists unique sites within IS*Rl1* (note that some of these sites are found within *rpsL*). In cases of isoschizomers only one enzyme is reported and where the recognition site of an enzyme contains a smaller site recognized by another enzyme, only the former is reported.

Table 9. Unique Restriction Sites in IS*Rl1*.

Enzyme	Target	Pos.	Enzyme	Target	Pos.
Spo I	(TCGCGA)	156	*Alw*N I	($CAGN_3CTG$)	1767
*Eco*N I	($CCTN_5AGG$)	332	*Hae* I	(WGGCCW)	1795
Bgl II	(AGATCT)	410	*Bal* I	(TGGCCA)	1797
*Bse*P I	(GCGCGC)	451	*Apa*L I	(GTGCAC)	1914
Dra II	($CACN_3GTG$)	687	*Rsa* I	(GTAC)	1947
Asu II	(TTCGAA)	1355	*Pvu* II	(CAGCTG)	2061
Sph I	(GCATGC)	1371	*Eco*R V	(GATATC)	2076
Xba I	(TCTAGA)	1497	*Afl* II	(ACRYGT)	2128
Bcl I	(TGATCA)	1620	*Bsa*A I	(YACGTR)	2129
*Ppu*M I	(RGGWCCY)	1661	*Fsp* I	(TGCGCA)	2248
Sgr AI	(CRCCGGYG)	1714	*Dra* I	(TTTAAA)	2304
Sal I	(GTCGAC)	1749	*Nde* I	(CATATG)	2375

Search for Coding Sequences in IS*Rl1*.

There are two approaches to finding coding regions from nucleotide sequence data, gene search by signal and gene search by content.

Gene Search by Signal.

The first method is based on finding all start codons and their corresponding (in frame) stop codons. The sequence between these signals constitutes an open reading frame (ORF). Usually only ORFs above a certain size are considered. Then control sequences, such as those described above, are sought and tested for association with the ORFs, with regard to orientation, being in frame and positioning relative to the ORF.

Galas and Chandler, in a survey of all available fully

sequenced IS elements from gram-negative bacteria;[124] considered all ORFs over 50 codons in length beginning with either ATG or GTG. This approach was followed with IS*R11*, in this study. Table 10 lists the start and stop sites, the reading frame (RF) and the length (in codons) of the ORFs found on both strands of IS*R11*, where two start codons are in frame with the same stop codon, the longer ORF is reported.

Table 10. Open Reading Frames of IS*R11*.

Plus Strand.				Minus Strand			
ORF	Start-Stop	RF	Codons	ORF	Start-Stop	RF	Codons
a	45-318	3	91	k	2412-2028	3	128
b	382-541	1	53	l	2312-2123	1	63
c	531-870	3	113	m	2029-1714	2	105
d	1031-1250	2	73	n	1919-1691	1	76
e	1054-1666	1	204	o	1872-1659	3	71
f	1389-1569	3	60	p	1382-1229	1	51
g	1550-1778	2	76	q	1230-924	3	102
h	1991-2156	2	55	r	1151-860	1	97
i	2130-2304	3	58	s	922-718	2	68
j	2234-2429	2	65	t	861-621	3	80
				u	722-206	1	172
				v	655-25	2	210

Figure 17 (page 125) shows a map of all these ORFs relative to IS*R11*, ORFs from the plus strand are represented by arrows above the bar representing IS*R11*, while ORFs from the minus strand are drawn below this bar. ORFs from reading frame 3 are drawn furthest from IS*R11* and those from reading frame 1 are drawn nearest. Where an ORF contains an in frame frameshift motif this is indicated by an asterisk under the arrow.

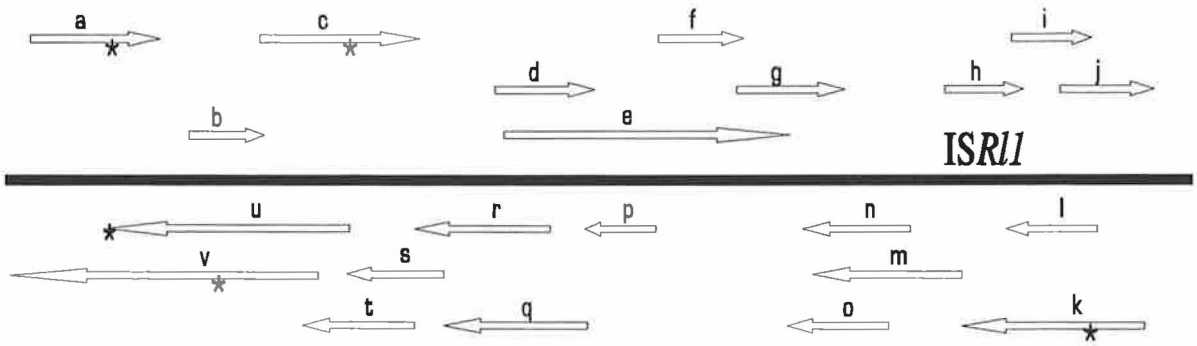

Figure 17. Map of ORFs in IS*Rl1*.

The sequence of each ORF was translated to protein, using the SEQAID package and submitted to the BLAST network server for comparison to its protein sequence database. This database includes translated coding regions from nucleotide sequences as well as simple protein sequences.

Interesting matches were reported to four ORFs, these ORFs and the sequences with matches to them are listed in table 11 along with the lengths of both sequences. The sequences producing the match are presented below table 11, homologous residues are rewritten between the two segments and the plus symbols reflect the scoring scheme used by the search algorithm, a base substitution at these symbols contributes favourably to the overall alignment score.[5]

Table 11. Matches to the Protein Sequences of the ORFs from IS*Rl1*.

ORF	Length	Target	Length
g	76	ATUACV2_1	156
h	55	RHAFIXA_2	34
r	102	MPSAFTERQ_2	260
v	210	ATACH5_14	104
		ARVIRCD_8	824

```
ATUACV2_1
 Score = 37, P = 0·65,
 Identities = 9/24 (37%), Positives = 12/24 (50%).

Query:  43 RRSGFTQQSREASPARIELASSPF  66
           RRSG    +R +S A +   S  F
Sbjct: 112 RRSGTVSAARCSSTANVSTTSRDF 135

 Score = 37, P = 0·65,
 Identities = 12/34 (35%), Positives = 17/34 (50%).

Query:   2 ARTMALRASTSSGSSISGRVMGMIRAYFAALRPC  35
           A   A R  +SS +S S      ++A  +A RPC
Sbjct:  70 ADARAARWISSSSASCSICARRDMKASISAWRPC 103
```

ATUACV2 (accession D13800) is an unpublished sequence (Wirawan, I. *et al.*) of an *A. tumefaciens* gene encoding a putative membrane protein. ORFg produced two matches to this sequence both of which were not to the gene itself but to an 470bp ORF (ORF1) which begins nearly 500bp upstream of the gene. There is no information about this ORF in the GENBANK entry, (accession D13800) other than the presence of a ribosome binding site 10bp upstream of the start codon.

```
RHAFIXA_2
 Score = 40, P = 0·99,
 Identities = 7/21 (33%), Positives = 13/21 (61%).

Query:  25 CASPISASARIMTSPITSSII 45
           C   +  SA+I   P+T++I+
Sbjct:   6 CIKQVPDSAQIRVHPVTNTIM 26
```

RHAFIXA_2 (accession M35122) is part of the sequence of a nitrogen fixation protein (FixA) from *Azorhizobium caulinodans*.[203]

```
MPSAFTERQ_2
 Score = 91, P = 2·1e-05,
 Identities = 22/57 (38%), Positives = 31/57 (54%).

Query:  18 IGEDVSERLDVVAAKFRVIVTRRPKYAFRNEDGVVQAAAPAHIIEGGI
           +GEDV+E L  +  +F+VI T R K+  R+ + + Q  AP H    G
Sbjct: 106 LGEDVTETLKEIPRRFKVIETVREKFTCRDCEAISQPPAPFHATPRGF
```

```
Query:  66 PTEALLAQI  74
              LLA I
Sbjct: 154 IGPHLLATI 162
```

Score = 41, P = 0·12,
Identities = 6/14 (42%), Positives = 11/14 (78%).

```
Query:  33 FRVIVTRRPKYAFR 46
           F+ +V R PK++F+
Sbjct:  41 FKTLVLREPKHSFK 54
```

MPSAFTERQ_2 is the sequence of an ORF of unknown function downstream of *nodQ1* and is part of the GENBANK entry MPSAFTERQ (accession L08667) to which nucleotide sequence homology with IS*R11* is reported above (page 117). The high scoring segment reported above does not overlap with any part of ORFr.

ATACH5_14

Score = 153, P = 2·4e-15,
Identities = 35/79 (44%), Positives = 46/79 (58%).

```
Query: 115 ISRRDIFERFLTDGRIELDSNIVERAIRPQAITRKNSLFAGSDGGGRT
           + R D    FL DG I L +N  ER +R   + RK+ LFAGSD
Sbjct:   2 LKRWDGITSFLDDGPICLTNNAAERTLRGYVLGRKSWLFAGSDRCAER

Query: 163 WATIATLLQTAKMNAVDPQAWLTQTLEALAN 193
            A +ATL+ +AK+N +DPQAWL      LA+
Sbjct:  50 AAFMATLIMSAKLNNIDPQAWLADVRADLAD  80
```

Score = 38, P = 1·0,
Identities = 5/15 (33%), Positives = 10/15 (66%).

```
Query: 196 PNSEIDALMPWNYAA 210
           P S ++  +PWN+ +
Sbjct:  82 PISRLEQQLPWNWTS  96
```

ATACH5_14 corresponds to ORF14 from the GENBANK entry ATACH5 (accession X00493 J05108 X00282). The entry is the sequence of the T-DNA region from *Agrobacterium tumefaciens* Ti plasmid pTi15955. Homology to ORF14 was also found at the nucleotide level (see above, page 118) though the scoring segment from that match does not overlap with either segment here and in fact the scoring segment from the nucleotide search came from the plus strand while ORFv comes

from the minus strand.

```
ARVIRCD_8
 Score = 48, P = 0·13,
 Identities = 9/21 (42%), Positives = 14/21 (66%).

Query: 169 LLQTAKMNAVDPQAWLTQTLE 189
           L+  A+++A+D  AWL Q  E
Sbjct: 221 LILKARLDAIDSSAWLPQLNE 241


 Score = 48, P = 0·13,
 Identities = 10/27 (37%), Positives = 16/27 (59%).

Query: 159 GGRTWATIATLLQTAKMNAVDPQAWLT 185
           GG    T   L+  A+++A+D  AWL+
Sbjct:  76 GGSGRPTWKDLVVKARLDAIDSSAWLS 102
```

ARVIRCD_8[115a] is the translated protein sequence of the gene *virD5* from *Agrobacterium rhizogenes* pRiA4b plasmid, reported as part the sequence ARVIRCD (accession X12867).

Gene Search By Content.

The second method of finding coding regions is through gene search by content. The basis of this method is to use a codon bias table for the organism (in this case *R. leguminosarum* - all biovars) and to calculate and then plot the probability that the stretch of sequence around each base codes for protein.

Table 12 shows the codon bias table composed for *R. leguminosarum* from protein coding regions found in GENBANK sequences. The accession numbers of the entries used, along with a short description of each and the location of the protein coding region in the GENBANK entry, are listed in appendix B. The total number of codons used was 22,647.

Table 12. Codon Bias Table for *R. leguminosarum*.

Res.	Codon	Count	Freq.	Res.	Codon	Count	Freq.
Ala	GCA	408	0·161	Pro	CCA	191	0·163
	GCC	936	0·368		CCC	277	0·236
	GCG	836	0·329		CCG	540	0·461
	GCT	362	0·142		CCT	164	0·140
Cys	TGC	275	0·797	Gln	CAA	203	0·301
	TGT	70	0·203		CAG	471	0·699
Asp	GAC	705	0·569	Arg	AGA	84	0·057
	GAT	533	0·431		CGA	173	0·118
Glu	GAA	560	0·460		CGC	571	0·390
	GAG	657	0·540		AGG	119	0·081
Phe	TTC	615	0·710		CGG	326	0·223
	TTT	251	0·290		CGT	192	0·131
Gly	GGA	243	0·134	Ser	TCA	132	0·098
	GGC	1032	0·571		TCC	273	0·204
	GGG	237	0·131		AGC	314	0·234
	GGT	295	0·163		TCG	422	0·315
His	CAC	266	0·512		TCT	112	0·084
	CAT	254	0·488		AGT	88	0·066
Ile	ATA	126	0·095	Thr	ACA	173	0·143
	ATC	915	0·690		ACC	446	0·368
	ATT	286	0·216		ACG	425	0·351
Lys	AAA	244	0·295		ACT	167	0·138
	AAG	583	0·705	Val	GTA	137	0·081
Leu	CTA	126	0·056		GTC	758	0·449
	TTA	72	0·032		GTG	482	0·286
	CTC	681	0·303		GTT	310	0·184
	CTG	675	0·301	Trp	TGG	284	1·000
	TTG	312	0·139	Tyr	TAC	263	0·503
	CTT	379	0·169		TAT	260	0·497
Met	ATG	591	1·000	Ter	TAA	17	0·221
Asn	AAC	386	0·561		TGA	50	0·649
	AAT	302	0·439		TAG	10	0·130

The results of these calculations are shown in figures 18 (plus strand) and 19 (minus strand). Each graph represents one reading frame and the vertical scale within each box is log[P/(1-P)] so that, for example, two points up the scale from mid-height corresponds to 90% probabil-

ity. The horizontal scale marks every 50th base and the lines above each graph indicate the extent of the open reading frames found above (page 124 and 125).

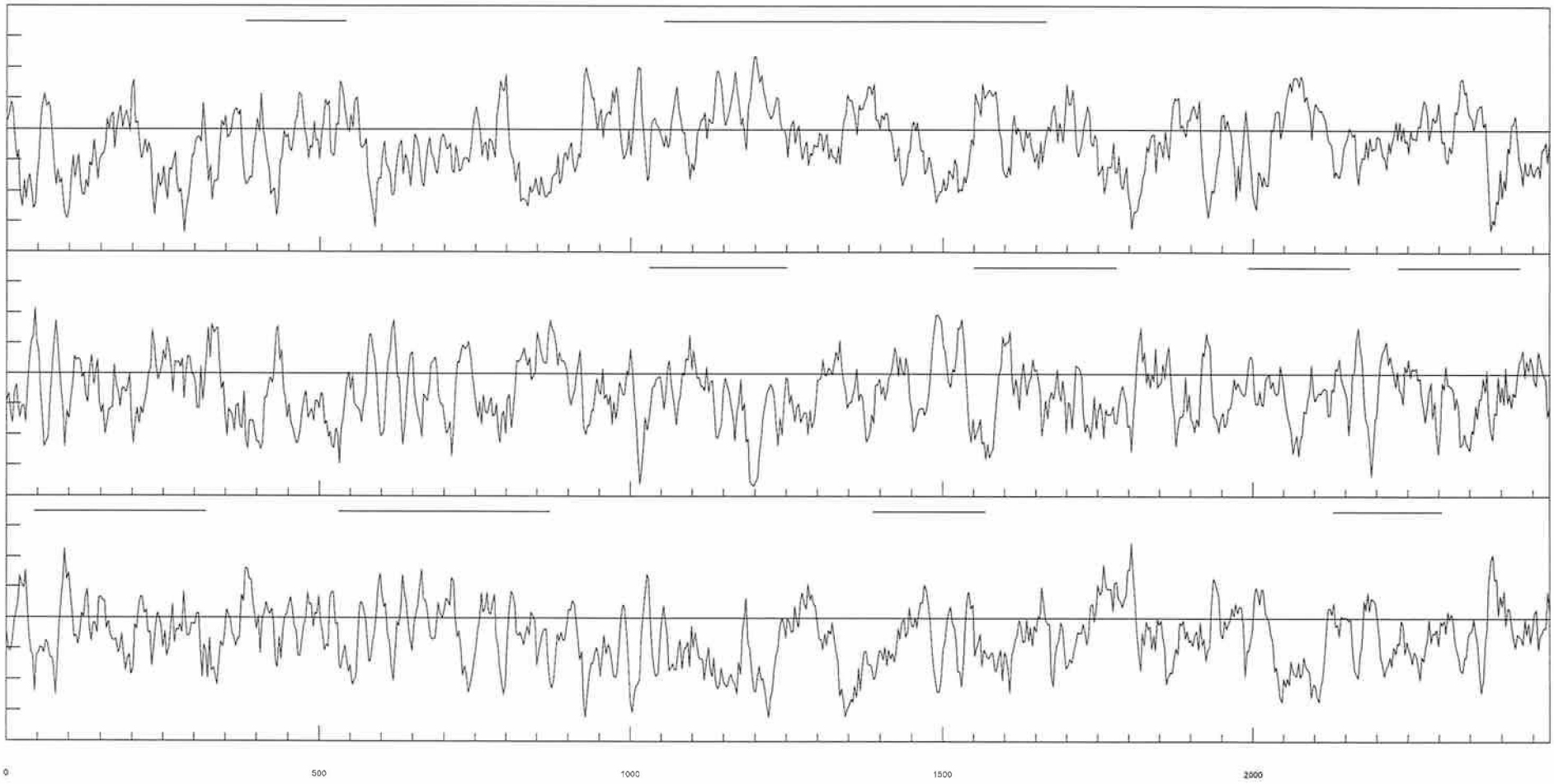

Figure 18.
Protein Coding Region Prediction for + Strand of IS*Rl1*.

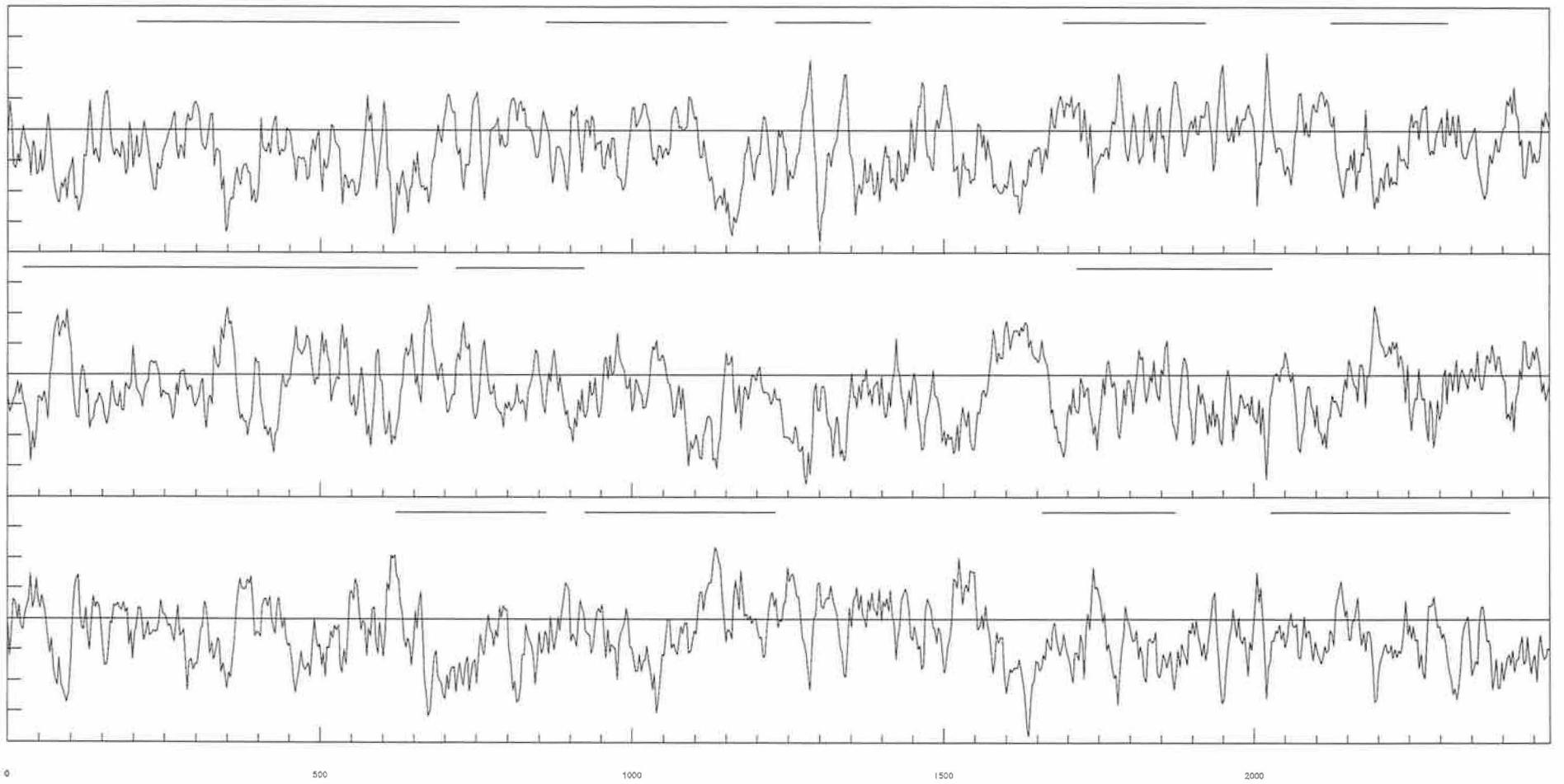

Figure 19.
Protein Coding Region Prediction for - Strand of IS*Rl1*.

Discussion.

When all the sequence homologies found through database searches are mapped to positions on IS*R11*, there are four regions of interest.

Region 1, extends from nucleotides 30 to 377. This region was found to have homology to two insertion sequences and an unidentified open reading frame from *A. tumefaciens* (IS*866* TIPIS866A, IS*1131* ATUIS22A and ORF14 ATACH5-_14), to a lipopolysaccharide gene from *R. meliloti* (*lpsZ* RHMLPSZ) and to a virulence gene from *A. rhizogenes* (*virD5* ARVIRCD_8).

Region 2, extends from position 1008 to 1180 and shows homology to another insertion sequence from *A. tumefaciens* (IS*66* TIPISA66) and to an unidentified ORF from *R. meliloti* (ORF2 MPSAFTERQ).

Region 3, (1556 - 1748) has homology to a putative membrane protein from *A. tumefaciens* (ATUACV2) and region 4 (2066 - 2143) has similarities to a putative nitrogen-regulatory gene from *B. japonicum* (*glnB* RHBGLNB) and to a nitrogen fixation gene from *A. caulinodans* (*fixA* RHA-FIXA_2).

All of these sequences come from species of *Rhizobiaceae* which is a strong argument against the observed similarities being any kind of artefact. However, there is a possibility that the homologies observed are due to local similarities of GC content. To test this the sequence data was edited, substituting G for C and C for G. In this way the GC content of the sequence was unaltered and so, if this is what produced the observed homologies, repeated searches would be expected to produce homologies at the same identities as before (although to different genes). When these searches were performed no homologies of any kind were reported, reinforcing the conclusion that the similarities observed above are real.

Altogether these regions account for 788bp which is just under a third of IS*R11*. Since only a very small fraction of any of the genomes of these organisms has been sequenced it seems probable that better, more extensive homologies remain to be discovered.

The observed homology to the various genes provides plenty of support of the theory that IS*R11* encodes accessory functions. *glnB* is only 336bp long so IS*R11* could easily encode a similar gene and still have sufficient space to encode transposition functions.

Region 3, which corresponds roughly to ORFg is preceded by a possible promoter sequence at position 1543 (see page 122 above). Since there is no published information on the ATUACV2 sequence it is difficult to make any comparisons with ORF1 from this sequence and ORFg from IS*R11*. There is room to speculate that this ORF is functionally associated with the downstream gene and that its homology with ORFg may, in that case, be quite significant.

None of the matches to IS elements were to transposases, putative or otherwise, although the homology between IS*R11* and IS*66* was located in open reading frames in both elements (ORFr and ORF2 respectively). The possibility remains that IS*R11* is deficient in one or more transposition functions, for which it depends on other elements. However, if this were the case, it would be reasonable to expect a reduced rate of transposition in this element as transposase are known to act efficiently only *in cis*.[138] IS*R11* has been shown to transpose at a frequency of $1{\cdot}5\times10^{-5}$ which compares favourably with other elements.[124]

That IS*866* should be the best match to IS*R11* comes as little surprise in light of the similarities already observed between these elements. A weak match was also reported to IS*66*, which was also already known to have important similarities to IS*R11*.

IS*868*, another IS element from *A. tumefaciens*, is 65% homologous to IS*51*, an IS element from *Pseudomonas savastanoi* and is associated with the IS*3* family.[286] (Another *A. tumefaciens* element IS*426* also shows homology to IS*51*[90] while IS*866* shows none[268]). This element shows no similarity to IS*R11*.

Of the three elements IS*R11*, IS*866* and IS*66* (the sequence of IS*Rm2* is not known) only IS*R11* and IS*866* have an IHF site. In both cases, this site is approximately half way along the element (positions 1188 and 1123, respectively). IHF is a histone-like protein and induces a strong bend on binding. The presence of this site halfway along

both of these long insertion sequences, suggests that it may be there to bring the ends of the elements together and that this may be an important step in their transposition.

The DnaA protein is involved in the initiation of chromosomal replication. If the transposition mechanism is replicative in nature (see page 46) the presence of recognition sites for this protein has obvious advantages to insertion sequences.

As the Dam methylase site is only four bases long it is not significant that IS*R11* contains so many of them (26). These sites, at the end of IS elements, have been shown to be involved in both the expression of IS-encoded transposition functions and to determine the intrinsic activity of the ends in transposition.[215] In the case of IS*R11*, however, there are no Dam sites in the inverted repeats and so these sites, if they have any role, are probably confined to regulating the expression of genes.

The 24 unique restriction endonuclease sites along the length of IS*R11* would lend themselves very well to studies on the element, for example by disrupting the reading frame of ORFs or by inserting marker genes.

The number of ORFs found in IS*R11* is consistent with the number found in IS*66*[124] (22 and 16 respectively). Like IS*R11*, IS*66* is unlike typical insertion sequence in that it lacks a long ORF spanning the length of the element. The longest ORF in IS*66* (258 codons) is longer than its counterpart in IS*R11* (ORFv - 210 codons). No pattern in the spacing or arrangement of ORFs was discernible between the two elements, although if the elements use translational frameshifting such a pattern could be difficult to detect. Two of the four ORFs in IS*868* are almost the full length of the element (1·3kb),[286] IS*866* also has a lower number of ORFs (five), one of which is 322 codons long.[38]

Five of the ORFs carry in-frame translational -1 frameshifting motifs. One of these (in ORFu) is at the end of the element (and ORF) and so cannot have any significance to the transposition process of the element but may have implications for the region into which it transposes. Use of the motif in ORFv would result in termination at the ORFu stop codon, yielding a shorter protein than ORFv on its own would produce. The other motifs, however, would all

cause fusion of the product of two ORFs. In the case of ORFa, this fusion product would be particularly long (more than 400 codons spanning ORFa and ORFd, assuming that there are no intervening in frame transcriptional terminator sequences).

Open reading frames, ORFu and ORFv, are among the longest ORFs in IS*R11* and also overlap region 1, described above, which showed homology to two different insertion sequences. This is an indication that these ORFs may be involved in the transposition of IS*R11*.

The attempt to determine the protein coding regions in IS*R11* through the use of a codon bias table failed. There are no stretches of any of the six graphs that stay consistently above the 90% mark. There may be several reasons for this. Firstly, as this method is a statistical method, it may work well only for searches for longer genes in longer sequences. No other attempt to analyze an IS element in this way has been published. Secondly, the method works on the assumption that all the genes in a particular organism have a similar codon usage and that this usage is sufficiently strong to be used as a means of discrimination. It is possible that IS elements favour rarer codons, as one means of self regulation. Furthermore, another assumption of the method is that the sequence is coding in only one frame, which is unlikely to be the case for IS elements.* Lastly, given the number of sequences from *A. tumefaciens* to which IS*R11* shows homology, it is possible that IS*R11* does not originate in *R. leguminosarum*. If this were the case then a different codon bias table would be necessary for this method to work.

* References 50,124,138,212

Chapter 5. Distribution of ISRL1.

Introduction.

Leguminous crops, such as pea, soybean and alfalfa, in symbiosis with rhizobia, can fix all the nitrogen required for their own needs and may enrich soil nitrogen for associated or subsequent crops. The use of rhizobia in this way has several advantages over the application of chemical fertilizers. They are cheaper and cleaner, (much of the chemical fertilizers applied is lost through runoff and volatilization); they are more efficient (as all the nitrogen produced is available directly to the plant) and the practice of crop rotation contributes greatly to pest and pathogen control.

For these and other reasons there is considerable interest in the rhizobia/legume symbiosis and in engineering improved rhizobial strains. The symbiont of soybean, *B. japonicum*, is attracting a lot of interest in this regard, not only because of its association with soybeans, a very important crop commercially, but also because it nodulates *Parasponia*, a non-legume.[345]

One obstacle to such research is the difficulty of distinguishing easily among different strains and thus tracing introduced strains in field tests, particularly since it has been shown that indigenous strains easily outcompete introduced ones.[147] This problem is even more complex in the tropics where rhizobia are less specific as to the plant species they nodulate, than rhizobia in temperate zones.

The principal approaches to identifying bradyrhizobial strains are intrinsic antibiotic resistance,[262] bacteriophage typing,[223] serology[339] and DNA fingerprinting.[149,323] The first two of these, antibiotic resistance and phage typing are very limited, serology is a better method but is impaired by numerous cross reactions (serogroup 123, one of the most important types, has been shown to be very diverse).[340] A limitation that all three of these methods share is that there is very little correlation between the distinctions they make and the N_2-fixing or competitive properties of the strains.

The last technique, DNA fingerprinting, seems to be capable both of providing much better resolution between strains and of bearing a direct relation to their symbiotic properties. Previous work with IS*RI1* (A. Ryan, M.Sc. thesis, DCU, see page 64 above) suggested that this element might make a good probe for use in such a protocol. To test this theory the distribution of IS*RI1* among strains of *B. japonicum* and some rhizobial species was examined. Also, in the hope of obtaining more information on the properties of this insertion sequence, its distribution among a wide variety of bacterial species was examined.

The plasmids pSVB30-*rpsL*::IS*RI1* and pSVB30-*rpsL* were used for the following Southern work. pSVB30-*rpsL*::IS*RI1* was also used in the sequencing of IS*RI1* (see chapter 4, page 94 where the construction of this plasmid is described) pSVB30-*rpsL* was constructed in a similar manner.

Detection of IS*RI1* in *R. leguminosarum* bv. *viciae* and *R. meliloti*.

When IS*RI1* was first discovered several rhizobial species were examined for its presence by Southern hybridization with pSUP104-*rpsL*::IS*RI1* as a probe. These experiments were performed at a hybridization temperature of 68°C and no control experiments were performed (i.e. hybridization with pSVB30-*rpsL*. If IS*RI1* is to be a useful DNA probe it must show several bands in Southern blots. To maximize this, the hybridization experiments in this study were performed at 64°C. Also, since *rpsL* encodes a bacterial ribosomal protein (S12) and is likely to be conserved across a wide range of species, it was necessary to repeat each experiment using pSVB30-*rpsL* as a probe to ensure that each band detected was due to IS*RI1*.

To extend the survey of the *Rhizobiaceae*, two strains of *R. leguminosarum* bv. *viciae* and nine strains of *R. meliloti* were probed with pSVB30-*rpsL*::IS*RI1* and pSVB30-*rpsL*. The gels and autoradiograms are shown in figure 20. The order in which samples were loaded was the same in each case and is listed in table 13. The only difference being the control samples (lane 15) which were total DNA from *R. leguminosarum* bv. *viciae* 897 for the IS*RI1* hybridization

and pSVB30- *rpsL*::IS*Rl1* for the control hybridization.

Table 13. Key to Figure 20.

Lane	Strain/Marker
1	1kb DNA ladder.
2	Empty.
3	*R. leguminosarum* VF39
4	*R. leguminosarum* DCU300
5	*R. meliloti* 65
6	*R. meliloti* 65C
7	*R. meliloti* 220-3
8	*R. meliloti* 220-7
9	*R. meliloti* 220-8
10	*R. meliloti* 220-15
11	*R. meliloti* 220-16
12	*R. meliloti* 220-20
13	*R. meliloti* 2011
14	Empty.
15	Control.

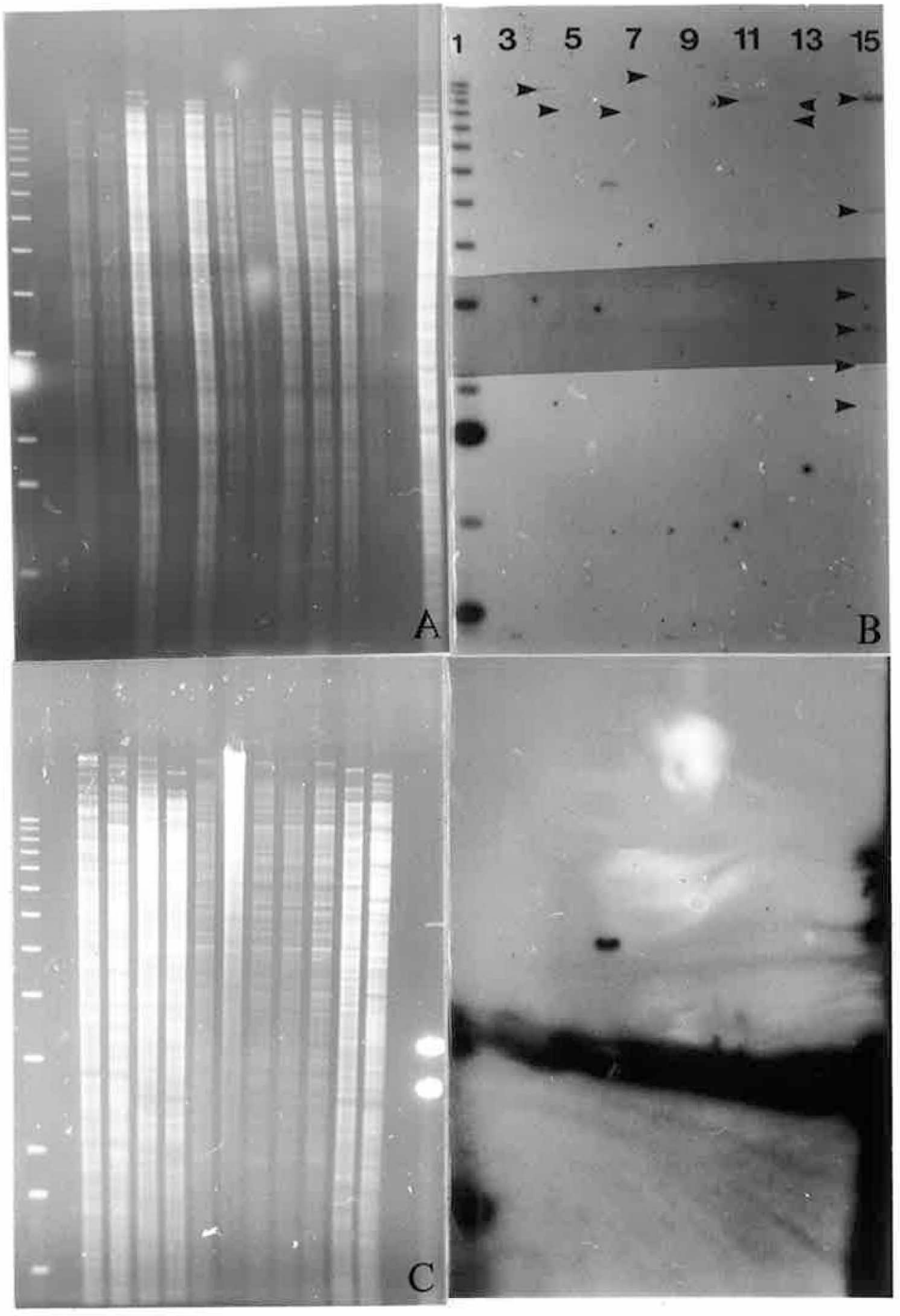

Figure 20. Detection of IS*R11* in *R. leguminosarum* bv. *viciae* and *R. meliloti*.

A. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from two *R. leguminosarum* bv. *viciae* strains, nine *R. meliloti* strains and *R. leguminosarum* bv. *viciae* 897.

B. Autoradiogram of the agarose gel in figure 20A probed with pSVB30-*rpsL*::IS*R11*.

C. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from two *R. leguminosarum* bv. *viciae* strains and nine *R. meliloti* strains, and pSVB30-*rpsL*::IS*R11*.

D. Autoradiogram of the agarose gel in figure 20C probed with pSVB30-*rpsL*.

The inclusion of 1kb DNA ladder in the gels and in the reaction solution for ^{32}P-labelling allowed the convenient size-estimation of the hybridizing bands. The size and mobility of each marker is given in table 14. The DNA ladder was omitted from the labelling reaction for the control experiments, however, as explained in the discussion, this presented no real obstacle.

Table 14. Size Markers in Figure 20B.

Size (bp)	Mobility (mm)	%Deviation
12216	45·50	1·93
11198	47·50	0·31
10180	50·00	-0·87
9162	53·00	-2·01
8144	57·50	-0·69
7126	62·50	-0·69
6108	69·00	-0·20
5090	77·50	0·47
4072	89·00	1·31
3054	105·00	1·74
2036	128·50	0·74
1636	140·50	-1·52
1018	169·00	-1·04

Std. Dev. = 92·73bp

The dark smear across figure 20D is caused by nonspecific hybridization of the probe to the nitrocellulose. The most likely cause of this is that the heatsealer bag stuck to the nitrocellulose during the prehybridization stage thereby preventing, or at least impeding, the salmon sperm DNA from blocking the filter over this area. However, as no bands appear in the autoradiogram in figure 20B below the 5·090kb marker, this smear cannot be obscuring anything of interest and so can be ignored.

To confirm that a band on the control autoradiogram matches one or another band on the main autoradiogram, it is possible to take advantage of the way the size estimation program works (see above, page 82) by re-entering the sizes and mobilities of the standards and including the

band in question as if it was a standard. When the program returns the standard deviation for each fragment, the accuracy of the size chosen for the band can be determined.

Table 15 lists all the IS*R11* hybridizing fragments present in figure 20B.

Table 15. Fragment Sizes Hybridizing to IS*R11* in Figure 20B.

Strain	Mobility	Size (bp)
R. leguminosarum DCU300 (Lane 4)	47·00	11358
R. meliloti 65 (Lane 5)	53·00	9346
R. meliloti 220-3 (Lane 7)	54·00	9070
R. meliloti 220-7 (Lane 8)	46·00	11766
R. meliloti 220-16 (Lane 11)	51·50	9789
R. meliloti 220-20 (Lane 12)	56·50	8434
	59·00	7869
R. leguminsarum bv. *viciae* 897 (Lane 15)	51·50	9789
	78·50	4960
	101·50	3192
	112·00	2660
	122·00	2249
	132·00	1908
Band #1.	102·00	3164
Band #2.	112·50	2637

A smaller band is present above the two expected bands in lane 15 of figures 20C and 20D (pSVB30-*rpsL*::IS*R11* cut with *Eco*R I). The two expected bands are the *rpsL*::IS*R11* fragment (3·2kb) and the pSVB30 fragment (2·7kb). This third band is likely to be partially restricted pSVB30-*rpsL*::IS*R11*, as indicated by its size (~5·9kb) and the fact that it shows up on the autoradiogram (figure 20D).

There are two faint bands present in all the sample lanes (3 - 13). These bands are indicated in table 15 as bands #1 and #2 and are also present in the other autoradiograms presented in this chapter, though this is the clearest example.

Despite the fact that the 1kb DNA ladder was left out of the labelling experiment three bands show up in this lane in figure 20B. This is because although these size markers are largely prepared from λ, some vector fragments are included (see Gibco BRL catalogue #520-5615SA). Therefore these fragments, along with the bands in lane 15 (pSVB30-*rpsL*::IS*Rl1*), can be used to estimate the size of fragments in the negative control (figure 20D).

Detection of ISR*l1* in *R. leguminosarum* bv. *trifolii*.

R. leguminosarum bv. *trifolii*, a symbiont of clover, was formerly classified as a species in its own right but has been reclassified as a biovar of *R. leguminosarum* (as has *R. phaseoli*). Because of its close relationship to the original source of IS*Rl1* (*R. leguminosarum* bv. *viciae* 897), and because a large collection of strains was available, *R. leguminosarum* bv. *trifolii* was subjected to the same experiments as *R. leguminosarum* bv. *viciae* and *R. meliloti*.

Two experiments were carried out (hybridization with IS*Rl1* and a control hybridization with the vector only) for reasons explained on page 138. The samples were loaded in the same order in each case (listed in table 16). Total DNA from *R. leguminosarum* bv. *viciae* 897 and pSVB30-*rpsL*::-IS*Rl1* were used as the controls (lane 22) in the IS*Rl1* and control hybridizations, respectively. Figure 21 shows the agarose gels and autoradiograms.

Table 16. Key to Figure 21.

Lane	Strain/Marker	Lane	Strain/Marker
1	1kb DNA ladder	12	S12-19
2	SB-H1	13	S12-H3
3	SB-H2	14	S30-2
4	SB-H4	15	S30-3
5	SB-H9	16	S30-7
6	S10-H5	17	S30-11
7	S10-H6	18	S34-H3
8	S10-H8	19	S34-13
9	S10-H2	20	S34-H9
10	S12-11	21	S34-16
11	S12-14	22	Control.

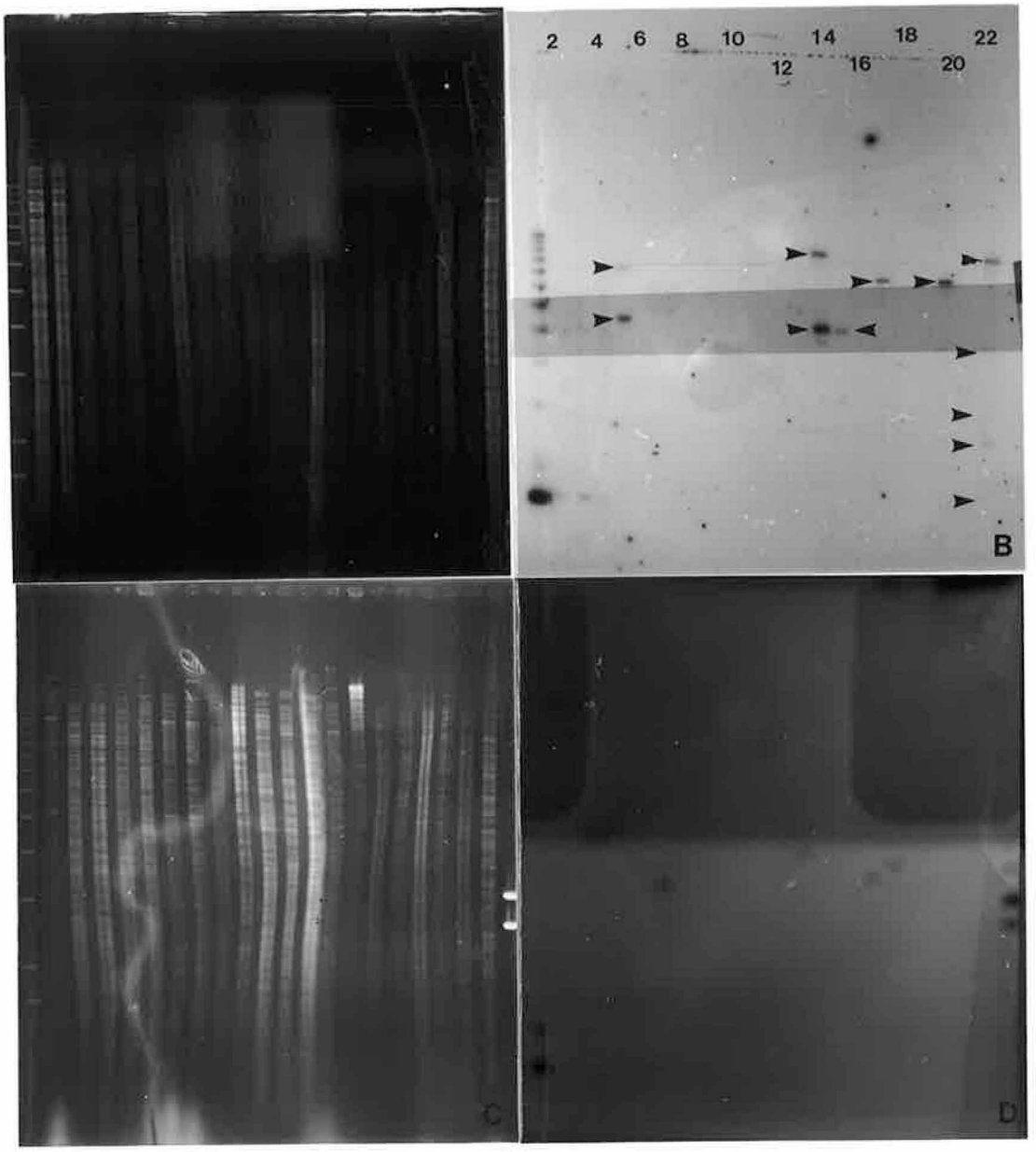

Figure 21. Detection of IS*Rl1* in *R. leguminosarum* bv. *trifolii*.

A. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from twenty *R.leguminosarum* bv. *trifolii* strains and from *R. leguminosarum* bv. *viciae* 897.

B. Autoradiogram of the agarose gel in figure 21A probed with pSVB30-*rpsL*::IS*Rl1*.

C. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from twenty *R. leguminosarum* bv. *trifolii* strains and pSVB30-*rpsL*::IS*Rl1*.

D. Autoradiogram of the agarose gel in figure 21C probed with pSVB30-*rpsL*.

Table 17 lists the size and mobility of the 1kb DNA ladder fragments. As in the previous experiment, the size markers were omitted from the labelling reaction for the control experiment.

Table 17. Size Markers in Figure 21B.

Size (bp)	Mobility (mm)	%Deviation
12216	57·50	1·09
11198	60·00	0·29
10184	63·00	-0·24
9162	66·50	-0·99
8144	71·00	-0·93
7126	76·50	-0·83
6108	83·50	-0·40
5090	92·50	0·17
4072	105·00	1·88
3054	121·00	1·70
2036	144·00	0·45
1636	155·00	-2·62

Std. Dev. = 63·30bp

The same points made above (pages 142 and 143) about the marker and control lanes in the *R. leguminosarum* bv. *viciae/R. meliloti* control autoradiogram apply here (to lanes 1 and 22, figure 21D) and the two faint bands across all the lanes are also visible in figure 21B.

These strains of *R. leguminosarum* bv. *trifolii* were isolated from clover plants taken from various locations throughout Leinster (G. Reigh, personal communication). No information is available about the relatedness of these strains but some degree of relatedness can be inferred in several cases by the similarity of the restriction banding patterns in figure 21B (see, for example, lanes 16 and 19).

Those bands present in figure 21A but not in figure 21D (i.e. bands produced by *ISRl1* hybridization) are listed in table 18.

Table 18. Fragment Sizes Hybridizing to IS*R11* in Figure 21B.

Strain	Mobility	Size (bp)
SB-H9 (Lane 5)	69·50	8543
	88·50	5514
S30-2 (Lane 14)	62·50	10359
	93·00	5031
S30-3 (Lane 15)	93·50	4981
S30-11 (Lane 17)	74·00	7629
S34-H9 (Lane 20)	74·00	7629
R. leguminosarum	64·50	9780
bv. *viciae* 897	93·50	4981
(Lane 22)	123·00	2900
	133·00	2443
	144·50	2010
Band #1.	120·50	3028
Band #2.	129·50	2594

Detection of IS*R11* in *B. japonicum*.

In earlier work A.Ryan (M.Sc. thesis, 1987) found that IS*R11* hybridized to ten *B. japonicum* strains, although control blots with pSVB30-*rpsL* were not performed in that case. In this study, to confirm that those results were due to hybridization to IS*R11*, those strains were reprobed (at the lower temperature of 64°C) and a number of previously untested strains were screened.

Figure 22 shows the agarose gels and autoradiograms from the hybridization experiments. The DNA samples were loaded in the same order in each case, the only difference being the control samples in lane 15, (total DNA from *R. leguminosarum* bv. *viciae* 897, the source strain of the element, for the IS*R11* hybridization, and pSVB30-*rpsL*::-IS*R11* for the control experiment). The order in which samples were loaded is given in table 19.

Table 19. Key to Figure 22.

Lane	Strain/Marker
1	1kb DNA ladder.
2	Empty
3	27-2
4	40-2
5	NJ-2
6	5-7
7	USDA110
8	USDA123
9	AR6
10	AR14
11	AR34
12	U
13	SD
14	Empty
15	Control.

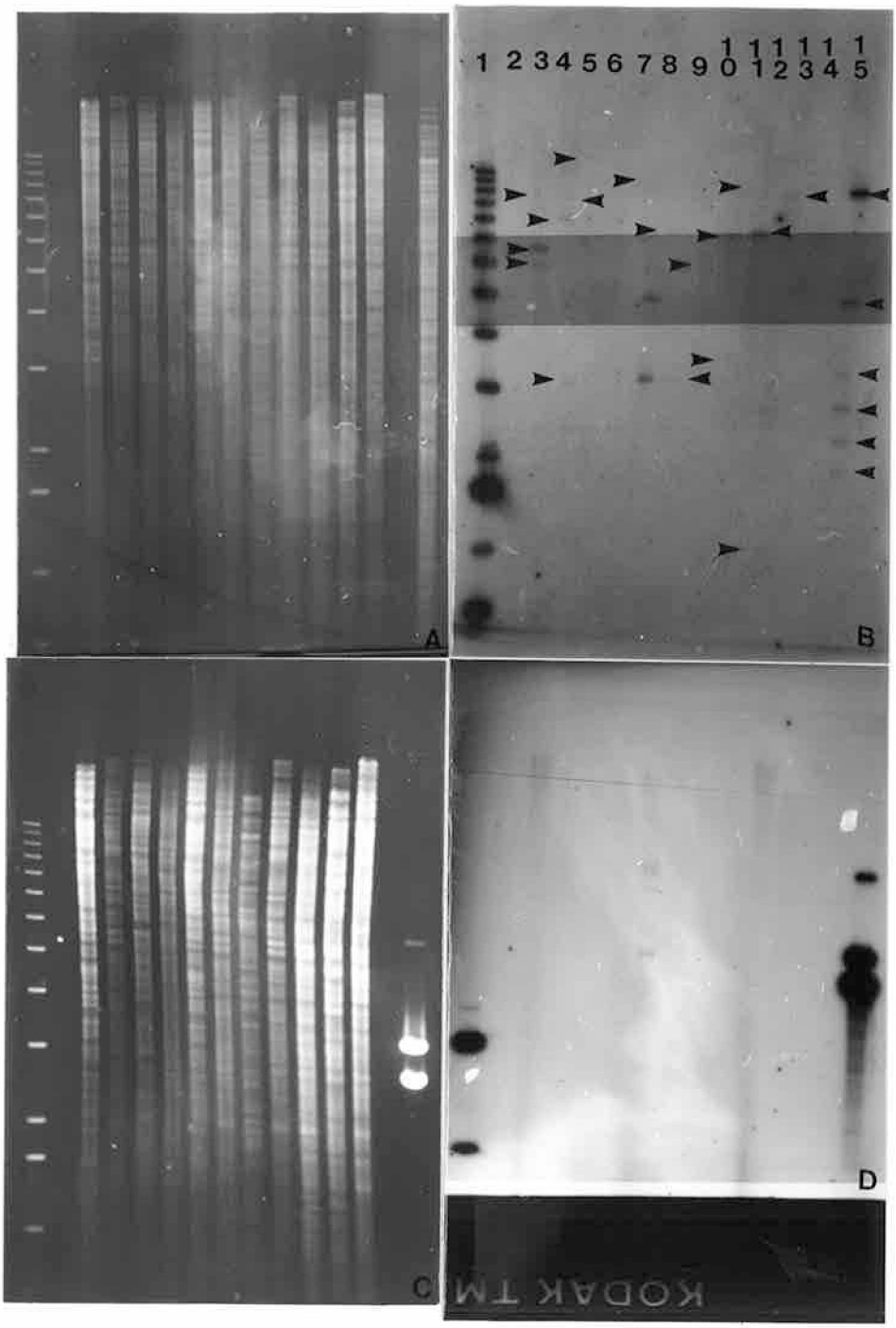

Figure 22. Detection of ISR*l1* in *B. japonicum*.

A. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from eleven *B. japonicum* strains and from *R. leguminosarum* bv. *viciae* 897.

B. Autoradiogram of the agarose gel in figure 22A probed with pSVB30-*rpsL*::IS*Rl1*.

C. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from eleven *B. japonicum* strains and pSVB30-*rpsL*::IS*Rl1*.

D. Autoradiogram of the agarose gel in figure 22C probed with pSVB30-*rpsL*.

The size and mobility of the 1kb DNA size markers for the IS*R11* hybridization are listed in table 20, the size markers were omitted from the control experiment.

Table 20. Size Markers in Figure 22B.

Size (bp)	Mobility (mm)	%Deviation
12216	54·50	0·94
11198	57·50	0·84
10180	60·50	-0·50
9162	64·50	-0·65
8144	69·00	-1·39
7126	75·00	-0·84
6108	82·00	-0·99
5090	91·50	0·29
4072	103·50	1·30
3054	119·50	2·47
2036	141·00	2·38
1636	150·00	-1·74
1018	171·00	-3·45

Std. Dev. = 68·65bp

During the course of this study more than 30 strains of *B. japonicum* were probed in this way for the presence of IS*R11*. The strains listed in table 19 are only those which showed positive results. Although strain 5-7 above (lane 6) shows no positive band it was included as it had shown a faint band in earlier experiments with lower stringency.

Table 21 lists those bands found in figure 22B and not in 22D (i.e. bands due to hybridization with IS*R11*). In only one case was a band repeated in both autoradiograms (lane 7 - USDA110), in spite of the fact that, for most of the strains the concentration of DNA in the control gel (figure 22C) is clearly greater than in the main gel (figure 22A).

Strains USDA123, AR6, AR14, AR34, U and SD (lanes 8 - 13) are all from serogroup 123. No information is available for the other strains except for USDA110 which is from serogroup 110.

Table 21. Fragment Sizes Hybridizing to IS*R11* in Figure 22B.

Strain	Mobility	Size (bp)
27-2 (Lane 3)	62·00	9833
	77·50	6797
	82·00	6168
40-2 (Lane 4)	62·50	9706
	68·00	8458
	118·50	3035
NJ-2 (Lane 5)	51·00	13459
USDA110 (Lane 7)	53·00	12654
USDA123 (Lane 8)	70·50	7968
	116·00	3179
AR6 (Lane 9)	81·00	6301
AR14 (Lane 10)	72·00	7694
	116·00	3179
AR34 (Lane 11)	72·00	7694
	116·00	3179
U (Lane 12)	60·50	10230
	166·00	1184
S.D. (Lane 13)	58·00	10950
R. leguminsarum	62·00	9833
bv. *viciae* 897	92·00	5025
(Lane 15)	114·00	3299
	124·00	2740
	132·00	2359
	141·50	1968
Band #1.	128·50	2519
Band #2.	132·00	2359

Detection of ISR*l1* in Bacteria Other Than Rhizobia.

To examine the distribution of IS*Rl1* outside *Rhizobium* and *Bradyrhizobium* species a range of gram positive and gram negative species were tested, including one more member of the *Rhizobiaceae*, *A. tumefaciens*.

As before, IS*Rl1* and control hybridizations were performed. The gels and autoradiograms are shown in figure 23. The order in which samples were loaded is listed in table 22. The control samples (lane 22) were total DNA from *R. leguminosarum* bv. *viciae* 897 for the IS*Rl1* hybridization and pSVB30-*rpsL*::IS*Rl1* for the control hybridization.

Table 22. Key to Figure 23.

Lane	Strain/Marker	Lane	Strain/Marker
1	1kb DNA Ladder.	12	*P. aeruginosa*
2	Empty.	13	*P. flourescens*
3	*B. megaterium*	14	*P. putida*
4	*B. subtilis*	15	*A. tumefaciens*
5	*B. thuringiensis*	16	*E. coli*
6	*L. amylovorus*	17	*P. vulgaris*
7	*L. plantarum*	18	*S. typhimurium*
8	*S. faecalis*	19	*S. marcescens*
9	*S. lactis*	20	*V. natriegens*
10	*P. acidilactici*	21	Empty.
11	*A. globiformis*	22	Control.

Inspection of the hybridization results, showed that every band in the IS*Rl1* hybridization was also present in the control hybridization and in fact there are some bands in the control which are absent from the IS*Rl1* hybridization. The autoradiogram of the control hybridization and the corresponding agarose gel are shown in figure 23. There is no possibility of the two blots having been confused. The most likely explanation is that the presence of IS*Rl1* in the *rpsL* gene reduces the homology between this sequence and the target sequences thus reducing the intensity of some hybridzation signals below the detection limit. Whatever the reason, there are clearly no bands caused by IS*Rl1*

hybridization in the strains tested.

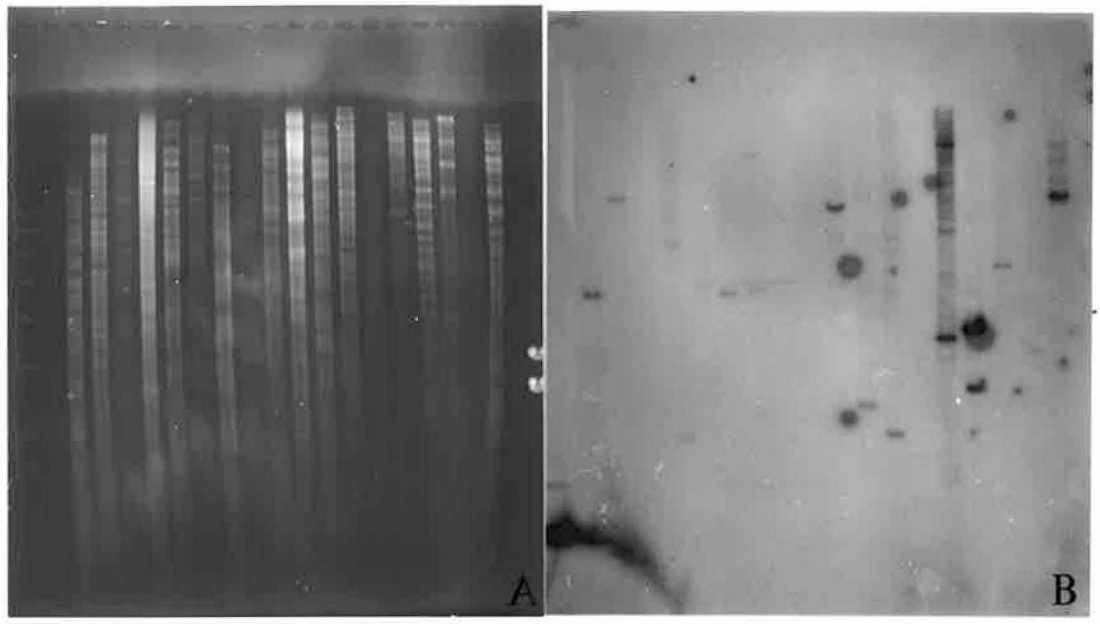

Figure 23. Detection of IS*R11* in Bacteria Other Than Rhizobia.

A. Agarose gel of 1kb DNA ladder and *Eco*R I restricted total DNA from eighteen bacterial species and pSVB30-*rpsL*::IS*R11*.

B. Autoradiogram of agarose gel in figure 23A probed with pSVB30-*rpsL*.

Discussion.

The effect of lowering the stringency of the hybridization experiments is clearly seen in the source organism *R. leguminosarum* bv. *viciae* 897. When IS*Rl1* was first isolated it was found to hybridize to only three bands in *R. leguminosarum* bv. *viciae* 897 (at a temperature of 68°C). In this study, it was found to hybridize to six. Averaging over three autoradiograms (figures 15B, 18B and 21B), the sizes of these bands are c. 9·8, 5·0, 3·1, 2·6, 2·2 and 1·9, where the first, second and fifth bands are those previously detected.

The other three bands are probably related elements or, as one of them is smaller than IS*Rl1*, derivatives of the same element in deleted or reduced form. As *R. leguminosarum* bv. *viciae* 897 was not included in any of the control gels it is possible that one or more of these bands hybridized to *rpsL*.

The results from the probing experiments with *B. japonicum* (figure 22, page 149) show that when IS*Rl1* is present in strains it shows a fair degree of variation in the number and size of *Eco*R I bands to which it hybridizes. But it has been found in only ten strains out of more that thirty screened over the course of this study.

In those strains to which it does hybridize, it seems to provide a means of differentiating between strains that may be very closely related. In two instances from figure 22A, there are two strains showing very similar restriction banding patterns (27-2 and USDA110, 40-2 and USDA123) but with different hybridization patterns (see table 21, page 151). No other information is available about the relatedness of these pairs of strains but the close similarities between their restriction patterns indicates that the relationship is a close one. If this is the case then IS*Rl1* is exactly the kind of probe required for fingerprinting.

The drawback to the use of IS*Rl1* in this way is that the fraction of strains in which it is found, though large, is limited. A more thorough approach may be that taken by Simon *et al.* who isolated several IS elements from *R. meliloti* and used all of them to create a highly distinctive

fingerprint for a number of *R. meliloti* strains.[357]

All the *B. japonicum* strains screened with IS*R11* in a previous study (A. Ryan, M.Sc. thesis, DCU) were members of serogroup 123. The serological placing of all the strains screened in this study is not known, though it is likely that many are from serogroup 123, but at least one (USDA110) is from serogroup 110. The presence of IS*R11* in this strain shows that it is not restricted to just one group of strains, but is likely to be found in a diverse range of isolates. Though this is hardly surprising, considering the established diversity of serogroup 123.[340]

The homology of IS*R11* to RHBGLNB (page 119 above) raises the possibility that this sequence may have at least contributed to the hybridization patterns seen in this chapter (page 149), the percentage homology of the scoring segment is 72% but if the alignment of the full sequence is examined this falls to 20%. So, although the homology of the scoring segment is quite high, it is too short to form a stable duplex with IS*R11* under the conditions used, therefore it is unlikely that *glnB* can account for any of the bands in figure 22B.

The spread of IS*R11* in *R. leguminosarum* bv. *trifolii* seems more limited (figure 21, page 145). It is present in five out of twenty strains screened, and shows a different banding pattern in four of these strains. The strains S30-11 and S34-H9 have a single band at c. 7·6kb. The restriction banding pattern of these strains (lanes 16 and 19, figure 21, page 103) seem identical and it is likely that they are the same strain isolated twice.

If other IS elements were found in *R. leguminosarum* bv. *trifolii*, it is likely that they could be used for fingerprinting this organism in a similar fashion to the study by Simon *et al.*[357] already mentioned. To date, however, there has been no report of any IS element in *R. leguminosarum* bv. *trifolii*, and apart from the uncharacterized elements found by Simon *et al.*, IS*R11* is the only known IS element from *R. leguminosarum* as a whole.

When IS*R11* was first isolated it was found to hybridize to five out of five *R. leguminosarum* strains. In this study it was found to be present in one additional strain but not in another (figure 20, page 140). In strain DCU300,

IS*R11* hybridized to a single band at c. 11·3kb, this compares with a similar band found previously (A.Ryan, M.Sc. thesis, DCU) in another strain. That IS*R11* was not found in strain VF39 implies that the element is not ubiquitous in this species.

IS*R11* was found in five out of nine *R. meliloti* strains tested. Several of the strains tested seem to be related (see restriction banding patterns in figure 20C). For example strains 220-8, 220-15 and 220-16 all have very similar banding patterns and IS*R11* hybridized to only one of these strains (220-16). Several IS elements have been well characterized in *R. meliloti*[*] and as the bands in figure 20B are all faint, even though the sequence comparisons in chapter 4 (pages 114 - 119) show that IS*R11* has no significant homology to any of the characterised elements from *R. meliloti*, it would seem possible that IS*R11* has hybridized to other uncharacterized elements.

The study by Simon *et al.*[357] shows that *R. meliloti* can easily be fingerprinted through the use of a range of IS elements and a separate report describes the use of IS*Rm2* on its own to the same end. IS*R11* does not seem to show the same degree of variation as the elements used in these experiments.

The strain 65C is a spontaneous deletion mutant of a cryptic plasmid observed in strain 65 (M. O'Connell, personal communication). A faint band is observed at c. 9·3kb in 65 but not in 65C. The loss of this band in strain 65C must be connected to the loss of this plasmid. The simplest explanation is that the sequence which hybridized to IS*R11* is on the plasmid or, if this sequence is an IS element, it may have caused the loss of the plasmid by inducing deletion of an essential region, such as the origin of replication. The presence of a bright band in 65C (in both the IS*R11* and control blots) but not in 65 is more difficult to explain.

Of the bacterial species screened in figure 23 (page 111) the one most likely to harbour IS*R11* was *A. tumefaciens*, as this species is a member of the *Rhizobiaceae* and given the fact that IS*R11* was found to be so widespread in *B. japonicum* (*R. leguminosarum* is much more closely related

* References 101,357,344,395

to *A. tumefaciens* than to *B. japonicum*[6]) and also particularly in light of the demonstrated sequence homology between IS*Rl1* and three IS elements from *A. tumefaciens* (pages 116, 118 and 119). Homology between IS*Rm1* (from *R. meliloti*) and IS*426* (from *A. tumefaciens*) and between IS*66* (from *A. tumefaciens*) and *R. fredii* has been reported.[301,391]

A. Ryan did find homology between IS*Rl1* and *A. tumefaciens* 1060 (one band at c.9·6kb) but no control experiment with pSUP104-*rpsL* was done in that case. No hybridization to strain C58 was found in this study, perhaps screening a wider range of strains would reveal the presence of related sequences in this organism.

None of the bacterial species outside the *Rhizobiaceae* showed any homology to IS*Rl1*.

The faint bands across all lanes in figures 20 to 22 (band #1 and band #2) are probably due to hybridization of *rpsL* to traces of RNA in the blot. Similar bands were seen by a colleague who was probing total DNA samples, cut with different restriction enzymes, with rDNA sequences (F. Duffner, personal communication), so these bands cannot be due to DNA:DNA hybridization. Any RNA molecules in the agarose gels would be folded and would not be cut by *Eco*R I, so there would be little or no separation of these molecules and no inferences can be made as to the size of the bands.

The size markers were omitted from the probe for the control blots. Most of the positive bands from the various Southern blots can be recognized by a simple visual comparison with the control blots, but for an objective comparison, some bands are available for use as size markers. The 2·036, 1·636 and 0·517kb fragments hybridize to the pSVB30 vector and the sizes of the bands in the control lanes are known - 2·7kb and 3·2kb (and 5·9kb, see above page 142).

There is little danger of picking false bands because of the omission of the size markers, as these markers are used routinely in this manner in this laboratory and have been used in preliminary blots in this study. Hybridization of these fragments to total DNA has never been observed.

It is difficult to quantify the amount of sequence homology required to produce the positive bands seen in

figures 20B, 21B and 22B. An approximation can be calculated with the following formula[394]

$$T_m^{\infty} = 81{\cdot}5 + 16{\cdot}6\left[\log_{10}\left(\frac{M}{1+0{\cdot}7M}\right)\right] + 0{\cdot}41(\%GC) - \left(\frac{500}{D}\right) - P$$

Where T_m^{∞} = Melting temperature of polynucleotide duplex.
M = Concentration of monovalent cation.
%GC = The G+C content of the DNA.
D = The length of the duplex (bp).
P = The percentage mismatch.

The hybridization experiments were performed at 64°C, so any duplexes with a T_m below this temperature will not be stable. The G + C content of the *Rhizobiaceae* is in the range 57 - 65%,[197] and the concentration of Na^+ is known, (1·17M). Therefore, taking an average GC value of 61%, this formula reduces and can be rearranged to

$$P = 39{\cdot}32 - (500/D)$$

So this means that if all of IS*R11* (D = 2500bp) is hybridizing to a positive band, the maximum percentage mismatch is c.39%, (or the minimum homology is 61%). This percentage will not change significantly for smaller duplexes, until the value for D falls below 500bp. Therefore even if only part of the IS element (a transposase gene, for example) is producing the positive band the percentage homology is still significant.

It cannot safely be assumed that every positive band found in these experiments is an IS element as the Southern blot technique is sensitive enough to detect matches of only 50bp.[244] For such short duplexes, since T_m increases by 1°C for each percentage mismatch, the presence of even one mismatched basepair (T_m falls by 2°C) may prevent the duplex from forming.

So each of the positive bands represents a stretch of at least 50bp of homology to some part of IS*R11*, with up to 61% matches for the longer duplexes.

Another possibility is that IS*R11* may be a transposon (2·5kb is large for a simple insertion sequence) in which case it will carry a gene unconnected to transposition

functions and at least some of the positive bands may be due to hybridization to this gene.

Chapter 6. Detection of ISRL1 by Polymerase Chain Reaction.

Introduction.

As discussed above, there is a possibility that at least some of the bands in the Southern blots in chapter 5 may be due to homology between a gene carried by IS*Rl1* and a nontransposable gene in the genomes of the organisms probed. Also the results from the computer homology searches seem to show similarities between stretches of IS*Rl1* and DNA sequences where there is no evidence of an IS element. One way of determining whether IS*Rl1* is indeed present in these genomes is through the application of the polymerase chain reaction (PCR).

Primer Design and Reaction Conditions.

The inverted repeats of IS*Rl1* are ideal for use as primers in PCR experiments. They are identical, on opposite strands, a suitable distance apart, contain no significant secondary structure or unusual sequence and have no potential to form a duplex with each other.

One possible drawback to their use would be their size, 13bp, making it unlikely that the primers would anneal at the temperature usually used for PCR, (~ 55-65°C), though it should be possible to overcome this by choosing a low annealing temperature, T. This was chosen as 42°C, based on the following formula.[394]

$$T = 81{\cdot}5 + 16{\cdot}6[\log_{10}(M)] + (0{\cdot}41\times GC) - (0{\cdot}72\times F) - 650/L$$

Where M = Concentration of monovalent cation = 0·2M
GC = Percentage G+C content = 53·85%
F = Concentration of formamide = 0%
L = Length of primer = 13bp

Having decided on an annealing temperature the next danger is that the primer will not remain annealed as the reaction temperature rises to the extension temperature (72°C). To avoid this, the reaction was heated slowly at 2sec/°C. This is called a ramping time. The rest of the

procedure presents no obstacles and is fully detailed in chapter 2 (page 85).

To test the conditions chosen the reaction was performed on pSVB30-*rpsL*::IS*Rl1* and total DNA prepared from *R. leguminosarum* bv. *viciae* 897. Also, as a control, the reaction was performed on one sample without adding any DNA, this eliminates the possibility of any positive results being due to contaminating DNA in one of the reaction components. The results of this experiment can be seen in figure 24.

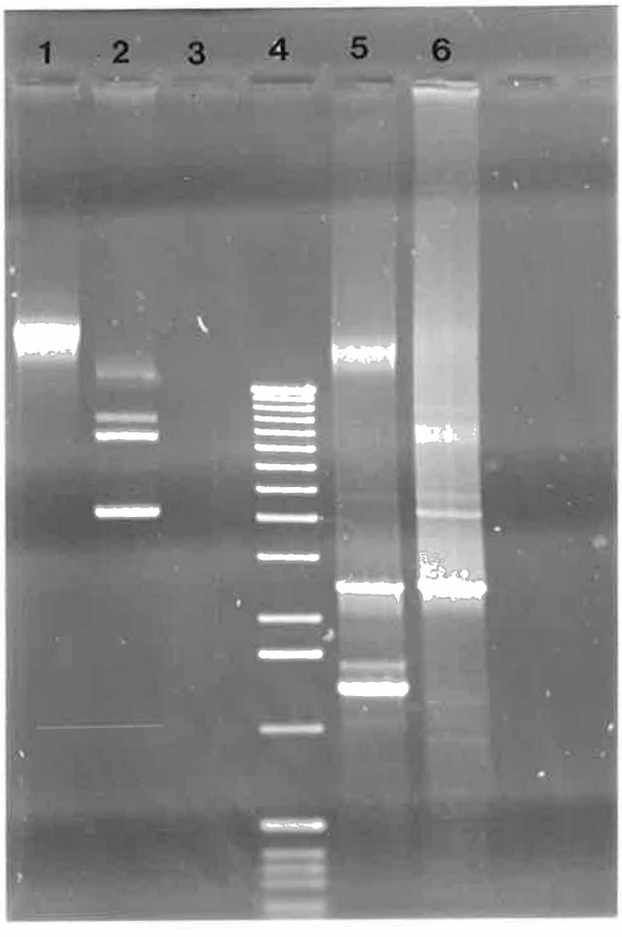

Figure 24. PCR on pSVB30-*rpsL*::IS*Rl1* and Total DNA from *R. leguminosarum* bv. *viciae*.

Lane 1. Uncut Total DNA from *R. leguminosarum* bv. *viciae* 897.

Lane 2. Uncut pSVB30-*rpsL*::IS*Rl1*.

Lane 3. PCR Control.

Lane 4. 1kb DNA Ladder.

Lane 5. PCR on Total DNA from *R. leguminosarum* bv. *viciae* 897.

Lane 6. PCR on pSVB30-*rpsL*::IS*Rl1*.

The two DNA samples tested were included on the gel diluted to the same concentration as was used in the reaction (i.e. diluted 1/50). This is to show that any bands present in lanes 5 and 6 were due to the PCR reaction and were not already present and, in fact, bands corresponding to those present in lanes 1 and 2 can be seen faintly in the upper parts of lanes 5 and 6. These bands are fainter because, having passed through 31 successive denaturation steps, a fraction of the duplexes have not reannealed.

Also a certain amount of smearing is evident in lanes 5 and 6. This is due to the high concentration of template in the reaction (~0·025μg/μl) which allows the Taq polymerase to proceed beyond the ends of IS*R11*.

Clearly, the PCR on pSVB30-*rpsL*::IS*R11* and *R. leguminosarum* bv. *viciae* 897 was a success with a clear, bright band present at c.2·5kb in lanes 5 and 6, and absent from the three control lanes. Unexpectedly, a second band is present in lane 5, at about 1·25kb. Having established the conditions for the reaction, similar experiments were performed on total DNA from other organisms.

Detection of IS*R11* in *B. japonicum*.

Figure 25 shows the results of the PCR on total DNA from *B. japonicum* strains. All the strains which were tested in the hybridization experiments to IS*R11*, reported in chapter 5, were tested, only those strains which gave positive results are shown. Included in the gel is total DNA from each strain, diluted to the same degree as the PCR sample (1/50) to show that the positive bands are not due to plasmids or fragments already present in the sample.

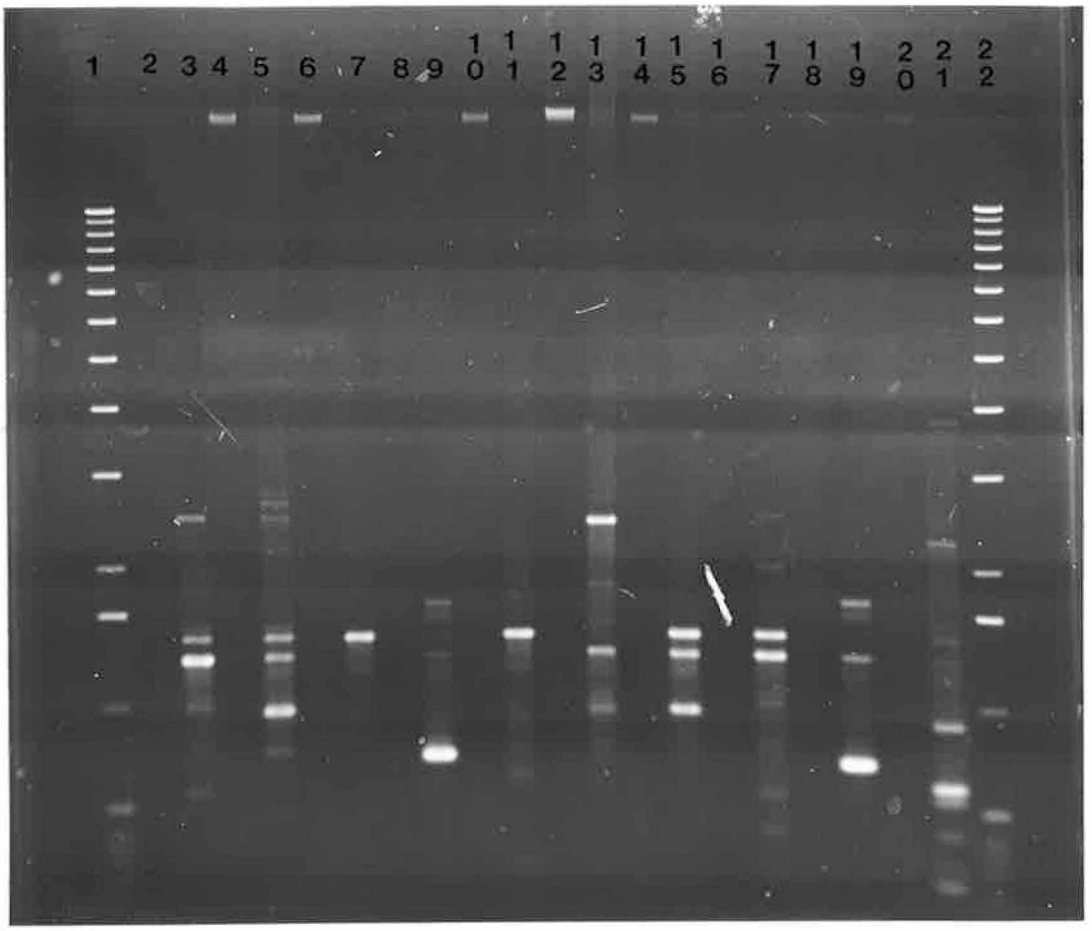

Figure 25. PCR on Total DNA from *B. japonicum* Strains.

Lane 1. 1kb DNA Ladder.
Lane 2. Diluted Total DNA from Strain 27-2.
Lane 3. PCR on " " " " "
Lane 4. Diluted Total DNA from Strain 40-2.
Lane 5. PCR on " " " " "
Lane 6. Diluted Total DNA from Strain NJ-2.
Lane 7. PCR on " " " " "
Lane 8. Diluted Total DNA from Strain 5-7.
Lane 9. PCR on " " " " "
Lane 10. Diluted Total DNA from Strain USDA123.
Lane 11. PCR on " " " " "
Lane 12. Diluted Total DNA from Strain AR6.
Lane 13. PCR on " " " " "
Lane 14. Diluted Total DNA from Strain AR14.
Lane 15. PCR on " " " " "
Lane 16. Diluted Total DNA from Strain AR34.
Lane 17. PCR on " " " " "
Lane 18. Diluted Total DNA from Strain U.
Lane 19. PCR on " " " " "
Lane 20. Diluted Total DNA from Strain SD.
Lane 21. PCR on " " " " "
Lane 22. 1kb DNA Ladder.

The sizes and mobilities of the size markers are presented in table 23, and from these the size of each band in the sample lanes was estimated, these are also presented in table 23. Only the bright bands are considered, the fainter ones are likely to be non-specific PCR products.

In the control lanes 4, 6, 10, 12 and 14 (strains 40-2, NJ-2, USDA 123, AR6 and AR14) a single band is clearly visible, near the top of the gel. These are probably total DNA which is running slowly (unrestricted and undenatured).

Table 23. Size Estimation for Bands in Figure 25.

Standards			Unknowns		
Size	Dist.	Dev.	Lane	Dist.	Est. Size
12·216	8·70	+0·259	3	18·70	2·627
11·198	9·00	-0·014		23·60	1·389
10·180	9·40	-0·150		25·45	1·063
9·162	10·00	-0·039	5	23·60	1·389
8·144	10·65	-0·037		25·45	1·063
7·126	11·45	-0·020		26·95	0·836
6·108	12·35	-0·090	7	23·60	1·389
5·090	13·65	-0·025	9	28·45	0·787
4·072	15·30	-0·002	11	23·60	1·389
3·054	17·70	+0·071	13	18·70	2·627
2·036	20·95	+0·062		25·45	1·063
1·636	22·50	+0·023		26·95	0·836
1·018	25·60	-0·022	15	23·60	1·389
0·506	29·30	-0·027		25·45	1·063
				26·95	0·836
			17	23·60	1·389
				25·45	1·036
			19	29·85	0·690
			21	26·40	0·921
				31·20	0·650

Detection of IS*R11* in *R. leguminosarum* bv. *trifolii*.

Figure 26 shows the results of the PCR on total DNA from *R. trifolii*. All the strains screened by Southern

hybridization were tested, only those strains which came up positive are shown, and a diluted sample of DNA from each is included on the gel.

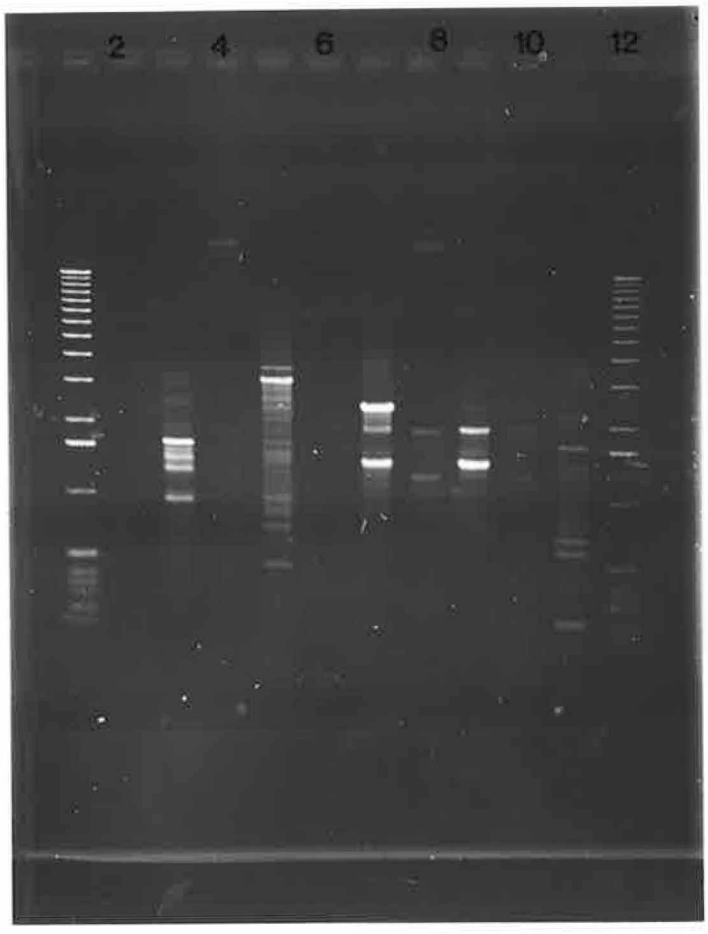

Figure 26. PCR on Total DNA from *R. leguminosarum* bv. *trifolii* Strains.

Lane 1. 1kb DNA Ladder.
Lane 2. Diluted Total DNA from Strain S12-14.
Lane 3. PCR on " " " " "
Lane 4. Diluted Total DNA from Strain S12-H3.
Lane 5. PCR on " " " " "
Lane 6. Diluted Total DNA from Strain S30-3.
Lane 7. PCR on " " " " "
Lane 8. Diluted Total DNA from Strain S30-11.
Lane 9. PCR on " " " " "
Lane 10. Diluted Total DNA from Strain S34-H9.
Lane 11. PCR on " " " " "
Lane 12. 1kb DNA Ladder.

Estimated sizes of bands are presented in table 24, along with the sizes and mobilities of the size markers which were used to calculate them.

Table 24. Size Estimation for Bands in Figure 26.

Standards			Unknowns		
Size	Dist.	Dev.	Lane	Dist.	Est. Size
12·216	8·25	+0·311	3	14·90	3·830
11·198	8·60	+0·223		15·50	3·534
10·180	8·85	-0·202		15·90	3·353
9·162	9·15	-0·576		17·10	2·872
8·144	9·50	-0·921	5	12·45	5·439
7·126	10·30	-0·653	7	13·55	4·621
6·108	12·50	+0·711		14·45	4·076
5·090	14·10	+0·815		15·75	3·419
4·072	15·05	+0·319	9	14·45	4·076
3·054	16·85	+0·089		15·75	3·419
2·036	19·30	-0·141	11	15·00	3·778
1·636	20·00	-0·359		18·70	2·347
				19·15	2·218
				21·90	1·568

Detection of IS*Rl1* in Other *Rhizobiaceae*.

Figure 27 shows the results of the PCR on total DNA from strains of *R. leguminosarum* bv. *viciae*, *R. meliloti* and *A. tumefaciens*. All the strains screened in chapter 5 were tested but only those which gave a positive result are shown.

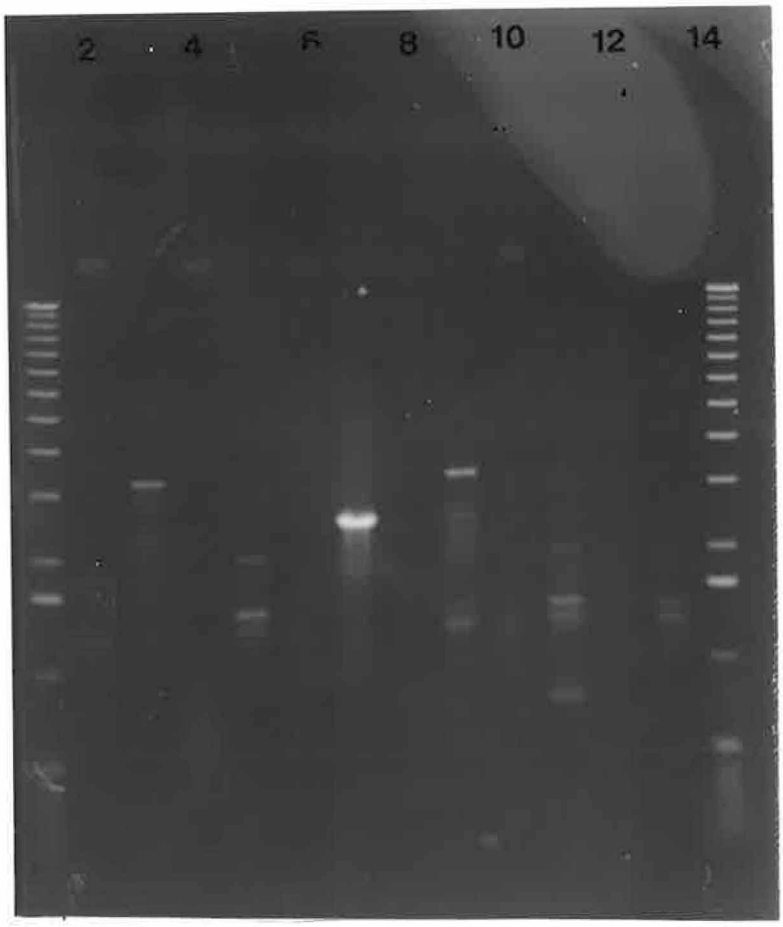

Figure 27. PCR on Total DNA from *Rhizobiaceae* Strains.

Lane 1. 1kb DNA Ladder.
Lane 2. Diluted Total DNA from Strain VF39.
Lane 3. PCR on " " " " "
Lane 4. Diluted Total DNA from Strain DCU300.
Lane 5. PCR on " " " " "
Lane 6. Diluted Total DNA from Strain 220-3.
Lane 7. PCR on " " " " "
Lane 8. Diluted Total DNA from Strain 220-7.
Lane 9. PCR on " " " " "
Lane 10. Diluted Total DNA from Strain 220-16.
Lane 11. PCR on " " " " "
Lane 12. Diluted Total DNA from Strain C58.
Lane 13. PCR on " " " " "
Lane 14. 1kb DNA Ladder.

Table 25 lists the mobilities and sizes of the PCR bands along with the sizes and mobilities of the standard fragments which were used to calculate them.

Table 25. Size Estimation for Bands in Figure 27.

Standards			Unknowns		
Size	Dist.	Dev.	Lane	Dist.	Est. Size
12·216	7·25	-0·310	3	11·75	3·274
11·198	7·55	+0·056	5	13·70	1·933
10·180	7·80	+0·017		14·75	1·429
9·162	8·10	+0·011		15·15	1·262
8·144	8·50	+0·125	7	12·80	2·463
7·126	8·95	+0·154	9	11·75	3·274
6·108	9·45	+0·092		15·15	1·262
5·090	10·05	+0·007	11	14·75	1·429
4·072	10·85	-0·027		15·15	1·262
3·054	11·85	-0·103		16·95	0·641
2·036	13·35	-0·091	13	14·75	1·429
1·636	14·25	-0·020		15·15	1·262
1·018	16·00	+0·073			

Discussion.

The smaller band found in the control experiment on *R. leguminosarum* bv. *viciae* 897 above (page 161) points to the existence of a second IS element, related to IS*RI1* in this strain. This smaller IS element (~1·25kb) may be a derivative of IS*RI1* which has lost internal sequences, or if IS*RI1* does encode accessory functions, this band may represent the 'original' element and IS*RI1* the derivative.

It is unlikely that any of the bands in the other photographs are IS*Rm2*, IS*66* or IS*866* as the inverted repeats of these elements are only homologous to the first 6 - 8bp of IS*RI1*, and while the PCR can tolerate mismatches in the 5′ ends of the primers, mismatches in the 3′ ends will prevent the reaction from proceeding.

Several strains which showed no hybridization to IS*RI1* in the Southern blot experiments, gave positive results in the PCR experiments (*R. leguminosarum* bv. *trifolii* S12-14, S12-H3, *R. leguminosarum* bv. *viciae* VF39, *R. meliloti* 65

and *A. tumefaciens* C58). Conversely, three strains which showed homology to IS*Rl1* through DNA:DNA hybridization yielded no band in the PCR experiments (*B. japonicum* USDA110, *R. leguminosarum* bv. *trifolii* SB-H9 and *R. meliloti* 220-20). The former case indicates the presence of insertion sequences with little internal homology to IS*Rl1* but terminating in the same inverted repeats. The latter case indicates the presence of sequences in the various genomes bearing homology to internal regions of IS*Rl1*, these sequences may not be IS elements (homology to accessory genes?). In no case did the PCR produce a band from DNA from a species outside the *Rhizobiaceae*.

No correlation can be made between the sizes of the bands found in the two experiments, as in the Southern work these bands are *Eco*R I fragments from the genomes of the various strains only part of which may be homologous to IS*Rl1*, while in the bands from the PCR work represent only those sequences bound by the inverted repeats of IS*Rl1*. This is well illustrated in the case of the source organism *R. leguminosarum* bv. *viciae* 897, in which 6 bands were found by DNA:DNA hybridization (9·8, 5·0, 3·1, 2·6, 2·2 and 1·9kb) while through the PCR only two bands were found (2·5 and 1·25kb). If it can be assumed that all these bands are due to IS*Rl1* or derivatives of it, then clearly this element is present in two forms in different locations throughout the genome.

A simple calculation shows that it is reasonable to assume that all the bands obtained through the PCR are, in fact, IS*Rl1* or closely related elements. The probability of finding a particular 13bp sequence in inverted repeats (no matter how far apart) is the same as the probability of the random occurrence of a 26bp sequence, $4^{26} = 4{\cdot}5\times10^{15}$. Bacterial genomes are typically in the range 10^6 to 10^7bp and so the possibility of these bands being anything other than IS*Rl1* or its relatives can be dismissed.

All of the *B. japonicum* strains, except 5-7 and SD show one or both of a pair of bands found at ~1·4kb and 1·0kb. Similar bands are found in *R. leguminosarum* bv. *trifolii* S12-14, S30-3, S30-11, *R. leguminosarum* bv. *viciae* DCU300, *R. meliloti* 220-7, 220-17 and *A. tumefaciens* C58. Smaller bands are present in *B. japonicum* 40-2, 5-7, AR14,

U, *R. leguminosarum* bv. *trifolii* S12-14 and *R. meliloti* 220-16.

Bands of 2·5kb are only found in *B. japonicum* 27-7, AR6, *R. trifolii* S30-3, and *R. meliloti* 220-3. Larger bands are found in *R. leguminosarum* bv. *trifolii* S30-3, *R. leguminosarum* bv. *viciae* VF39 and *R. meliloti* 220-7.

Thus the members of the IS*RI1* family show a good deal of diversity with regard to genomic location (which would be expected) and sequence length. The PCR would be an excellent tool for the study of these elements as it is highly specific and lends itself well to use in other techniques such as cloning and sequencing.

SEQMAN.BAS

The following is the listing of a program called SEQMAN which was written during the course of this project. The program is written in the computer language BASIC (VAX version) and is a crude method of sequence input, with error checking. The corrected sequence is written to a file called ONE.OUT for later use in various sequence analysis programs. A description of the program and its use is printed when the program is run. These instructions are stored in a file called TEXT.DAT and are listed here below the program. Also, another short program is listed called CONVERT.COM, which inserts the name of the sequence at the start of the file and slightly alters the format of the file.

```
 10 LET GH$=STRING$(80,291) \PRINT GH$
 20 FOR T=1 TO 5 \PRINT "#";TAB(79);"#" \NEXT T
 30 PRINT "#";TAB(32);"SEQUENCE MANAGER";TAB(79);"#"
 40 PRINT "#";TAB(32);"----------------";TAB(79);"#"
 50 PRINT "#";TAB(79);"#"
 60 PRINT "#";TAB(27);"DEVELOPED BY JOHN O'BRIEN";TAB(79);"#"
 70 PRINT "#";TAB(79);"#"
 80 PRINT "#";TAB(33);"BIOLOGY DEPT.";TAB(79);"#"
 90 PRINT "#";TAB(79);"#"
100 PRINT "#";TAB(28);"DUBLIN CITY UNIVERSITY";TAB(79);"#"
110 PRINT "#";TAB(79);"#"
120 PRINT "#";TAB(35);"05-07-1991";TAB(79);"#"
130 FOR T=1 TO 6 \PRINT "#";TAB(79);"#" \NEXT T
140 PRINT GH$
150 L$=INKEY$(0,WAIT 255) \PRINT
160 !
170 !                          INSTRUCTIONS
180 !
190 OPEN "TEXT.DAT" FOR INPUT AS FILE #7
200 FOR Y=1 TO 2
210 FOR T=1 TO 22
220 INPUT LINE #7;TT$ \PRINT TT$
230 NEXT T
240 PRINT "Hit any key when ready. "
250 Q$=INKEY$(0,WAIT 255) \PRINT \PRINT \PRINT
260 NEXT Y
270 CLOSE #7
280 !
290 !                   MOVING BETWEEN MODULES
300 !
310 PRINT
320 PRINT TAB(31);"----------------" \PRINT
330 PRINT TAB(31);"SEQUENCE MANAGER" \PRINT
340 PRINT TAB(31);"----------------" \PRINT
350 FOR T=1 TO 10
360 IF T=5 THEN PRINT TAB(22);"Make sure that
```

```
                                the CAPSLOCK is on."  ELSE PRINT
370 NEXT T
380 PRINT "INPUT Module (I)."
390 PRINT "COMPARE Module (C)."
400 PRINT "EDIT Module (E)."
410 PRINT "EXIT (X)."
420 PRINT \PRINT "Enter your choice. "
430 OPT$=INKEY$ (0, WAIT 255) \PRINT
440 IF OPT$="I" THEN GOTO 490
450 IF OPT$="C" THEN GOTO 730
460 IF OPT$="E" THEN GOTO 1520
470 IF OPT$="X" THEN GOTO 3150
480 GOTO 310
490 !
500 !                        INPUT MODULE
510 !
520 PRINT \PRINT TAB(18);"-------------- INPUT MODULE
                                        --------------" \PRINT
530 INPUT "Copy No. (1, 2 or 3) [0 to escape]";C$ \PRINT
540 IF C$="1" THEN OPEN "ONE.DAT" FOR OUTPUT AS #1 ELSE 560
550 GOTO 600
560 IF C$="2" THEN OPEN "TWO.DAT" FOR OUTPUT AS #1 ELSE 580
570 GOTO 600
580 IF C$="3" THEN OPEN "THR.DAT" FOR OUTPUT AS #1 ELSE 585
581 GOTO 590
585 IF C$="0" THEN GOTO 310 ELSE GOTO 520
590 LET O=0
600 PRINT "  Using capital letters, enter the sequence
                                       one base at a time. Do"
610 PRINT "not try to correct mistakes, corrections can
                                            be made later."
620 PRINT "  When finished hit 'Z'." \PRINT \PRINT
630 B$=INKEY$(0,WAIT 255)
640 IF INSTR(1,"ACGTZ",B$)<>0 THEN GOTO 650 ELSE GOTO 630
650 LET O=O + 1
660 IF O>60 THEN LET O=O - 60
670 PRINT B$; \PRINT #1;B$
680 IF O=60 THEN PRINT
690 IF B$<>"Z" THEN GOTO 630 ELSE O=0
700 PRINT \PRINT TAB(7);"------------ END OF INPUT MODULE
                                        ------------" \PRINT
710 CLOSE #1
720 GOTO 310
730 !
740 !                       COMPARE MODULE
750 !
760 PRINT \PRINT TAB(17);"-------------- COMPARE MODULE
                                          --------------"
770 PRINT \PRINT "Please wait." \PRINT
780 REM Set up parameters.
790 LET BIG=0
800 OPEN "ONE.DAT" FOR INPUT AS #1 \LET A=1 \GOSUB 1270
810 OPEN "TWO.DAT" FOR INPUT AS #2 \LET A=2 \GOSUB 1270
820 OPEN "THR.DAT" FOR INPUT AS #3 \LET A=3 \GOSUB 1270
830 RESTORE #1 \RESTORE #2 \RESTORE #3
840 REM Input of sequences from data files.
850 LET ARR=BIG + 3
860 DIM ONE$(ARR),TWO$(ARR),THR$(ARR)
870 FOR P=1 TO 3
880 ONE$(BIG+P)="Z" \TWO$(BIG+P)="Z" \THR$(BIG+P)="Z"
890 NEXT P
900 FOR Q=1 TO BIG
```

```
 910 IF ONE$(Q-1)="Z" THEN ONE$(Q)="Z" ELSE INPUT #1; ONE$(Q)
 920 IF TWO$(Q-1)="Z" THEN TWO$(Q)="Z" ELSE INPUT #2; TWO$(Q)
 930 IF THR$(Q-1)="Z" THEN THR$(Q)="Z" ELSE INPUT #3; THR$(Q)
 940 NEXT Q
 950 REM                          --SUBUNIT 1--
 960 REM Comparison of sequences and presentation of results.
 970 LET QW=(BIG/60)
 980 IF QW=INT(QW) THEN LET LNS=QW ELSE LET LNS=INT(QW)+1
 990 FOR W=1 TO LNS
1000 LET L=1+(60*(W-1))
1010 IF ONE$(L)="Z" AND TWO$(L)="Z" AND THR$(L)="Z" THEN 1230
1020 LET Y=(W/2) \LET Z=60*(W-1)
1030 LET R$=STR$(1+Z) \LET U$=STR$(60+Z)
1040 IF LEN(U$)=2 THEN LET U$=" "+U$
1050 PRINT R$;TAB(17);(15+Z);TAB(37);(30+Z);
                                    TAB(57);(45+Z);TAB(76);U$
1060 PRINT TAB(0);"|";TAB(18);"|";TAB(38);"|";
                                      TAB(58);"|";TAB(78);"|"
1070 FOR A=1 TO 3
1080 FOR Q=3 TO 60 STEP 3
1090 LET N=Q+(60*(W-1))
       \REM N is the position or number of the nucleotide.
1100 IF N>BIG+1 THEN GOTO 1140
1110 IF A=1 THEN PRINT ONE$(N-2);ONE$(N-1);ONE$(N);" ";
1120 IF A=2 THEN PRINT TWO$(N-2);TWO$(N-1);TWO$(N);" ";
1130 IF A=3 THEN PRINT THR$(N-2);THR$(N-1);THR$(N);" ";
1140 NEXT Q
1150 PRINT
1160 IF A<>3 THEN GOSUB 1370
1170 NEXT A
1180 IF FLG=0 THEN GOTO 1210
1190 PRINT \PRINT "First mismatch at position no.";FLG \PRINT
1200 GOTO 1220
1210 PRINT \PRINT "No mismatches." \PRINT
1220 IF Y<>INT(Y) THEN GOTO 1250
1230 PRINT \PRINT "Hit any key for the next page."
1240 WT$=INKEY$(0,WAIT 255)
1250 NEXT W
1260 CLOSE #1,#2,#3 \GOTO 310
1270 REM                          --SUBUNIT 2--
1280 REM To find the size of the longest sequence.
1290 LET SZE=0
1300 LET SZE=SZE + 1
1310 IF A=1 THEN INPUT #1;Q$
1320 IF A=2 THEN INPUT #2;Q$
1330 IF A=3 THEN INPUT #3;Q$
1340 IF Q$<>"Z" THEN GOTO 1300
1350 IF SZE>BIG THEN LET BIG=SZE
1360 RETURN
1370 REM                          --SUBUNIT 3--
1380 REM To flag mismatches.
1390 LET FLG=0
1400 FOR Q=1 TO 60
1410 LET N=Q+(60*(W-1))
1420 IF N>BIG THEN GOTO 1490
1430 IF ONE$(N)<>TWO$(N) OR ONE$(N)<>THR$(N) THEN 1440
                                                    ELSE 1490
1440 LET K=INT(Q/3)
1450 IF K(Q/3) THEN LET K=K+1
1460 LET J=Q+K-2
1470 IF FLG=0 THEN LET FLG=N
1480 PRINT TAB(J);"*";
```

```
1490 NEXT Q
1500 PRINT
1510 RETURN
1520 !
1530 !                    EDIT MODULE
1540 !
1550 REM                         --SET UP--
1560 PRINT
1570 PRINT TAB(19);"------------- EDIT MODULE ------------"
1580 PRINT \PRINT "Please wait." \PRINT
1590 OPEN "ONE.DAT" FOR INPUT AS #1
1600 OPEN "TWO.DAT" FOR INPUT AS #2
1610 OPEN "THR.DAT" FOR INPUT AS #3
1620 LET BIG=0 \LET PSN=0
1630 FOR A=1 TO 3 \GOSUB 1270 \NEXT A
1640 LET GAP=BIG + 50
1650 RESTORE #1 \RESTORE #2 \RESTORE #3
1660 DIM ONE$(GAP),TWO$(GAP),THR$(GAP)
1670 FOR B=1 TO BIG
1680 IF ONE$(B-1)="Z" THEN ONE$(B)="Z" ELSE INPUT #1;ONE$(B)
1690 IF TWO$(B-1)="Z" THEN TWO$(B)="Z" ELSE INPUT #2;TWO$(B)
1700 IF THR$(B-1)="Z" THEN THR$(B)="Z" ELSE INPUT #3;THR$(B)
1710 IF ONE$(B)=TWO$(B) AND TWO$(B)=THR$(B) THEN GOTO 1730
1720 IF PSN=0 THEN LET PSN=B
1730 NEXT B
1740 REM                         --MAIN PROGRAM--
1750 IF PSN<>0 THEN GOTO 1790
1760 INPUT "There are no more mismatches.
                          Do you wish to continue (Y/N)";Q$
1770 IF Q$<>"Y" THEN GOTO 3010
1780 INPUT "Enter the position of the error
                                   to be corrected";PSN
1790 IF PSN23 THEN LET CNTR=23
1800 IF PSN=>23 AND PSN<=(BIG-22) THEN LET CNTR=PSN
1810 IF PSN>(BIG-22) THEN LET CNTR=(BIG-22)
1820 PRINT \GOSUB 2740
1830 IF PSN=0 THEN LET K$="" ELSE LET K$="First error
                                     at position "+STR$(PSN)
1840 PRINT \PRINT K$ \PRINT
1850 PRINT "Copy (C), delete (D), insert (I),"
1860 PRINT "alter (A), swap (S) or exit (E)."
1870 PRINT "What action is to be taken ?"
1880 LET CHO$=INKEY$(0,WAIT 255) \PRINT
1890 IF CHO$="Q" THEN GOTO 1780
1900 IF CHO$<>"E" THEN GOTO 1910 ELSE GOTO 3010
1910 IF CHO$<>"C" THEN GOTO 1920 ELSE GOSUB 2030 \GOTO 1960
1920 IF CHO$<>"D" THEN GOTO 1930 ELSE GOSUB 2140 \GOTO 1960
1930 IF CHO$<>"I" THEN GOTO 1940 ELSE GOSUB 2270 \GOTO 1960
1940 IF CHO$<>"A" THEN GOTO 1950 ELSE GOSUB 2460 \GOTO 1960
1950 IF CHO$<>"S" THEN GOTO 1850 ELSE GOSUB 2590
1960 LET J=0
1970 FOR L=1 TO GAP
1980 IF ONE$(L)=TWO$(L) AND TWO$(L)=THR$(L) THEN GOTO 2000
1990 IF J=0 THEN LET J=L
2000 NEXT L
2010 LET PSN=J
2020 GOTO 1740
2030 REM                         --SUBUNIT 1--
2040 REM To correct one copy by reading from another.
2050 PRINT "Copy mode.";TAB(50);"To escape, enter '0'."
2060 PRINT \INPUT "Which copy is correct ";COP$ \PRINT
2070 IF COP$="0" THEN GOTO 2130
```

```
2080 IF INSTR(1,"123",COP$)<>0 THEN 2090 ELSE 2060
2090 IF COP$="1" THEN LET NEW$=ONE$(PSN)
2100 IF COP$="2" THEN LET NEW$=TWO$(PSN)
2110 IF COP$="3" THEN LET NEW$=THR$(PSN)
2120 ONE$(PSN)=NEW$ \TWO$(PSN)=NEW$ \THR$(PSN)=NEW$
2130 RETURN
2140 REM                          --SUBUNIT 2--
2150 REM To delete bases.
2160 PRINT "Delete mode.";TAB(50);"To escape, enter '0'."
                                                    \PRINT
2170 PRINT "Deleting the base at position no. ";STR$(PSN);
                                        ", from which copy";
2180 INPUT COP$ \PRINT
2190 IF COP$="0" THEN GOTO 2260
2200 IF INSTR(1,"123",COP$)<>0 THEN 2210 ELSE 2170
2210 FOR DLT=PSN TO (BIG-1)
2220 IF COP$="1" THEN LET ONE$(DLT)=ONE$(DLT+1)
2230 IF COP$="2" THEN LET TWO$(DLT)=TWO$(DLT+1)
2240 IF COP$="3" THEN LET THR$(DLT)=THR$(DLT+1)
2250 NEXT DLT
2260 RETURN
2270 REM                          --SUBUNIT 3--
2280 REM To insert bases.
2290 PRINT "Insert mode.";TAB(50);"To escape, enter '0'."
                                                    \PRINT
2300 PRINT "Insert a base at position no. ";STR$(PSN);
                                        ", into which copy";
2310 INPUT COP$
2320 IF COP$="0" THEN GOTO 2450
2330 IF INSTR(1,"123",COP$)<>0 THEN GOTO 2340 ELSE GOTO
                                                        2300
2340 INPUT "Enter the new base ";NEW$
2350 IF INSTR(1,"ACGT",NEW$)<>0 THEN 2360 ELSE 2340
2360 LET BIG=BIG+1
2370 FOR NSRT=(BIG+1) TO (PSN+1) STEP -1
2380 IF COP$="1" THEN LET ONE$(NSRT)=ONE$(NSRT-1)
2390 IF COP$="2" THEN LET TWO$(NSRT)=TWO$(NSRT-1)
2400 IF COP$="3" THEN LET THR$(NSRT)=THR$(NSRT-1)
2410 NEXT NSRT
2420 IF COP$="1" THEN LET ONE$(PSN)=NEW$
2430 IF COP$="2" THEN LET TWO$(PSN)=NEW$
2440 IF COP$="3" THEN LET THR$(PSN)=NEW$
2450 RETURN
2460 REM                          --SUBUNIT 4--
2470 REM To change bases.
2480 PRINT "Alter mode.";TAB(50);"To escape, enter '0'."
                                                    \PRINT
2490 PRINT "Change the base at position no. ";STR$(PSN);
                                          ", in which copy";
2500 INPUT COP$
2510 IF COP$="0" THEN GOTO 2580
2520 IF INSTR(1,"123",COP$)<>0 THEN 2530 ELSE 2490
2530 INPUT "Enter the replacement base ";NEW$
2540 IF INSTR(1,"ACGT",NEW$)<>0 THEN 2550 ELSE 2530
2550 IF COP$="1" THEN LET ONE$(PSN)=NEW$
2560 IF COP$="2" THEN LET TWO$(PSN)=NEW$
2570 IF COP$="3" THEN LET THR$(PSN)=NEW$
2580 RETURN
2590 REM                          --SUBUNIT 5--
2600 REM To swap bases around.
2610 PRINT "Swap mode.";TAB(50);"To escape, enter '0'."
                                    \PRINT \PRINT "Swap ";
```

```
2620 PRINT "the bases at positions";PSN;"and ";STR$(PSN+1);
                                          ", in which copy";
2630 INPUT COP$ \PRINT
2640 IF COP$="0" THEN GOTO 2730
2650 IF COP$<>"1" THEN GOTO 2680
2660 NEW$=ONE$(PSN) \ONE$(PSN)=ONE$(PSN+1)
                                          \ONE$(PSN+1)=NEW$
2670 GOTO 2730
2680 IF COP$<>"2" THEN GOTO 2720
2690 NEW$=TWO$(PSN)  \TWO$(PSN)=TWO$(PSN+1)
                                          \TWO$(PSN+1)=NEW$
2700 GOTO 2730
2710 IF COP$<>"3" THEN GOTO 2620
2720 NEW$=THR$(PSN) \THR$(PSN)=THR$(PSN+1)
                                          \THR$(PSN+1)=NEW$
2730 RETURN
2740 REM                          --SUBUNIT 6--
2750 REM To print the relevant fragment from all 3 copies.
2760 LET RT=INT((CNTR+1)/3) \IF RT=((CNTR+1)/3) THEN 2790
2770 LET RT=INT(CNTR/3)
2780 IF RT=(CNTR/3) THEN CNTR=CNTR-1 ELSE CNTR=CNTR+1
2790 FOR L=1 TO 10 \PRINT \NEXT L
2800 PRINT STR$(CNTR-22);TAB(20);(CNTR-5);
                        TAB(40);(CNTR+10);TAB(56);(CNTR+22)
2810 PRINT "|";TAB(22);"|";TAB(42);"|";TAB(58);"|"
2820 FOR A=1 TO 3
2830 FOR S=(CNTR-20) TO (CNTR+22) STEP 3
2840 IF A=1 THEN PRINT ONE$(S-2);ONE$(S-1);ONE$(S);" ";
2850 IF A=2 THEN PRINT TWO$(S-2);TWO$(S-1);TWO$(S);" ";
2860 IF A=3 THEN PRINT THR$(S-2);THR$(S-1);THR$(S);" ";
2870 NEXT S
2880 PRINT TAB(68);"Copy no.";A
2890 IF A=3 THEN GOTO 2970
2900 FOR S=(CNTR-22) TO (CNTR+22)
2910 LET T=(S-(CNTR-22)) \LET R=T+1
2920 LET G=INT(R/3) \IF G(R/3) THEN LET G=G+1
2930 LET O=T+G-1
2940 IF ONE$(S)<>TWO$(S) OR ONE$(S)<>THR$(S)
                                       THEN PRINT TAB(O);"*";
2950 NEXT S
2960 PRINT
2970 NEXT A
2980 PRINT TAB(10);"|";TAB(30);"|";TAB(50);"|"
2990 PRINT TAB(8);(CNTR-14);TAB(28);(CNTR+1);
                                          TAB(48);(CNTR+16)
3000 RETURN
3010 REM                          --END OF MODULE--
3020 PRINT \PRINT "Please wait" \PRINT
3030 REM To write new sequences to files.
3040 CLOSE #1,#2,#3
3050 OPEN "ONE.DAT" FOR OUTPUT AS FILE #1
3060 OPEN "TWO.DAT" FOR OUTPUT AS FILE #2
3070 OPEN "THR.DAT" FOR OUTPUT AS FILE #3
3080 FOR CNT=1 TO BIG
3090 IF ONE$(CNT-1)="Z" THEN 3100 ELSE PRINT #1;ONE$(CNT)
3100 IF TWO$(CNT-1)="Z" THEN 3110 ELSE PRINT #2;TWO$(CNT)
3110 IF THR$(CNT-1)="Z" THEN 3120 ELSE PRINT #3;THR$(CNT)
3120 NEXT CNT
3130 CLOSE #1,#2,#3
3140 GOTO 310
3150 REM                          --END OF PROGRAM--
3160 FOR T=1 TO 10 \PRINT \NEXT T
```

```
3170 PRINT TAB(20);"**************************************"
3180 PRINT
3190 PRINT TAB(20);"Don't forget to rename the data files."
3200 PRINT
3210 PRINT TAB(20);"**************************************"
3220 FOR T=1 TO 10 \PRINT \NEXT T
3230 END
```

TEXT.DAT

This package is used to help ensure the error free reading of sequencing autoradiograms. It works as follows....

First the user reads and enters the sequence three times. The computer then compares and reprints these sequences, with an asterisk flagging each position where any of the sequences differs from either of the others. Then the user has the option of editing these errors until one or all of the sequences is a fully correct transcription of the data on the autoradiogram.

The program is made up of three modules called "INPUT", "COMPARE" and "EDIT". Unsurprisingly, these modules input, compare and edit the data. The user can access any of these modules at any stage once there are three copies of the sequence present for the computer to work with.

The "INPUT" module must be run three times for the program to work, but the user can overwrite any of these entries at any stage. When you are finished with one sequence be sure to rename the data files containing the correct copy of the sequence, otherwise they will be overwritten when you start on another. The data files are called "ONE.DAT", "TWO.DAT" and "THR.DAT". You can find out how to rename them by typing "HELP RENAME" when you have exited from the program.

The "COMPARE" module prints the three sequences side by side in 60bp lengths. Markers are printed showing the position of each length and all discrepancies are flagged with an asterisk. The position of the first error in each length is given. No input is required from the user.

The "EDIT" module is made up of several subunits. The various editing operations are performed on a 45bp fragment of the sequence which the user defines by telling the computer the approximate position of the midpoint. The three copies of this fragment are then printed (in much the same way as the "COMPARE" module works), with all the errors flagged. There are five different edit modes. Any base in any sequence can be deleted or changed to any other base, a new base can be inserted at any point, any pair of adjacent bases on one sequence can be swapped around or an error on one sequence can be corrected by copying from another. The names of the subunits (or edit modes) that perform these functions are "DELETE "ALTER", "INSERT", "SWAP" and "COPY" respectively.

CONVERT.BAS

```
 10 REM To convert data files from type handled by sequence
 20 REM manager to type handled by NCB programs.
 30 INPUT "Enter the name of the sequence";NME$
```

```
 40 LET NME$ = ">" + NME$
 50 OPEN "ONE.DAT" FOR INPUT AS FILE #1
 60 OPEN "ONE.CNV" FOR OUTPUT AS FILE #2
 70 PRINT #2;NME$
 80 INPUT #1;BBASE$
 90 IF BBASE$ = "Z" THEN GOTO 120
100 PRINT #2;BBASE$;
110 GOTO 80
120 CLOSE #1,#2
130 PRINT "Rename ONE.CNV."
140 END
```

FRAME.BAS

This program searches DNA sequences (from a file called "FRAME.SEQ") for translational frameshifting motifs. These motifs take the form XXXYYYZ and when the program finds one it reports the position of the first base and the sequence of the motif, both on the screen and to a file called "FRAME.OUT". The program was written in GW-BASIC and runs under MSDOS.

```
10 CLEAR :OPEN "I",1,"FRAME.SEQ" :OPEN "O",2,"FRAME.OUT"
                                              :DIM BASE$(2500)
20 IF EOF(1) THEN CLOSE #1 :GOTO 70
30 LINE INPUT #1, ROW$
40 IF LEFT$(ROW$,1)=">" GOTO 20
50 FOR A=1 TO LEN(ROW$) :B=B+1 :BASE$(B) = MID$(ROW$,A,1)
                                                    :NEXT A
60 GOTO 20
70 FOR A=1 TO 2495
80 IF BASE$(A)=BASE$(A+1) AND BASE$(A)=BASE$(A+2) GOTO 90
                                          ELSE GOTO 140
90 IF BASE$(A+3)=BASE$(A+4) AND BASE$(A+3)=BASE$(A+5)
                                       GOTO 100 ELSE GOTO 140
100 FOR B=0 TO 6 :B$=B$+BASE$(A+B) :NEXT B
110 PRINT "Possible site at";A;TAB(30);"Sequence =";B$
120 PRINT #2, A;B$
130 B$=""
140 NEXT A
150 CLOSE #2 :END
```

APPENDIX B. R. LEGUMINOSARUM CODING SEQUENCES.

These are part of the GENBANK entries for the sequences which were used to build a codon bias table for *R. leguminosarum*. All of these sequences code for protein and the information listed includes the accession number, the coordinates of the coding regions and a brief description of each.

```
Organism:  Rhizobium leguminosarum
Genes:     90
Codons:    22647

LOCUS       S114388     1456 bp    DNA  BCT  05-FEB-1993
DEFINITION glnT=glutamine synthetase III [Rhizobium legum-
           inosarum, bv. phaseoli, Genomic, 1456 nt]
ACCESSION   S48357 M88676
FEATURES             Location/Qualifiers
     CDS             106. .1413
                     /gene="glnT"

LOCUS       RHMFIXA     1045 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum nitrogen fixation prot-
           ein (fixA) gene, complete cds.
ACCESSION   L11081
FEATURES             Location/Qualifiers
     CDS             161. .1000
                     /gene="fixA"

LOCUS       RHMFIXCX    600 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum nitrogen fixation prot-
           ein (fixC) gene, 3' end; nitrogen fixation
           protein (fixX) gene, complete cds.
ACCESSION   L11082
FEATURES             Location/Qualifiers
     CDS             <1. .279
                     /gene="fixC"
     CDS             292. .591
                     /gene="fixX"

LOCUS       RHMNIFAX   2898 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum transcriptional
           activator (nifA) gene, complete cds; positive
           regulator (nifB2) gene, 5' end.
ACCESSION   L11084
FEATURES             Location/Qualifiers
     CDS             416. .2167
                     /gene="nifA"
     CDS             2380. .>2898
                     /gene="nifB2"

LOCUS       RHMNIFB2X  1746 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum positive regulator
           (nifB2) gene, 5' end.
ACCESSION   L11083
FEATURES             Location/Qualifiers
     CDS             1180. .>1746
                     /gene="nifB2"
```

```
LOCUS       RHMNIFH1X   461 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum Fe protein (component
           II) of nitrogenase (nifH1) gene, 5' end.
ACCESSION  L11085
FEATURES                Location/Qualifiers
     CDS               201. .>461
                       /gene="nifH1"

LOCUS       RHMNIFH2X   461 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum Fe protein (component
           II) of nitrogenase (nifH2) gene, 5' end.
ACCESSION  L11086
FEATURES                Location/Qualifiers
     CDS               201. .>461
                       /gene="nifH2"

LOCUS       RHMNIFH3X   461 bp ds-DNA  BCT  19-FEB-1993
DEFINITION Rhizobium leguminosarum Fe protein (component
           II) of nitrogenase (nifH3) gene, 5' end.
ACCESSION  L11087
FEATURES                Location/Qualifiers
     CDS               201. .>461
                       /gene="nifH3"

LOCUS       P4DNODA    1296 bp ds-DNA  BCT  21-MAY-1991
DEFINITION Plasmid p42d (from Rhizobium leguminosarum,
           strain phaseoli) nodulation protein (nodA) gene,
           complete cds.
ACCESSION  M58625
FEATURES                Location/Qualifiers
     CDS               290. .877
                       /gene="nodA"

LOCUS       P4DNODBC   3260 bp ds-DNA  BCT  05-MAR-1991
DEFINITION Plasmid p42d (from Rhizobium leguminosarum,
           strain phaseoli) nodulation protein (nodB and
           nodC) genes, complete cds.
ACCESSION  M58626
FEATURES                Location/Qualifiers
     CDS               549. .722
                       /note="ORF1"
     CDS               732. .1007
                       /note="ORF2"
     CDS               1004. .1693
                       /gene="nodB"
     CDS               1798. .3084
                       /gene="nodC"

LOCUS       RHMMODD    1438 bp ds-DNA  BCT  15-SEP-1990
DEFINITION R.leguminosarum nodD gene, complete cds.
ACCESSION  J03671
FEATURES                Location/Qualifiers
     CDS               complement(<1. .145)
                       /note="NodA protein"
     CDS               406. .1374
                       /note="NodD protein"

LOCUS       RHMOMPIIIA 1658 bp ds-DNA  BCT  18-MAR-1992
DEFINITION R.leguminosarum outer membrane protein IIIA
           (OMPIIIA) gene, complete cds.
ACCESSION  M69214
```

```
FEATURES              Location/Qualifiers
     CDS             407. .1507
                     /gene="OMPIIIA"

LOCUS       RHMPSS4A   1468 bp ds-DNA  BCT  07-MAY-1992
DEFINITION Rhizobium leguminosarum pss4 gene encoding a
           protein involved in polysaccharide production
ACCESSION  M93042
FEATURES              Location/Qualifiers
     CDS             471. .1262
                     /gene="pss4"

LOCUS       RHMRLV     4184 bp ds-DNA  BCT  18-AUG-1992
DEFINITION Rhizobium leguminosarum nodulation-associated
           protein (rhiA, rhiB, rhi,C rhiR) genes, complete
           cds.
ACCESSION  M98835
FEATURES              Location/Qualifiers
     CDS             186. .881
                     /gene="rhiA"
     CDS             931. .1590
                     /gene="rhiB"
     CDS             1679. .2128
                     /gene="rhiC"
     CDS             complement(2347. .3090)
                     /gene="rhiR"

LOCUS       RLDCTA     5820 bp    DNA  BCT  23-SEP-1992
DEFINITION R.leguminosarum dctA gene encoding C4-
           dicarboxylate permease.
ACCESSION  Z11529
FEATURES              Location/Qualifiers
     CDS             complement(874. .2208)
                     /gene="dctA"
     CDS             2433. .4301
                     /gene="dctB"
     CDS             4306. .5652
                     /gene="dctD"

LOCUS       RLFIXW     1360 bp    DNA  BCT  23-JUL-1991
DEFINITION R.leguminosarum fixW gene
ACCESSION  X16521
FEATURES              Location/Qualifiers
     CDS             164. .379
                     /note="ORF 71;  (AA 1-71)"
     CDS             325. .564
                     /note="ORF 79;  (AA 1-79)"
     CDS             585. .1250
                     /note="fixW;  (AA 1-218)"
     CDS             1307. .>1360
                     /note="fixA;  (AA 1-18)"

LOCUS       RLGLNA     2217 bp    DNA  BCT  06-JUL-1989
DEFINITION Rhizobium leguminosarum glnA gene for glutamine
           synthetase I (GSI) (EC 6.3.1.2)
ACCESSION  X04880
FEATURES              Location/Qualifiers
     CDS             251. .586
                     /note="ORF 111;  (AA 1-111)"
     CDS             665. .2074
                     /note="GSI;  (AA 1-469)"
```

```
LOCUS       RLGLNII    1617 bp    DNA  BCT  15-OCT-1992
DEFINITION R.leguminosarum gln II gene for glutamine synth-
           etase II
ACCESSION  X67296
FEATURES                 Location/Qualifiers
     CDS                 569. .1549
                         /gene="glnII"

LOCUS       RLHUPGENE 14781 bp    DNA  BCT  21-NOV-1992
DEFINITION Rhizobium leguminosarum hydrogenase stuctural
           genes hupA to hupO
ACCESSION  X52974
FEATURES                 Location/Qualifiers
     CDS                 156. .1238
                         /gene="hupS"
     CDS                 1262. .3052
                         /note="hupL"
     CDS                 3066. .3785
                         /gene="hupC"
     CDS                 3786. .4394
                         /gene="hupD"
     CDS                 4406. .4981
                         /gene="hupE"
     CDS                 5078. .5374
                         /gene="hupF"
     CDS                 5543. .5992
                         /gene="hupG"
     CDS                 5989. .6837
                         /gene="hupH"
     CDS                 6834. .7046
                         /gene="hupI"
     CDS                 7036. .7560
                         /gene="hupJ"
     CDS                 7557. .8669
                         /gene="hupK"
     CDS                 8662. .9003
                         /gene="hypA"
     CDS                 9003. .9902
                         /gene="hypB"
     CDS                 9906. .12185
                         /gene="hypF"
     CDS                 12176. .12403
                         /gene="hypC"
     CDS                 12400. .13557
                         /gene="hypD"
     CDS                 13554. .14606
                         /gene="hypE"

LOCUS       RLNIFA     3301 bp    DNA  BCT  06-JUL-1989
DEFINITION Rhizobium leguminosarum plasmid pRL6JI nifA
           gene, URF and flanking regions (fixC gene / nifB
           gene)
ACCESSION  X05049
FEATURES                 Location/Qualifiers
     CDS                 <1. .492
                         /note="fixC gene product;  (163 AA)"
     CDS                 505. .801
                         /note="ORF;  (AA 1-98) CDS"
     CDS                 989. .2548
                         /note="nifA gene product;  (AA1-519)"
     CDS                 2816. .>3301
                         /note="nifB gene product;  (162 AA)"
```

```
LOCUS       RLNIFAB    1980 bp    DNA  BCT  08-JUN-1990
DEFINITION Rhizobium leguminosarum biovar trifolii nifA
           gene and nifB gene 5' region
ACCESSION  X16311
FEATURES             Location/Qualifiers
     CDS             225. .1286
                     /note="nifA protein;  (AA 1-353)"
     CDS             1522. .>1980
                     /note="N-terminal nifB protein;
                                    (AA 1-153)"

LOCUS       RLNIFPRE   2076 bp    DNA  BCT  25-SEP-1990
DEFINITION Rhizobium leguminosarum fixX (partial), nifA,
           and nifB (partial) genes
ACCESSION  X17073
FEATURES             Location/Qualifiers
     CDS             <1. .35
                     /note="fixX protein;  (10 AA) (1 is 2nd
base in codon)"
     CDS             1987. .>2076
                     /note="nifB protein;  (AA 1-30)"

LOCUS       RLNOD     12055 bp    DNA  BCT  06-JUL-1989
DEFINITION Rhizobium leguminosarum plasmid pRL1JI nodul-
           ation genes
ACCESSION  Y00548 M13658 X01650 X03785
FEATURES             Location/Qualifiers
     CDS             complement(143. .922)
                     /note="nodJ"
     CDS             complement(926. .1861)
                     /note="nodI"
     CDS             complement(1998. .3272)
                     /note="nodC"
     CDS             complement(3295. .3945)
                     /note="nodB"
     CDS             complement(3942. .4532)
                     /note="nodA"
     CDS             4792. .5703
                     /note="nodD"
     CDS             6647. .7858
                     /note="nodE"
     CDS             8452. .9024
                     /note="nodL"
     CDS             9594. .11420
                     /note="nodM"
     CDS             11498. .11983
                     /note="nodN"

LOCUS       RLNODABC   2627 bp    DNA  BCT  06-JUL-1989
DEFINITION Rhizobium leguminosarum nodulation genes nodAB
           and C for induction of root hair curling
ACCESSION  X01650
FEATURES             Location/Qualifiers
     CDS             157. .630
                     /note="put. nodA protein"
     CDS             653. .1303
                     /note="put. nodB protein"
     CDS             1326. .2600
                     /note="put. nodC protein"

LOCUS       RLNODD1    2162 bp    DNA  BCT  14-DEC-1992
```

```
DEFINITION R.leguminosarum nodD1 and nolE genes
ACCESSION  X54214
FEATURES             Location/Qualifiers
     CDS             506. .838
                     /gene="nolE"
     CDS             1005. .1949
                     /gene="nodD1"

LOCUS      RLNODD2    3016 bp    DNA  BCT  14-DEC-1992
DEFINITION R.leguminosarum nodD2, nodD3 and nolP genes
ACCESSION  X54215
FEATURES             Location/Qualifiers
     CDS             complement(14. .301)
                     /gene="nolP"
     CDS             677. .1633
                     /gene="nodD2"
     CDS             complement(2042. .2950)
                     /gene="nodD3"

LOCUS      RLNODIJT   2668 bp    DNA  BCT  21-JUN-1990
DEFINITION Rhizobium leguminosarum nodI, nodJ and nodT
           genes involved in nodulation
ACCESSION  X51411
FEATURES             Location/Qualifiers
     CDS             <1. .221
                     /note="nodI protein (73 AA) (1 is 2nd
                      base in codon)"
     CDS             218. .1006
                     /note="nodJ protein;  (AA 1-262)"
     CDS             1131. .2534
                     /note="nodT protein;  (AA 1-467)"

LOCUS      RLNODX     2152 bp    DNA  BCT  06-OCT-1988
DEFINITION Rhizobium leguminosarum sym plasmid pRL5JI nodX
           gene
ACCESSION  X07990
FEATURES             Location/Qualifiers
     CDS             <1. .132
                     /note="nodJ;  (43 AA) C-term"
     CDS             332. .1435
                     /note="nodX protein;  (AA 1-367)"

LOCUS      RLPNOD     4327 bp    DNA  BCT  25-APR-1990
DEFINITION Rhizobium leguminosarum biovar viciae nodO and
           nodT genes
ACCESSION  X17285
FEATURES             Location/Qualifiers
     CDS             101. .1549
                     /note="nodT;  (AA 1-482)"
     CDS             complement(3225. .4079)
                     /note="nodO;  (AA 1-284)"

LOCUS      RLPSS12    1664 bp    DNA  BCT  02-FEB-1989
DEFINITION R. leguminosarum pss gene required for exopoly-
           saccharide synthesis and nodulation of peas.
ACCESSION  X12568
FEATURES             Location/Qualifiers
     CDS             113. .289
                     /note="ORF1;  (AA -1 to 57)"

LOCUS      RLRECA     1841 bp    DNA  BCT  23-JUL-1991
DEFINITION R.leguminosarum recA gene and alaS gene 5'end
```

```
            for the RecA protein and alanyl-tRNA synthetase,
            respectively
ACCESSION   X59956
FEATURES                Location/Qualifiers
     CDS                32. .1087
                        /gene="recA"
     CDS                1239. .>1841
                        /gene="alaS"

LOCUS       S108191     2067 bp    DNA  BCT  10-NOV-1992
DEFINITION dctA1=NGRDCTA1, dctA2=NGRDCTA2 [Rhizobium legum-
            inosarum, NGR234, symbiotic megaplasmid pSym,
            Plasmid, 3 genes, 2067 nt]
ACCESSION   S38912 J03708
FEATURES                Location/Qualifiers
     CDS                complement(<1. .230)
                        /note="Description: orf 5' of dctA1"
     CDS                449. .1819
                        /gene="dctA1"
     CDS                470. .1819
                        /gene="dctA2"

LOCUS       RHMNIFHAA   524 bp ds-DNA  BCT  15-SEP-1990
DEFINITION R.leguminosarum nitrogen fixation protein nifH
            gene, 5' end.
ACCESSION   M36435
FEATURES                Location/Qualifiers
     CDS                435. .>524
                        /note="nifH protein"
```

APPENDIX C. REFERENCES.

1. **Adhya, S.L. and Shapiro, J.A.** (1969). The Galactose Operon on *E. coli* K-12. I. Structural and Pleiotropic Mutations of the Operon. *Genetics* **62**:231-247.
2. **Ahmed, A., Bidwell, K. and Musso, R.** (1980). Internal Rearrangements of IS2 in *Escherichia coli. Cold Spring Harbor Symposia on Quantitative Biology* **45**:141-151.
3. **Akhverdyan, V.Z., Khrenova, E.A., Bogush, V.G., Gerasimova, T.V. Kirsanov, N.B. and Krylov, V.N.** (1984). Frequent Occurrence of Transposable Phages in Natural Populations of *Pseudomonas aeruginosa*. *Genetika* **20**:1612-1619.
4. **Allet, B.** (1979). Phage Mu Insertion Duplicates a 5 Basepair Sequence at the Host Inserted Site. *Cell* **16**:123-129.
5. **Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.** (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**:403-410.
6. **Appelbaum, E.** (1990). The *Rhizobium/Bradyrhizobium*-Legume Symbiosis, p131-158. *In* Grasshoff, P.M. (ed.), *The Molecular Biology of Symbiotic Nitrogen Fixation*, CRC Press Inc., Boca Raton.
7. **Anderson, R.P. and Roth, J.R.** (1977). Tandem Genetic Duplications in Phage and Bacteria. *Annual Review of Microbiology* **31**:473-505.
8. **Arber, W., Hümbelin, M., Caspers, P., Reif, H.-J., Iida, S. and Meyer, J.** (1980). Spontaneous Mutations in *Escherichia coli* Prophage P1 and IS-mediated Processes. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:38-40.
9. **Arnold, W. and Pühler, A.** (1988). A Family of High-copy-number Plasmid Vectors with Single End Label Sites for Rapid Nucleotide Sequencing. *Gene* **70**:171-179.
10. **Arthur, A. and Sherratt, D.** (1979). Dissection of the Transposition Process: A Transposon-encoded Site-specific Recombination System. *Molecular and General Genetics* **175**:267-274.
11. **Auerswald, E.A., Ludwig, G and Schaller, H.** (1980). Structural Analysis of Tn*5*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:107-114.

12. **Barker, R.F., Idler, K.B., Thomson, D.V. and Kemp, J.D.** (1983). Nucleotide Sequence Analysis of the T-DNA Region from the *Agrobacterium tumefaciens* Octopine Ti Plasmid pTi15955. *Plant Molecular Biology* **2**:335-350.

13. **Barran, L.R., Bromfield, E.S.P., Rastogi, V., Whitwill, S.T. and Wheatcroft, R.** (1991). Transposition and Copy Number of Insertion Sequence IS*Rm1* Are Not Correlated with Symbiotic Performance of *Rhizobium meliloti* from Two Field Sites. *Canadian Journal of Microbiology* **37**:576-579.

14. **Beck, C.F.** (1979). A Genetic Approach to Analysis of Transposons. *Proceedings of the National Academy of Sciences USA* **76**:2376-2380.

15. **Bennett, P.M., Grinsted, J. and Richmond, H.** (1977). Transposition of Tn*A* Does Not Generate Deletions. *Molecular and General Genetics* **154**:205-212.

16. **Berg, D.E.** (1977). Insertion and Excision of the Transposable Kanamycin Resistance Determinant Tn*5*, p205-212. *In* Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. (ed.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

17. **Berg, D.E.** (1980). Control of Gene Expression by a Mobile Recombinational Switch. *Proceedings of the National Academy of Sciences USA* **77**:4880-4884.

18. **Berg, D.E.** (1983). Structural Requirement for IS*50*-mediated Gene Transposition. *Proceedings of the National Academy of Sciences USA* **80**:792-796.

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

20. **Berg, D.E.** (1989). Transposon Tn*5*, p185-210. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington D.C.

21. **Berg, D.E., Davies, J., Allet, B and Rochaix, J.D.** (1975). Transposition of R Factor Genes to Bacteriophage Lambda. *Proceedings of the National Academy of Sciences USA* **72**:3628-3632.

22. **Berg, D.E., Egner, C., Hirschel, B.J., Howard, J., Jorgensen, R.A. and Tlsty, T.D.** (1980). Insertion, Excision and Inversion of Tn*5*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:115-124.

23. **Berg, D.E., Egner, C. and Lowe, J.B.** (1983). Mechanism of F Factor-enhanced Excision of Transposon Tn*5*. *Gene* **22:**1-7.

24. **Berg, D.E., Sasakawa, C., Hirschel, B.J., Johnsrud, L., McDivitt, L., Egner, C. and Ramabhadran, R.** (1981). Transposition of the Inverted Repeats of Tn*5*, p380-391. *In* Levy, S.B. (ed.), *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids*, Plenum, N.Y.

25. **Berg, D.E., Schmandt, M.A. and Lowe, J.B.** (1983). Specificity of Transposon Tn*5* Insertion. *Genetics* **105:**813-828.

26. **Beringer, J.E.** (1974). R Factor Transfer in *Rhizobium leguminosarum*. *Journal of General Microbiology* **84:**188-198.

27. **Bernardi, A. and Bernardi, F.** (1981). Site-specific Deletions in the Recombinant Plasmid pSC101 Containing the *redB-ori* Region of Phage Lambda. *Gene* **13:**103-109.

28. **Bernardi, A. and Bernardi, F.** (1981). Complete Sequence of an IS Element Present in pSC101. *Nucleic Acids Research* 9:2905-2911.

29. **Bernardi, F. and Bernardi, A.** (1986). Intramolecular Transposition of IS*102*. *Gene* **42:**11-19.

30. **Besemer, J. and Herpers, M.** (1977). Suppression of Polarity of Insertion Mutations within the *gal* Operon of *E. coli*. *Molecular and General Genetics* **151:**295-304.

31. **Biek, D. and Roth, J.R.** (1980). Regulation of Tn*5* Transposition in *Salmonella typhimurium*. *Proceedings of the National Academy of Sciences USA* **77:**6047-6051.

32. **Biek, D. and Roth, J.R.** (1980). Regulation of Tn*5* Transposition. *Cold Spring Harbor Symposia on Quantitative Biology* **45:**189-191.

33. **Biel, S.W., Adelt, G. and Berg, D.E.** (1984). Transcriptional Control of IS*1* Transposition in *Escherichia coli*. *Journal of Molecular Biology* **174:**251-264.

34. **Biel, S.W. and Berg, D.E.** (1984). Mechanism of IS*1* Transposition in *E. coli*: Choice between Simple Insertion and Cointegration. *Genetics* **108:**319-330.

35. **Binns, A.N., Sciaky, D. and Wood, H.N.** (1982). Variation in Hormone Autonomy and Regenerative Potential of Cells Transformed by Strain A66 of *Agrobacterium tumefaciens*. *Cell* **31:**605-612.

36. **Birnboim, H.C. and Doly, J.** (1979). A Rapid Alkaline Extraction Procedure for Screening Recombinant Plasmid DNA. *Nucleic Acids Research* **7**:1513-1523.

37. **Bonnard, G., Tinland, B., Paulus, F., Szegedi, E. and Otten, L.** (1989). Nucleotide Sequence, Evolutionary Origin and Biological Role of a Rearranged Cytokinin Gene Isolated from a Wide Host Range Biotype III *Agrobacterium* Strain. *Molecular and General Genetics* **216**:428-438

38. **Bonnard, G., Vincent, F. and Otten, L.** (1989). Sequence and Distribution of IS*866*, a Novel T Region-associated Insertion Sequence from *Agrobacterium tumefaciens*. *Plasmid* **22**:70-81.

39. **Bossi, L. and Ciampi, M.S.** (1981). DNA Sequences at the Sites of Three Insertions of the Transposable Element Tn*5* in the Histidine Operon of *Salmonella*. *Molecular and General Genetics* **183**:406-408.

40. **Botstein, D., and Kleckner, N.** (1977). Translocation and Illegitimate Recombination by the Tetracycline Resistance Element Tn*10*, p185-204. *In* Bukhari, A. I., Shapiro, J. A. and Adhya, S. L. (ed.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

41. **Brachet, P., Eisen, H. and Rambach, A.** (1970). Mutations of Coliphage Lambda Affecting the Expression of Replicative Functions O and P. *Molecular and General Genetics* **108**:266-276.

42. **Braedt, G.** (1988). Different Reading Frames Are Responsible for IS*1*-dependent Deletions and Recombination. *Genetics* **118**:561-570.

43. **Brewster, J.M. and Morgan, E.A.** (1981). Tn*9* and IS*1* Inserts in a Ribosomal Ribonucleic Acid Operon of *Escherichia coli* Are Incompletely Polar. *Journal of Bacteriology* **148**:897-903.

44. **Brown, A.M.C., Coupland, G.M. and Willets, N.S.** (1984). Characterization of IS*46*, an Insertion Sequence Found on Two IncN Plasmids. *Journal of Bacteriology* **159**:472-481.

45. **Bukhari, A.I.** (1975). Reversal of Mutator Phage Mu Integration. *Journal of Molecular Biology* **96**:87-99.

46. **Bukhari, A.I. and Metlay, M.** (1973). Genetic Mapping

of Prophage Mu. *Virology* **54**:109-116.

47. **Burkardt, H.J., Priefer, U., Pühler, A., Rieß, G. and Spitzbarth, P.** (1979). Naturally Occurring Insertion Mutants of Broad Host Range Plasmids RP4 and R68, p387-398. *In* Timmis, K.N. and Pühler, A. (ed.), *Plasmids of Medical, Environmental and Commercial Importance*, Elsevier/North-Holland Biomedical Press, Amsterdam.

48. **Brzoska, P.M. and Signer, E.R.** (1991). *lpsZ*, a Lipopolysaccharide Gene Involved in Symbiois of *Rhizobium meliloti*. *Journal of Bacteriology* **173**:3235-3237.

49. **Calos, M.P., Johnsrud, L. and Miller, J.H.** (1978). DNA Sequence at the Integration Sites of the Insertion Element IS*1*. *Cell* **13**:411-418.

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

51. **Calos, M.P. and Miller, J.H.** (1980). Molecular Conseqeuences of Deletion Formation Mediated by the Transposon Tn*9*. *Nature* **285**:38-41.

52. **Campbell, A., Berg, D.E., Botstein, D., Lederberg, E.M., Novick, R.P., Starlinger, P. and Szybalski, W.** (1979). Nomenclature of Transposable Elements in Prokaryotes. *Gene* **5**:197-206.

53. **Carlson, T.A., Guerinot, M.L. and Chelm, B.K.** (1985). Characterizaiton of the Gene Encoding Glutamine Synthetase I (*glnA*) from *Bradyrhizobium japonicum*. *Journal of Bacteriology* **162**:698-703.

54. **Caspers, P., Dalrymple, B., Iida, S. and Arber, W.** (1984). IS*30*, a New Insertion Sequence of *Escherichia coli* K12. *Molecular and General Genetics* **196**:68-73.

55. **Chakrabarti, S.K., Halder, A.K. and Chakrabartty, P.K.** (1990). Relatedness among *Rhizobium* Species. *Journal of General and Applied Microbiology* **36**:47-54.

56. **Chandler, M., Allet, B., Gallay, E., Boy de la Tour, E. and Caro, L.** (1977). Involvement of IS*1* in the Dissociation of the R-determinant and RTF Components of the Plasmid R100.1. *Molecular and General Genetics* **153**:289-295.

57. **Chandler, M., Boy de la Tour, E., Willems, D. and Caro, L.** (1979). Some Properties of the Chloramphenicol Resistance Transposon Tn*9*. *Molecular and General Genetics* **176**:221-231.

58. **Chandler, M., Clerget, M. and Caro, L.** (1980). IS*1*-promoted Events Associated with Drug-resistance Plasmids. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:157-165.

59. **Chandler, M. and Fayet, O.** (1993). Translational Frameshifting in the Control of Transposition in Bacteria. *Molecular Microbiolgy* **7**:497-503.

60. **Chandler, M. and Galas, B.J.** (1983). IS*1*-mediated Tandem Duplication of Plasmid pBR322. Dependence on RecA and on DNA Polymerase I. *Journal of Molecular Biology* **165**:183-190.

61. **Chandler, M. and Galas, D.J.** (1983). Cointegrate Formation Mediated by Tn*9*. II. Activity of IS*1* Is Modulated by External DNA Sequences. *Journal of Molecular Biology* **170**:61-91.

62. **Chandler, M. and Galas, D.J.** (1985). Studies on the Transposition of IS*1*, p53-77. *In* Helinski, D.R., Cohen, S.N., Clewell, D.B., Jackson, D.A. and Hollaender, A. (ed.), *Plasmids in Bacteria*. Plenum Publishing Corporation, N.Y.

63. **Charlebois, R.L. and Doolittle, W.F.** (1989). Transposable Elements and Genome Structure in Halobacteria, p297-307. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington D.C.

64. **Clements, M.B. and Syvanen, M.** (1980). Isolation of a *polA* Mutation That Affects Transposition of Insertion Sequences and Transposons. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:201-204.

65. **Clerget, M., Chandler, M. and Caro, L.** (1980). Isolation of an IS*1* Flanked Kanamycin Resistance Transposon from Rldrd19. *Molecular and General Genetics* **180**:123-127.

66. **Clewell, D.B.** (1981). Conjugation and Resistance Transfer in Streptococci and Other Gram Positive Species: Plasmids, Sex Pheromones and Conjugative Transposons, p199-206. *In* Levy, S.B. (ed.), *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids*, Plenum, N.Y.

67. **Cohen, S.N. and Shapiro, J.A.** (1990). Transposable Genetic Elements. *Scientific American* **242**:40-49.

68. **Cole, M.A. and Elkan, G.H.** (1973). Transmissible Re-

sistance to Penicillin G, Neomycin and Chloramphenicol in *Rhizobium japonicum*. *Antimicrobial Agents and Chemotherapy* **14**:248-253.

69. **Collins, J.** (1980). Instability of Palindromic DNA in *Escherichia coli*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:409-416.

70. **Colonna, B., Nicoletti, M., Visca, P., Casalino, M., Valenti, P. and Maimone, F.** (1985). Composite IS*1* Elements Encoding Hydroxamate-mediated Iron Uptake in F1*me* Plasmids from Epidemic *Salmonella* spp. *Journal of Bacteriology* **162**:307-316.

71. **Cornelis, G.** (1980). Transposition of Tn*951* (Tn*lac*) and Cointegrate Formation Are Thermosensitive Processes. *Journal of General Microbiology* **117**:243-247.

72. **Cornelis, G., Ghosal, D. and Saedler, H.** (1978). Tn*951*: A New Transposon Carrying a Lactose Operon. *Molecular and General Genetics* **160**:215-24.

73. **Cornelis, G. and Saedler, H.** (1980). Deletions and Inversions Induced by a Resident IS*1* of the Lactose Transposon Tn*951*. *Molecular and General Genetics* **178**:367-374.

74. **Courvalin, P. and Carlier, C.** (1986). Transposable Multiple Antibiotic Resistance in *Streptococcus pneumoniae*. *Molecular and General Genetics* **205**:291-297.

75. **Couturier, M.** (1976). The Integration and Excision of Bacteriophage Mu-1. *Cell* **7**:155-163.

76. **Craig, N.L.** (1989). Transposon Tn7, p211-226. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington D.C.

77. **Craigie, R. and Mizuuchi, K.** (1985). Mechanism of Transposition of Bacteriophage Mu: Structure of a Transposition Intermediate. *Cell* **41**:867-876.

78. **Craigie, R., Mizuuchi, M. and Mizuuchi, K.** (1984). Site-specific Recognition of the Bacteriophage Mu Ends by the Mu A Protein. *Cell* **39**:387-394.

79. **Cullum, J. and Broda, P.** (1979). Chromosome Transfer and Hfr Formation by F in *rec*$^+$ and *recA* Strains of *Escherichia coli* K12. *Plasmid* **2**:358-365.

80. **Dalrymple, B. and Arber, W.** (1986). The Characterization of Terminators of RNA Transcription on IS*30* and an Analysis of Their Role in IS Element-mediated Polar-

ity. *Gene* **44**:1-10.

81. **Daniell, E., Roberts, R. and Abelson, J.** (1972). Mutations in the Lactose Operon Caused by Bacteriophage Mu. *Journal of Molecular Biology* **69**:1-8.

82. **Das, A., Court, D., Gottesman, M. and Adhya, S.** (1977). Polarity of Insertion Mutations Is Caused by Rho-mediated Termination of Transcription, p93-97. *In* Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. (ed.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

83. **Datta, N. and Hughes, V.M.** (1983). Plasmids of the Same Inc Groups in Enterobacteria before and after the Medical Use of Antibiotics. *Nature* **306**:616-617.

84. **Datta, A.R. and Rosner, J.L.** (1987). Reduced Transposition in *rho* Mutants of *Escherichia coli* K-12. *Journal of Bacteriology* **169**:888-890.

85. **Davidson, N., Deonier, R.C., Hu, S. and Ohtsubo, E.** (1975). Electron Microscopic Heteroduplex Studies of Sequence Relations among Plasmids of *Escherichia coli*. X. Deoxyribonucleic Acid Sequence Organization of F and of F-primes, and the Sequences Involved in Hfr Formation, p56-65. *In* Schlessinger, D. (ed.), *Microbiology - 1974*, American Society for Microbiology, Washington, D.C.

86. **Dean, D.** (1981). A Plasmid Cloning Vector for the Direct Selection of Strains Carrying Recombinant Plasmids. *Gene* **15**:99-102.

87. **De Bruijn F.J. and Bukhari, A.I.** (1978). Analysis of Transposable Elements Inserted in the Genomes of Bacteriophages Mu and P1. *Gene* **3**:315-331.

88. **de Combrugghe, B., Adhya, S., Gottesman, M. and Pastan, I.** (1973). Effect of Rho on Transcription of Bacterial Operons. *Nature New Biology* **241**:260-264.

89. **De Meirsman, C., Croes, C., Desair, J., Verreth, C., Van Gool, A. and Vanderleyden, J.** (1989). Identification of Insertion Sequence Element IS*427* in pTiT37 Plasmid DNA of an *Agrobacterium tumefaciens* T37 Isolate. *Plasmid* **21**:129-137.

90. **De Meirsman, C., Desair, J., Vanderleyden, J., van Gool, A.P. and Jen, G.C.** (1987). Similarities between Nucleotide Sequences of Insertion Elements of *Agrobact-*

erium tumefaciens and *Pseudomonas savastanoi* in Relation to *Agrobacterium tumefaciens* T_C-DNA. *Nucleic Acids Research* **15**:10591.

91. **De Meirsman, C., Van Soom, C., Verreth, C., Van Gool, A. and Vanderleyden, J.** (1990). Nucleotide Sequence Analysis of IS*427* and Its Target Sites in *Agrobacterium tumefaciens* T37. *Plasmid* **24**:227-234.
92. **Demerec, M., Adelberg, E.A., Clark, A.J. and Hartman, P.E.** (1966). A Proposal for Uniform Nomenclature in Bacterial Genetics. *Genetics* **54**:61-76.
93. **Deonier, R.C.** (1987). Locations of Native Insertion Sequence Elements, p982-989. *In* Neidhart, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, American Society for Microbiology, Washington, D.C.
94. **Deonier, R.C., Hadley, R.G. and Hu, M.** (1979). Enumeration and Identification of IS*3* Elements in *Escherichia coli* Strains. *Journal of Bacteriology* **137**:1421-1424.
95. **Derbyshire, K.M., Hwang, L. and Grindley, N.D.F.** (1987). Genetic Analysis of the Insertion Sequence IS*903* Transposase with Its Terminal Inverted Repeats. *Proceedings of the National Academy of Sciences USA* **84**:8049-8053.
96. **Diaz-Aroca, E., de la Cruz, F., Zabala, J.C. and Ortiz, J.M.** (1984). Characterization of the New Insertion Sequence IS*91* from an Alpha-hemolysin Plasmid of *Escherichia coli*. *Molecular and General Genetics* **193**:493-499.
97. **Diaz-Aroca, E., Mendiola, M.V., Zabala, J.C. and de la Cruz, F.** (1987). Transposition of IS*91* Does Not Generate a Target Duplication. *Journal of Bacteriology* **169**:442-443.
98. **Dickerson, R.E.** (1983). Base Sequence and Helix Structure Variation in B and A DNA. *Journal of Molecular Biology* **166**:419-441.
99. **DiNardo, S., Voelkel, K.A., Sternglanz, R., Reynolds, A.E. and Wright, A.** (1982). *Escherichia coli* DNA Topoisomerase I Mutants Have Compensatory Mutations in DNA Gyrase Genes. *Cell* **31**:43-51.
100. **Diver, W.P., Grinsted, J., Fritzinger, D.C., Brown,**

N.L., Altenbuchner, J., Rogowsky, P. and Schmitt, R. (1983). DNA Sequences of and Complementation by the *tnpA* Genes of Tn*21*, Tn*501* and Tn*1721*. *Molecular and General Genetics* **191**:189-193.

101. Dusha, I., Kovalenko, S., Banfalvi, Z. and Kondorosi, A. (1987). *Rhizobium meliloti* Insertion Element IS*Rm2* and Its Use for Identification of the *fixX* Gene. *Journal of Bacteriology* **169**:1403-1409.

102. Eckhardt, T. (1978). A Rapid Method for the Identification of Plasmid Deoxyribonucleic Acid in Bacteria. *Plasmid* **1**:584-588.

103. Edlund, T. and Normark, S. (1981). Recombination between Short DNA Homologies Causes Tandem Duplication. *Nature* **292**:269-271.

104. Egner, C. and Berg, D.E. (1981). Excision of Transposon Tn*5* Is Dependent on the Inverted Repeats But Not on the Transposase Function of Tn*5*. *Proceedings of the National Academy of Sciences USA* **78**:459-463.

105. Engler, J.A. and van Bree, M.P. (1981). The Nucleotide Sequence and Protein-coding Capacity of the Transposable Element IS*5*. *Gene* **14**:155-163.

106. Faelen, M. and Toussaint, A. (1978). Stimulation of Deletions in the *Escherichia coli* Chromosome by Partially Induced Muc*ts*62 Prophages. *Journal of Bacteriology* **136**:477-483.

107. Faelen, M. and Toussaint, A. (1980). Chromosomal Rearrangements Induced by Temperate Bacteriophage D108. *Journal of Bacteriology* **143**:1029-1030.

108. Faelen, M., Toussaint, A. and de Lafonteyne, J. (1975). Model for the Enhancement of λ-*gal* Integration into Partially Induced Mu-1 Lysogens. *Journal of Bacteriology* **121**:873-882.

109. Farabaugh, P.J., Schmeissner, U., Hofer, M. and Miller, J.H. (1978). Genetic Studies of the *lac* repressor VII. On the Molecular Nature of Spontaneous Hotspots in the *lacI* Gene of *E. coli*. *Journal of Molecular Biology* **126**:847-863.

110. Fennewald, M.A. and Shapiro, J.A. (1979). Transposition of Tn7 in *Pseudomonas aeruginosa* and Isolation of *alk*-Tn7 Mutations. *Journal of Bacteriology* **139**:264-269.

111. Fiandt, M., Szybalski, W. and Malamy, M.H. (1972).

Polar Mutations in *lac*, *gal*, and phage lambda consist of a few IS-DNA sequences inserted in either orientation. *Molecular and General Genetics* **119**:223-231.

112. Flores, M., González, V., Brom, S., Martínez, E., Piñero, D., Romero, D., Dávila, G. and Palacios, R. (1987). Reiterated DNA Sequences in *Rhizobium* and *Agrobacterium* spp. *Journal of Bacteriology* **169**:5782-5788.

113. Foster, T.J. (1977). Insertion of the Tetracycline Resistance Translocation Unit Tn*10* in the *lac* Operon of *Escherichia coli* K12. *Molecular and General Genetics* **154**:305-309.

114. Foster, T.J., Davis, M.A., Roberts, D.E., Takeshita, K. and Kleckner, N. (1981). Genetic Organization of Transposon Tn*10*. *Cell* **23**:201-213.

115. Foster, T.J., Howe, T.G.B. and Richmond, K.M.V. (1975). Translocation of the Tetracycline Resistance Determinant from R100-1 to the *Escherichia coli* K-12 Chromosome. *Journal of Bacteriology* **124**:1153-1158.

116. Foster, T.J., Lundblad, V., Hanley-Way, S., Halling, S.M. and Kleckner, N. (1981). Three Tn*10*-associated Excision Events: Relationship to Transposition and Role of Direct and Inverted Repeats. *Cell* **23**:215-227.

117. Franke, A.E. and Clewell, D.B. (1981). Evidence for a Chromosome-borne Resistance Transposon (Tn*916*) in *Streptococcus faecalis* That Is Capable of "Conjugal" Transfer in the Absence of a Conjugative Plasmid. *Journal of Bacteriology* **145**:494-502.

118. Franke, A.E. and Clewell, D.B. (1980). Evidence for Conjugal Transfer of a *Streptococcus faecalis* Transposon (Tn*916*) from a Chromosomal Site in the Absence of Plasmid DNA. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:77-80.

119. Franklin, N.C. and Luria, S.E. (1961). Transduction by Bacteriophage P1 and the Properties of *lac* Genetic Region in *E. coli* and *S. dysenteriae*. *Virology* **15**:299-311.

120. Fuller, R.S., Funnel, B.E. and Kornberg, A. (1984). The dnaA Protein Complex with the *E. coli* Chromosomal Replication Origin (*oriC*) and other DNA Sites. *Cell* **38**:889-900.

121. Galas, D.J., Calos, M.P. and Miller, J.H. (1980). Sequence Analysis of Tn*9* Insertions in the *lacZ* Gene.

Journal of Molecular Biology **144**:19-41.

122. Galas, D.J. and Chandler, M. (1981). On the Molecular Mechanisms of Transposition. *Proceedings of the National Academy of Sciences USA* **78**:4858-4862.

123. Galas, D.J. and Chandler, M. (1982). Structure and Stability of Tn*9*-mediated Cointegrates. Evidence for Two Pathways of Transposition. *Journal of Molecular Biology* **154**:245-272.

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

125. Gamas, P., Chandler, M.G., Prentki, P. and Galas, D.J. (1987). *Escherichia coli* Integration Host Factor Binds Specifically to the Ends of the Insertion Sequence IS*1* and to Its Major Insertion Hot-spot in pBR322. *Journal of Molecular Biology* **195**:261-272.

126. Gamas, P., Galas, D. and Chandler. M. (1985). DNA Sequence at the End of IS*1* Required for Transposition. *Nature* **317**:458-460.

127. Georgopolous, C., McKittrick, N., Herrick, G. and Eisen, H. (1982). An IS4 Transposition Causes a 13-bp Duplication of Phage Lambda DNA and Results in the Constitutive Expression of the *cI* and *cro* Gene Products. *Gene* **20**:83-90.

128. Ghosal, D., Sommer, H. and Saedler, H. (1979). Nucleotide Sequence of the Transposable DNA-element IS*2*. *Nucleic Acids Research* **6**:1111-1122.

129. Gill, R., Heffron, F., Dougan, G. and Falkow, S. (1978). Analysis of Sequences Transposed by Complementation of Two Classes of Transposition-deficient Mutants of Tn*3*. *Journal of Bacteriology* **136**:742-756.

130. Glansdorff, N., Charlier, D. and Zafarullah, M. (1980). Activation of Gene Expression by IS*2* and IS*3*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:153-156.

131. Giphart-Gassler, M. and van de Putte, P. (1978). Early Gene Products of Bacteriophage Mu: Identification of the B Gene Product. *Journal of Molecular Biology* **120**:1-12.

132. Gouy, M., Gautier, C., Attimunelli, M., Lanuve, C. and

Di Paola, G. (1985). ACNUC - a Portable Retrieval System for Nucleic Acid Sequence Databases: Logical and Physical Designs and Usage. *Computer Applications in the Biosciences* **1**:167-172.

133. Green, L., Miller, R.D., Dykhuizen, D.E. and Hartl, D.L. (1984). Distribution of DNA Insertion Element IS*5* in Natural Isolates of *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* **81**:4500-4504.

134. Grindley, N.D. (1978). IS*1* Insertion Generates Duplication of a Nine Base Pair Sequence at Its Target Site. *Cell* **13**:419-426.

135. Grindley, N.D.F. (1983). Transposition of Tn*3* and Related Transposons. *Cell* **32**:3-5.

136. Grindley, N.D.F. and Joyce, C.M. (1980). Genetic and DNA Sequence Analysis of the Kanamycin Resistance Transposon Tn*903*. *Proceedings of the National Academy of Sciences USA* **77**:7176-7180.

137. Grindley, N.D. and Joyce, C.M. (1980). Analysis of the Structure and Function of the Kanamycin-resistance Transposon Tn*903*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:125-134.

138. Grindley, N.D.F. and Reed, R.R. (1985). Transpositional Recombination in Prokaryotes. *Annual Review of Biochemistry* **54**:863-896.

139. Grindley, N.D.F. and Sherratt, D.J. (1978). Sequence Analysis at IS*1* Insertion Sites: Models for Transposition. *Cold Spring Harbor Symposia on Quantitative Biology* **43**:1257-1261.

140. Grindley, N. and Wiater, L.A. (1988). $\gamma\delta$ Transposase and Integration Host Factor Bind Cooperatively at Both Ends of $\gamma\delta$. *EMBO Journal* **7**:1907-1912.

141. Grinsted, J., Bennett, P.M., Higginson, S. and Richmond, M.H. (1978). Regional Preference of Insertion of Tn*501* and Tn*802* into RP1 and Its Derivatives. *Molecular and General Genetics* **166**:313-320.

142. Grinsted, J., De La Cruz, F., Altenbuchner, J. and Schmidt, R. (1982). Complementation of *tnpA* Mutants of Tn*3*, Tn*21*, Tn*501* and Tn*1721*. *Plasmid* **8**:276-286.

143. Grinsted, J., De La Cruz, F and Schmitt, R. (1990). The Tn*21* Subgroup of Bacterial Transposable Elements. *Plasmid* **24**:163-189.

144. Guyer, M.S. (1978). The $\gamma\delta$ of F Is an Insertion Sequence. *Journal of Molecular Biology* **126**:347-365.

145. Guyer, M.S., Reed, R.R., Steitz, J.A. and Low, K.B. (1980). Identification of a Sex Factor Affinity Site in *E. coli* as Gamma Delta. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:135-140.

146. Halling, S.M. and Kleckner, N. (1982). A Symmetrical Six-base-pair Target Site Sequence Determines Tn*10* Insertion Specificity. *Cell* **28**:155-163.

147. Ham, G.E., Cardwell, V.B. and Johnson, H.W. (1971). Evaluation of *Rhizobium japonicum* Inoculants in Soils Containing Naturalized Populations of Rhizobia. *Agronomy Journal* **63**:301-303.

148. Harshey, R.M. and Bukhari, A.I. (1981). A Mechanism of DNA Transposition. *Proceedings of the National Academy of Sciences USA* **78**:1090-1094.

149. Hartmann, A, Catroux, G. and Amarger, N. (1992). *Bradyrhizobium japonicum* Strain Identification by RFLP Analysis Using the Repeated Sequence RSα. *Letters in Applied Microbiology* **15**:15-19.

150. Hashimoto, H. and Rownd, R.H. (1975). Transition of the R Factor *NR1* in *Proteus mirabilis*: Level of Drug Resistance of Nontransitioned and Transitioned Cells. *Journal of Bacteriology* **123**:56-68.

151. Hedges, R. and Jacob, A. (1974). Transposition of Ampicillin Resistance from RP4 to Other Replicons. *Molecular and General Genetics* **132**:31-40.

152. Heffron, F., Bedinger, P., Champoux, J. and Falkow, S. (1977). Deletions Affecting the Transposition of an Antibiotic Resistance Gene. *Proceedings of the National Academy of Sciences USA* **74**:702-706.

153. Heffron, F., McCarthy, B.J., Ohtsubo, H. and Ohtsubo, E. (1979). DNA Sequence Analysis of the Transposon Tn*3*: 3 Genes and 3 Sites Involved in Transposition of Tn*3*. *Cell* **18**:1153-1164.

154. Heffron, F., Ruebens, C. and Falkow, S. (1975). The Transposition of a Plasmid DNA Sequence which Mediates Ampicillin Resistance: Molecular Nature and Specificity of Insertion. *Proceedings of the National Academy of Sciences USA* **72**:3623-3627.

155. Herman, R., Neugebaur, K., Zentraf, H. and Schaller,

H. (1978). Transposition of a DNA Sequence Determining Kanamycin Resistance into the Single-stranded Genome of Bacteriophage Fd. *Molecular and General Genetics* **159:**171-178.

156. Herrero, M., De Lorenzo, V. and Timmis, K.N. (1990). Transposon Vectors Containing Non-antibiotic Resistance Selection Markers for Cloning and Stable Chromosomal Insertion of Foreign Genes in Gram-negative Bacteria. *Journal of Bacteriology* **172:**6557-6567.

157. Herschey, A.D. (ed.), *The Bacteriophage Lambda*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1971).

158. Hinton, D.M. and Musso, R.E. (1983). Specific *in vitro* Transcription of the Insertion Sequence IS2. *Journal of Molecular Biology* **169:**53-81.

159. Hirayama, T. and Oka, A. (1988). The *virE* and *virD3* Genes are Nonessential for Induction of Hairy Roots on Plants by *Agrobacterium rhizogenes* A4. *Bulletin of the Institute of Chemical Research, Kyoto University* **67:**227-238.

160. Hirsch, H.-J., Starlinger, P. and Brachet, P. (1972). Two Kinds of Insertions in Bacterial Genes. *Molecular and General Genetics* **119:**191-206.

161. Hirschel, B.J., Galas, D.J., Berg, D.E. and Chandler, M. (1982). Structure and Stability of Transposon *5*-mediated Cointegrates. *Journal of Molecular Biology* **159:**557-580.

162. Hirschel, B.J., Galas, D.J. and Chandler, M. (1982). Cointegrate Formation by Tn*5*, But Not Transposition, Is Dependent on *recA*. *Proceedings of the National Academy of Sciences USA* **79:**4530-4534.

163. Hobom, G. and Lusky, M. (1980). Origin and Inceptor of DNA Replication in Bacteriophage Lambda, p231-256. *In* Alberts, B. (ed.), *Mechanistic Studies of DNA Replication and Genetic Recombination*, Academic Press, N.Y.

164. Holmes, D.S. and Quigley, M. (1981). A Rapid Boiling Method for the Preparation of Bacterial Plasmids. *Analytical Biochemistry* **114:**193-197.

165. Hopkins, J.D., Clements, M.B., Liang, T.Y., Isberg, R.R. and Syvanen, M. (1980). Recombination Genes on the *Escherichia coli* Sex Factor Specific for Transposable

Elements. *Proceedings of the National Academy of Sciences USA* **77:**2814-2818.

166. Hopkins, J.D., Clements, M. and Syvanen, M. (1983). New Class of Mutations in *Escherichia coli* (*uup*) That Affect Precise Excision of Insertion Elements and Bacteriophage Mu Growth. *Journal of Bacteriology* **153:**384-389.

167. Howe, M. and Bade, E. (1975). Molecular Biology of Bacteriophage Mu. *Science* **190:**624-632.

168. Hu, S., Ohtsubo, E. and Davidson, N. (1975). Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of *Escherichia coli*: Structure of F13 and Related F-primes. *Journal of Bacteriology* **122:**749-763.

169. Hu, S., Ohtsubo, E., Davidson, N. and Saedler, H. (1975). Electron Microscope Heteroduplex Studies of Sequence Relations Among Bacterial Plasmids: Identification and Mapping of the Insertion Sequences IS*1* and IS*2* in F and R Plasmids. *Journal of Bacteriology* **122:**764-775.

170. Hu, S., Ptashne, K., Cohen, S.N. and Davidson, N. (1975). $\alpha\beta$ Sequence of F Is IS*3*. *Journal of Bacteriology* **123:**687-692.

171. Huber, P., Iida, S. and Arber, W. (1987). A Transcriptional Terminator Sequence in the Prokaryotic Transposable Element IS*1*. *Molecular and General Genetics* **206:**485-490.

172. Ikeda, H. and Koyabashi, I. (1977). Involvement of DNA-dependent RNA Polymerase in a *recA*-independent Pathway of Genetic Recombination in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* **74:**3932-3936.

173. Ikeda, H. and Matsumoto, T. (1979). Transcription Promotes *recA*-independent Recombination Mediated by DNA-dependent RNA Polymerase in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* **76:**4571-4575.

174. Iida, S., Heistand-Nauer, R. and Arber, W. (1985). Transposable Element IS*1* Intrinsically Generates Target Duplications of Variable Length. *Proceedings of the National Academy of Sciences USA* **82:**839-843.

175. Iida, S., Meyer, J. and Arber, W. (1978). The Insertion Element IS*1* Is a Natural Constituent of Coliphage P1 DNA. *Plasmid* **1**:357-365.

176. Iida, S., Meyer, J. and Arber, W. (1983). Prokaryotic IS Elements, p159-221. *In* Shapiro, J.A. (ed.), *Mobile Genetic Elements*, Academic Press Inc., N.Y.

177. Iida, S., Meyer, J., Kennedy, K.E., and Arber, W. (1982). A Site-specific, Conservative Recombination System Carried by Bacteriophage P1. Mapping the Recombinase Gene *cin* and the Crossover Site *cix* for the Inversion of the C Segment. *EMBO Journal* **1**:1445-1453.

178. Iida, S., Meyer, J., Linder, P., Goto, N., Nakaya, R., Reif, H.-J. and Arber, W. (1982). The Kanamycin Resistance Transposon Tn*2680* Derived from the R Plasmid Rts1 and Carried by Phage P1Km Has Flanking 0·8kb Long Direct Repeats. *Plasmid* **8**:187-198.

179. Iino, T. and Kutsukake, K. (1980). *Trans*-acting Genes of Bacteriophages P1 and Mu Mediate Inversion of a Specific DNA Segment Involved in Flagellar Phage Variation of *Salmonella*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:11-16.

180. Ilyina, T.S., Nechaeva, E.V., Romanova, Y.M. and Smirnov, G.B. (1981). Isolation and Mapping of *Escherichia coli* K-12 Mutants Defective in Tn*9* Transposition. *Molecular and General Genetics* **181**:384-389.

181. Ilyina, T.S., Romanova, Y.M. and Smirnov, G.B. (1981). The Effect of *tnm* Mutations of *Escherichia coli* K-12 on Transposition of Various Mobile Genetic Elements. *Molecular and General Genetics* **183**:376-379.

182. Isberg, R.R., Lazaar, A.L. and Syvanen, M. (1982). Regulation of Tn*5* by the Right-repeat Proteins: Control at the Level of the Transposition Reaction. *Cell* **30**:883-892.

183. Isberg, R.R. and Syvanen, M. (1981). Replicon Fusions Promoted by the Inverted Repeats of Tn*5*: The Right Stem Is an Insertion Sequence. *Journal of Molecular Biology* **150**:15-32.

184. Isberg, R.R. and Syvanen, M. (1982). DNA Gyrase Is a Host Factor Required for Transposition of Tn*5*. *Cell* **30**:9-18.

185. Ishiguro, N, and Sato, G. (1984). Spontaneous Deletion

of Citrate-utilizing Ability Promoted by Insertion Sequences. *Journal of Bacteriology* **160**:642-650.

186. Jacob, F. and Monod, J. (1961). On the Regulation of Gene Activity. *Cold Spring Harbor Symposia on Quantitative Biology* **26**:193-209.

187. Jacowec, M., Prentki, P., Chandler, M. and Galas, D.J. (1988). Mutational Analysis of the Open Reading Frames in Transposable Element IS*1*. *Genetics* **120**:47-55.

188. Jaurin, B. and Normark, S. (1983). Insertion of IS2 Creates a Novel *ampC* Promoter in *Escherichia coli*. *Cell* **32**:809-816.

189. Johnson, R.C., Bruist, M.F. and Simon, M.I. (1986). Host Protein Requirements for *in vitro*, Site-specific DNA Inversion. *Cell* **46**:531-539.

190. Johnson, R.C. and Reznikoff, W.S. (1983). DNA Sequences at the Ends of Transposon Tn*5* Required for Transposition. *Nature* **304**:280-282.

191. Johnson, R.C. and Reznikoff, W.S. (1984). Copy Number Control of Tn*5* Transposition. *Genetics* **107**:9-18.

192. Johnson, R.C. and Reznikoff, W.S. (1984). Role of the IS*50* R Proteins in the Promotion and Control of Tn*5* Transposition. *Journal of Molecular Biology* **177**:645-661.

193. Johnson, R.C., Yin, J.C. and Reznikoff, W.S. (1982). Control of Tn*5* Transposition in *Escherichia coli* Is Mediated by Protein from Right Repeat. *Cell* **30**:873-882.

194. Johnson, S.R., Liu, B.C.S. and Romig, W.R. (1981). Auxotrophic Mutations Induced by *Vibrio cholerae* Mutator Phage VcA1. *FEMS Microbiology Letters* **11**:13-16.

195. Johnston, A.W.B. and Beringer, J.E. (1976). Pea Root Nodules Containing More than One *Rhizobium* Species. *Nature* **263**:502-504.

196. Johnsrud,L., Calos, M.P. and Miller, J.H. (1978). The Transposon Tn*9* Generates a 9bp Repeated Sequence During Integration. *Cell* **15**:1209-1219.

197. Jordan, D.C. (1984). Family III. *Rhizobiaceae* Conn 1938, 321.[AL] p234-244. *In* Krieg, N.R. and Holt, J.G. (ed.), *Bergey's Manual of Systematic Bacteriology Vol. I*, Williams and Wilkins, Baltimore, USA.

198. Jordan, E., Saedler, H. and Starlinger, P. (1967). Strong Polar Mutations in the Transferase Gene of the

Galactose Operon in *E. coli*. *Molecular and General Genetics* **100**:296-306.

199. Jordan, E., Saedler, H. and Starlinger, P. (1968). 0-zero and Strong Polar Mutations in the *gal* Operon Are Insertions. *Molecular and General Genetics* **102**:353-363.

200. Jorgensen, R.A. and Reznikoff, W.S. (1979). Organization of Structural and Regulatory Genes That Mediate Tetracycline Resistance in Transposon Tn*10*. *Journal of Bacteriology* **138**:705-714.

201. Kahmann, R. and Kamp, D. (1979). Nucleotide Sequences of the Attachment Sites of Bacteriophage Mu DNA. *Nature* **280**:247-250.

202. Kaluza, K., Hahn, M. and Hennecke, H. (1985). Repeated Sequences Similar to Insertion Elements Clustered around the *nif* Region of the *Rhizobium japonicum* Genome. *Journal of Bacteriology* **162**:535-542.

203. Kaminski, P.A., Norel, F., Desnoues, N., Kush, A., Salzano, G. and Elmerich, C. (1988). Characterization of the *fixABC* Region of *Azorhizobium caulinodans* ORS571 and Identification of a New Nitrogen Fixation Gene. *Molecular and General Genetics* **214**:496-502.

204. Kamp, D. and Kahmann, R. (1980). Two Pathways in Bacteriophage Mu Transposition? *Cold Spring Harbor Symposia in Quantitative Biology* **45**:329-336.

205. Karlin, S and Altschul, S.F. (1990). Methods for Assessing the Statistical Significance of Molecular Sequence Features by Using General Scoring Schemes. *Proceeding of the National Academy of Sciences USA* **87**:2264-2268.

206. Khan, S. and Novick, R. (1980). Terminal Nucleotide Sequence of Tn*551*, a Transposon Specifying Erythromycin Resistance in *S. aureus*: Homology with Tn*3*. *Plasmid* **4**:148-154.

207. Kitts, P., Symington, L., Burke, M., Reed, R. and Sherratt, D. (1982). Transposon-specified Site-specific Recombination. *Proceedings of the National Academy of Sciences USA* **79**:46-50.

208. Klaer, R., Kühn, S., Fritz, H.-J., Tillmann, E., Saint-Girons, I., Habermann, P., Pfeifer, D. and Starlinger, P. (1980). Studies on Transposition Mechanisms and Specificity of IS*4*. *Cold Spring Harbor Symposia on*

Quantitative Biology **45**:215-224.

209. Klaer, R., Kühn, S., Tillmann, E., Fritz, H.J. and Starlinger, P. (1981). The Sequence of IS*4*. *Molecular and General Genetics* **181**:169-175.

210. Klaer, R., Pfeifer, D. and Starlinger, P. (1980). IS*4* Is Still Found at Its Chromosomal Site after Transposition to *galT*. *Molecular and General Genetics* **178**:281-284.

211. Klaer, R. and Starlinger, P. (1980). IS*4* at Its Chromosomal Site in *E. coli*. *Molecular and General Genetics* **178**:285-291.

212. Kleckner, N. (1977). Translocatable Elements in Procaryotes. *Cell* **11**:11-23.

213. Kleckner, N. (1979). DNA Sequence Analysis of Tn*10* Insertions: Origin and Role of 9 bp Flanking Repetitions during Tn*10* Translocation. *Cell* **16**:711-720.

214. Kleckner, N. (1981). Transposable Elements in Prokaryotes. *Annual Review of Genetics* **15**:341-404.

215. Kleckner, N. (1989). Transposon Tn*10*, p227-268. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington D.C.

216. Kleckner, N., Chan, R.K., Tye, B.K. and Botstein, D. (1975). Mutagenesis by Insertion of a Drug Resistance Element Carrying an Inverted Repetition. *Journal of Molecular Biology* **97**:561-575.

217. Kleckner, N., Reichardt, K. and Botstein, D. (1979). Inversions and Deletions of the *Salmonella* Chromosome Generated by the Translocatable Tetracycline Resistance Element Tn*10*. *Journal of Molecular Biology* **127**:89-115.

218. Kleckner, N. and Ross, D.G. (1980). *recA*-dependent Genetic Switch Generated by Transposon Tn*10*. *Journal of Molecular Biology* **144**:215-221.

219. Kopecko, D.J. (1980). Specialized Genetic Recombination Systems in Bacteria: Their Involvement in Gene Expression and Evolution. *Progress in Molecular and Subcellular Biology* **7**:135-234.

220. Kopecko, D.J., Brevet, J. and Cohen, S.N. (1976). Involvement of Multiple Translocating DNA Segments and Recombinational Hot Spots in the Structural Evolution of Bacterial Plasmids. *Journal of Molecular Biology* **108**:333-360.

221. Kornberg, A. (1980). *DNA Replication*. Freeman, San Francisco.

222. Kothary, R.K., Jones, D. and Candido, E.P.M. (1985). IS*186*: An *Escherichia coli* Insertion Element Isolated from a cDNA Library. *Journal of Bacteriology* **164**:957-959.

223. Kowalski, M., Ham, G.E., Frederick, L.R. and Anderson, I.E. (1974). Relationship between Strains of *Rhizobium japonicum* and Their Bacteriophages from Soil and Nodules of Field-grown Soybeans. *Soil Science* **118**:221-228.

224. Kröger, M. and Hobom, G. (1982). Structural Analysis of Insertion Sequence IS*5*. *Nature* **297**:159-162.

225. Labes, G. and Simon, R. (1990). Isolation of DNA Insertion Elements from *Rhizobium meliloti* Which Are Able to Promote Transcription of Adjacent Genes. *Plasmid* **24**:235-239.

226. Labigne-Roussel, A., Briaux-Gerbaud, S. and Courvalin, P. (1983). Tn*1525*, a Kanamycin R Determinant Flanked by Two Direct Copies of IS*15*. *Molecular and General Genetics* **189**:90-101.

227. Labinge-Roussel, A., Gerbaud, G. and Courvalin, P. (1981). Translocation of Sequences Encoding Antibiotic Resistance from the Chromosome to a Receptor Plasmid in *Salmonella ordonez*. *Molecular and General Genetics* **182**:390-408.

228. Lam, S. and Roth, J.R. (1983). Genetic Mapping of IS*200* Copies in *Salmonella typhimurium* Strain LT2. *Genetics* **105**:801-811.

229. Landy, A. and Ross, W. (1977). Viral Integration and Excision: Structure of the Lambda *att* Sites. *Science* **197**:1147-1160.

230. Lee, Y. and Schmidt, F.J. (1985). Characterization of the *in vivo* RNA Product of the pOUT Promoter of IS*10*$_R$. *Journal of Bacteriology* **164**:556-562.

231. Lindahl, G., Hirota, Y. and Jacob, F. (1971). On the Process of Cellular Division in *Escherichia coli*: Replication of the Bacterial Chromosome under Control of Prophage P2. *Proceedings of the National Academy of Sciences USA* **68**:2407-2412.

232. Lowe, J.B. and Berg, D.E. (1983). A Product of the Tn*5* Transposase Gene Inhibits Transposition. *Genetics*

103:605-615.

233. **Lundblad, V., Taylor, A.F., Smith, G.R. and Kleckner, N.** (1984). Unusual Alleles of *recB* and *recC* Stimulate Excision of Inverted Repeat Transposons Tn*10* and Tn*5*. *Proceedings of the National Academy of Sciences USA* **81**:824-828.

234. **Lupski, J.R., Gershon, P., Ozaki, L.S. and Godson, G.N.** (1984). Specificity of Tn*5* Insertions into a 36-bp DNA Sequence Repeated in Tandem Seven Times. *Gene* **30**:99-106.

235. **Lusky, M., Kröger, M. and Hobom, G.** (1980). Detection of Replication Inceptor Signals in IS*5*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:173-176.

236. **Lydiate, D.J., Ikeda, H. and Hopwood, D.A.** (1986). A 2·6kb DNA Sequence of *Streptomyces coelicolor* A3(2) which Functions as a Transposable Element. *Molecular and General Genetics* **203**:79-88

237. **MacHattie, L.A. and Jackowski, J.B.** (1977). Physical Structure and Deletion Effects of the Chloramphenicol Resistance Element Tn*9* in Phage Lambda, p219-228. *In* Bukhari, A.I., Shapiro, J.A. and Adhya, S. (ed.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

238. **Machida, Y., Machida, C., and Ohtsubo, E.** (1982). A Novel Type of Transposon Generated by Insertion Element IS*102* Present in a pSC*101* Derivative. *Cell* **30**:29-36.

239. **Machida, Y., Machida, C., Ohtsubo, H. and Ohtsubo, E.** (1982). Factors Determining Frequency of Plasmid Cointegration Mediated by Insertion Sequence IS*1*. *Proceedings of the National Academy of Sciences USA* **79**:277-281.

240. **Machida, Y., Sakurai, M. Kiyokawa, S., Ubasawa, A., Suzuki, Y. and Ikeda, J.-E.** (1984). Nucleotide Sequence of the Insertion Sequence Found in the T-DNA Region of Mutant Ti Plasmid pTiA66 and Distribution of Its Homologues in Octopine Ti Plasmid. *Proceedings of the National Academy of Sciences USA* **81**:7495-7499.

241. **Magazin, M. and Allet, B.** (1978). Synthesis of Bacteriophage Mu Proteins *in vitro*. *Virology* **85**:84-97.

242. **Malamy, M.** (1966). Frameshift Mutations in the Lactose Operon of *E. coli*. *Cold Spring Harbor Symposia on Quan-*

titative Biology **31**:189-201.

243. Malamy, M.H. (1970). Some Properties of Insertion Mutations in the *lac* Operon, p359-373. *In* Beckwith, J.R. and Zipser, D. (ed.), *The Lactose Operon*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

244. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular Cloning.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

245. Martin, G.B., Thomashow, M.F. and Chelm, B.K. (1989). *Bradyrhizobium japonicum glnB*, a Putative Nitrogen-regulatory Gene, Is Regulated by NtrC at Tandem Promoter. *Journal of Bacteriology* **171**:5638-5645.

246. Martin, R.G. (1969). Control of Gene Expression. *Annual Review of Genetics* **3**:181-216.

247. Mattes, R., Altenbuchner, J. and Schmitt, R. (1978). The Translocation of the Raffinose Operon Occurs by IS*1*-mediated, Site-specific Recombination. *Hoppe-Seylers Zeitschrift fur Physiologische Chemie* **359**:1118-1119.

248. Matsutani, S., Ohtsubo, H., Maeda, Y. and Ohtsubo, E. (1987). Isolation and Characterization of IS Elements Repeated in the Bacterial Genome. *Journal of Molecular Biology* **196**:445-455.

249. Mayaux, J.F., Springer, M., Graffe, M., Fromant, M. and Fayat, G. (1984). IS*4* Transposition in the Attenuator Region of the *Escherichia coli pheS, T* Operon. *Gene* **30**:137-146.

250. McClintock, B. (1951). Chromosome Organization and Genic Expression. *Cold Spring Harbor Symposia on Quantitative Biology* **16**:13-47.

251. McClintock, B. (1965). The Control of Gene Action in Maize. *Brookhaven Symposia in Biology* **18**:162-184.

252. McCommas, S.A. and Syvanen, M. (1988). Temporal Control of Transposition in Tn*5*. *Journal of Bacteriology* **170**:889-894.

253. McDougall, S. and Neilands, J.B. (1984). Plasmid- and Chromosome-coded Aerobactin Synthesis in Enteric Bacteria: Insertion Sequences Flank Operon in Plasmid-mediated Systems. *Journal of Bacteriology* **159**:300-305.

254. Meyer, J. and Iida, S. (1979). Amplification of Chloramphenicol Resistance Transposons Carried by Phage P1CM

in *Escherichia coli*. *Molecular and General Genetics* **176**:209-219.

255. Meyer, J., Iida, S. and Arber, W. (1980). Does the Insertion Element IS*1* Transpose Preferentially into A+T-rich DNA Segments. *Molecular and General Genetics* **178**:471-473.

256. Miller, C.A. and Cohen, S.N. (1980). F Plasmid Provides a Function That Promotes *recA*-independent Site-specific Fusions of pSC101 Replicon. *Nature* **285**:577-579.

257. Mollet, B., Iida, S. and Arber, W. (1985). Gene Organization and Target Specificity of the Prokaryotic Mobile Genetic Element IS*26*. *Molecular and General Genetics* **201**:198-203

258. Morgan, E.A. (1980). Insertions of Tn*10* Are Incompletely Polar. *Cell* **21**:257-265.

259. Morisato, D. and Kleckner, N. (1987). Tn*10* Transposition and Circle Formation *in vitro*. *Cell* **51**:101-111.

260. Morisato, D., Way, J.C., Kim, H.-J. and Kleckner, N. (1983). Tn*10* Transposase acts Preferentially on Nearby Transposon Ends *in vivo*. *Cell* **32**:799-807.

261. Moss, M.T., Malik, Z.P., Tizard, M.L.V., Green, E.P., Sanderson, J.D. and Hermon-Taylor, J. (1992). IS*902*, an Insertion Element of the Chronic-enteritis-causing *Mycobacterium avium* subsp. *silvaticum*. *Journal of General Microbiology* **138**:139-145.

262. Mueller, J.G., Skipper, H.D., Shipe, E.R., Grimes, L.W. and Wagner, S.C. (1988). Intrinsic Antibiotic Resistance in *Bradyrhizobium japonicum*. *Soil Biology and Biochemistry*. **20**:879-882.

263. Murphy, E. (1989). Transposable Elements in Gram-positive Bacteria, p269-288. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington D.C.

264. Murphy, E. and Löfdahl, S. (1984). Transposition of Tn*554* Does Not Generate a Target Duplication. *Nature* **307**:292-294.

265. Musso, R., die Lauro, R., Rosenberg, M. and de Crombrugghe, B. (1977). Nucleotide Sequence of the Operator-promoter Region of the Galactose Operon of *Escherichia coli*. *Proceedings of the National Academy of*

Sciences USA. **74**:106-110.

266. Nash, H.A. (1981). Integration and Excision of Bacteriophage λ: The Mechanism of Conservative Site Specific Recombination. *Annual Review of Genetics* **15**:143-167.

267. Nevers, P., Reif, H.-J. and Saedler, H. (1977). On Mutations Affecting IS*1*-mediated Deletion Formation in *E. coli*, p125-128. *In* Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. (ed.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

268. Nevers, P. and Saedler, H. (1978). Mapping and Characterization of an *E.coli* Mutant Defective in IS*1*-mediated Deletion Formation. *Molecular and General Genetics* **160**:209-214.

269. Nisen, P., Purucker, M. and Shapiro, L. (1979). Deoxyribonucleic Acid Sequence Homologies Among Bacterial Insertion Sequence Elements and Genomes of Various Organisms. *Journal of Bacteriology* **140**:588-596.

270. Noel, K.D. and Ames, G. (1978). Evidence for a Common Mechanism for the Insertion of the Tn*10* Transposon and for the Generation of Tn*10*-stimulated Deletions. *Molecular and General Genetics* **166**:217-223.

271. Nomura, N., Yomagishi, H. and Oka, A. (1978). Isolation and Characterization of Transducing Coliphage Fd Carrying a Kanamycin Resistance Determinant. *Gene* **3**:39-51.

272. Novick, R.P., Clowes, R.C., Cohen, S.N., Curtiss III, R., Datta, N. and Falkow, S. (1976). Uniform Nomenclature for Bacterial Plasmids: A Proposal. *Bacteriological Reviews* **40**:168-189.

273. Nyman, K., Nakamura, K., Ohtsubo, H. and Ohtsubo, E. (1981). Distribution of the Insertion Sequence IS*1* in Gram-negative Bacteria. *Nature* **289**:609-612.

274. Ogawa, J., Brierley, H.L. and Long, S.R. (1991). Analysis of *Rhizobium meliloti* Nodulation Mutant WL131: Novel Insertion Sequence IS*Rm3* in *nodG* and Altered *nodH* Protein Product. *Journal of Bacteriology* **173**:3060-3065.

275. Ohtsubo, E., Ohtsubo, H., Doroszkiewicz, W. Nyman, K., Allen, D. and Davison, D. (1984). An Evolutionary Analysis of Iso-IS*1* Elements from *Escherichia coli* and

Shigella strains. *Journal of General and Applied Microbiology* **30**:359-376.

276. Ohtsubo, E., Zenilman, M. and Ohtsubo, H. (1980). Plasmids Containing Insertion Elements Are Potential Transposons. *Proceedings of the National Academy of Sciences USA* **77**:750-754.

277. Ohtsubo, E., Zenilman, M, Ohtsubo, H., McCormick, M., Machida, C. and Machida, Y. (1980). Mechanism of Insertion and Cointegration Mediated by IS*1* and Tn*3*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:283-296.

278. Ohtsubo, H., Nyman, K., Doroszkiewicz, W. and Ohtsubo, E. (1981). Multiple Copies of Iso-insertion Sequences of IS*1* in *Shigella dysenteriae* Chromosome. *Nature* **292**:640-643.

279. Ohtsubo, H., Ohmori, H. and Ohtsubo, E. (1978). Nucleotide Sequence Analysis of Tn*3* (Ap): Implications for Insertions and Deletions. *Cold Spring Harbor Symposia on Quantitative Biology* **43**:1269-1277.

280. Ohtsubo, H. and Ohtsubo, E. (1978). Nucleotide Sequence of an Insertion Element IS*1*. *Proceedings of the National Academy of Sciences USA* **75**:615-619.

281. Oka, A, Nomura, N., Sugimoto, K., Sugisaki, H. and Takanami, M. (1978). Nucleotide Sequence at the Insertion Sites of a Kanamycin Transposon. *Nature* **276**:845-847.

282. Otten, L., Canaday, J., Gérard, J.-C., Fournier, P., Crouzet, P. and Paulus, F. (1992). Evolution of Agrobacteria and Their Ti Plasmids - A Review. *Molecular Plant-microbe Interactions* **5**:279-287.

283. Ow, D.W., Sundaresan, V., Rothstein, D.M., Brown, S.E. and Ausubel, F.M. (1983). Promoters Regulated by the *glnG* (*ntrC*) and *nifA* Gene Products Share a Heptameric Consensus Sequence in the -15 Region. *Proceedings of the National Academy of Sciences USA* **80**:2524-2528.

284. Patee, P.A. (1981). Distribution of Tn*551* Insertion Sites Responsible for Auxotrophy on the *Staphylococcus aureus* Chromosome. *Journal of Bacteriology* **145**:479-488.

285. Pato, M.L. (1989). Bacteriophage Mu, p23-52. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington, D. C.

286. Paulus, F., Canaday, J. and Otten, L. (1991). Limited Host Range Ti Plasmids: Recent Origin from Wide Host Range Ti Plasmids and Involvement of a Novel IS Element, IS*868*. *Molecular Plant-microbe Interactions* **4**:190-197.

287. Phadnis, S.H. and Berg, D.E. (1987). Identification of Base Pairs in the Outside End of Insertion Sequence IS*50* That are Needed for IS*50* and Tn*5* Transposition. *Proceedings of the National Academy of Sciences USA* **84**:9118-9122.

288. Platt, T. (1986). Transcription Termination and the Regulation of Gene Expression. *Annual Review of Biochemistry* **55**:339-372.

289. Post, L.E., Arfsten, A.E., Reusser, F. and Nomura, M. (1978). DNA Sequences of Promoter Regions for the *str* and *spc* Ribosomal Protein Operons in *E. coli*. *Cell* **15**:215-229.

290. Post, L.E. and Nomura, M. (1980). DNA Sequences from the *str* Operon of *Escherichia coli*. *The Jounal of Biological Chemistry* **255**:4660-4666.

291. Prentki, P., Chandler, M. and Galas, D. (1987). *Escherichia coli* Integration Host Factor Bends the DNA at the Ends of IS*1* and in an Insertion Hotspot with Multiple IHF Binding Sites. *EMBO Journal* **6**:2479-2487.

292. Prentki, P., Teter, B., Chandler, M. and Galas, D.J. (1986). Functional Promoters Created by the Insertion of Transposable Element IS*1*. *Journal of Molecular Biology* **191**:383-393.

293. Prestridge, D.S. (1991). Signal Scan: A Computer Program That Scans DNA Sequences for Eukaryotic Transcriptional Elements. *Computer Applications in Biosciences* **7**:203-206.

294. Priefer, U.B., Burkardt, H.J., Klipp, W. and Pühler, A. (1980). IS*R1*: An Insertion Element Isolated from the Soil Bacterium *Rhizobium lupini*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:87-91.

295. Priefer, U.B., Kalinowski, J., Rüger, B., Heumann, W. and Pühler, A. (1989). IS*R1*, a Transposable DNA Sequence Resident in *Rhizobium* Class IV Strains, Shows Structural Characteristics of Classical Insertion Elements. *Plasmid* **21**:120-128.

296. **Pruss, G.J., Manes, S.H. and Drlica, K.** (1982). *Escherichia coli* DNA Topoisomerase I Mutants: Increased Supercoiling Is Corrected by Mutations Near Gyrase Genes. *Cell* **31**:35-42.

297. **Ptashne, K. and Cohen, S.N.** (1975). Occurrence of Insertion Sequence (IS) Regions on Plasmid Deoxyribonucleic Acid as Direct and Inverted Nucleotide Sequence Duplications. *Journal of Bacteriology* **122**:776-781.

298. **Raibaud, O., Gutierrez, C. and Schwartz, M.** (1985). Essential and Nonessential Sequences in *malPp*, a Positively Controlled Promoter in *Escherichia coli*. *Journal of Bacteriology* **161**:1201-1208.

299. **Rak, B., Lusky, M. and Hable, M.** (1982). Expression of Two Proteins from Overlapping and Oppositely Oriented Genes on Transposable DNA Insertion Element IS*5*. *Nature* **297**:124-128.

300. **Rak, B. and von Reutern, M.** (1984). Insertion Element IS*5* Contains a Third Gene. *EMBO Journal* **3**:807-811.

301. **Ramakrishnan, N., Prakash, R.K. and Atherly, A.G.** (1986). Conservation of IS*66* Homologue of Octopine Ti Plasmid DNA in *Rhizobium fredii* Plasmid DNA. *Plant Molecular Biology* **7**:177-188.

302. **Read, H.A., Das Sarma, S. and Jaskunas, S.R.** (1980). Fate of Donor Insertion Sequence IS*1* during Transposition. *Proceedings of the National Academy of Sciences USA* **77**:2514-2418.

303. **Read, H.A. and Jaskunas, S.R.** (1981). Isolation of *E. coli* Mutants Containing Multiple Transpositions of IS Sequences. *Molecular and General Genetics* **180**:157-164.

304. **Reed, R.R., Young, R.A., Steitz, J.A., Grindley, N.D.F. and Guyer, M.S.** (1979). Transposition of the *Escherichia coli* Insertion Element $\gamma\delta$ generates a Five-Base-Pair Repeat. *Proceedings of the National Academy of Sciences USA* **76**:4882-4886.

305. **Reif, H.-J. and Saedler, H.** (1975). IS*1* Is Involved in Deletion Formation in the *gal* Region of *E. coli* K12. *Molecular and General Genetics* **137**:17-28.

306. **Reif, H.-J. and Saedler, H.** (1977). Chromosomal Rearrangements in the *gal* Region of *E. coli* K12 after Integration of IS*1*, p81-91. *In* Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. (ed.), *DNA Insertion Elements,*

Plasmids and Episomes, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

307. **Reyes, O., Gottesman, M. and Adhya, S.** (1976). Suppression of Polarity of Insertion Mutants in the *gal* Operon and *N* Mutations in Bacteriophage Lambda. *Journal of Bacteriology* **126**:1108-1112.

308. **Reynolds, A.E., Felton, J. and Wright, A.** (1981). Insertion of DNA Activates the Cryptic *bgl* Operon in *E. coli* K12. *Nature* **293**:625-629.

309. **Riess, G., Holloway, B.W. and Pühler, A.** (1980). R68.45, a Plasmid with Chromosome Mobilizing Ability (*Cma*) Carries a Tandem Duplication. *Genetical Research* **36**:99-109.

310. **Riess, G., Masepohl, M. and Pühler, A.** (1983). Analysis of IS*21*-mediated Mobilization of Plasmid pACYC184 by R68.45 in *Escherichia coli*. *Plasmid* **10**:111-118.

311. **Roberts, D., Hoopes, B.C., McClure, W.R. and Kleckner, N.** (1985). IS*10* Transposition Is Regulated by DNA Adenine Methylation. *Cell* **43**:117-130.

312. **Robinson, M.K., Bennett, P.M. and Richmond, M.H.** (1977). Inhibition of TnA Translocation by TnA. *Journal of Bacteriology* **129**:407-414.

313. **Rogowsky, P. and Schmidt, R.** (1984). Resolution of a Hybrid Cointegrate between Transposons Tn*501* & Tn*1721* Defines the Recombination Site. *Molecular and General Genetics* 193:162-166.

314. **Romecker, H.-J. and Rak, B.** (1987). Genetic Organization of Insertion Element IS2 Based on a Revised Nucleotide Sequence. *Gene* **59**:291-296.

315. **Ross, D.G., Grisafi, P., Kleckner, N. and Botstein, D.** (1979). The Ends of Tn*10* Are Not IS*3*. *Journal of Bacteriology* 139:1097-1101.

316. **Ross, D.G., Swan, J. and Kleckner, N.** (1979). Physical Structures of Tn*10* Promoted Deletions and Inversions: Role of 1400 bp Inverted Repetitions. *Cell* **16**:721-731.

317. **Ross, D.G., Swan, J. and Kleckner, N.** (1979). Nearly Precise Excision: A New Type of DNA Alteration Associated with the Translocatable Element Tn*10*. *Cell* **16**:733-738.

318. **Rothstein, S.J., Jorgensen, R.A., Postle, K. and Resnikoff, W.S.** (1980). The Inverted Repeats of Tn*5* Are

Functionally Different. *Cell* **19**:795-805.

319. **Rothstein, S.J., Jorgensen, R.A., Yin, J.C.-P., Yongdi, Z., Johnson, R.C. and Reznikoff, W.S.** (1980). Genetic Organization of Tn*5*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:99-106.

320. **Rothstein, S.J. and Reznikoff, W.S.** (1981). The Functional Differences in the Inverted Repeats of Tn*5* Are Caused by a Single Base Pair Nonhomology. *Cell* **23**:733-738.

321. **Rownd, R. and Mickel, S.** (1971). Dissociation and Reassociation of RTF and r-determinants of the R-factor *NR1* in *Proteus mirabilis*. *Nature New Biology* **234**:40-43.

322. **Ruvkun, G.B., Long, S.R., Meade, H.M., van den Bos, R.C. and Ausubel, F.M.** (1982). IS*Rm1*: A *Rhizobium meliloti* Insertion Sequence That Transposes Preferentially into Nitrogen Fixation Genes. *Journal of Molecular and Applied Genetics* **1**:405-418.

323. **Sadowsky, M.J., Cregan, P.B. and Keyser, H.H.** (1990). DNA Hybridization Probe for Use in Determining Restricted Nodulation among *Bradyrhizobium japonicum* Serocluster 123 Field Isolates. *Applied and Environmental Microbiology* **56**:1768-1774.

324. **Saedler, H.** (1977). IS*1* and IS*2* in *E. coli*: Implications for the Evolution of the Chromosome and Some Plasmids, p65-72. *In* Bukhari, A.I., Shapiro, J.A. and Adhya, S.L. (ed.), *DNA Insertion Elements, Plasmids and Episomes*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

325. **Saedler, H., Besemer, J., Kemper, B., Rosenwirth, B. and Starlinger, P.** (1972). Insertion Mutations in the Control Region of the *gal* Operon of *E. coli*. *Molecular and General Genetics* **115**:258-265.

326. **Saedler, H., Cornelis, G., Cullum, J., Schumacher, B. and Sommer, H.** (1980). IS*1*-mediated DNA Rearrangements. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:93-98.

327. **Saedler, H. and Heiss, B.** (1973). Multiple Copies of the Insertion-DNA Sequences IS*1* and IS*2* in the Chromosome of *E. coli* K12. *Molecular and General Genetics* **122**:267-277.

328. **Saedler, H., Reif, H.-J., Hu, S. and Davidson, N.**

(1974). IS2, a Genetic Element for Turn-off and Turn-on of Gene Activity in *E. coli*. *Molecular and General Genetics* **132:**265-289.

329. **Saedler, H. and Starlinger, P.** (1967). 0^{o}-mutations in the Galactose Operon in *E. coli*. I. Genetic Characterization. *Molecular and General Genetics* **100:**178-189.

330. **Saedler, H. and Starlinger, P.** (1967). 0^{o}-mutations in the Galactose Operon in *E. coli*. II. Physiological Characterization. *Molecular and General Genetics* **100:**190-202.

331. **Saint-Girons, I, Fritz, H.J., Shaw, C., Tillmann, E. and Starlinger, P.** (1981). Integration Specificity of an Artificial Kanamycin Transposon Constructed by the *in vitro* Insertion of an Internal Tn*5* Fragment into IS*2*. *Molecular and General Genetics* **183:**45-50.

332. **Sakaguchi, K.** (1990). Invertrons, a Class of Structurally and Functionally Related Genetic Elements That Includes Linear DNA Plasmids, Transposable Elements and Genomes of Adeno-type Viruses. *Microbiological Reviews* **54:**66-74.

333. **Sanger, F., Nicklen, S. and Coulson, A.R.** (1977). DNA Sequencing with Chain-terminating Inhibitors. *Proceedings of the National Academy of Sciences USA* **74:**5463-5467.

334. **Sasakawa, C. and Berg, D.E.** (1982). IS*50*-mediated Inverse Transposition. Discrimination between the Two Ends of an IS Element. *Journal of Molecular Biology* **159:**257-271.

335. **Sasakawa, C., Uno, Y. and Yoshikawa, M.** (1981). The Requirement for Both DNA Polymerase and 5′ to 3′ Exonuclease Activities of DNA Polymerase I during Tn*5* Transposition. *Molecular and General Genetics* **182:**19-24.

336. **Sawyer, S.A., Dykhuizen, D.E., DuBose, R.F., Green, L., Mutangadura Mhlanga, T., Wolczyk, D.F. and Hartl, D.L.** (1987). Distribution and Abundance of Insertion Sequences among Natural Isolates of *Escherichia coli*. *Genetics* **115:**51-63.

337. **Schaffer, H.E. and Sederoff, R.R.** (1981). Improved Estimation of DNA Fragment Lengths from Agarose Gels. *Analytical Biochemistry* **115:**113-122.

338. **Schaller, H.** (1978). The Intergenic Region and the Origins for Filamentous Phage DNA Replication. *Cold Spring Harbor Symposia on Quantitative Biology* **43**:401-408.

339. **Schmidt, E.L., Bankole, R.O., and Bohlool, B.B.** (1968). Flourescent-antibody Approach to Study of Rhizobia in Soil. *Journal of Bacteriology* **95**:1987-1992.

340. **Schmidt, E.L., Zidwick, M.J. and Abebe, H.M.** (1986). *Bradyrhizobium japonicum* Serocluster 123 and Diversity among Member Isolates. *Applied and Environmental Microbiology* **51**:1212-1215.

341. **Schmitt, R., Altenbuchner, J., Weibaur, K., Arnold, W., Pühler, A. and Schoffl, F.** (1980). Basis of Transposition and Gene Amplification by Tn*1721* and Related Tetracycline-resistance Transposons. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:59-63.

342. **Schwartz, E., Herberger, C. and Rak, B.** (1988). Second Element Turn-on of Gene Expression in an IS*1* Insertion Mutant. *Molecular and General Genetics* **211**:282-289.

343. **Schwartz, E., Kroeger, M. and Rak, B.** (1988). IS*150*: Distribution, Nucleotide Sequence and Phylogenetic Relationship of a New *E. coli* Insertion Element. *Nucleic Acids Research* **16**:6789-6802.

344. **Sciaky, D. and Thomashow, M.F.** (1984). The Sequence of the *tms* Transcript 2 Locus of the *A. tumefaciens* Plasmid pTiA6 and Characterization of the Mutation in pTiA66 That Is Responsible for Auxin Attenuation. *Nucleic Acids Research* **12**:1447-1461.

345. **Scott, K.F. and Bender, G.L.** (1990). The *Parasponia-Bradyrhizobium* Symbiosis, p231-251. *In* Gresshoff, P.M. (ed.), *The Molecular Biology of Symbiotic Nitrogen Fixation*, CRC Press Inc., Boca Raton.

346. **Senghas, E., Jones, J.M., Yamamoto, M., Gawron-Burke, C. and Clewell, D.B.** (1988). Genetic Organization of the Bacterial Conjugative Transposon Tn*916*. *Journal of Bacteriology* **170**:245-249.

347. **Sengstag, C. and Arber, W.** (1987). A Cloned DNA Fragment from Bacteriophage P1 Enhances IS*2* Insertion. *Molecular and General Genetics* **206**:344-351.

348. **Shapiro, J.A.** (1969). Mutations Caused by the Insertion of Genetic Material into the Galactose Operon of

Escherichia coli. Journal of Molecular Biology **40:**93-105.

349. Shapiro, J.A. (1979). Molecular Model for the Transposition and Replication of Bacteriophage Mu and Other Transposable Elements. *Proceedings of the National Academy of Sciences USA* **76:**1933-1937.

350. Shapiro, J.A. (1983). Genomic Reorganization in Cell Lineages, pxi-xvi. *In* Shapiro, J.A. (ed.), *Mobile Genetic Elements*, Academic Press Inc., N.Y.

351. Shapiro, J.A. and Adhya, S.L. (1969). The Galactose Operon of *E. coli* K12. II. A Deletion Analysis of Operon Structure and Polarity. *Genetics* **62:**249-264.

352. Shapiro, J.A. and MacHattie, L. (1978). Integration and Excision of Prophage λ Mediated by the IS*1* Element. *Cold Spring Harbor Symposia on Quantitative Biology* **43:**1135-1142.

353. Sharp, P.A., Cohen, S.N. and Davidson, N. (1973). Electron Microscope Heteroduplex Studies of Sequence Relations among Plasmids of *Escherichia coli. Journal of Molecular Biology* **75:**235-255.

354. Sherratt, D. (1989). Tn*3* and Related Transposable Elements: Site Specific Recombination and Transposition, p163-184. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington, D. C.

355. Siehnel, R.J. and Morgan, E.A. (1983). Efficient Readthrough of Tn*9* and IS*1* by RNA Polymerase Molecules That Initiate at rRNA Promoters. *Journal of Bacteriology* **153:**672-684.

356. Simon, R. (1984). High Frequency Mobilization of Gram-negative Bacterial Replicons by the *in vivo* Constructed Tn*5-mob* Transposon. *Molecular and General Genetics* **196:**413-420.

357. Simon, R., Hötte, B., Klauke, B. and Kosier, B. (1991). Isolation and Characterization of Insertion Sequence Elements from Gram-negative Bacteria by Using New Broad-host-range, Positive Selection Vectors. *Journal of Bacteriology* **173:**1502-1508.

358. Simon, R., O'Connell, M., Labes, M. and Pühler, A. (1986). Plasmid Vectors for the Genetic Analysis and Manipulation of Rhizobia and Other Gram-negative Bac-

teria. *Methods in Enzymology* **118**:640-659.

359. **Simons, R.W. and Kleckner, N.** (1983). Translational Control of IS*10* Transposition. *Cell* **34**:683-691.

360. **Smirnov, G.B., Ilyina, T.S., Romanova, Y.M., Markov, A.P. and Nechaeva, E.V.** (1980). Mutants of *Escherichia coli* Affected in the Processes of Transposition and Genomic Rearrangements. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:193-200.

361. **Smith, G.S., Armundsen, S., Chaudhury, A., Cheng, K., Ponticelli, A. and Taylor, A.** (1984). Roles of RecBC Enzyme and Chi Sites in Homologous Recombination. *Cold Spring Harbor Symposia on Quantitative Biology* **49**:485-495.

362. **So, M., Heffron, F. and McCarthy, B.J.** (1979). The *E. coli* Gene Encoding Heat Stable Toxin Is a Bacterial Transposon Flanked by Inverted Repeats of IS*1*. *Nature* **277**:453-456.

363. **So, M. and McCarthy, B.J.** (1980). Nucleotide Sequence of the Bacterial Transposon Tn*1681* Encoding a Heat Stable Enterotoxin and Its Identification in Enterotoxigenic *E. coli* Strains. *Proceedings of the National Academy of Sciences USA* **77**:4011-4015.

364. **Sommer, H., Schumacher, B. and Saedler, H.** (1981). A New Type of IS*1*-mediated Deletion. *Molecular and General Genetics* **184**:300-307.

365. **Southern, E.** (1975). Detection of Specific Sequences among DNA Fragments Separated by Gel Electrophoresis. *Journal of Molecular Biology* **98**:503-517.

366. **Staden, R.** (1990). Finding Protein Coding Regions in Genomic Sequences. *Methods in Enzymology* **183**:163-180.

367. **Staden, R. and McLachlan, A.D.** (1982). Codon Preference and Its Use in Identifying Protein Coding Regions in Long DNA Sequences. *Nucleic Acids Research* **10**:141-156.

368. **Starlinger, P.** (1980). IS Elements and Transposons. *Plasmid* **3**:241-259.

369. **Starlinger, P. and Saedler, H.** (1972). Insertion Mutations in Microorganisms. *Biochimie* **54**:177-185.

370. **Starlinger, P. and Saedler, H.** (1976). IS-elements in Microorganisms. *Current Topics in Microbiology and Immunology* **75**:111-152.

371. **Staudenbauer, W.L.** (1976). Replication of Small Plasmids in Extracts of *Escherichia coli*: Requirements for Both DNA Polymerases I and III. *Molecular and General Genetics* **149**:151-158.

372. **Stibitz, S. and Davies, J.E.** (1978). Tn*602*: A Naturally Occurring Relative of Tn*903* with Direct Repeats. *Plasmid* **17**:202-209.

373. **Stellwagen, N.** (1983). Anomalous Electrophoresis of DNA Restriction Fragments on Polyacrylamide Gels. *Biochemistry* **22**:6186-6193.

374. **Sternglanz, R., DiNardo, S., Voelkel, K.A., Nishimura, Y., Hirota, Y., Becherer, K., Zumstein, L. and Wang, J.C.** (1981). Mutations in the Gene Coding for *Escherichia coli* DNA Topoisomerase I Affect Transcription and Transposition. *Proceedings of the National Academy of Sciences USA* **78**:2747-2751.

375. **Sullivan, K.M. and Lilley, D.M.J.** (1986). A Dominant Influence of Flanking Sequences on a Local Structural Transition in DNA. *Cell* **47**:817-827.

376. **Syvanen, M., Hopkins, J.D. and Clements, M.** (1982). A New Class of Mutants in DNA Polymerase I That Affects Gene Transposition. *Journal of Molecular Biology* **158**:203-212.

377. **Syvanen, M., Hopkins, J.D., Griffin, T.J., Liang, T.-Y., Ippen-Ihler, K. and Kolodner, R.** (1986). Stimulation of Precise Excision and Recombination by Conjugal Proficient F′ Plasmids. *Molecular and General Genetics* **203**:1-7.

378. **Takeya, T., Nomiyama, H., Miyoshi, J., Shimada, K. and Takagi, Y.** (1979). DNA Sequences of the Integration Sites and Inverted Repeated Structures of Transposon Tn*3*. *Nucleic Acids Research* **6**:1831-1841.

379. **Taylor, A.L.** (1963). Bacteriophage-induced Mutations in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* **50**:1043-1051.

380. **Tetsuro, N.** (1991). Mini-maxi Prep: An Efficient and Inexpensive Procedure for Medium Scale Preparation of High Quality Supercoiled Plasmid DNA. *Technique* **3**:33-36.

381. **Toussaint, A. and Resibois, A.** (1983). Phage Mu: Transposition as a Life-style, p105-158. *In* Shapiro,

J.A. (ed.), *Mobile Genetic Elements*, Academic Press Inc., N.Y.

382. **Trieu Cuot, P. and Courvalin, P.** (1984). Nucleotide Sequence of the Transposable Element IS*15*. *Gene* **30**:113-120.

383. **Trieu Cuot, P. and Courvalin, P.** (1985). Transposition Behaviour of IS*15* and Its Progenitor IS*15*-delta: Cointegrates Are Exclusive End Products. *Plasmid* **14**:80-89.

384. **van de Putte, P., Cramer, S. and Giphart-Gassler, M.** (1980). Invertible DNA Determines Host Specificity of Bacteriophage Mu. *Nature* **286**:218-222.

385. **Vanderleyden, J., Desair, J., De Meirsman, C., Michiels, K., Van Gool, A.P., Dell Chitton, A.P. and Jen, G.C.** (1986). Nucleotide Sequence of an Insertion Sequence (IS) Element Identified in the T-DNA Region of a Spontaneous Variant of the Ti-Plasmid pTiT37. *Nucleic Acid Research* **14**:6699-6709.

386. **Varmus, H. and Brown, P.** (1989). Retroviruses, p53-108. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington, D. C.

387. **Vieira, J. and Messing, J.** (1982). The pUC Plasmids, an M13mp7-derived System for Insertion Mutagenesis and Sequencing with Synthetic Universal Primers. *Gene* **19**:259-268.

388. **Wabiko, H.** (1992). Sequence Analysis of an Insertion Element, IS*1131*, Isolated from the Nopaline-type Ti Plasmid of *Agrobacterium tumefaciens*. *Gene* **114**:229-233.

389. **Wallace, L.J., Ward, J.M., Bennett, P.M., Robinson, M.K. and Richmond, M.H.** (1980). Transposition Immunity. *Cold Spring Harbor Symposia on Quantitative Biology* **45**:183-188.

390. **Watanabe, T. and Ogata, Y.** (1970). Genetic Stability of Various Resistance Factors in *Escherichia coli* and *Salmonella typhimurium*. *Journal of Bacteriology* **102**:363-368.

391. **Watson, R.J. and Wheatcroft, R.** (1991). Nucleotide Sequence of *Rhizobium meliloti* Insertion Sequence IS*Rm1*: Homology to IS*2* from *Escherichia coli* and IS*426* from *Agrobacterium tumefaciens*. *DNA Sequence* **2**:163-172.

392. **Weinert, T.A., Schaus, N.A. and Grindley, N.D.F.** (1983). Insertion Sequence Duplication in Transpos-

itional Recombination. *Science* **222**:755-765.

393. **Wells, R.G. and Grindley, N.D.F.** (1984). Analysis of the *γδ res* Sites Required for Site-specific Recombination and Gene Expression. *Journal of Molecular Biology* **179**:667-681.

394. **Wetmur, J.G.** (1991). DNA Probes: Applications of the Principles of Nucleic Acid Hybridization. *Critical Reviews in Biochemistry and Molecular Biology* **26**:227-259.

395. **Wheatcroft, R. and Laberge, S.** (1991). Identification and Nucleotide Sequence of *Rhizobium meliloti* Insertion Sequence IS*Rm3*: Similarity between the Putative Transposase Encoded by IS*Rm3* and Those Encoded by *Staphylococcus aureus* IS*256* and *Thiobacillus ferrooxidans* IS*T2*. *Journal of Bacteriology* **173**:2530-2538.

396. **Wheatcroft, R. and Watson, R.J.** (1987). Identification and Characterization of Insertion Sequence IS*Rm1* in *Rhizobium meliloti* JJ1c10. *Canadian Journal of Microbiology* **33**:314-321.

397. **Wheatcroft, R. and Watson, R.J.** (1988). Distribution of Insertion Sequence IS*Rm1* in *Rhizobium meliloti* and Other Gram-negative Bacteria. *Journal of General Microbiology* **134**:113-121.

398. **Wheatcroft, R. and Watson, R.J.** (1988). A Positive Strain Identification Method for *Rhizobium meliloti*. *Applied and Environmental Microbiology* **54**:574-576.

399. **Willets, N.S., Crowther, C. and Holloway, B.W.** (1981). The Insertion Sequence IS*21* of R68.45 and the Molecular Basis for Mobilization of the Bacterial Chromosome. *Plasmid* **6**:30-52.

400. **Willets, N. and Wilkins, B.** (1984). Processing of Plasmid DNA during Bacterial Conjugation. *Microbiological Reviews* **48**:24-41.

401. **Winter, R., Berg, O. and von Hippel, P.** (1981). Diffusion-driven Mechanisms of Protein Translocation on Nucleic Acids. 3: The *Escherichia coli lac* Repressor-operator Interaction: Kinetic Measurements and Conclusions. *Biochemistry* **20**:6961-6969.

402. **Wolfe, R.E. Jr.** (1980). Integration of Specialized Transducing Bacteriophage λ *c*I857 *St68 h80 dgno his* by an Unusual Pathway Promotes Formation of Deletions and

Generates a New Translocatable Element. *Journal of Bacteriology* **142**:588-602.

403. Yin, J.C.P., Krebs, M.P. and Reznikoff, W.S. (1988). The Effect of Dam Methylation on Tn*5* Transposition. *Journal of Molecular Biology* **199**:35-45.

404. Yin,, J.C.P. and Reznikoff, W.S. (1987). *dnaA*, an Essential Host Gene, and Tn*5* Transposition. *Journal of Bacteriology* **169**:4637-4645.

405. York, M.K. and Stodolsky, M. (1981). Characterization of P1 *argF* Derivatives from *E. coli* K12 Transduction. I. IS*1* Elements Flank the *argF* Gene Segment. *Molecular and General Genetics* **181**:230-240.

406. Zambryski, P. (1989). *Agrobacterium*-Plant Cell DNA Transfer, p309-333. *In* Berg, D.E. and Howe, M.M. (ed.), *Mobile DNA*, American Society for Microbiology, Washington D.C.

407. Zambryski, P., Depicker, A., Kruger, K. and Goodman, H.M. (1982). Tumor Induction by *Agrobacterium tumefaciens*: Analysis of the Boundaries of T-DNA. *Journal of Molecular and Applied Genetics* **1**:361-370.

408. Zambryski, P., Holsters, M., Kruger, K., Depicker, A., Schell, J., Van Montagu, M. and Goodman, H.M. (1980). Tumor DNA Structure in Plant Cells Transformed by *A. tumefaciens*. *Science* **209**:1385-1391.

409. Zerbib, D., Gamas, P., Chandler, M., Prentki, P., Bass, S. and Galas, D. (1985). Specificity of Insertion of IS*1*. *Journal of Molecular Biology* **185**:517-524.

410. Zerbib, D., Jakowec, M., Prentki, P., Galas, D. and Chandler, M. (1987). Expression of Proteins Essential for IS*1* Transposition: Specific Binding of InsA to the Ends of IS*1*. *EMBO Journal* **6**:3163-3169.

411. Zerbib, D., Polard, P., Escoubas, J.M., Galas, D. and Chandler, M. (1990). The Regulatory Role of the IS*1*-encoded InsA Protein in Transposition. *Molecular Microbiology* **4**:471-477.