ISOLATION AND DNA SEQUENCE ANALYSIS OF A *RHIZOBIUM LOTI* GENE REQUIRED FOR EFFECTIVE NODULATION OF *LOTUS PEDUNCULATUS*

A thesis presented in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy in Microbiology at
Massey University, Palmerston North, New Zealand

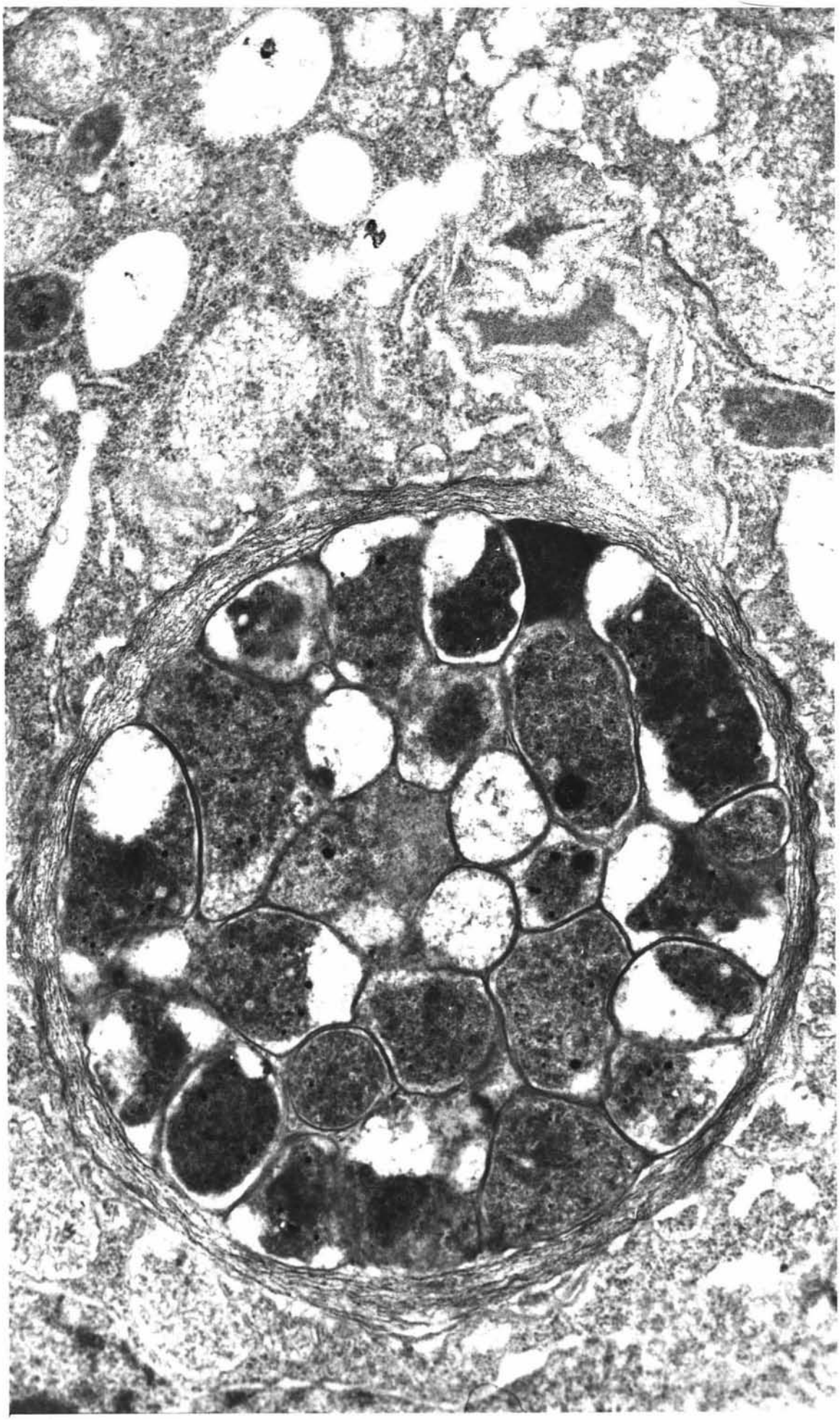
LAWRENCE JAMES HENRY WARD

1989

FRONTISPIECE

Transverse section through a swollen infection thread in a *Lotus pedunculatus* nodule. The plant was inoculated with the *Rhizobium loti* mutant strain PN239.

Magnification approximately 3500 times
Photograph courtesy of Clive Pankhurst



ERRATA

p4	line 17	'Spanik' (twice) should be 'Spaink'
p11	line 7	'is indicated' should be 'is indicated by X'
p17	line 19	'Sodium chloride, 0.5' should be 'Sodium chloride, 5.0'
p18	line 14	'*lac*I^9' should be '*lac*I^q'
p20	line 17	'$MnCl_2.4Cl_2O$' should be '$MnCl_2.4H_2O$'
p22	line 21	'Tn5' should be 'Tris'
p27	line 12	'0.5 µg/ml' should be '0.5 mg/ml'
p33	line 14	'BC1G' should be 'BCIG'
p36	line 1	'Whitfield' should be 'Whitfeld'
p51	line 4	'*bar*$^-$' should be '*bar*'
p54	line 6	'(2 µl/ml)' should be '(2 µg/ml)'
p54	line 21	Insert '(Fig. 14)' after '*Xho*I'
p66	Fig. 18	Add 'Vertical lines indicate the location of Tn5 insertions which result in a Fix^+ (longer vertical line) and Fix^- (shorter vertical line) phenotype on *Lotus pedunculatus*'
p76	Fig. 22	'(arrowed)' should be '(two examples of which are arrowed)'
p77	Fig. 23	'the sequence read from them (B)' should be 'part of the sequence read from them (B)' Add 'The sequences given in B (left to right) read from bottom to top in the segments of gels from the + strand and from top to bottom in the segments of gels from the - strand. Due to poor reproduction not all bands are readily visible.'
p81	line 7	'cosmids' should be 'plasmids'
p81	line 33	'base repeat' should be 'base direct repeat'
p83	Fig. 26	'sequence of two Tn5 - *Rhizobium* junctions.' should be 'sequences of Tn5 - *Rhizobium* DNA junctions from two different mutants.'

ABSTRACT

A *Rhizobium loti* gene required for effective nodulation of the host *Lotus pedunculatus* has been identified by transposon Tn5 mutagenesis. Cosmids from a *R. loti* gene library which complemented a previously isolated mutant strain, PN239, (Chua *et al* 1985; J. Bacteriol. 162; 335-343) at this locus were identified by *in planta* complementation. A physical map of these cosmids was constructed and the site of insertion of the Tn5 was mapped to a 7.5 kb *Eco*RI fragment common to all cosmids which complemented the mutation. This 7.5 kb *Eco*RI fragment was subcloned into pBR328 and pLAFR1 and a more detailed physical map constructed. The 7.5 kb *Eco*RI fragment in pLAFR1 was able to complement the Tn5 mutation when introduced into strain PN239.

Further Tn5 mutagenesis of the 7.5 kb *Eco*RI fragment was carried out in *E. coli* and the mutations were homogenotised into *R. loti* NZP2037. Three additional mutations were isolated which caused a Fix^- phenotype on *Lotus pedunculatus*. The Tn5 inserts which caused a Fix^- phenotype were mapped to positions adjacent to the position of the original mutation in strain PN239. All other Tn5 insertions isolated in the 7.5 kb *Eco*RI fragment gave a Fix^+ phenotype on *Lotus pedunculatus*.

A region was sequenced which was involved in effective nodulation of *Lotus pedunculatus* as indicated by the position of the Tn5 insertions. Analysis of the consensus sequence of 2307 bases identified a potential open reading frame (ORF) of 576 base pairs, coding for a putative protein of 21.2 kD. The positions of the Tn5 insertions causing a Fix^- phenotype and the adjacent Tn5 insertions which did not affect fixation were determined in the sequence. The position and orientation of the ORF identified was consistent with the sequenced positions of these Tn5 insertions.

A fragment containing most of the ORF identified from the sequence was used as a hybridization probe to various strains of rhizobia. Homology was only demonstrated with DNA from other *R. loti* strains. *R. loti* strains containing Tn5 insertions which were Fix$^-$ on *Lotus pedunculatus* were found to be fully effective on *Lotus corniculatus*. These observations suggest that the gene characterised in this investigation may be involved in the host specificity of *R. loti* for *Lotus pedunculatus.*

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Assoc. Prof. B.D.W. Jarvis and Prof. D.B. Scott for their encouragement and guidance throughout this project.

I would also like to thank Prof. D.F. Bacon for his interest and encouragement.

Thanks are also due to the academic, technical and secretarial staff (past and present) of the Department of Microbiology and Genetics for their help and interest; in particular Assoc. Prof. T.J. Brown for use of space in his laboratory and Dr B. Mansfield for assistance with the Department's computers.

The co-operation, assistance and encouragement of my fellow postgraduate students throughout the course of this project has been appreciated (I'm not going to mention names - I'm sure to miss somebody!).

Thanks are also due to the staff of the DSIR computer unit and to Mr Mark Pritchard (DSIR) for help with the VAX computer and Staden programs and to the staff of the Massey University Computer Centre, in particular Mr Glen Eustace, for their co-operation in transferring data and assistance with using the Prime computer.

I would also like to thank my friends and colleagues at the N.Z. Dairy Research Institute for their interest and encouragement during the final stages of this thesis.

The co-operation and assistance of Mr Paul Le Ceve (DRI) and Mrs Veronica Fieldsend with the photographs and typing respectively is gratefully acknowledged.

TABLE OF CONTENTS

Table of Contents Cont'd.

Table of Contents Cont'd.

Table of Contents Cont'd.

Table of Contents Cont'd.

LIST OF TABLES

List of Tables Cont'd.

LIST OF FIGURES

List of Figures Cont'd.

List of Figures Cont'd.

List of Figures Cont'd.

INTRODUCTION

1.1 BIOLOGICAL NITROGEN FIXATION AND ITS IMPORTANCE

The process by which atmospheric nitrogen is fixed by microorganisms is important in increasing the availability of fixed nitrogen for plant growth. In New Zealand agriculture the most important aspect of biological nitrogen fixation is the symbiotic association between the genus *Rhizobium* and the roots of leguminous plants, especially clover and lotus species. Mackinnon *et al.*, (1975) estimated that 97% of the nitrogen used in New Zealand is fixed biologically. It is clear that the agricultural economy of countries such as New Zealand and Australia depends largely on the process of biological nitrogen fixation. In the absence of this process the increased cost of producing and using extra nitrogenous fertilisers would be prohibitive. Additionally, the extension of agricultural practices involving symbiotic nitrogen fixation to underdeveloped countries has the potential to substantially improve soil fertility and thus animal and human nutrition. Furthermore, such improvements tend to be self-sustaining and dispense with the need for continued high cost inputs to maintain productivity. Comparatively small increases in the efficiency of the *Rhizobium*-legume symbiosis could therefore increase the productive capacity and profitability of some agricultural economies.

The genetics of *Rhizobium* and the genetic basis of the symbiotic association with legumes have been the subject of considerable research; with much of the effort focussed on understanding the mechanism by which symbiotic nitrogen fixation occurs and how it is controlled.

1.2 SIGNIFICANCE OF *LOTUS* spp. IN NEW ZEALAND AGRICULTURE

The *Rhizobium* strain studied in this thesis nodulates plants of the genus *Lotus*. There are five main species of *Lotus* in New Zealand (Smetham, 1973). Two are annual (*Lotus angustissimus* and *Lotus subbiflorus*), while the three perennial species are of more importance in New Zealand agriculture. These are *L. tenuis* (slender birdsfoot trefoil), *L. corniculatus* (common birds-foot trefoil) and *L. pedunculatus* (Lotus major, big trefoil). *Lotus pedunculatus* Cav. "Grasslands Maku" is the most commonly sown cultivar of *Lotus* in New Zealand.

Under conditions of low fertility especially low levels of phosphate and potassium which are common in hill country pastures, *Lotus* species persist and can compete more successfully than other legume species (Smetham, 1973). There is a range of tolerance to acidity or alkalinity, depending on the *Lotus* species, which affects both plant growth and nodulation (Smetham, 1973). Under conditions of high fertility and good drainage, white and red clovers and lucerne grow better than *Lotus* sp., however where conditions such as high water table, poor drainage, high salinity or low nutrient levels prevail *Lotus* sp. frequently give superior yields in addition to superior persistence (Seaney and Henson, 1970).

1.3 CHARACTERISTICS OF RHIZOBIA WHICH NODULATE *LOTUS* spp.

Rhizobia which nodulate *Lotus* species include both fast and slow growing strains (Greenwood and Pankhurst, 1977). The fast growing strains are classified as *Rhizobium loti* (Jarvis *et al.*, 1982) while the slow growing strains are classified as *Bradyrhizobium* sp. (*Lotus*) (Jordan, 1982). Most *R. loti* strains, including the type strain NZP2213, form effective nitrogen fixing nodules (Nod^+ Fix^+) on *Lotus corniculatus* and

Lotus tenuis, but ineffective nodules (Nod^+ Fix^-) on *Lotus pedunculatus*. However some *R. loti* strains, including strain NZP2037, are able to nodulate all three *Lotus* species effectively (Greenwood and Pankhurst, 1977; Pankhurst, 1977). The slow-growing *Bradyrhizobium* sp. (*Lotus*), such as strain CC814S, form effective nodules on *L. pedunculatus* but ineffective nodules on *L. corniculatus* and *L. tenuis* (Greenwood and Pankhurst, 1977). It is interesting to note that *R. loti* and *Bradyrhizobium* sp. (*Lotus*) share less than 5% total DNA homology (Crow *et al.*, 1981) even though they are able to nodulate similar plant species.

In most of the *Rhizobium* strains examined (reviewed by Kondorosi *et al.*, 1984), the genetic material for nitrogen fixation (*nif*) and nodulation (*nod*) is carried on a large plasmid. However, in *Rhizobium loti* strains it is localised on the chromosome (Pankhurst *et al.*, 1983; Chua *et al.*, 1985)

1.4 NODULE DEVELOPMENT

The formation of a nodule and establishment of a nitrogen fixing symbiosis between rhizobia and leguminous plants is a complex developmental process involving the expression of both rhizobial and plant genes. This process has been extensively reviewed (Dart, 1977; Goodchild, 1978; Robertson and Farnden, 1980; Newcomb, 1981).

The first stage of nodule formation is the infection of the root by rhizobia. This is influenced by environmental conditions (e.g. temperature, pH, nutrients), microbial factors (e.g. number of rhizobia, competition between rhizobial strains, specificity of rhizobia) and plant factors (e.g. cultivar, stage of growth) (Newcomb, 1981).

Lectins (carbohydrate-binding proteins) are produced by the roots of many leguminous plants and it has been proposed that they have a role in cell surface interaction between the

plant and the bacteria. Several *Rhizobium* species produce cell surface molecules which specifically bind to lectin molecules secreted by their host plants (Dazzo *et al.*, 1985; Halverson and Stacey, 1985; Halverson and Stacey, 1986(a); Diaz *et al.*, 1989; Philip-Hollingsworth *et al.*, 1989). While cell surface interaction involving bacteria and plant lectins may have a role in the nodulation process this is not the only basis of host specificity.

Flavonoid type compounds exuded by the host roots can act as inducers of nodulation genes, as a result of their interaction with the *nod*D gene product which regulates the expression of other *nod* genes (Innes *et al.*, 1985; Mulligan and Long, 1985; Rossen *et al.*, 1985; Peters *et al.*, 1986; Redmond *et al.*, 1986; Zaat *et al.*, 1987).

Thus the host specificity of the symbiosis is partly determined by the *nod*D gene and the inducers present in the root exudate (Spanik *et al.*, 1987; Spanik *et al.*, 1989; De Maagd *et al.*, 1989).

Once rhizobia have attached to the surface of root hairs, deformation (curling or branching) (Dart, 1977) of the root hairs is observed. Infection proceeds by means of an infection thread which develops within the root hair at the point of contact with attached rhizobia. Observations that *R. trifolii* cell surface polysaccharides (Capsular polysaccharide, lipopolysaccharide, ß-glucans, exopoly-saccharides) enhance the number of infection threads and nodules formed in white clover (Abe *et al*, 1984; Dazzo *et al.*, 1984(b)) suggest that *Rhizobium* cell surface polysaccharides may function as signal molecules which induce host responses for infection thread formation (Halverson & Stacey, 1986). ß-2-glucan has been shown to be involved in infection thread formation in white clover (Abe *et al.*, 1982; Dylan *et al.*, 1986; Higashi & Abe, 1980) and it is possible that ß-2-glucan may be one of the signal molecules necessary for infection thread formation (Halverson & Stacey, 1986(b)).

Extracellular polysaccharides may have roles in the invasion of the nodule by rhizobia through their interaction with plant genes or gene products causing root hair deformation and curling. Lipopolysaccharide has been postulated to have a role in infection thread development and in the release of rhizobia from the infection thread (De Maagd *et al.*, 1989). Mutants with altered lipopolysaccharides have been isolated which are ineffective (Carlson *et al.*, 1987; Noel *et al.*, 1986), however in some rhizobia alterations to the lipopolysaccharide structure do not appear to affect nitrogen fixation (Clover *et al.*, 1989). This is dependent on the plant genetic background (Clover *et al.*, 1989) and reinforces the concept that the formation of an effective nodule is a complex process involving interaction of both bacterial and plant genes.

Degradation of the plant cell wall is observed at the site of infection thread formation (Callaham and Torrey, 1981). The infection thread, which is of plant origin, grows toward the base of the root hair. The infection thread wall encloses a matrix containing the invading rhizobia (Newcomb, 1981). Infection thread growth and wall deposition are the result of a tightly regulated balance between the plant and bacteria; any imbalance may result in infection thread abortion (Callaham and Torrey, 1981). The presence of rhizobia on the surface of the root hair induces division of the root outer cortical cells (Calvert *et al.*, 1984). Penetration of the root hair-cortical cell wall junction by the growing infection thread involves localized cell wall degradation followed by infection thread cell deposition such that the infection thread continues into the cortical cell (Newcomb, 1981). Within the root cortical cells the infection thread may grow both intracellularly and intercellularly; where growth is intercellular, the invading rhizobia and infection thread matrix occupy the intercellular spaces and utilize the existing cell walls as the infection thread wall, hence no new infection thread wall deposition is required (Newcomb, 1981).

During the growth of the infection thread through the root outer cortex, meristematic activity is initiated in the root cortex directly in front of the tip of the infection thread. Infection thread growth continues into this meristematic region where the invading rhizobia are released from the tip of the infection thread into the innermost meristematic cells (Goodchild, 1978). Mutants of *B. japonicum* in which bacteria are not released from the infection thread but are degraded within the infection thread (Noel *et al.,* 1982) have lead to the suggestion (Halverson and Stacey, 1986) that bacterial release is not a passive process but requires a mechanism initiated by either the plant or bacteria for proper release into the cortical cell.

Rhizobia are released from the tip of the infection thread by endocytosis (Dart, 1977; Goodchild, 1978) which results in the bacteria being enclosed within the infection thread membrane (at this stage called the peribacteroid membrane (Robertson *et al.*, 1978)) in the cytoplasm of the meristematic cells. The tip of the infection thread contains little or no infection thread wall material (Goodchild and Bergersen, 1966; Robertson *et al.,* 1978). Robertson *et al.*, (1978) proposed that this lack of wall material in the tips of infection threads results from a change in the function of the Golgi body system for the synthesis of plant cell walls and plasma membranes such that the rate of membrane synthesis remained constant or increased while the rate of cell wall material synthesis and deposition decreased (Goodchild and Bergersen, 1966). Increased cellulolytic and pectolytic activity in the tips of infection threads in meristematic cells has also been suggested (Verma *et al.*, 1978 a,b; Robertson *et al.*, 1978). The regulatory mechanism controlling these processes is not known.

After the rhizobia have been released from the tip of the infection thread they continue to divide until the cytoplasm of the plant cortical cells is filled with bacteroids which

are extracellular to the cortical cell cytoplasm since they are enclosed within the plant derived peribacteroid membrane.

All legumes form a meristematic zone in the root cortex following infection of the root hairs by rhizobia, however the subsequent development of this zone differs between legume species resulting in differences in nodule shape (Goodchild, 1978; Robertson & Farnden, 1980; Newcomb, 1981).

1.4.1 Nodule Formation in *Lotus*

The nitrogen-fixing root nodules formed on *Lotus* spp. by *R. loti* strains are spherical and initially appear on the root 10-12 days after inoculation (Pankhurst *et al.*, 1979). The nodules contain a central zone of plant cells, many of which are invaded by infection threads and subsequently contain bacteroids. Surrounding the central zone are several layers of vacuolate cells which become differentiated into a nodule cortex and nodule vascular bundles.

Ineffective tumour-like structures (for example those induced by NZP2213 on *Lotus pedunculatus*) contain a central zone of actively dividing cells surrounded by a cortex (Pankhurst *et al.*, 1979), however the rhizobia do not infect the cells.

1.5 GENETICS OF NODULATION AND NITROGEN FIXATION

Development of a nodule and the subsequent fixation of nitrogen is a complex multistep process which requires the induction of many *Rhizobium* genes. Most work has been done on the genetics of *R. leguminosarum* and *R. meliloti* (reviewed by Kondorosi *et al.*, 1984). In these organisms most of the genes involved in the symbiotic process have been located on large plasmids, termed symbiotic plasmids (pSym) (Higashi,

1967; Prakash *et al.*, 1978; Casse *et al.*, 1979; Nuti *et al.*, 1979; Denarie *et al.*, 1981; Prakash *et al.*, 1981; Nuti, 1977), which range in size from 180 kilobase pairs (kb) in *R. trifolii* to 1600 kb in *R. meliloti* (Burkardt and Burkardt, 1984). The symbiotic genes in *R. loti* are not carried on plasmids, instead they are located on the chromosome (Pankhurst *et al.*, 1983; Chua *et al.*, 1985). The symbiotic genes in *R. loti* obviously have a similar function and have structural homology with the symbiotic genes in *R. leguminosarum* and *R. meliloti* (D.B. Scott, personal communication). Therefore, I will briefly review the genetics of nodulation and nitrogen fixation in *R. leguminosarum* and *R. meliloti* with the assumption that much of this material is relevant to *R. loti*.

Symbiotic mutants of *Rhizobium* can be divided phenotypically into two groups; those which are nodulation deficient (Nod^-) and those which are able to form nodules but unable to fix atmospheric nitrogen (Fix^-). Nod^- mutants fail to form nodules due to defects affecting processes which occur before the induction of a visible nodule meristem. Fix^- mutants are inhibited in later steps of the symbiotic process - they nodulate the host plant but fail to fix nitrogen, or do so at a lower level than the wild type. Symbiotic genes in *Rhizobium* fall into many different genotypic classes depending on the stage at which nodulation is blocked. Most well characterized are the *nod* genes (required in the early steps of nodule initiation) and the *nif* and *fix* genes (required for the process of nitrogen fixation). However, many other genes are involved at other stages; for a review see Triplett *et al.*, (1989). Vincent, (1980) proposed a phenotypic code for the detailed classification of symbiotic mutants. This proposed code is outlined in Table 1. Mutations up to and including step 4 would result in a Nod^- phenotype whereas mutations from step 5 on would be Nod^+ but either fully or partially Fix^-.

Table 1: Proposed steps in Rhizobium-legume symbiosis (Vincent, 1980)

Sequence of events	Phenotypic code
(I) Preinfection	
(1) Root colonization	Roc
(2) Root adhesion	Roa
(3) Hair branching	Hab
(4) Hair curling	Hac
(II) Infection and nodule formation	
(5) Infection	Inf
(6) Nodule initiation	Noi
(7) Bacterial release	Bar
(8) Bacteroid development	Bad
(III) Nodule function	
(9) Nitrogen fixation	Nif
(10) Complementary function	Cof
(11) Nodule persistence	Nop

The number of symbiotic genes which are actually required for the successful development of a nitrogen fixing symbiosis by any *Rhizobium* strain is not known. New genes and loci essential for symbiotic nitrogen fixation continue to be identified (Aguilar *et al.*, 1987; Renalier *et al.*, 1987). In *R. leguminosarum* bv *trifolii* it has been established that in addition to the chromosomal genes, less than 14kb of the 180kb symbiotic plasmid are required for nodule formation on clover roots (Schofield *et al.*, 1984). Strains of *R. leguminosarum* bv *trifolii* without pSym but containing this 14kb nodulation region can form nodules to the stage where bacteria are released from infection threads into the cortical cells. However the released bacteria do not form

bacteroids and cannot fix nitrogen since the *nif* structural genes are absent. Innes *et al.*, (1988) have subcloned a 32kb region of pSym which is able to induce nodulation and fixation, but at a lesser level than the wild type strain.

1.5.1 Nodulation Genes

Bacterial genes involved in nodulation can be divided into two broad groups; firstly those genes involved in determining the basic properties of nodulation which could be expected to be conserved between *Rhizobium* species, and secondly, those genes involved in determining the high degree of host specificity observed in rhizobia. Most of the work in this area has been on *R. meliloti* and *R. leguminosarum*. In these species at least, the nodulation genes form two clusters (Kondorosi *et al.*, 1984). The first of these includes the so called "common" *nod* genes. The arrangement of these genes in various *Rhizobium* species studied is shown in Figure 1. *Nod* A, B and C genes are required for root hair curling (Hac) (Rossen *et al.*, 1984). These genes are conserved and are functionally equivalent in different species since mutations in *nod* A, B and C can be corrected by the corresponding genes from other species (Fisher *et al.*, 1985).

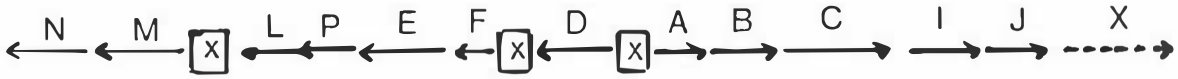
Figure 1:

Arrangement and direction of transcription of nodulation (*nod*) genes in:

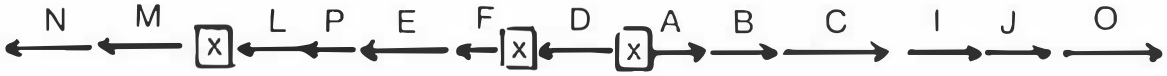
1. *Rhizobium leguminosarum* bv. *viciae*
2. *Rhizobium leguminosarum* bv. *trifolii*
3. *Rhizobium meliloti*

The position of "*nod* box" sequences is indicated.

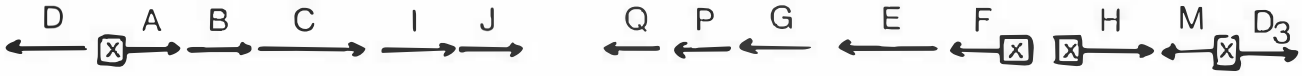
1


N
M
X
L
P
E
F
X
D
X
A
B
C
I
J
X

2


N
M
X
L
P
E
F
X
D
X
A
B
C
I
J
O

3


D
X
A
B
C
I
J
Q
P
G
E
F
X
X
H
M
X
D3

Downstream of *nod*C are two genes, *nod*I and J. *Nod*I is similar to inner membrane ATP-dependent transport proteins of

E. coli and *S. typhimurium* (Evans and Downie, 1986) and *nod*J is hydrophobic suggesting that it is a membrane protein (Evans and Downie, 1986). It has been postulated that the *nod*CIJ gene products may form a membrane associated transport system (Evans and Downie, 1986). The nature of the molecule transported has yet to be established, however Evans and Downie, (1986) have suggested that it may be a plant metabolite exuded from plant roots which is taken up by the *Rhizobium* and modified so that it can interact with plant root hairs and cause root hair curling.

*Nod*D is adjacent to and transcribed divergently from *nod*A. Mutations in *nod*D in *R. leguminosarum* prevent nodulation (Downie *et al.,* 1985; Schofield and Watson, 1986), *nod*D mutations in *R. meliloti* cause only a delay in nodulation because *R. meliloti* has multiple copies of *nod*D on the sym plasmid (Gottfert *et al.,* 1986). *Nod*D has a regulatory function and is responsible for activating transcription of other *nod* genes. Upstream of *nod* genes A, F, M and H is a 35 base pair conserved sequence, the "nod-box" which is involved in the activation of *nod* genes by the *nod*D gene product (Rostas *et al.,* 1986). *Nod*D has an important role in the determination of host specificity (see Section 1.4 Nodule Development).

Mutations in *nod*E and F inhibit nodulation but cannot be corrected by the corresponding genes from other species (Schofield and Watson, 1986; Shearman *et al.,* 1986; Horvath *et al.,* 1986), which implies that they are involved in the determination of host specificity. These genes have also been termed *hsn*A and *hsn*B (host specific nodulation, Kondorosi *et al.*, 1984 and Bachem *et al.*, 1986). Transfer of these genes also transfers host range specificity (Horvath *et al.,* 1986). Mutations in *nod*E and F inhibit nodulation of

clover by *R. leguminosarum* bv. *trifoli* but allow nodulation of peas (Djordjevic *et al.*, 1985). One explanation suggested is that the products of *nod*E and F modify some cell surface component of *R. leguminosarum* strains which enhances recognition of the normal host but causes non-host legumes to recognize the *Rhizobium* as foreign and thus reject them (Rossen *et al.*, 1987). Two other genes, *nod*G and H (*hsn*C and D, Kondorosi and Kondorosi, 1986) from *R. meliloti*, are involved in host range specificity. These genes do not appear to have homologous genes in *R. leguminosarum* which implies that they are required specifically for nodulating *R. meliloti* hosts (Horvath *et al.*, 1986). Host range specific genes have also been identified in *R. leguminosarum*. *Nod*X confers the ability to nodulate primitive peas (cv. Afghanistan) (Gotz *et al.*, 1985; Davis *et al.*, 1988). *Nod*K is found in the slow-growing rhizobia which nodulate Parasponia and is absent from fast growing *Rhizobium* strains. Other *nod* genes L, M and N have been identified from analysis of DNA sequences. The significance and function of these three genes is not known.

1.5.2 Nitrogen Fixation Genes

Nitrogen fixation (*nif*) genes in *Rhizobium* were identified by hybridization with hetrologous DNA containing the *nif* genes from *Klebsiella pneumoniae,* a free-living nitrogen-fixing bacterium. In *K. pneumoniae* there are 20 genes whose products appear to be required for fully active nitrogen fixation (Roberts and Brill, 1981; Triplett *et al.*, 1989). These genes are organised into eight transcriptional units and occupy a contiguous region of about 23kb in the *K. pneumoniae* genome.

Nitrogenase protein 1 is encoded by *nif*DK and protein 2 by *nif*H. *Nif*Q is involved in molybdenum transport. Five genes (*nif*B, *nif*V, *nif*N, *nif*E and *nif*H) are involved in processing protein 1 while *nif*M plays a role in the processing of protein 2. Electron transport to the nitrogenase is mediated by the products of *nif*F and *nif*J. Regulation is mediated by the *nif*LA gene product. The roles of *nif*T, *nif*Y, *nif*X, *nif*U, *nif*S, *nif*W and *nif*Z have not yet been established. Genes homologous to many of these *nif* genes have been identified in various *Rhizobium* strains (Aguilar *et al.*, 1987; Corbin *et al.*, 1982; Rossen *et al.*, 1984; Szeto *et al.*, 1984). By convention these *Rhizobium* genes have been given the same name as their homologs in *K. pneumoniae.*

In addition to these *nif* genes, a number of adjacent genes, which do not appear to have homologous sequences in *K. pneumoniae,* are also required for nitrogen fixation in *Rhizobium*. These genes have been designated *fix* genes and have been identified by site-directed transposon mutagenesis of fragments known to contain *nif* genes (Ruvkun *et al.*, 1982) or by random Tn5 mutagenesis of entire Sym plasmids (Buchanan-Wollaston *et al.*, 1980). Both of these approaches could easily overlook essential *fix* genes. Site-directed mutagenesis only examines a relatively small part of pSym, while mutagenising the entire symbiotic plasmid makes it impracticable to saturate the DNA with insertions. Additionally, if *fix* or *nif* genes are repeated within the genome (Quinto *et al.*, 1985; Renalier *et al.*, 1987) a single Tn5 insertion will not be effective in inactivating and thus identifying the gene. In *R. loti* where the *nif* and *fix* genes are chromosomal rather than plasmid-borne these problems are magnified.

A number of Fix$^-$ mutants have been isolated which form empty nodules which have either no infection threads (Chua *et al.*, 1985; Finan *et al.*, 1985; Leigh *et al.*, 1985; Vandenbosch *et al.*, 1985; Dylan *et al.*, 1986) or contain aborted infection threads (Noel *et al.*, 1986). This demonstrates that the process of nodule formation can take place without invasion of the plant host by the rhizobia. These Fix$^-$ mutants have an altered composition of exopolysaccharides (Finan *et al.*, 1985; Leigh *et al.*, 1985), lipopolysaccharides (Noel *et al.*, 1986) or altered ß-glucan synthesis (Dylan *et al.*, 1986). The role of these polysaccharide fractions in the infection process is still unknown, although they appear to have a role in bacterial invasion of the nodule.

Genes required for bacterial release from the infection thread and genes involved in the subsequent steps culminating in the differentiation of the rhizobia into the bacteroid state are not well described. A number of mutants have been isolated in which bacterial release into the developing nodule fails to occur (Pankhurst *et al.*, 1972; Pankhurst, 1974; Truchet *et al.*, 1980; Chua *et al.*, 1985; Morrison and Verma, 1987). One of these genes, from *Bradyrhizobium japonicum* has recently been cloned and sequenced (Morrison and Verma, 1987).

1.6 BACKGROUND AND AIMS OF THIS INVESTIGATION

Chua et al., (1985) used Tn5 mutagenesis to isolate a number of mutants in *R. loti* strain NZP2037 which were defective in aspects of nodulation and/or nitrogen fixation. One of these mutants, PN239, appeared to be blocked in bacterial release (Bar$^-$) of the rhizobia from the infection thread into the plant cell. Swollen infection threads packed with rhizobia were observed by electron microscopy but there was no evidence for release into the plant cortical cells and hence

no nitrogen fixation (Chua *et al.,* 1985).

Chua *et al.,* (1985) isolated from a pLAFR1 gene library of NZP2037 three cosmids which complemented the above mutation. The aim of this investigation was to examine the gene(s) involved in the process of bacterial release by;

(1) Mapping the cosmids which complemented the mutation and selecting a smaller common restriction fragment spanning the site of the original Tn5 insertion.

(2) Determining the position of the original Tn5 insertion and further Tn5 insertions made in the smaller restriction fragment.

(3) Subcloning fragments of the gene region defined by the Tn5 insertions for DNA sequencing.

(4) Sequencing the *bar* gene region.

(5) Determining the sequenced position of Tn5 inserts affecting expression of the *bar* gene region.

(6) Analysing the DNA sequence data and comparing the sequence to other known DNA sequences.

MATERIALS AND METHODS

2.1 BACTERIAL STRAINS AND PLASMIDS

The strains and plasmids used in this investigation are described in Table 2.

2.2 GROWTH OF BACTERIA

Rhizobium loti cultures were grown at 28°C in TY medium (Section 2.3.2) supplemented where necessary with neomycin (Neo; 400ug/ml), tetracycline (Tet; 2ug/ml) or streptomycin (Str; 200ug/ml). *E. coli* cultures were grown in either TY medium or LB medium (Section 2.3.1). Antibiotic concentrations used for *E. coli* were 15ug/ml for Tet and 25ug/ml for Kan.

2.3 MEDIA

All media were sterilized at 121°C for 15 minutes unless otherwise indicated. Solid media were cooled to 55°C before pouring.

2.3.1 Luria Broth (LB) (Miller, 1972) contained (g/l): Tryptone (Difco), 10.0; Yeast extract (Difco), 5.0; Sodium chloride, 0.5. The pH was adjusted to 7.0. Solid media was obtained by adding agar (15g/l, Davis).

2.3.2 TY Medium (Beringer, 1974) contained (g/l): Tryptone (Difco), 5.0; Yeast extract (Difco), 3.0; Calcium Chloride ($CaCl_2.6H_2O$), 1.3. Agar (Davis, 15g/l) was added for solid media.

Table 2: Bacterial strains and plasmids used in this investigation

Strain or Plasmid	Relevant Characteristics[a]	Source or Reference[b]
Strains		
Rhizobium loti		
NZP2037	Nod$^+$ Fix$^+$ (*L.pedunculatus, L.corniculatus*)	DSIR Culture Collection
PN184	NZP2037 *str* 1	Chua *et al.*, 1985
PN239	PN184*bar*239::Tn5	Chua *et al.*, 1985
PN1093	PN184*bar*41::Tn5	Ward *et al.*, 1989
PN1094	PN184*bar*78::Tn5	Ward *et al.*, 1989
PN1095	PN184*bar*18::Tn5	Ward *et al.*, 1989
Escherichia coli		
JM101	*sup*E,*thi*,Δ(*lac-pro*AB), [F',*tra*D36,*pro*AB,*lac*I^{q}ZΔM15]	Yanish-Peron *et al.*, 1985
HB101	*pro leu thi gal lac*Y *rec*A *str hsd*D *hsd*M	Boyer & Roulland-Dussoix, 1969
C2110	*pol*A Nal^R	Leong *et al.*, 1983
PN372	HB101/pPN318	Chua, 1984
PN374	HB101/pPN320	Chua, 1984
PN1000	HB101/pPN27	this study
PN1001	HB101/pPN28	this study
PN1120	C2110/pPN28::Tn5 (56)	Ward *et al.*, 1989
PN1121	C2110/pPN28::Tn5 (41)	Ward *et al.*, 1989
PN1123	C2110/pPN28::Tn5 (18)	Ward *et al.*, 1989
PN1124	C2110/pPN28::Tn5 (63)	Ward *et al.*, 1989
PN1135	C2110/pPN28::Tn5 (239)	Ward *et al.*, 1989

Table 2: continued

Strain or Plasmid	Relevant Characteristics[a]	Source or Reference[b]
Plasmids		
pRK2073	Km^S derivative of pRK2013	Leong *et al.*, 1983
pPH1JI	Gm^R IncP	Beringer *et al.*, 1978
pBR328	$Ap^R Tc^R Cm^R$	Bolivar *et al.*, 1977
pLAFR1	Lambda *cos* derivative of pRK290	Friedman *et al.*, 1982
pPN318	pLAFR1 cosmid containing *bar* gene from NZP2037	Chua, 1984
pPN319	pLAFR1 cosmid containing *bar* gene from NZP2037	Chua, 1984
pPN320	pLAFR1 cosmid containing *bar* gene from NZP2037	Chua, 1984
pPN27	pBR328 containing 7.5 kb *Eco*RI *bar* fragment from pPN318	this study
pPN28	pLAFR1 containing 7.5 kb *Eco*RI *bar* fragment from pPN318	this study
Bacteriophages		
Lambda467		de Bruijin and Lupski, (1984)
M13mp8		Messing and Vieira, (1982)
M13mp9		Messing and Vieira, (1982)

[a] abbreviations: Nal, Naladixic Acid; Km, Kanamycin; Gm, Gentamycin; Ap, Ampicillin; Tc, Tetracycline; Cm, Chloramphenicol.

[b] DSIR, Department of Scientific and Industrial Research, New Zealand.

2.3.3 M9 Medium (Miller, 1972) contained (g/l): Disodium hydrogen phosphate, 6.0; potassium dihydrogen phosphate, 3.0; sodium chloride, 0.5; ammonium chloride, 1.0. After autoclaving the following (separately sterilized) were added (ml/l): 10% Magnesium sulphate ($MgSO_4.7H_2O$), 2.5; 1% calcium chloride ($CaCl_2.2H_2O$) 1.5; 20% w/v glucose, 12.0; 0.5% thymine, 10.0.

2.3.4 Thornton's Medium (Thornton, 1930) was prepared from the following stock solutions:

Potassium phosphate stock solution which contained (g/l): KH_2PO_4, 54.6; K_2HPO_4, 30.0.

Magnesium sulphate/sodium chloride stock solution which contained (g/l): $MgSO_4.7H_2O$, 40.0; NaCl, 20.0.

Ferric chloride stock solution which contained (g/l): $FeCl_3$, 100,0.

Hoaglands trace element solution which contained (g/l): H_2BO_3, 2.86; $MnCl_2.4Cl_2O$, 1.81; $ZnSO_4.H_2O$, 0.22; $CuSO_4.5H_2O$, 0.08; $CoSO_4.7H_2O$, 0.095; $NaMoO_4.2H_2O$, 0.054.

Thornton's Medium consisted of (ml/l): Potassium phosphate stock solution, 2.0; magnesium sulphate/sodium chloride stock solution, 5.0; ferric chloride stock solution, 0.1; calcium orthophosphate, 1.0g/l; ferric phosphate, 1.0g/l; Hoaglands trace element solution, 1.0 and agar, 12.0g/l.

2.3.5 Top Agar contained (g/l): Tryptone (Difco) 10.0; sodium chloride, 8.0; agar (Davis) 8.0.

2.4 BUFFERS AND SOLUTIONS

2.4.1 Electrophoresis (E) buffer contained 40mM Tris(hydroxymethyl) aminomethane (Tris), 5mM sodium acetate and 1mM disodium ethylenediaminetetraacetic acid (EDTA), pH7.8. A 10 times concentrated stock solution was prepared by dissolving 96.8g Tris, 13.6g sodium acetate and 7.4g EDTA in 2 litres of deionized water. The pH was adjusted to 7.8 with glacial acetic acid.

2.4.2 Tris-Borate - EDTA (TBE) buffer contained 89mM Tris, 2.5mM EDTA and 8.9mM boric acid, pH 8.3. A 10 times concentrated stock solution was prepared by dissolving 107.7g Tris, 9.3g EDTA and 5.5g boric acid in 1 litre of deionized water.

2.4.3 Standard Saline Citrate (SSC) contained 0.15M sodium chloride and 0.015M sodium citrate. A 20 times concentrated stock solution was prepared by dissolving 877g sodium chloride and 441g sodium citrate in 5 litres of deionized water.

2.4.4 Phenol/chloroform solution

Solid phenol, melted in a water bath at 50°C, was equilibrated with Tris buffer (10mM, pH 8.0) and stored at 4°C. Phenol/chloroform solution was made by mixing an equal volume of Tris- equilibrated phenol with a 24:1 mixture of chloroform/isoamyl alcohol.

2.4.5 STET Buffer contained 8% sucrose, 5% Triton X-100, 50mM EDTA pH 8.0, 50mM Tris-HCl pH 8.0.

2.4.6 Solutions for plasmid preparation (Ish-Horowicz and Burke method, Section 2.7.2).

2.4.6.1 Solution I contained 50mM glucose, 25mM Tris-HCl pH 8.0, 10mM Na_2 EDTA.

2.4.6.2 Solution II was freshly prepared and contained 0.2N sodium hydroxide and 1% sodium dodecyl sulphate.

2.4.6.3 Solution III Potassium acetate solution.
Potassium acetate solution (pH 4.8) consisted of 29.44g potassium acetate and 11.5ml glacial acetic acid made up to a total volume of 100ml. This solution was 3M with respect to potassium and 5M with respect to acetate.

2.4.7 *Hae*III Buffer was prepared at 10 times the working concentration. Ten times *Hae*III buffer contained 14mM mercaptoethanol, 60mM Tris and 100mM magnesium chloride. The pH was adjusted to 7.6 with hydrochloric acid.

2.4.8 SDS Dye Mixture contained 0.1% sodium dodecyl sulphate and 0.05% bromophenol blue.

2.4.9 TE Buffer was prepared to the required concentration from 1M Tris and 250mM EDTA stock solutions. A working concentration of 50mM Tn5, 20mM EDTA was commonly used. The solution was adjusted to the required pH with hydrochloric acid.

2.4.10 Scintillation Fluid contained per litre; 2,5-diphenyloxazole (PPO), 4.0g; 2,2'paraphenylene-bis-5 phenyloxazole (POPOP), 100mg; toluene, 667ml; Triton X-100, 333ml.

2.4.11 TEC Buffer contained 10mM Tris, pH 8.0; 0.25mM EDTA, pH 8.0 and 30mM $CaCl_2$.

2.4.12 Hybridization Buffer contained (per litre) 50ml 1M Hepes pH 7.0, (Sigma); 150ml 20xSSC; 6ml phenol extracted herring sperm DNA (3mg/ml Sigma); 2ml *E. coli* transfer RNA, (10mg/ml, Sigma or Boehringer); 5ml 20% SDS; 2g Ficoll, (Sigma type 70); 2g Bovine Serum Albumin (0.2%); 2g Polyvinyl pyrolidine (0.2%, Sigma PVP-10).

2.4.13 Formamide Dye Mixture Formamide was deionized by gentle stirring with Amberlite MB1 resin for 30 minutes. The resin was removed by filtration and 0.3g/l xylene cyanol FF, 0.3g/l bromophenol blue and EDTA to 20mM were added.

2.5 PLANT NODULATION TESTS

Seeds of *Lotus pedunculatus* Cav. Grasslands Maku were surface sterilized by immersion for 5 minutes in a mixture of equal volumes of 30% hydrogen peroxide and 95% ethanol. The seeds were drained, rinsed 3 times in sterile distilled water, placed on 1% water agar plates and germinated in the dark. Seedlings were transferred aseptically to 20x150mm test tubes containing 6ml of Thorntons seedling agar (Thornton, 1930) (Section 2.3.4). Seedlings were inoculated with a drop of the broth culture to be tested, grown under artificial illumination (12 hour photoperiod, 180w/m^2 light intensity) and examined for nodulation and growth at regular intervals.

2.5.1 Reisolation of bacteria from Nodules

Nodules were excised, often with a small section of root material to aid handling, and surface sterilised in the same way as the seeds (Section 2.5). Surface sterilised nodules were squashed between flamed microscope slides and the released fluid plated on TY plates (Section 2.3.2).

2.6 ISOLATION OF TOTAL DNA FROM *RHIZOBIUM*

Rhizobium cultures (25ml) were grown overnight in TY medium at 28°C. The cells were harvested by centrifugation in a Sorvall SS34 rotor at 8000rpm for 10 minutes (ambient temperature). The pellet was washed once in 10ml of TE buffer (50mM Tris-HCl, 20mM EDTA pH 8.0, Section 2.4.9) containing 0.1% Sarkosyl (N-lauroyl sarcosine, Sigma), once in 10ml of TE buffer, then resuspended in 10ml TE buffer. Lysozyme was added to a final concentration of 300ug/ml and the mixture incubated at 37°C for 30 minutes. Proteinase K (Boehringer) was added to a final concentration of 300ug/ml and Sarkosyl to a final concentration of 1%. This mixture was incubated overnight at 50°C, then cooled and extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1), and twice with chloroform. The DNA was precipitated by the addition of 1/10 volume of 3M sodium acetate, and 2.5 volumes of 95% ethanol (-20°C). After standing at -20°C overnight, the DNA was recovered by centrifuging in an Eppendorf centrifuge for 15 minutes, washed with 95% Ethanol and dried under vacuum. The DNA was resuspended in TE buffer or water to a final concentration of 1ug/ml.

2.7 PLASMID ISOLATION METHODS

2.7.1 Rapid Boiling Method (Holmes and Quigley, 1981)

An overnight *E. coli* culture (1.5ml) was pelleted in an Eppendorf tube. The supernatant was drained thoroughly and the pellet resuspended in 350ul of STET buffer (Section 2.4.5). Freshly prepared lysozyme (25ul of a 10mg/ml solution)was added and the tube placed in a boiling water bath for exactly 40 seconds and then centrifuged for 10 minutes in an Eppendorf centrifuge.

The gelatinous pellet was removed with a toothpick and the plasmid DNA precipitated from the supernatant by the addition of 400ul of cold (-20°C) isopropanol. The contents of the tube were mixed by inversion and the tube allowed to stand at -20°C for 10-30 minutes. The plasmid DNA was pelleted by centrifuging for 15 minutes in an Eppendorf centrifuge. The DNA pellet was washed with 95% ethanol at room temperature, dried under vacuum for 15-30 minutes and resuspended in 20-50ul of deionized distilled water.

2.7.2 Cosmid Miniprep Method (Ish-Horowicz and Burke, 1981)

Cells from 5ml of an overnight culture were pelleted in an Eppendorf tube by sequential centrifugations. The supernatant was removed and the pellet resuspended in 100ul of solution I (Section 2.4.6.1). Lysozyme (10ul of a 50mg/ml solution in the above buffer) was added. The tube was allowed to stand at room temperature for 5 minutes then 200ul of solution II (Section 2.4.6.2) was added, mixed by inversion and the mixture allowed to stand on ice for 5 minutes. Solution III (Section 2.4.6.3, 150ul) was added, mixed by vigorous brief vortexing and the mixture allowed to stand on ice for 5 minutes. The resulting precipitate was pelleted by centrifugation in an Eppendorf centrifuge for 3 minutes and the supernatant was transferred to a new Eppendorf tube. The DNA was precipitated by the addition of 0.6 volumes (250ul) of cold (-20°C) isopropanol. The tube was allowed to stand at room temperature for 10-30 minutes and the recovered plasmid DNA washed and resuspended as described for the rapid boiling method (Section 2.7.1).

2.7.3 Preparative Plasmid Isolation

One litre of culture was grown overnight in LB or LB supplemented with the appropriate antibiotics. The cells were pelleted by centrifugation in a Sorvall GSA rotor at 8000 rpm for 10 minutes and resuspended in 30ml of Solution I (Section 2.4.6.1). Lysozyme (3ml of a 50mg/ml solution in the above buffer) was added and the GSA bottle allowed to stand at room temperature for 5 minutes. Solution II (Section 2.4.6.2; 60ml) was added, mixed by inversion and the mixture placed on ice for 5 minutes. Solution III (Section 2.4.6.3; 45ml) was added, mixed by vigorous brief vortexing and the bottle placed on ice for 5 minutes. The resulting precipitate was separated by centrifugation at 7,500 rpm for 10 minutes at 0°C in a Sorvall GSA rotor. The supernatant was transferred to a fresh GSA bottle and the DNA precipitated by the addition of 0.6 volumes of cold (-20°C) isopropanol. The bottle was allowed to stand at room temperature for 5 minutes then centrifuged at 8,000 rpm for 20 minutes at 0°C. The pellet was washed with 95% ethanol (room temperature) and dried under vacuum. The pellet was resuspended in 30ml of TE buffer (Section 2.4.9). Cesium chloride (30.7g, optical grade, Var Lac Oid Chem. Co. Inc, Bergenfield, New Jersey) and 1.9ml of a 10mg/ml solution of ethidium bromide (Sigma) were added and the cesium chloride allowed to dissolve with only gentle agitation. The solution was centrifuged (GSA rotor 8,000 rpm 15 minutes 10°C) to remove precipitated material, 4.5ml aliquots of the supernatant were transferred to TV865 Sorvall ultracentrifuge tubes and centrifuged overnight at 60,000 rpm, 15°C in a Beckman L5-75 ultracentrifuge (Sorvall TV865 rotor).

The plasmid band from the gradient was visualised by long wavelength UV light and removed by piercing the wall of the ultracentrifuge tube with a 25 gauge needle and drawing the fluorescent material into a 1ml syringe. Ethidium bromide was removed in subdued light by 3 extractions with

isopropanol (saturated with cesium chloride and TE buffer). The plasmid DNA was dialysed against TE buffer (2-3 changes, 24-36 hrs) to remove cesium chloride.

The optical density of the DNA was measured at 230, 258, 280 and 300nm to enable the purity and concentration of the DNA to be determined. Plasmid DNA was precipitated by the addition of 1/10 to 1/5 volume of 3M sodium acetate and 2.5 volumes of 95% Ethanol (-20°C). The mixture was allowed to stand overnight at -20°C and then centrifuged in an Eppendorf centrifuge for 15 minutes. The pellet was washed with 95% ethanol, dried under vacuum, and resuspended in TE buffer or water to a concentration of 0.5ug/ml.

2.8 PURIFICATION OF DNA SAMPLES BY EXTRACTION WITH PHENOL/CHLOROFORM

DNA to be purified by this method was extracted 1-2 times with an approximately equal volume of phenol/chloroform solution (Section 2.4.4) then 1-2 times with an approximate equal volume of chloroform before precipitation with ethanol (Section 2.9).

2.9 PRECIPITATION OF DNA WITH ETHANOL

DNA was precipitated from solution as follows; Sodium acetate (3M) equivalent to one fifth to one tenth volume of the DNA solution was added followed by 2 - 2.5 volumes of cold 95% ethanol. This mixture was allowed to stand at -20°C overnight or at -70°C for at least 2 hours before the DNA was precipitated by centrifugation for 15 to 30 minutes. The pellet was washed with ethanol and dried under vacuum before resuspending in water or buffer to the required concentration.

2.10 DETERMINATION OF DNA CONCENTRATION AND PURITY

DNA concentrations were determined from the absorbance at 258nm. Dilute solutions were measured undiluted in 0.5ml quartz cuvettes with a 1cm path length. More concentrated solutions were diluted 1:20 and measured in a standard 1cm quartz cuvette.

Absorbances were determined at 230, 258, 280 and 300nm. The ratio of the absorbances at 258/230 and 258/300nm gave an indication of protein contamination, while the ratio of the absorbances at 258 and 280nm was indicative of contaminating phenol. The DNA and RNA absorbance peak was at 258nm.

The final DNA concentration was calculated using the formula:

$$\frac{\text{Absorbance 258nm - Absorbance 300nm}}{20} \times \text{dilution factor} = \text{DNA concentration (mg/ml)}$$

2.11 RESTRICTION ENDONUCLEASE DIGESTS

Restriction endonuclease digestions were carried out in *Hae*III buffer (Section 2.4.7) with the salt concentration adjusted with 1M sodium chloride to the enzyme manufacturers recommendation. Digestions were performed at 37°C for 1-2 hours, stored at 4°C while an aliquot was checked on an agarose gel (Section 2.12) to ensure that the digestion had gone to completion, and stopped by adding 1/5th volume of SDS dye mixture (Section 2.4.8). If the DNA digestion was incomplete more enzyme was added and the mixture incubated for a further 1-2 hours or the DNA to be digested was further purified and the digestion repeated.

2.12 HORIZONTAL GEL ELECTROPHORESIS

Horizontal agarose gel electrophoresis was usually performed in a Biorad Mini Sub-Cell (100x60mm). Agarose concentrations from 0.7 to 1.6% were used and samples of DNA were usually separated by electrophoresis at 60V for periods of 1 to 2 hours, depending on the agarose concentration. After electrophoresis gels were stained for 15 to 30 minutes in ethidium bromide (approximately 1ug/l), destained briefly in approximately 0.1% magnesium sulphate, visualised on a U.V. transilluminator and photographed on Polaroid type 667 film or Kodak Tri-X 5'x4' film.

2.13 RECOVERY OF DNA FRAGMENTS FROM AGAROSE GELS

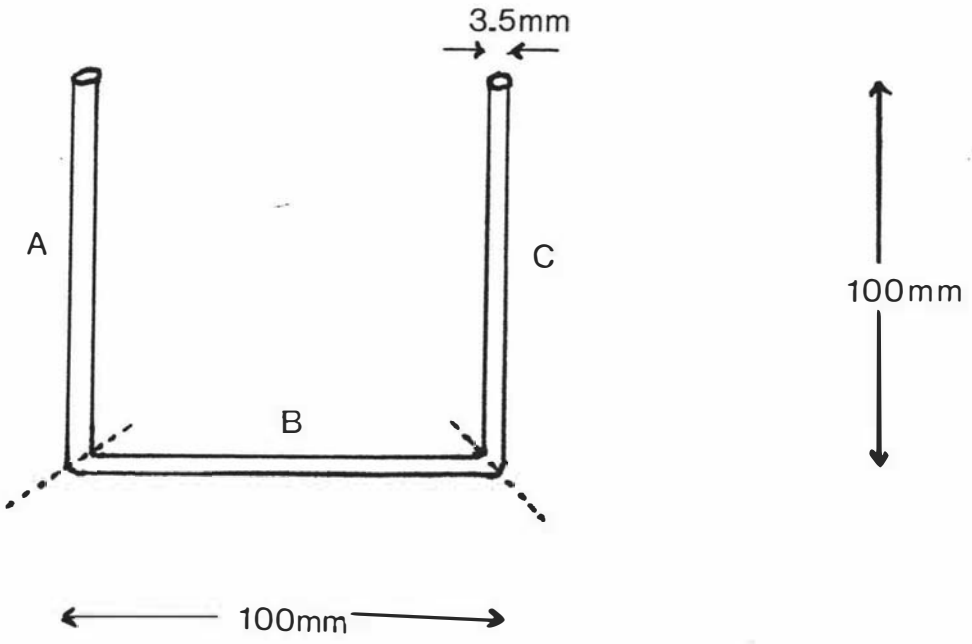
DNA was recovered from agarose gels by one of three methods:

2.13.1 U tube electroelution:

The DNA to be recovered was separated by electrophoresis through Seaplaque low melting temperature agarose. A glass U tube (Figure 2) was prepared by pouring molten agarose (0.8%) into the base region and allowing it to set. One arm (C) was similarly poured and set. Finally the DNA to be eluted was excised from the low melting temp agarose, placed in an Eppendorf tube and melted at 50°C. A drop of bromophenol blue dye solution (0.05% bromophenol blue in water) was placed at the junction of arm A and the base agarose (B). The molten agarose containing the DNA to be eluted was placed in arm A and allowed to set (if necessary this arm was filled as for arm C). The U tube was inverted over two tanks containing the same electrophoresis buffer as that used in the gels. Arm A was placed in the tank

Figure 2:

Glass "U" tube apparatus for recovery of DNA fragments from agarose gels (Section 2.13.1).



connected to the positive electrode and a piece of dialysis tubing about 20mm long was secured over the end of arm A. Air was excluded from the tubing and the open end closed with a dialysis closure (Spectrum Medical Industries, Los Angeles), enclosing the desired volume of buffer inside the dialysis tubing. Electrophoresis at 100V continued until the dye-front moved from the agarose into the buffer solution. The buffer in the dialysis tubing was recovered and the DNA in it extracted with phenol (Section 2.8) and precipitated with ethanol (Section 2.9).

2.13.2 Electroelution from a dialysis sac

After gel electrophoresis to separate DNA fragments a gel slice containing the DNA to be recovered was removed and placed in dialysis tubing with a minimal volume (about 200ul) of the appropriate electrophoresis buffer, sealed with dialysis clips and placed at right angles to the current flow in a standard gel electrophoresis system (e.g. Biorad Mini Sub-Cell). After electrophoresis for 1-2 hours at 100V, the current was reversed for 30 seconds, the liquid from the dialysis sac was recovered and the DNA in it extracted with phenol and precipitated with ethanol (Section 2.9).

2.13.3 Phenol freeze extraction

After gel electrophoresis to separate DNA fragments a gel slice containing the DNA of interest was excised from the gel with the minimum amount of excess agarose. The slice was cut into small fragments and placed in a 1.5ml Eppendorf tube which was filled with tris-saturated phenol and frozen at -20 or -70°C, usually overnight. The Eppendorf tube was centrifuged for 10 minutes and the aqueous phase recovered, extracted with phenol and precipitated with ethanol.

2.14 DNA LIGATIONS (King & Blakesley, 1986)

Ligation mixtures contained 200-1000ng of vector DNA, 200-500ng of insert DNA, 4ul of 5X concentrated ligation buffer, 1-2ul T4 DNA ligase, (Bethesda Research Laboratories) and sterile deionized water to make the final volume of 20ul. The mixtures were incubated at room temperature overnight before transformation (Section 2.15).

2.15 PREPARATION AND TRANSFORMATION OF COMPETENT CELLS

2.15.1 *E. coli* strain HB101

An overnight culture was diluted 1:50 in 25ml of Luria broth, then incubated with shaking at 37°C for 2 hours to an OD_{600} of 0.4-0.45. The culture was centrifuged at 6000 rpm for 10 minutes in a sterile glass Corex tube (Sorvall SS34 rotor) at 4°C. The pellet was resuspended without vortexing in 10ml cold (4°C) 60mM $CaCl_2$ and stored on ice for 20 minutes. The cells were then pelleted at 6000 rpm for 10 minutes at 4°C and resuspended in 250ul cold 60mM $CaCl_2$.

The DNA to be transformed (25-50ng DNA, 5ul), 45ul TEC buffer (Section 2.4.11) and 50ul of the competent cells were mixed in an Eppendorf tube and placed on ice for 1 hour. The tube was then heat shocked at 42°C for 2 minutes and placed on ice for a further 15 minutes. Luria broth (90ul) was added and the tube incubated at 37°C for 90 minutes. 100ul of the undiluted and 10^{-1} and 10^{-2} dilutions were plated on LB plates containing 100ug/ml Amp (pBR328) or 15ug/ml Tet (pLAFR1). Plates were incubated at 37°C overnight and putative transformants rechecked on LB plates supplemented with the appropriate antibiotic.

2.15.2 *E. coli* strain JM101

An overnight culture in YT media (2xTY) was diluted 1:100 in 25ml of YT and incubated with shaking at 37°C for 2 hours (OD_{550} ~0.3). The cells were harvested by centrifuging at 5000 rpm for 10 minutes in a sterile glass Corex tube (Sorvall SS34 rotor) at 4°C. The cells were resuspended without vortexing in 10ml ice cold 50mM $CaCl_2$, placed on ice for 20 minutes, then pelleted at 5000 rpm for 10 minutes at 4°C, resuspended in 2.5ml ice cold 50mM $CaCl_2$ and stored on ice.

Competent cells (200ul) and ligation mix (section 2.14) were mixed in a sterile 100mm x 13mm test tube, placed on ice for 40 minutes, and heat shocked for 2 minutes at 42°C. Three ml top agar (Section 2.3.5) 20ul BC1G, (5-bromo-4-chloro-3-indoyl-beta-galactoside; 20mg/ml in dimethyl formamide) 20ul IPTG (isopropyl-beta-D-thio-galactopyranoside; 24mg/ml in water) and 200ul exponential phase JM101 cells were added to the heat shocked mix in the above order. This was thoroughly mixed and plated onto M9 plates (Section 2.3.3). The plates were allowed to dry and incubated overnight at 37°C. White plaques were selected for the preparation of templates (Section 2.17.1).

2.16 SOUTHERN BLOT TECHNIQUE (Southern, 1975)

The DNA to be transferred was separated by electrophoresis, stained, visualised and photographed as previously described (Section 2.12).

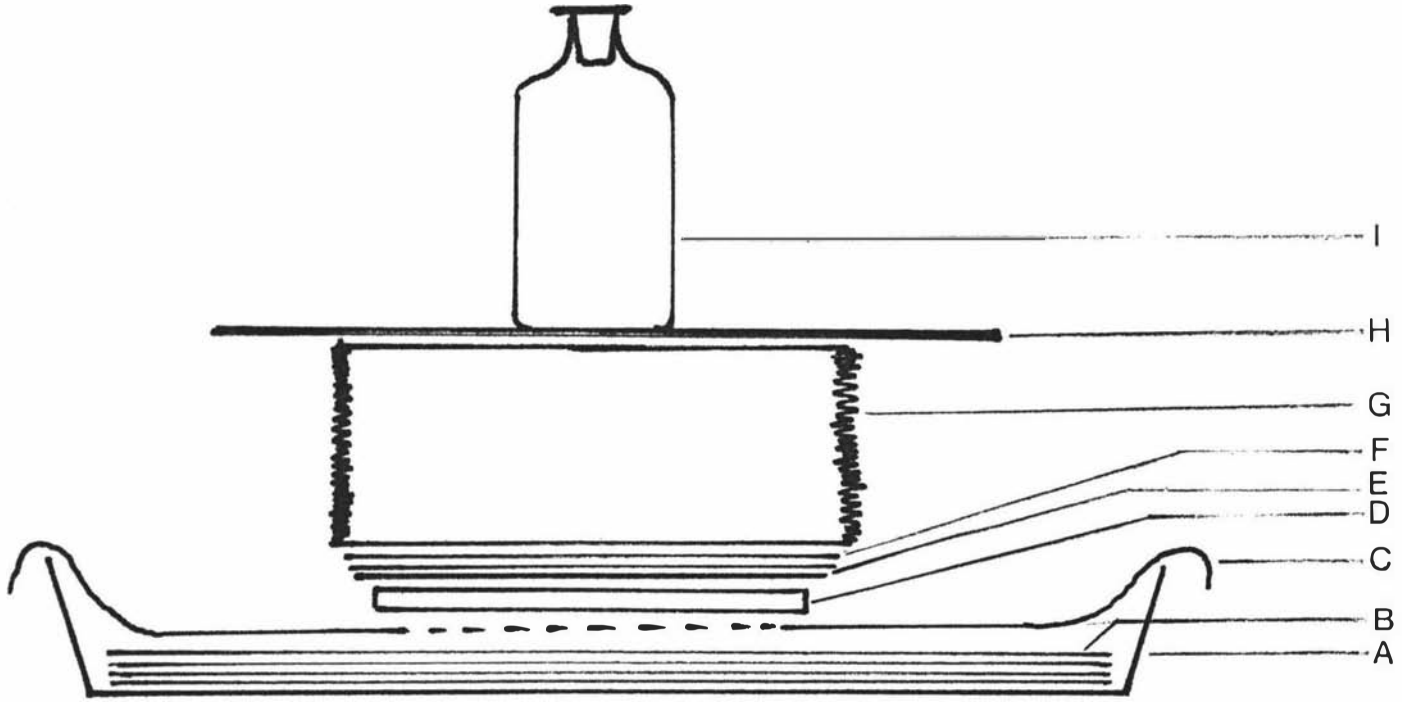
The gel was placed in a dish containing 0.25M HCl for 5 to 10 minutes after the bromophenol blue dye changed colour. The hydrochloric acid was poured off and the gel washed with distilled water. The gel was then immersed in 0.5M NaOH,

0.5M NaCl for 15 minutes, washed with distilled water as before and soaked in 0.5M Tris-HCl, 2.0M NaCl pH 7.2 for at least 15 minutes. During the treatment of the gel a pyrex baking tray large enough to take the gel flat was prepared by placing 4 layers of Whatman 3MM chromatography paper in it. A piece of Gladwrap was placed over the dish and pressed down on the paper. A square the same size as the gel was marked on the Gladwrap with a felt pen by drawing around a cardboard template the same size as the gel. The Gladwrap was cut away approximately 5mm inside this square with a scalpel. The paper in the dish was moistened with 20xSSC and the Gladwrap smoothed out so there were no bubbles between it and the paper. The treated gel was placed on the paper with its edges aligned with the marker pen lines on the Gladwrap. A piece of 0.45um nitrocellulose membrane (BA85, Schleicher and Schuell, W. Germany) was cut out which was slightly larger than the gel, wetted by floating in a dish of water and placed over the gel. One piece of 3MM paper wetted in 20xSSC was placed over the filter followed by two layers of dry 3MM paper. Care was taken to ensure that there were no air bubbles trapped between any of the layers. A stack of paper towelling about 50-100mm deep was placed on top of the 3MM paper followed by a glass plate and a weight (approximately 400g) (Figure 3).

Excess 20xSSC was placed in the dish and the Gladwrap sealed around the edges to prevent evaporation. The assembly was left overnight. The next day the nitrocellulose filter was removed and washed for 10 to 15 seconds in 2xSSC, then baked at 70-80°C under vacuum for 2 hours. After transfer the gel was restained in ethidium bromide and revisualised to verify transfer had taken place.

Figure 3: Apparatus for the transfer of DNA from agarose gels to nitrocellulose filters. (Southern blot, Section 2.16).

A - glass dish
B - 4 sheets Whatman 3mm paper
C - Gladwrap
D - agarose gel
E - nitrocellulose filter
F - 2 sheets Whatman 3MM paper
G - stack of paper towels
H - glass plate
I - weight


I
H
G
F
E
D
C
B
A

2.17 PREPARATION OF RANDOM PRIMERS (Whitfield *et al.*, 1982)

Calf thymus DNA (1g) was dissolved in 25ml 20mM Tris-HCl, pH 7.4, 10mM magnesium chloride to a final concentration of 40mg/ml. DNase I (2mg) was added and the mixture incubated at 37°C for 45 minutes. The DNA was cut into short fragments. DNase I was denatured by the addition of 1% SDS and 1mg/ml protease (Sigma type XV) and incubation for 45 minutes at 37°C. Protein was removed by two phenol chloroform extractions (Section 2.8) and the DNA was denatured by boiling for 10 minutes and then placing it in ice water. The DNA was loaded on a Whatman DE52 cellulose column (20x1cm) equilibrated with 5mM Tris-HCl pH 7.4, 1mM EDTA, 0.1M NaCl and the column was eluted with this buffer until the eluant did not absorb at 258nm. This treatment removed fragments less than 5 nucleotides in length. Fragments 5 to 12 nucleotides in length were eluted by washing with 5mM Tris pH 7.4, 1mM EDTA, 0.3M NaCl. This fraction was ethanol precipitated, resuspended in deionized water to a concentration of 50ug/ml and stored at -70°C.

2.18 PREPARATION OF-^{32}P-LABELLED PROBE DNA

The DNA to be labelled was digested with *Hae*III in a 25ul reaction mixture (as described in Section 2.11) for 30 minutes at 37°C. Random primers (Section 2.17, 4ul), and 10ul water were added and the mixture was boiled for 2 minutes then chilled rapidly on ice. The following reagents were added in the order listed; 2ul 10x *Hae*III buffer, (Section 2.4.7), 1ul dATP, 1ul dGTP, 1ul dTTP, 3ul [^{32}P]a dCTP, 2ul DNA polymerase I (Klenow fragment). The mixture was incubated at 37°C for 30 minutes and the reaction was stopped by the addition of 3ul of 0.2M Na_2 EDTA.

Free nucleotides and DNA were separated on a Sephadex G-50 "minispin" column constructed in a 1ml Tuberculin syringe (Figure 4) centrifuged at 1500 rpm for 5 minutes in a MSE bench centrifuge fitted with a swinging bucket rotor. The probe was collected in a 1.5ml Eppendorf tube. The efficiency of the labelling reaction was estimated by ascending chromatography of 2ul of the probe solution on a PEI-cellulose strip (10x50mm) with 2M HCl as the eluant. The strip was divided in two and each half was counted separately on a Beckman LS7000 scintillation counter in 10ml scintillation fluid (Section 2.4.10).

2.19 HYBRIDIZATION OF PROBE DNA TO SOUTHERN BLOTS

The southern blot to be labelled was prehybridized in a plastic bag with 10-20ml hybridization buffer (section 2.4.12), air excluded and the bag sealed and immersed in a waterbath at the hybridization temperature (usually 65°C) for 2 hours. After prehybridization most of the buffer was poured off and the boiled (2 minutes) and chilled ^{32}P labelled probe (not less than $1x10^6$ cpm) was added and mixed with the hybridization buffer in the bag. Air was excluded and the bag resealed and returned to the waterbath for 20-24 hours.

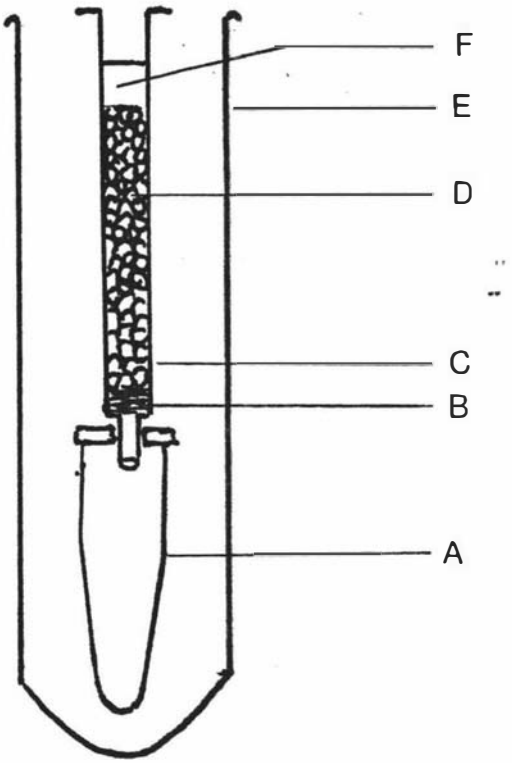
After hybridization the filter was removed from the bag, washed 3 times in 2xSSC for 15 minutes, taped to a sheet of filter paper and placed in an X-ray cassette with Agfa Curix RP2 X-ray film and an intensifying screen as shown in Figure 5. The film was exposed for 1-10 days at -70°C and then developed in a Kodak X-Omat automatic processer.

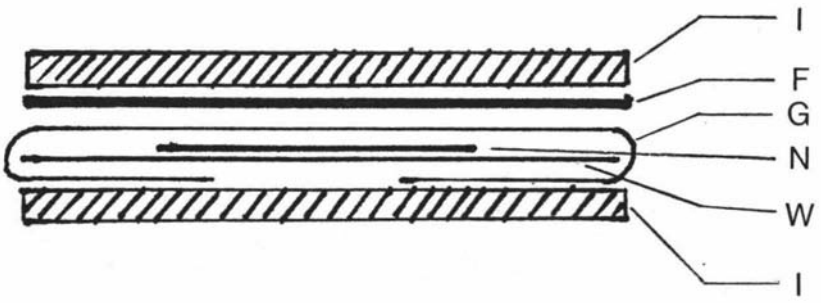
Figure 4: "Minispin" column apparatus (Section 2.18)

A = 1.5ml Eppendorf tube
B = glass wool plug (acid washed)
C = 1ml Tuberculin syringe
D = Sephadex G-50
E = carrier tube (50ml centrifuge tube)
F = probe to be separated

Figure 5: Exploded view of assembly of filters, film and intensifying screens in an X-ray cassette (not shown).

I = intensifying screen
F = Agfa Curix RP2 X-ray film
G = Gladwrap
N = nitrocellulose filter
W = Whatman 3MM paper


F
E
D
C
B
A


I
F
G
N
W
I

2.20 DNA SEQUENCING (Sanger *et al.*, 1977)

2.20.1 Preparation of Sequencing Templates

Usually 12 or 24 templates were made at the same time. 500ul of an overnight culture of JM101 was added to 25ml of YT medium which was then aliquoted into 24 13x100mm glass tubes. White plaques were selected from an M9 plate (Section 2.3.3) and picked with sterile toothpicks which were then dropped into the glass tubes. The tubes were incubated on a shaker at 200-250 rpm at 37°C for 5-6 hours and the contents transferred to 1.5ml Eppendorf tubes which were centrifuged for 5 minutes. The supernatants were transferred to a second series of Eppendorf tubes with care to ensure no cell debris was transferred. The pellets in the first series of tubes were drained thoroughly and plasmid DNA extracted by the rapid boil procedure (Section 2.7.1), 200ul 2.5M NaCl, 20% PEG 6000 was added to each tube in the second series containing the supernatants; these were allowed to stand at room temperature for 15 minutes, then centrifuged for 5 minutes in an Eppendorf centrifuge. The supernatants were removed, and the tubes recentrifuged for 10 seconds to bring all remaining liquid off the walls to the base of the tubes from where it was aspirated with a drawn-out pasteur pipette. 100ul 10mM Tris, 0.1mM EDTA and 50ul neutralized phenol was added to each PEG precipitate. The tubes were vortexed for 10 seconds, allowed to stand at room temperature for 5 minutes then vortexed again for 10 seconds and centrifuged for 1 minute. The upper aqueous layers were removed to another series of tubes and 10ul 3M sodium acetate pH 5.5 and 250ul ethanol (95%, -20°C) were added to each tube. The tubes were allowed to stand overnight at -20°C then centrifuged for 15 minutes and each template DNA pellet was resuspended in 30ul 10mM Tris, 0.1mM EDTA, pH 8.0, and stored at -20°C.

2.20.2 Annealing Template and Primer DNA

Four template DNA's were normally annealed and labelled at a time. Annealing buffer (100mM Tris, 100mM $MgCl_2$ pH 8.5; 5ul) 5ul of primer DNA (2.5ng/ml) and 15ul deionized distilled water were mixed and 5.5ul was aliquoted into each of four 500ul microcentrifuge tubes. 5.5ul of the DNA template to be labelled was added to each tube, mixed and the tubes placed in 12x100mm glass tubes full of water in a hot block (80-90°C) for 5 minutes. The glass tubes containing the microcentrifuge tubes were removed from the hot block and allowed to cool to room temperature (about 20 minutes) during which time annealing occurred. Finally the microcentrifuge tubes were briefly centrifuged.

2.20.3 Labelling the Annealed Template DNA

Stock solutions of d/ddATP, d/ddGNP, d/ddCTP and d/ddTTP (section 2.20.4) were thawed and kept on ice.

^{35}S dATP was aliquoted into four 500ul microcentrifuge tubes (1-1.5ul containing 8-10uCi per tube), the appropriate d/ddNTP stock solution (9ul) was added to each of the four tubes. DNA Polymerase I (Klenow fragment) (0.5ul) was added to each tube using a micropipette. Eppendorf tubes (1.5ml) without caps were arranged in a 4x4 array in two racks for an Eppendorf model 5413 Microfuge. Annealing mixtures (templates) (2.5ul) were loaded onto the side of each tube down a row and 2.5ul of the appropriate d/dd NTP mixture was loaded onto the side of each tube across a row. The racks were centrifuged for 1 minute to start the reaction, then incubated at 37°C for 20 minutes. During this time 1ul of dNTP chase mix (0.125mM of each dNTP) was added to the side of each tube. After 20 minutes incubation the racks were centrifuged to add this 1ul to the reaction mixture and then incubated for a further 15 minutes at 37°C. The products of the sequencing reactions were stored at -20°C.

2.20.4 Preparation of d/ddNTP Stock Solutions

All solutions were stored at -20°C unless otherwise stated. A 20mM stock solution of each of dATP, dGTP, dCTP and dTTP was made in distilled water.

Working solutions (0.5mM) were 1:40 dilutions of the stock solutions. dNTP° mixtures were made from these working solutions by mixing the volumes (in microlitres) shown in Table 3.

Table 3: Deoxynucleotide triphosphate mixtures

	dNTP° mixtures			
	dATP°	dGTP°	dCTP°	dTTP°
0.5 mM dGTP	200	10	200	200
0.5 mM dCTP	200	200	10	200
0.5 mM dTTP	200	200	200	10
Buffer (50 mM Tris-HCl 1 mM EDTA pH 8.0)	50	50	50	50

10mM stock solutions of dideoxy nucleotides were made in distilled water. These were diluted as shown in Table 4.

Table 4: Dideoxynucleotide triphosphate mixtures

	final concentration	ul of 10 mM stock per 1 ml water
ddATP	0.02 mM	2
ddGTP	0.1 mM	10
ddCTP	0.08 mM	8
ddTTP	0.5 mM	50

The d/ddNTP termination mixtures were prepared by mixing the ddNTP dilution and dNTP° mixtures as shown in Tables 5 and 6.

Table 5: Mixture of dideoxynucleotides used for sequencing with DNA polymerase I (Klenow fragment)

	dNTP° mix	ddNTP dilution
d/ddATP	50	50
d/ddGTP	55	45
d/ddCTP	60	40
d/ddTTP	50	50

Table 6: Deoxy/dideoxynucleotide triphosphate mixture used for sequencing with reverse transcriptase:

	volume of d/ddNTP mixture from table 4 (above): (ul)	Volume dNTP°
d/ddATP (rt)	1	149
d/ddGTP (rt)	1	119
d/ddCTP (rt)	2	38
d/ddTTP (rt)	1	119

2.20.5 Acrylamide Gel Electrophoresis of the Products of the Sequencing Reaction

A 40% acrylamide/bis acrylamide (38:2) stock solution was prepared. 288 g of urea was dissolved in 210ml of water and adjusted to a final volume of 450ml. The urea solution was mixed with 90ml of the acrylamide stock solution and 20g of Amberlite MB1 resin was added. This mixture was stirred on a magnetic stirrer for 30 minutes then the resin was filtered off by passing the solution through a number 3 porosity sintered glass funnel. Sixty ml of 10xTBE (pH 8.3) (section 2.4.2) was added to the filtrate and the ready to use gel mixture was stored at 4°C.

Formamide gels (Martin, 1987) were poured from 7M urea 40% formamide and 40% acrylamide/bis acrylamide (38:2) stock solutions. 252g of urea were dissolved in 240ml of dimethyl formamide (B.D.H.). The required quantity of acrylamide stock solution to give the desired gel concentration was added and the solution deionized as above. After removal of the ion exchange resin 60ml of 10xTBE (pH 8.3) was added to the filtrate and the ready to use gel mixture stored at 4°C.

Gels of two sizes were used, both of 0.4mm thickness. For 18x36cm gels 24ul TEMED (tetramethylethylenediamine) and 240ul 10% ammonium persulphate were added to 40ml of acrylamide mix. For 30x38cm gels 50ul TEMED and 500ul 10% ammonium persulphate were added to 60ml of acrylamide mix. For 7M urea 40% formamide gels 4 times the quantity of TEMED was used. The gels were poured between pretaped glass plates and allowed to set for about 90 minutes, after which the tape was removed from the bottom of the gel plates and they were assembled in the electrophoresis apparatus. The well area was flushed thoroughly with deionized water and a sharkstooth comb was inserted.

Samples to be separated by electrophoresis were mixed with 5ul of formamide dye mix (Section 2.4.13) and placed in a boiling water bath for 3 minutes. During this time the wells were flushed with buffer. The sample (3ul per well) (containing formamide dye mix) was loaded. Electrophoresis was carried out at 1500V constant voltage for 18x36cm gels or 60W constant power for 30x38cm gels for 1 1/2 hours (short run) or 4 hours (long run).

Gels were fixed in an aqueous solution containing 10% acetic acid and 10% ethanol or methanol, transferred to Whatman 3MM paper, covered with Gladwrap and dried under vacuum for 45 minutes at 80°C. The dried gel was then placed in a Cronex 35cmx43cm X-ray cassette with Cronex 4 X-ray film for 1-7 days at room temperature. The film was developed in a Kodak X-Omat automatic processor and the sequence manually read from the gels.

RESULTS

3.1 CONSTRUCTION OF A PHYSICAL MAP OF THE COSMIDS COMPLEMENTING THE BAR MUTATION

In order to characterize the gene involved in the *bar* mutation identified by Chua, (1984) a *Hin*dIII and *Eco*RI restriction enzyme map of the cosmids which complemented the mutation (pPN318, pPN319 and pPN320) was constructed. Cosmid DNA was purified by the preparative plasmid isolation method (Section 2.17.3) and subjected to partial, single and double digestions with *Eco*RI and *Hin*dIII enzymes (Section 2.11). The resulting fragments were separated by electrophoresis on a 0.7% agarose gel and the molecular weight of fragments estimated by reference to *Eco*RI and *Hin*dIII digests of lambda DNA. Photographs of the restriction gels are shown in figures 6A, 7A and 8A, and the fragment sizes are summarized in Table 7. Cosmids pPN318 and pPN319 gave identical fragment patterns. All cosmids contain an *Eco*RI fragment of 7.5kb which does not have a *Hin*dIII site since it remains unchanged in the double digest. In addition all cosmids have at least one 5.9kb *Eco*RI fragment which contains one *Hin*dIII site (since the 5.9kb fragment is not conserved in the double digest).

The 9.4kb and 7.2kb *Hin*dIII fragments from pPN320 and the 9.2kb *Hin*dIII fragment from pPN318 and pPN319 were extracted from agarose gels (Section 2.13), labelled with ^{32}P (Section 2.18), and hybridized to Southern blots of the single and double digests to elucidate the relationship between the fragments. The resulting autoradiographs are shown in Figures 6 B and C, 7 B and C and 8 B and C.

The 7.2kb *Hin*dIII fragment from pPN320 did not hybridize to pPN318 or pPN319 (result not shown). In pPN320 this fragment hybridized only to a 5.9kb fragment from the *Eco*RI digest and to a 5.0kb and 2.3kb fragment in the double digest (Figure 8D).

Figure 6:

Fragments obtained from the digestion of cosmid pPN318 hybridized to probes from pPN318 and pPN320.

A. Agarose gel (0.8%) of restriction endonuclease digests of:

Lane 1; lambda DNA digested with *Eco*RI.
Lane 2; lambda DNA digested with *Hin*dII.
Lane 3; pPN318 digested with *Eco*RI.
Lane 4; pPN318 digested with *Hin*dIII.
Lane 5; double digest of pPN318 with *Eco*RI and *Hin*dIII.

B. Southern blot of gel A probed with the 9.2kb *Hin*dIII fragment from pPN318.

C. Southern blot of gel A probed with the 9.4kb *Hin*dIII fragment from pPN320.

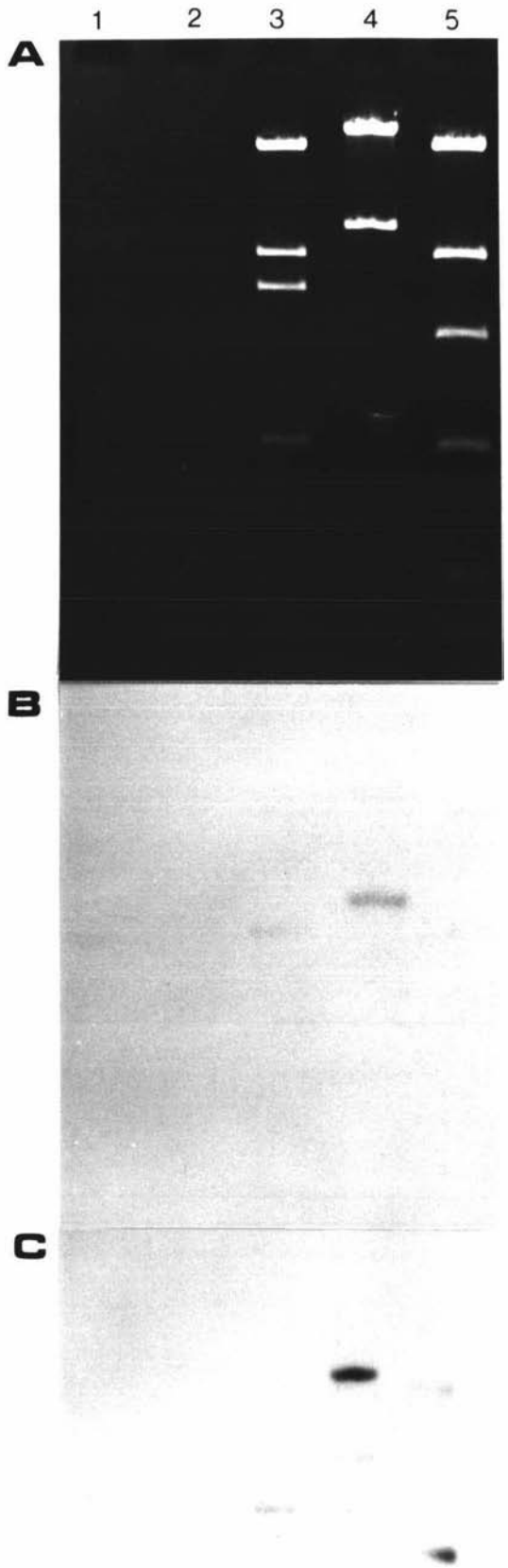

1
2
3
4
5
A
B
C

Figure 7:

Fragments obtained from the digestion of cosmid pPN319 hybridized to probes from pPN318 and pPN320.

A. Agarose gel (0.8%) of restriction endonuclease digests of:

Lane 1; lambda DNA digested with *Eco*RI.
Lane 2; lambda DNA digested with *Hin*dIII.
Lane 3; pPN319 digested with *Eco*RI.
Lane 4; pPN319 digested with *Hin*dIII.
Lane 5; double digest of pPN319 with *Eco*RI and *Hin*dIII.

B. Southern blot of gel A probed with the 9.2kb *Hin*dIII fragment from pPN318.

C. Southern blot of gel A probed with the 9.4kb *Hin*dIII fragment from pPN320.

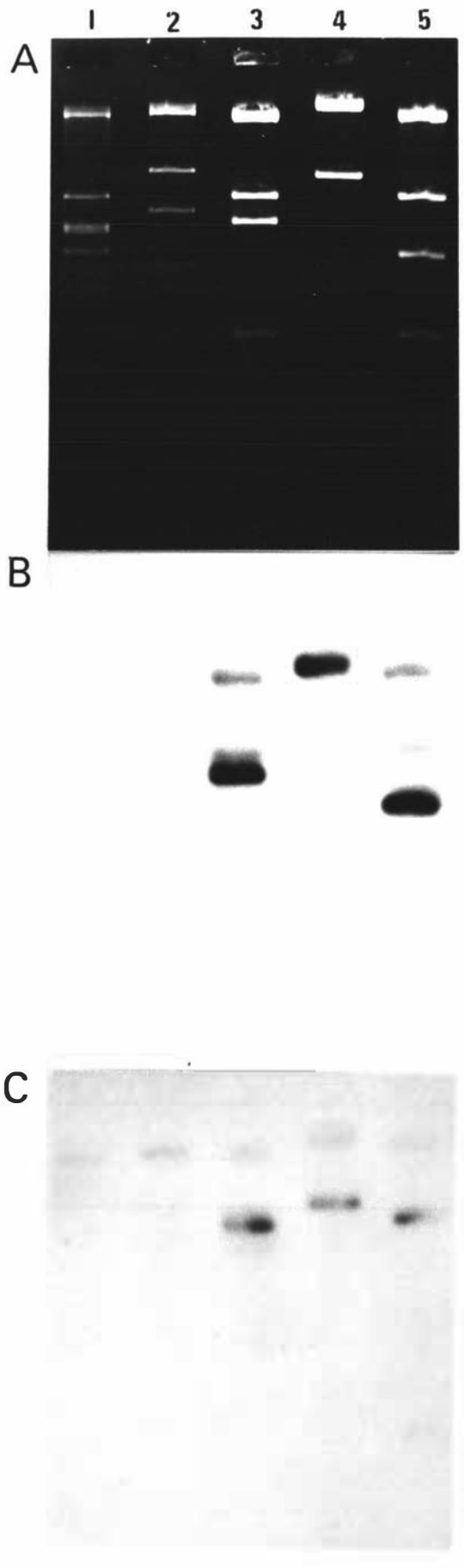

1
2
3
4
5
A
B
C

Figure 8:

Fragments obtained from the digestion of cosmid pPN320 hybridized to probes from pPN318 and pPN320.

A. Agarose gel (0.8%) of restriction endonuclease digests of:

Lane 1; lambda DNA digested with *Eco*RI.
Lane 2; lambda DNA digested with *Hin*dIII.
Lane 3; pPN320 digested with *Eco*RI.
Lane 4; pPN320 digested with *Hin*dIII.
Lane 5; double digest of pPN320 with *Eco*RI and *Hin*dIII.

B. Southern blot of gel A probed with the 9.2kb *Hin*dIII fragment from pPN318.

C. Southern blot of gel A probed with the 9.4kb *Hin*dIII fragment from pPN320.

D. Southern blot of gel A probed with the 7.2kb *Hin*dIII fragment from pPN320.

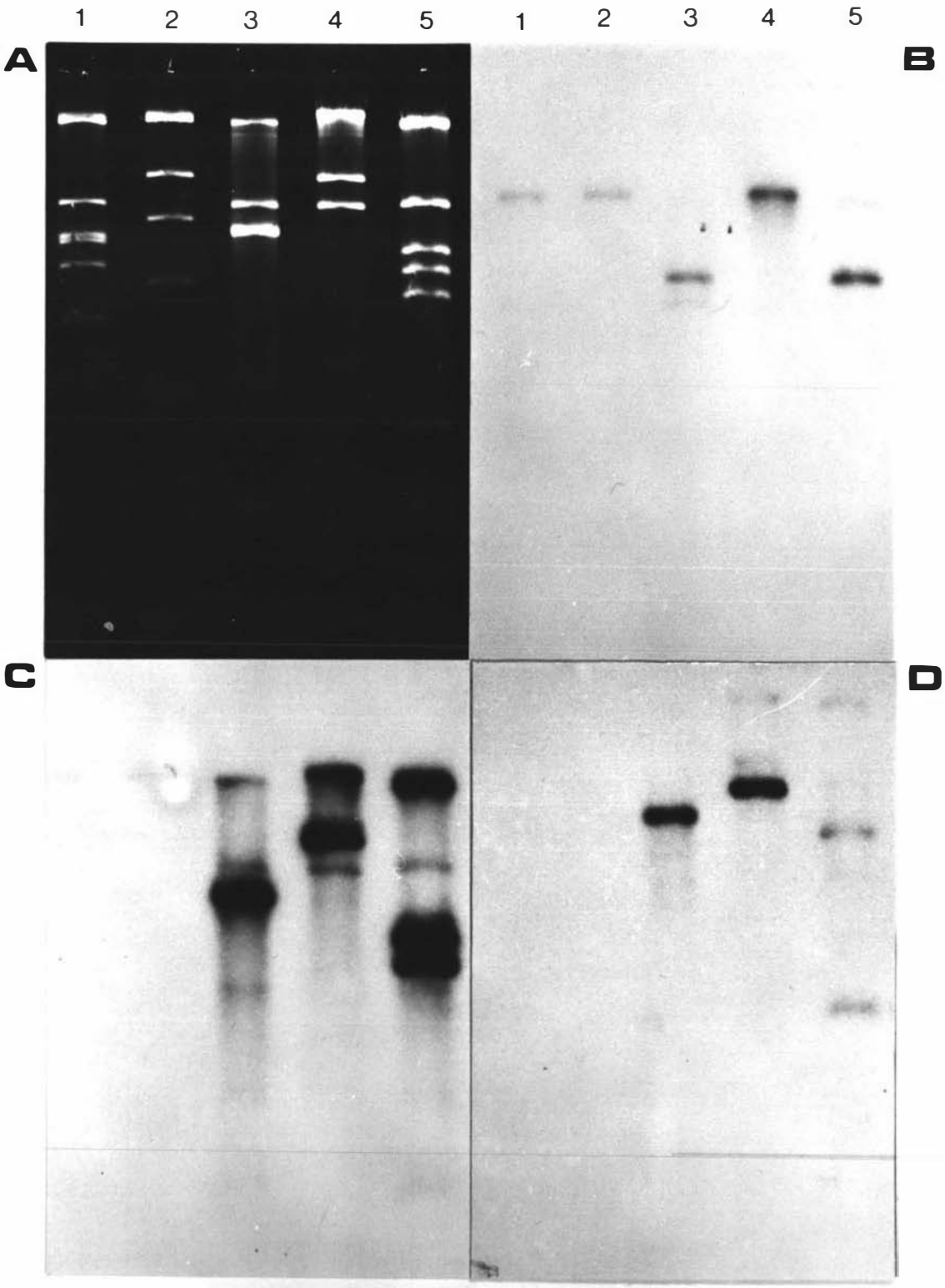

1
2
3
4
5
1
2
3
4
5
A
B
C
D

Table 7: Fragment sizes obtained from cosmids pPN318, pPN319 and pPN320 digested with *Eco*RI and *Hin*dIII.

	Fragment size (kb)[1]		
	*Eco*RI Digest	*Hin*dIII Digest	Double Digest
Cosmid pPN320			
	21	31	21
	7.5	9.4	7.5
	5.9 x 3	7.2	5.0
	1.8		4.5
			4.0
			2.3
			1.8
			1.5
	48.0	47.6	47.6
Cosmids pPN318 and pPN319	22	30	22
	7.5	9.2	7.5
	5.9		4.5
	2.9		2.8
	1.7		1.8
			1.6
	40.1	39.2	30.2

[1] Fragment sizes were determined by measuring the distance a fragment had migrated from the well in an agarose gel and calculating the molecular weight by comparison with a plot of the distance migrated in the same gel by standard lambda fragments against the logarithm of the molecular weight of the standard lambda fragments.

The 9.4kb *Hin*dIII fragment from pPN320 hybridized only to the 5.9kb fragment from the *Eco*RI digest of pPN320 and not to the 7.5kb *Eco*RI fragment (Figure 8C). In the double digest of pPN320, fragments of 4.0 and 4.5kb hybridized. The 9.4kb *Hin*dIII fragment from pPN320 hybridized only to the fragment containing the vector plus flanking sequences in the *Hin*dIII digest of pPN318 and pPN319, to a 5.9kb fragment in the *Eco*RI digest and to a 4.5kb fragment in the double digest of pPN318 and pPN319 (Figures 6C and 7C).

The 9.2kb *Hin*dIII fragment from pPN318 hybridized to the 7.5kb fragment in both the *Eco*RI digest and double digest of all three cosmids. This confirms the absence of a *Hin*dIII site in the 7.5kb *Eco*RI fragment and the overlap of the 9.2kb *Hin*dIII fragment with the 7.5kb *Eco*RI fragment. In pPN320 the 9.2kb *Hin*dIII fragment from pPN318 hybridized to the fragment containing the vector plus flanking sequences, indicating that this fragment is joined to the pLAFR1 vector in pPN320.

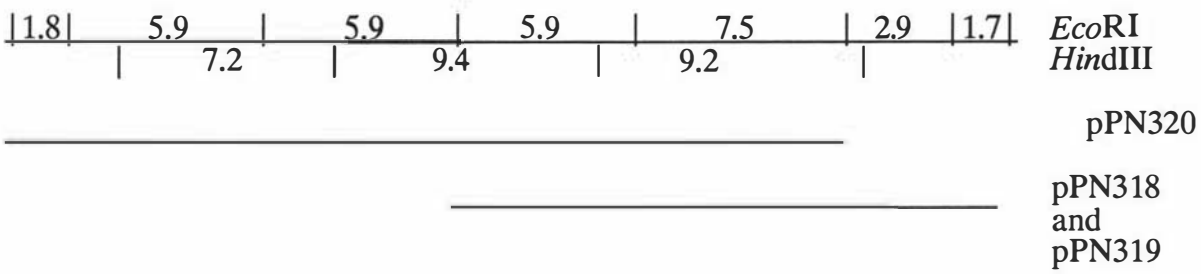
The *Eco*RI and *Hin*dIII restriction enzyme map deduced from the above observations is presented in Figure 9.

3.2 SUBCLONING OF THE COMMON 7.5KB *ECO*RI FRAGMENT

Chua, (1984) demonstrated that the Tn5 insertion which caused the Bar$^-$ phenotype was located in a wild type *Eco*RI fragment of molecular weight approximately 7.5kb. Section 3.1 demonstrates that a 7.5kb *Eco*RI fragment is common to each of the cosmids which can complement the *bar* mutation.

Figure 9:

Physical map of the *bar* region of *R. loti* strain NZP2037. The map was constructed by determining the position of *Eco*RI and *Hin*dIII sites in cosmids which complemented the *bar*$^-$ mutation.



Cosmid pPN318 was digested with *Eco*RI and the fragments separated on a seaplaque agarose gel. The 7.5kb fragment was removed from the gel (Section 2.13) and ligated into pBR328 and pLAFR1 (Section 2.14), to construct plasmids pPN27 and pPN28 respectively.

Competent HB101 cells were transformed with the ligation mixture and plated on LB agar plates supplemented with the appropriate antibiotics (Section 2.3.1), to produce strains PN1000 (pPN27) and PN1001 (pPN28).

The insertion of the 7.5kb *Eco*RI fragment into the vector was checked by digesting a rapid boil preparation (Section 2.7.1) of the plasmid with *Eco*RI and separating the resulting fragments by electrophoresis on a 0.7% agarose gel (Figure 10).

The presence in the *E. coli* vectors of the correct 7.5kb *Eco*RI fragment was confirmed by hybridization (data not shown).

3.3 COMPLEMENTATION OF THE *BAR* MUTATION BY THE SUBCLONED 7.5kb *ECO*RI FRAGMENT

The ability of the 7.5kb *Eco*RI fragment subcloned from pPN318 into pLAFR1 (plasmid pPN28, strain PN1001) to complement the *bar* mutation (PN239) was demonstrated as follows:

A triparental cross was performed with PN1001 (pPN28) (pLAFR1 containing the 7.5kb *Eco*RI fragment), *E. coli* HB101 containing pRK2013, a helper plasmid (Ditta *et al.*, 1980) to mobilize the pPN28, and PN239, the Bar$^-$ mutant. The resulting growth from the cross on M9 plates (Section 2.3.3) was used to inoculate *Lotus pedunculatus* seedlings (Section 2.5). Controls in the plant test experiment were an uninoculated control, *E. coli* HB101/pRK2013 by itself, PN1001 (pPN28) by itself, PN239, (Bar$^-$ mutant containing Tn5) and NZP2037 (wild type *R. loti*).

Figure 10:

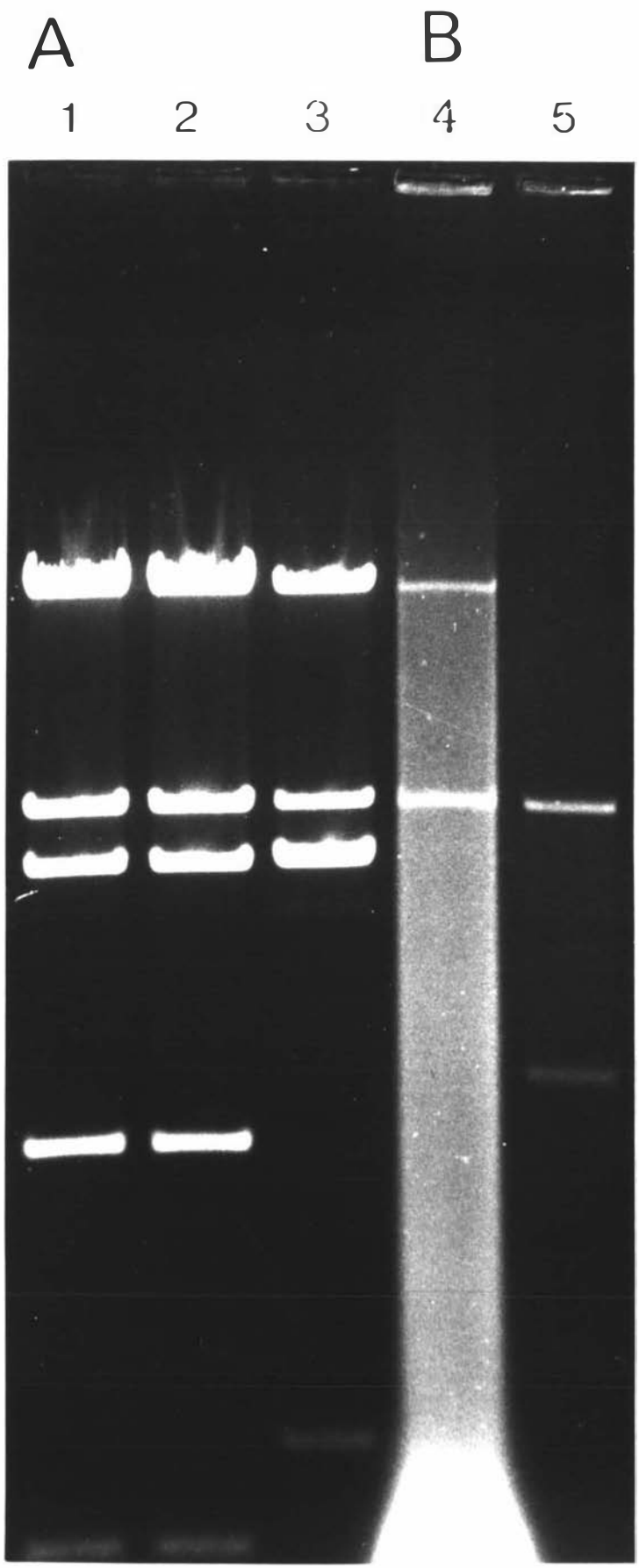
Agarose gel electrophoresis of *Eco*RI digests of:

A. The cosmids which complemented the Bar$^-$ mutant

Lane 1; pPN318.
Lane 2; pPN319.
Lane 3; pPN320.

B. Plasmids into which the common 7.5kb *Eco*RI fragment was subcloned from pPN318:

Lane 4; pLAFR1 + 7.5kb *Eco*RI fragment.
Lane 5; pBR328 + 7.5kb *Eco*RI fragment.


A
B
1
2
3
4
5

Effective nodules were observed on plants inoculated with the triparental cross and the seedlings appeared to grow as well as those inoculated with the wild type *R. loti* (Figure 11). Bacteria isolated from nodules formed by the complemented strain (Section 2.5.1) were resistant to tetracycline (2ul/ml), (from pLAFR1) and neomycin (500ug/ml), (from Tn5). Total genomic DNA obtained from single colony purified nodule isolates (Section 2.6) was digested with *Eco*RI and the fragments separated on a 1% agarose gel. A Southern blot of this gel was probed with the 7.5kb *Eco*RI fragment and two bands hybridized in each nodule reisolate track, one corresponding to 7.5kb and the other approximately 13kb, corresponding to the 7.5kb fragment containing Tn5 (5.7kb).

3.4 RESTRICTION ENZYME MAPPING OF THE 7.5KB *ECO*RI FRAGMENT

Plasmid DNA from strain PN1000 (pPN27) was prepared (Section 2.7.3) and digested with *Eco*RI. The vector and 7.5kb insert were separated on a 0.7% agarose gel and the 7.5kb fragment recovered by electroelution (Section 2.13). The 7.5kb fragment was subjected to digestion with *Bam*HI, *Bgl*II, *Kpn*I, *Pst*I, *Sal*I and *Xho*I. No sites were shown for *Bam*HI, *Bgl*II or *Kpn*I. *Pst*I cut the 7.5kb fragment into two fragments of 2.1 and 5.4kb. *Xho*I cut the 7.5kb fragment into three fragments of 1.8, 2.7 and 2.8kb. There were seven *Sal*I sites in the 7.5kb fragment giving rise to 8 fragments on digestion of the 7.5kb fragment with *Sal*I. Respective sizes were 0.2, 0.3, 0.7, 0.8, 1.0, 1.1, 1.2 and 2.2kb. By digesting the 9.2kb *Hin*dIII fragment from pPN318 with the same enzymes it was possible to determine which of the fragments were at the ends of the 7.5kb *Eco*RI fragment, since the 9.2kb *Hin*dIII

Figure 11:

Lotus pedunculatus plants:

A. Uninoculated negative control (no nitrogen).
B. Inoculated with wild type *R. loti* strain NZP2037.
C. Inoculated with *R. loti* strain PN239 (Bar^- mutant).
D. Inoculated with *R. loti* strain PN239/pPN28.

A B C D

fragment contains the 7.5kb *Eco*RI fragment, the internal fragments will remain unchanged but the end fragments will be longer as the *Hin*dIII fragment extends past the ends of the *Eco*RI fragment. The orientation of the ends was determined by using the 5.9kb *Eco*RI fragment from pPN318 as a hybridization probe to the enzyme digests of the *Hin*dIII fragment, as this *Eco*RI fragment overlaps the left end of the 9.2kb *Hin*dIII fragment (Figure 9). A double digest of the 9.2kb *Hin*dIII fragment with *Xho*I and *Pst*I demonstrated that the *Pst*I site was in a *Xho*I fragment of approximately 4.5kb which hybridized to the 5.9kb *Eco*RI probe (Table 8).

Table 8: Fragment sizes (kb) which result from digestion of the 9.2kb *Hin*dIII fragment.

*Pst*I	*Xho*I	*Xho*I/*Pst*I
5.6	*4.5	*3.6
*3.6	2.9	2.9
	1.8	1.8
		0.8

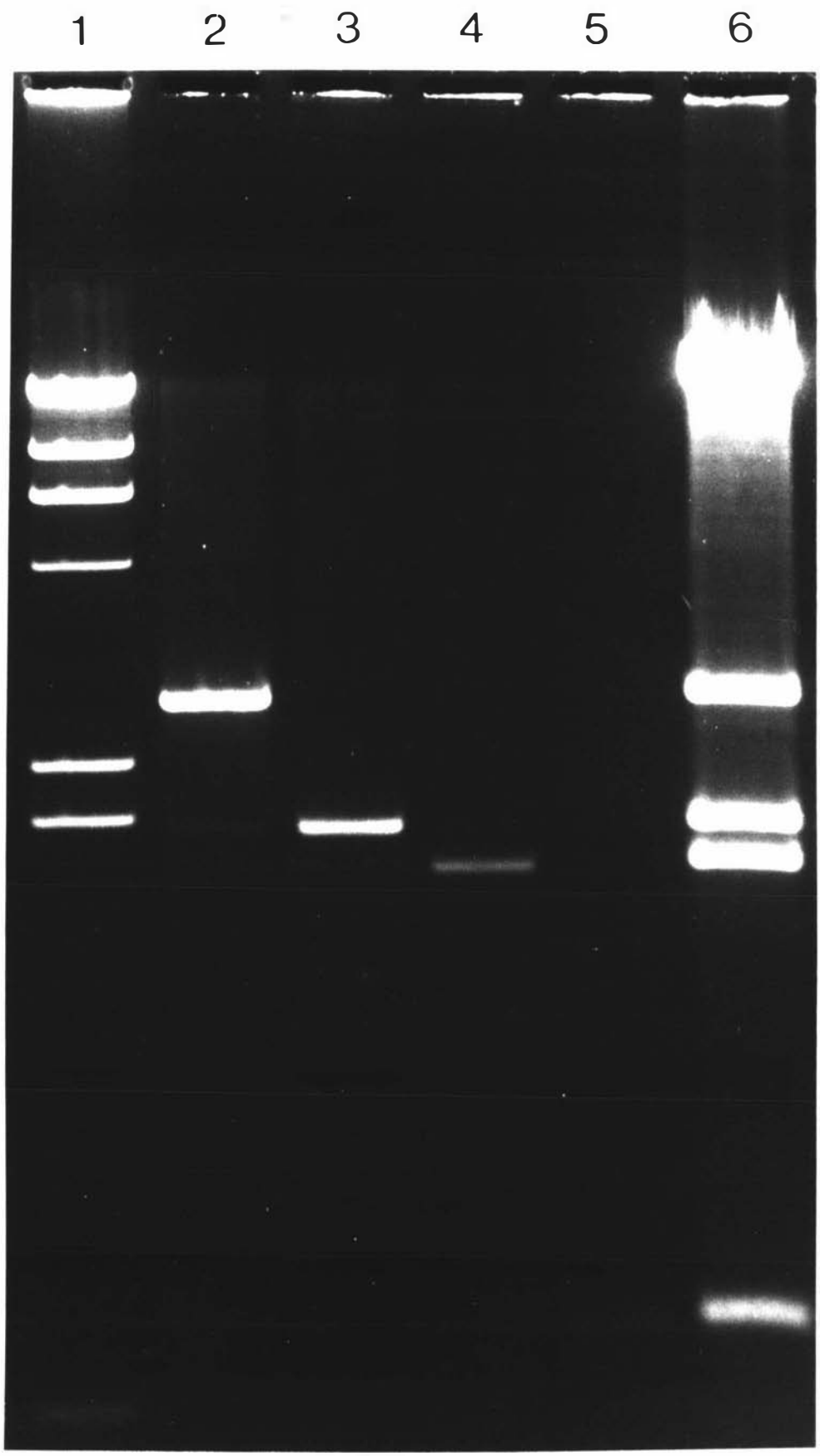
* denotes hybridization with the 5.9kb *Eco*RI fragment from pPN318.

Since there were no sites for *Xho*I or *Pst*I in pLAFR1 it was convenient to digest pPN28 with *Eco*RI, *Xho*I and *Pst*I in a triple digest in order to divide the 7.5kb fragment up into smaller pieces of such different sizes that they could be separated from each other in an agarose gel. Figure 12 shows the triple digest and the recovered fragments.

Figure 12:

Fragments electroeluted from a *Xho*I/*Pst*I/*Eco*RI triple digest of pPN28. The fragments have been separated by electrophoresis on a 0.8% agarose gel.

Lane 1; lambda DNA digested with *Hin*dIII.
Lane 2; 2.7kb fragment from triple digest.
Lane 3; 2.1kb fragment from triple digest.
Lane 4; 1.8kb fragment from triple digest.
Lane 5; 0.8kb fragment from triple digest.
Lane 6; *Xho*I/*Pst*I/*Eco*RI triple digest of pPN28.


1
2
3
4
5
6

The four fragments from the *Eco*RI/*Xho*I/*Pst*I triple digest were used as hybridization probes to *Sal*I, *Sal*I/*Xho*I and *Sal*I/*Pst*I digests of pPN28. In each case the plasmid was also digested with *Eco*RI to excise the 7.5kb insert from the vector. Figure 13A shows the fragments obtained from the above digest. Figure 13B and C show autoradiographs of these fragments probed with the 2.7kb *Xho*I - *Eco*RI and 0.8kb *Pst*I - *Xho*I fragments from the *Eco*RI/*Xho*I/*Pst*I triple digest. The fragment sizes and hybridization details are summarized in Table 9.

The *Sal*I/*Xho*I/*Pst*I restriction map resulting from the above observations is presented in Figure 14.

Table 9: Hybridization of four restriction fragments from the 7.5kb *Eco*RI fragment with double and triple digests of pPN28

pPN28 digested with: A. *Eco*RI and *Sal*I					B. *Eco*RI, *Sal*I and *Xho*I					C. *Eco*RI, *Sal*I and *Pst*I				
pPN28 digest fragment sizes (kb)	Hybridization to probes from the *Eco*RI/*Xho*I/*Pst*I triple digest of pPN28: 2.7	2.1	1.8	0.8	pPN28 digest fragment sizes (kb)	Hybridization to probes from the *Eco*RI/*Xho*I/*Pst*I triple digest of pPN28: 2.7	2.1	1.8	0.8	pPN28 digest fragment sizes (kb)	Hybridization to probes from the *Eco*RI/*Xho*I/*Pst*I triple digest of pPN28: 2.7	2.1	1.8	0.8
17.3	-	-	-	-	17.3	-	-	-	-	17.3	-	-	-	-
4.3	-	-	-	-	4.3	-	-	-	-	4.3	-	-	-	-
2.2	-	-	+	+	1.4	-	-	+	-	2.2	-	-	+	+
1.2	+	-	-	-	1.2	+	-	-	-	1.2	+	-	-	-
1.1	-	+	-	-	1.1	-	+	-	-	1.1	-	+	-	-
1.0	+	-	+/-	-	0.9	+	-	-	-	1.0	+	-	+/-	-
0.8	-	+	-	-	0.8)					0.8	-	+	-	-
0.7	+	-	-	-	0.8)	-	+	-	+	0.7	+	-	-	-
0.3	-	-	+	-	0.7	+	-	-	-	0.3	-	-	+	-
					0.3	-	-	+	-					

Figure 13:

Mapping of the 7.5kb *Eco*RI fragment by probing restriction endonuclease digests of pPN28 with fragments obtained from a *Xho*I/*Pst*I/*Eco*RI triple digest of pPN28.

A. Agarose gel (0.8%) electrophoresis of:
   Lanes 1 and 7; lambda DNA digested with *Eco*RI and *Hin*dIII.
   Lane 2; pPN28 digested with *Eco*RI and *Sal*I.
   Lane 3; pPN28 digested with *Eco*RI and *Sal*I and *Xho*I.
   Lane 4; pPN28 digested with *Eco*RI, *Sal*I and *Pst*I.
   Lane 5; pPN28 digested with *Eco*RI, *Pst*I and *Xho*I.
   Lane 6; 1.8kb *Xho*I fragment from pPN28 digested with *Sal*I.

B. Autoradiograph of a blot of the above gel after hybridization with the 2.7kb *Xho*I - *Eco*RI fragment (Figure 12).

C. Autoradiograph of a blot of the above gel after hybridization with the 0.8 kb *Pst*I - *Xho*I fragment (Figure 12).

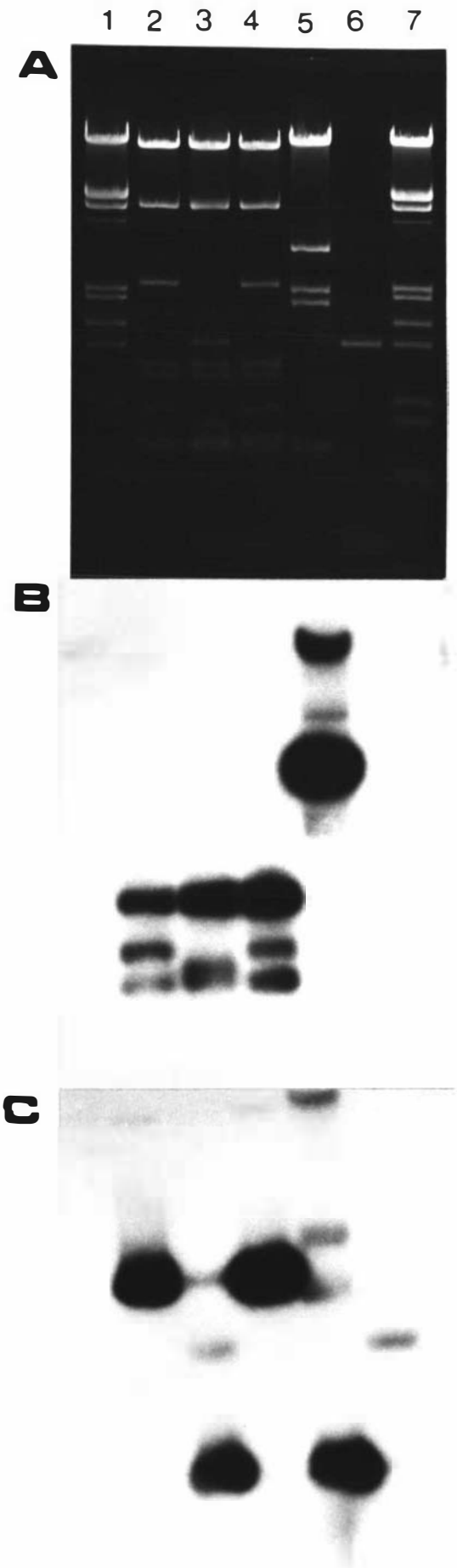
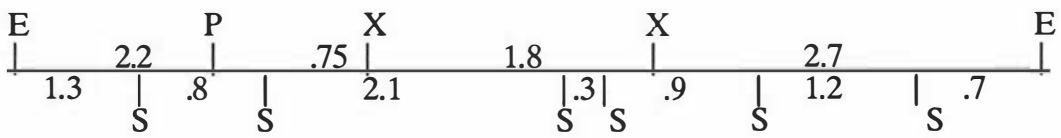

1 2 3 4 5 6 7
A
B
C

Figure 14:

Physical map of the 7.5kb *Eco*RI fragment containing the *bar* region of *R. loti* strain NZP2037.



E = *Eco*RI
X = *Xho*I
S = *Sal*I
P = *Pst*I

3.5 MAPPING THE POSITION OF THE Tn5 INSERT IN THE 7.5KB *ECO*RI FRAGMENT FROM STRAIN PN239

In the wild type *Rhizobium loti* NZP2037 the 7.5kb *Eco*RI fragment hybridizes to a 7.5kb *Eco*RI fragment (Figure 15, Lane 6) and a 9.2kb *Hin*dIII fragment (Figure 15, Lane 7). In the Bar$^-$ mutant PN239, the 7.5kb *Eco*RI fragment hybridizes to a 13.2kb *Eco*RI fragment (Figure 15, Lane 5) corresponding to 7.5kb plus 5.7kb from Tn5. Tn5 contains two *Hind* III sites, 1.2kb from each end (Jorgensen *et al*,. 1979). Thus in a *Hin*dIII digest of PN239 the 7.5kb fragment will hybridize to two *Hin*dIII fragments (Figure 16). These are of approximately 7.4 and 4.4kb in size (Figure 15, Lane 4). This determines the position of insertion of the Tn5 within the 7.5kb fragment that is 7.4 - 1.2 (6.2)kb from one *Hin*dIII site and 4.4 - 1.2 (3.2)kb from the other.

The orientation of the two pieces was determined by probing a *Hin*dIII digest of DNA from strain PN239 with the 5.9kb *Eco*RI fragment which overlapped the left end of the *Hin*dIII fragment. This resulted in two bands hybridizing in the *Hin*dIII digest of DNA from strain PN239 (Figure 17). The upper band of approximately 9.4kb corresponded to the *Hin*dIII fragments overlapping the left of the probe 5.9kb *Eco*RI fragment. The lower fragment of approximately 7.4kb indicated that the Tn5 insert is positioned as shown in Figure 16B.

The mapping of the Tn5 insert places it within the 1.8kb *Xho*I fragment in Figure 14.

Figure 15:

Hybridization of the 7.5kb *Eco*RI fragment from pPN28 to total genomic DNA from strains NZP2037 and PN239.

A. Agarose gel electrophoresis of:

Lane 1; lambda DNA digested with *Eco*RI.
Lane 2; empty.
Lane 3; lambda DNA digested with *Hin*dIII.
Lane 4; total genomic DNA from strain PN239 digested with *Hin*dIII.
Lane 5; total genomic DNA from strain PN239 digested with *Eco*RI.
Lane 6; total genomic DNA from strain NZP2037 digested with *Eco*RI.
Lane 7; total genomic DNA from strain NZP2037 digested with *Hin*dIII.

B. Autoradiograph of a blot of the above gel after hybridization with the 7.5kb *Eco*RI fragment from pPN318.

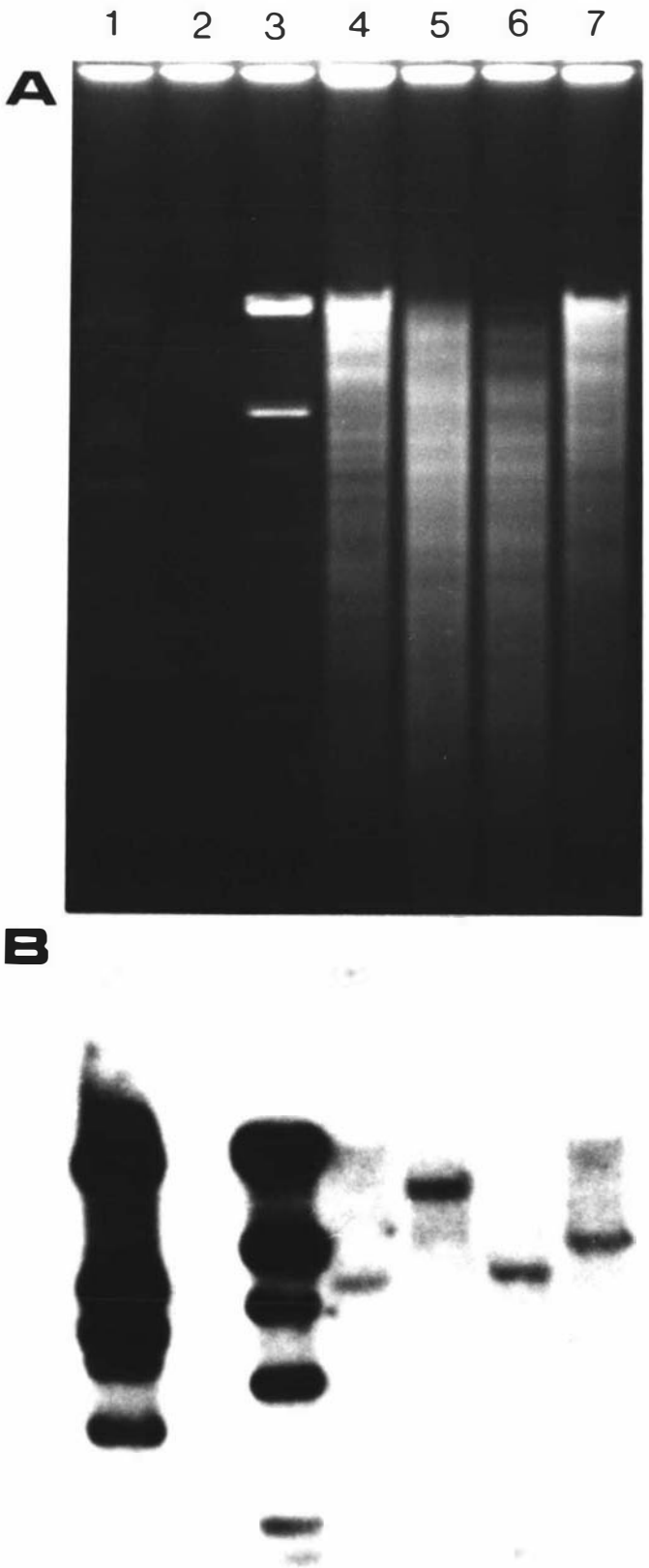
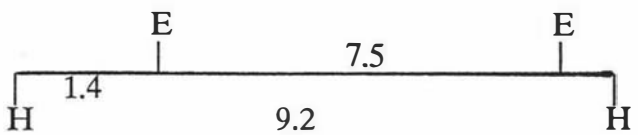

1
2
3
4
5
6
7
A
B

Figure 16:

Relationship of restriction fragments in *R. loti*:

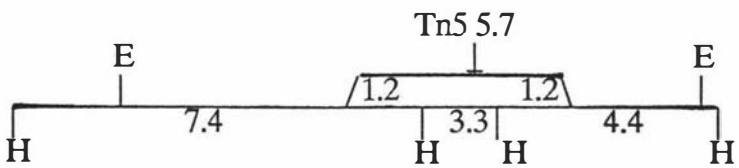
A; strain NZP2037 and B; strain PN239.

A.



WILD TYPE NZP2037

B.



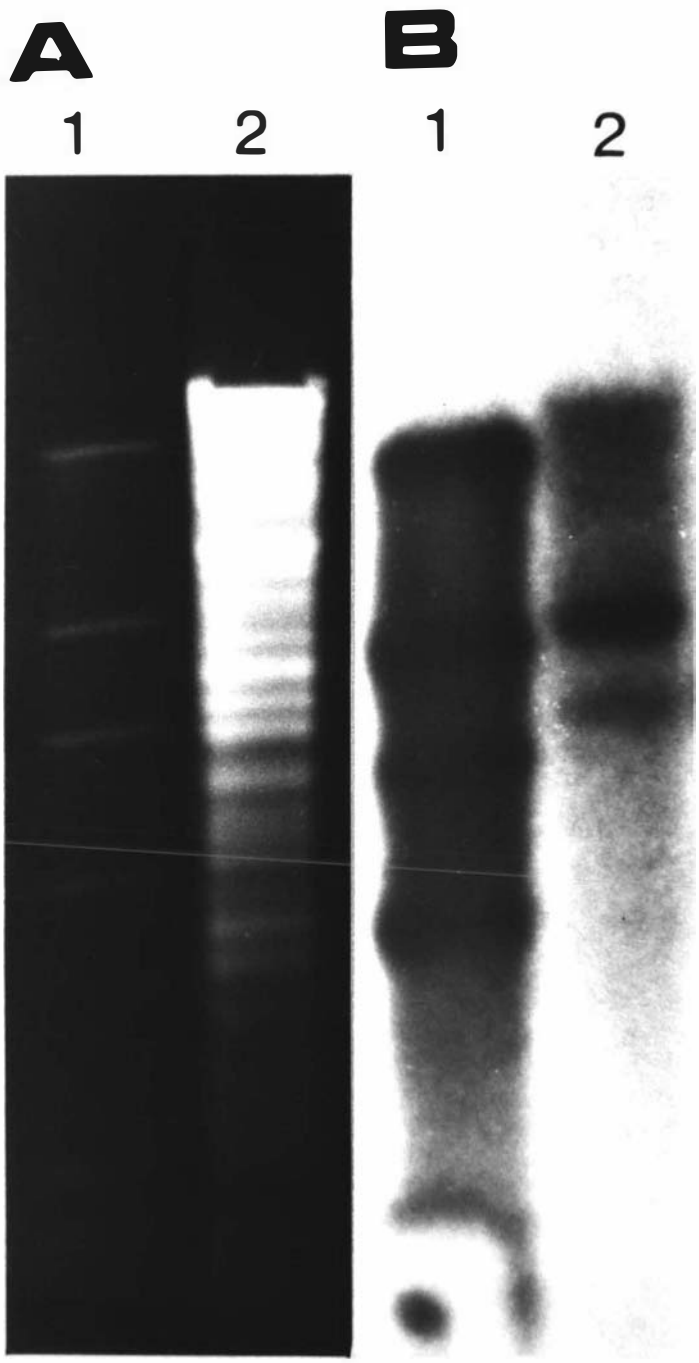
BAR MUTANT PN239

Figure 17:

Determination of the position of the Tn5 insert in strain PN239.

A. Agarose gel electrophoresis of:
   Lane 1; lambda DNA digested with *Hin*dIII.
   Lane 2; total genomic DNA from strain PN239 digested with *Hin*dIII.

B. Autoradiograph of a Southern blot of the above gel probed with the 5.9kb *Eco*RI fragment from pPN318.


A
B
1
2
1
2

3.5.1 Insertion of further Tn5's into the 7.5kb *Eco*RI fragment and the mapping of these insertions[1] (Ward *et al.*, 1989)

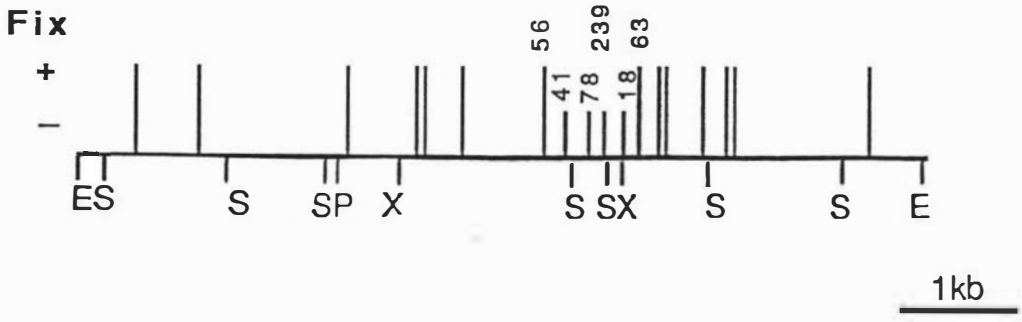
Further Tn5 mutagenesis of the 7.5kb fragment was carried out in *E. coli* strain PN1001(HB101/pPN28), using phage lambda 467 as the source of the Tn5. Transposon Tn5 insertions in pPN28 were selected by transferring this plasmid into *E. coli* strain C2110 by conjugation and selecting for $Kan^{R}Nal^{R}Tet^{R}$ transconjugants. Plasmid DNA was prepared from randomly selected colonies and cut with *Eco*RI to identify whether the Tn5 had inserted into the pLAFR1 vector or into the 7.5kb *Eco*RI fragment. Inserts in the 7.5kb *Eco*RI fragment were mapped by preparing *Hin*dIII/*Eco*RI double digestions of each plasmid and then probing a Southern blot of these digests with the 2.7kb *Eco*RI-*Xho* I fragment (Figure 14). The positions of the Tn5 insertions isolated are shown in Figure 18. Each of the pPN28::Tn5 cosmids was then transferred to *R. loti* strain PN184 in a further triparental cross, and the Tn5 homogenotized into the genome by plasmid incompatibility using pPH1JI. Total DNA was isolated from these $Str^{R}Gm^{R}Tet^{R}$ *R. loti* colonies, cut with *Eco*RI, separated on an agarose gel and a Southern blot of this gel probed with the 7.5kb *Eco*RI fragment to confirm that a double cross-over had occurred. Each confirmed mutant was then inoculated onto *Lotus pedunculatus* to seedlings to determine the symbiotic phenotype. Of the 17 Tn5 containing mutants tested, three (corresponding to insertions 18,41, and 78) gave a Fix^{-} phenotype and these all mapped close to the point of Tn5 insertion in strain PN239 (Figure 18). The outermost boundaries of this locus were defined by insertions 56 and 63 (Figure 18). When strains PN1093, PN1094 and PN1095 (containing insertions 41, 78, and 18 respectively) were tested on *Lotus corniculatus* they all formed fully effective nodules, as did strain PN239, indicating that the gene is host-specific. All mutants were prototrophic and formed mucoid colonies similar to the wild type.

[1] Note: The work described in Section 3.5.1 was mainly carried out by D.B. Scott and P. Ball, but is briefly described here because it draws on results already presented in this thesis and the information gained was used in the remainder of the thesis.

Figure 18:

The position of Tn5 insertions isolated in the 7.5kb *Eco*RI fragment (Section 3.5.1).

Restriction enzyme sites are abbreviated as follows:
E = *Eco*RI, S = *Sal*I, P = *Pst*I, X = *Xho*I.



3.6 HYBRIDIZATION OF RESTRICTION FRAGMENTS FROM THE *bar* REGION TO TOTAL GENOMIC BLOTS OF DNA FROM OTHER RHIZOBIA

3.6.1 Hybridization of the 7.5kb *Eco*RI Fragment from pPN318

The 7.5kb *Eco*RI fragment from pPN318 was electroeluted from an agarose gel (Section 2.13), labelled with ^{32}P (Section 2.18) and hybridized (Section 2.19) to Southern blots (Section 2.16) of total genomic DNA (Section 2.6) from strains of *R. leguminosarum* bv. *trifolii* (PN100, PN104, NZP561), *R. meliloti* (SU47 and NZP4009), slow-growing rhizobia which nodulate *Lotus* (CC814S and NZP2257), as well as to *R. loti* NZP2037 and NZP2213.

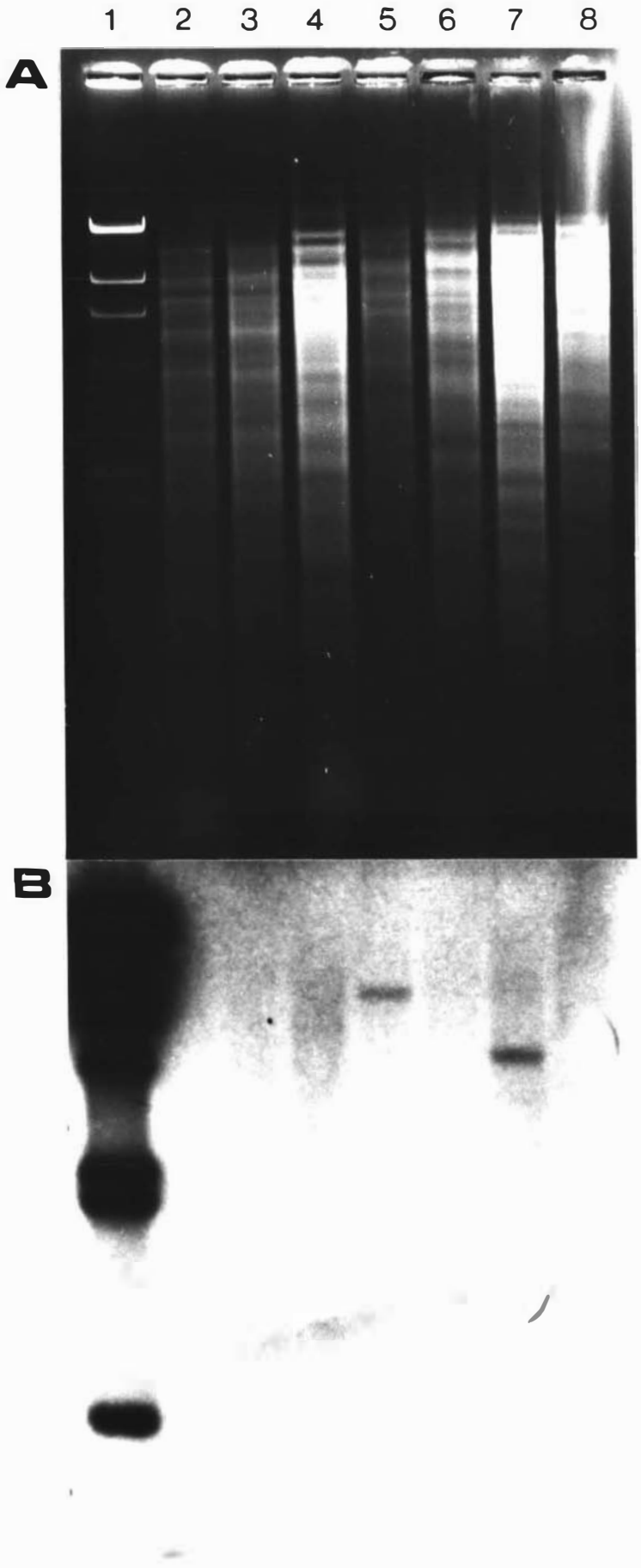
Under the stringency conditions used (hybrization at 65°C, washing with 2xSSC, ambient temperature), hybridization was observed between the 7.5kb *Eco*RI fragment from *R. loti* strain NZP2037 and DNA from *R. loti* strain NZP2213 another fast-growing *Lotus* strain. No hybridization was observed to DNA from strains of *R. leguminosarum*, *R. meliloti* or *Bradyrhizobium* sp. (*Lotus*) strain CC814S. This result was identical to that shown in Figure 19 with a more specific fragment of the *bar* gene (Section 3.6.2).

Figure 19:

Hybridization of the *bar* gene to total genomic DNA from various rhizobia.

A. Agarose gel (1.0%) of restriction endonuclease digests of DNA from:
Lane 1; lambda DNA digested with *Hin*dIII.
Lane 2; *Rhizobium leguminosarum* biovar *trifolii* strain PN100 digested with *Eco*RI.
Lane 3; *Rhizobium leguminosarum* biovar *trifolii* strain PN104 (a Sym plasmid cured derivative of strain PN100) digested with *Eco*RI.
Lane 4; *Rhizobium leguminosarum* biovar *trifolii* strain NZP561 digested with *Eco*RI.
Lane 5; *Rhizobium loti* strain NZP2037 digested with *Eco*RI.
Lane 6; *Rhizobium meliloti* strain SU47 digested with *Eco*RI.
Lane 7; *Rhizobium loti* strain NZP2213 digested with *Eco*RI.
Lane 8; *Bradyrhizobium* sp. (*Lotus*) strain CC814S digested with *Eco*RI.

B. An autoradiograph of a Southern blot from the above gel after hybridization with the *Sph*I-*Alu*I fragment described in Section 3.6.2.


1 2 3 4 5 6 7 8
A
B

3.6.2 Hybridization of the 0.5kb *Sph*I - *Alu*I Fragment Containing the *bar* gene

The restriction endonuclease *Sph*I recognizes the sequence GCATGC. This sequence occurs uniquely within the sequenced region (Section 3.8, Figure 24) at a point 2 base pairs upstream of the presumed ATG start codon for the *bar* gene. The fragment between this site to the *Alu*I site (AGCT) which starts at base 503 of the *bar* gene sequence, was eluted and used as a hybridization probe. This fragment was assumed to be specific for this gene and was missing only 76 bases from the 3' end of the sequence.

As was observed for the 7.5kb *Eco*RI probe (Section 3.6.1), under the same stringency conditions, hybridization was observed between the *Sph*I-*Alu*I fragment probe and DNA from *R. loti* strains NZP2037 (Figure 19, Lane 5) and NZP2213 (Figure 19, Lane 7). No hybridization was observed to DNA from strains of *R. leguminosarum* (Figure 19, Lanes 3 and 4), *R. meliloti* (Figure 19, Lane 6) or to DNA from *Bradyrhizobium* sp. (*Lotus*) strain CC814S (Figure 19, Lane 8).

3.7 SUBCLONING FRAGMENTS FROM THE 7.5KB *ECO*RI FRAGMENT INTO MI3 VECTORS FOR SEQUENCING

The fragments described in Table 10 were electroeluted from a 0.8% agarose gel (Section 2.13), extracted with phenol-chloroform (Section 2.8) and ethanol precipitated (Section 2.9). DNA recovered from the precipitates was ligated (Section 2.14) into the appropriate M13 vector (Table 10). Competent *E. coli* JM101 cells were transformed with these ligation mixtures (Section 2.15.2) and white plaques picked from the plates. The phage was grown in *E. coli* strain JM101 and phage DNA recovered (Section 2.20.1). The rapid boil method (Section 2.7.1) was used to verify the presence of an insert in the replicative form of M13. One microlitre (1ul) of phage DNA was electrophoresed on an agarose gel to ensure that the phage DNA had not been lost in any of the steps in the protocol. Fragments from a *Hae*III digest of the 1.8kb *Xho*I fragment and fragments from an *Alu*I digest of the 7.5kb *Eco*RI fragment which hybridized to the 1.8kb *Xho*I fragment, were shotgun cloned into *Sma*I digested M13mp9. The single-stranded M13 phage DNA containing the fragment of interest was then used as a template for sequencing the insert DNA.

Table 10: Fragments cloned into M13 vectors for determination of the DNA sequence.

Fragment	Source	Vector	Template numbers used in sequencing
1.8 kb *Xho*I fragment	*Xho*I digest of pBR328 - 7.5 kb *Eco*RI fragment	*Sal*I cut M13mp9.	54, 54A, 55
0.3 kb *Sal*I fragment	*Sal*I digest of 1.8 kb *Xho*I fragment	*Sal*I cut M13mp9	13, 22, 22A
2.7 kb *Eco*RI - *Xho*I fragment	*Xho*I/*Eco*RI/*Pst*I digest of pLAFR1 - 7.5 kb fragment	*Eco*RI-*Sal*I cut M13mp8	28, 28A
1.4 kb *Xho*I - *Sal*I fragment	*Sal*I digest of 1.8 kb *Xho*I fragment	*Sal*I cut M13mp9	191, 192, 193, 195
0.9 kb *Xho*I - *Sal*I fragment	*Sal*I digest of 2.7 kb *Xho*I - *Eco*RI fragment	*Sal*I cut M13mp9	1
Random *Hae*III fragments	*Hae*III digest of 1.8 kb *Xho*I fragment	*Sma*I cut M13mp9	183, 184, 177, 170, 174, 131, 162, 164, 132, 133, 115, 102, 124, 224, 166
Random *Alu*I fragments	*Alu*I digest of 7.5 kb *Eco*RI fragment or 1.8 kb *Xho*I fragment	*Sma*I cut M13mp9	370, 479, 366, 966, 362, 360

3.8 SEQUENCING OF THE SUBCLONED FRAGMENTS AND ASSEMBLY INTO A CONTIGUOUS SEQUENCE

Either the 17mer MI3 "universal primer" or the 15mer primer (New England Biolabs) were annealed to the single stranded phage template DNA (Section 2.18). The relationship between restriction sites used in cloning the insert fragments and the position in which the primers used in the sequencing reaction anneal to MI3 is shown in Figure 20. The sequencing reaction was carried out as described in Section 2.18. Polyacrylamide gels were prepared and electrophoresis and autoradiography carried out as described in Section 2.18. The sequence was manually read from the gel and entered using the Staden NBATIN program (Staden, 1980) on a VAX 750 computer. Overlapping regions of sequence data were identified using the comparison program DBCOMP and the data was assembled into a contiguous sequence using either the Staden VTUTIL programs or the University of Wisconsin Genetics Computer Group (UWGCG) GELASSEMBLE program. The overlapping templates and their relationships are shown in Figure 21. The origin of the various templates can be found by reference to Table 10 (Section 3.7).

Figure 20:

The region of M13mp8 and M13mp9 used for cloning fragments for sequencing. The 17mer "universal primer" is the complement of the boxed sequence.

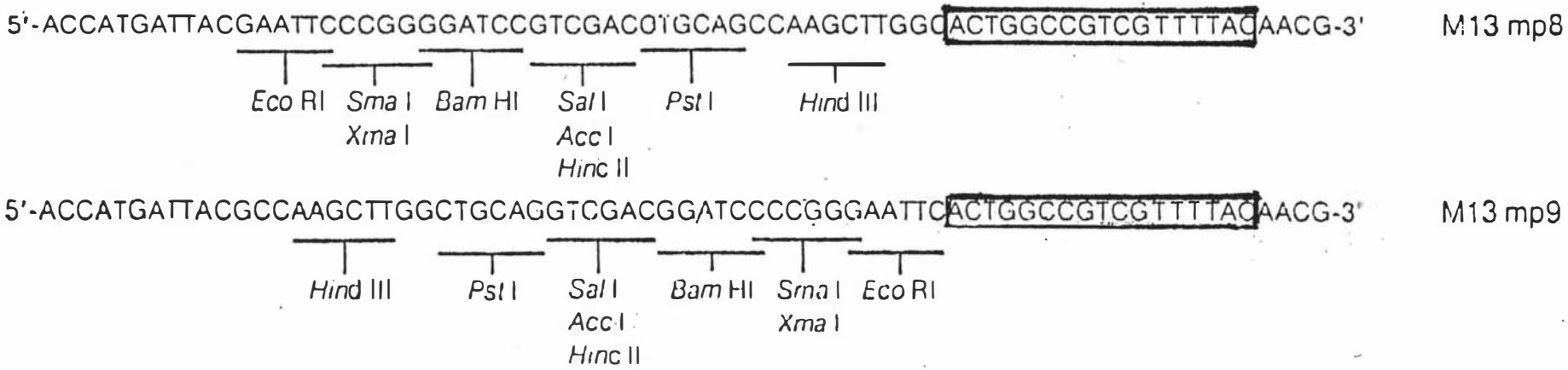
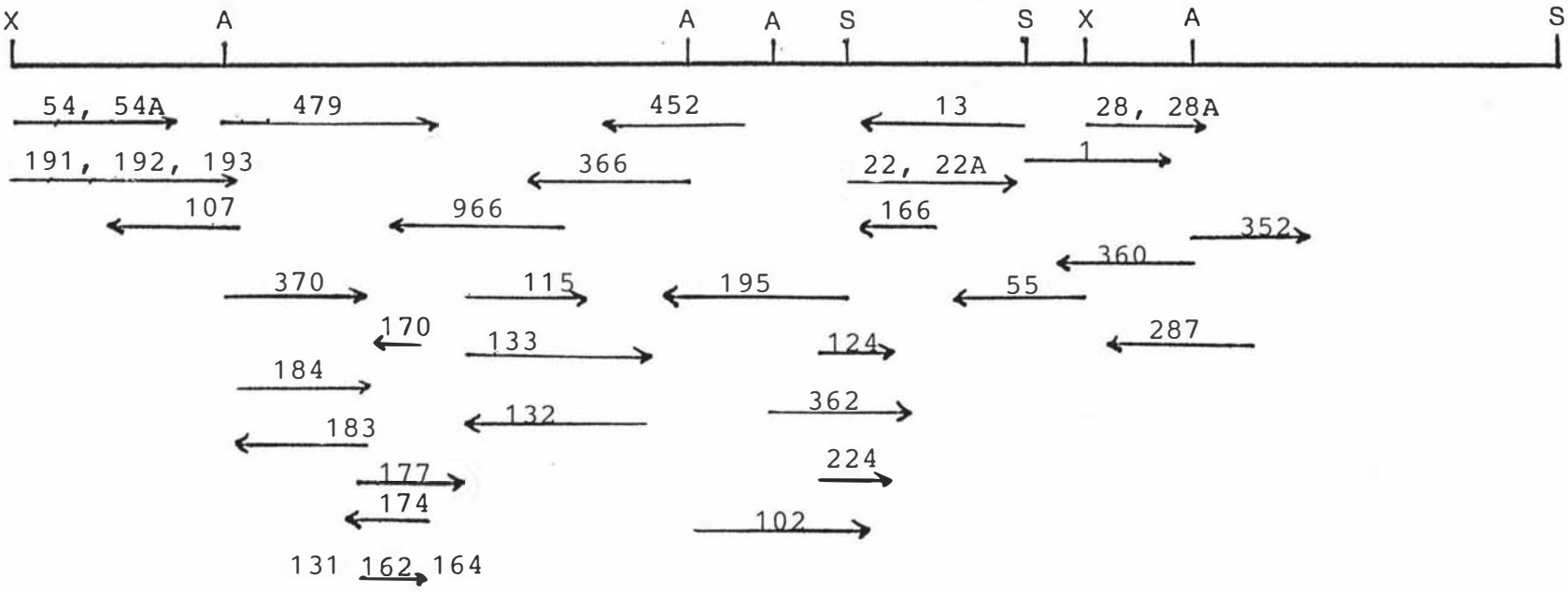

5'-ACCATGATTACGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTTGGCACTGGCCGTCGTTTTACAACG-3'
M13 mp8
Eco RI
Sma I
Xma I
Bam HI
Sal I
Acc I
Hinc II
Pst I
Hind III
5'-ACCATGATTACGCCAAGCTTGGCTGCAGGTCGACGGATCCCCGGGAATTCACTGGCCGTCGTTTTACAACG-3'
M13 mp9
Hind III
Pst I
Sal I
Acc I
Hinc II
Bam HI
Sma I
Xma I
Eco RI

Figure 21:

Strategy for sequencing the *bar* region of *R. loti* strain NZP2037 showing the location and direction of templates sequenced and used to construct the consensus sequence. Enzyme sites are abbreviated as follows; X, *Xho*I; A, *Alu*I; S, *Sal*I. Template numbers correspond to those in Table 10.


X
A
A
A
S
S
X
A
S
54, 54A
479
452
13
28, 28A
191, 192, 193
366
22, 22A
1
107
966
166
352
360
370
115
195
55
170
133
124
287
184
362
183
132
224
177
174
102
131
162
164

3.9 RESOLUTION OF COMPRESSIONS OBSERVED IN THE SEQUENCE

A number of regions in the sequence were found to give a compressed band pattern on standard 8M urea 6% acrylamide gels. In addition to sequencing the opposite strand of the DNA two methods were used to resolve these compressions:

3.9.1 Reverse Transcriptase

The template containing the compressed region was resequenced using avian myeloblastosis virus (AMV) reverse transcriptase instead of Klenow fragment in the sequencing reaction. This necessitated changing the deoxynucleotide triphosphate/ dideoxynucleotide triphosphate ratios as indicated in Section 2.20.4.

Use of reverse transcriptase provided a less ambiguous reading through the compressed areas however it was not a replacement for Klenow fragment since there were a number of places in the sequence where bands were observed in all four tracks of the sequencing gel (Figure 22) which would have made it impossible to determine the sequence if this had not already been found with Klenow fragment.

3.9.2 40% Formamide 7M Urea Gels

The cheapest and most effective way of resolving gel compressions was to run a polyacrylamide gel containing 7M urea and 40% formamide (Section 2.18). This method resolved all the gel compressions in the sequenced region, an example is shown in Figure 23.

The consensus sequence obtained for the region indicated in Figure 21 after resolution of the compressions (Section 3.9) is given in Figure 24. A probable protein translation and other features are described in Section 3.12 "Features of the sequence".

Figure 22:

Resolution of sequence compressions using reverse transcriptase in the sequencing reaction. The autoradiograph of an 8M urea sequencing gel demonstrates the presence of bands (arrowed) in all four tracks (AGCT) of the sequence as described in Section 3.8.1.

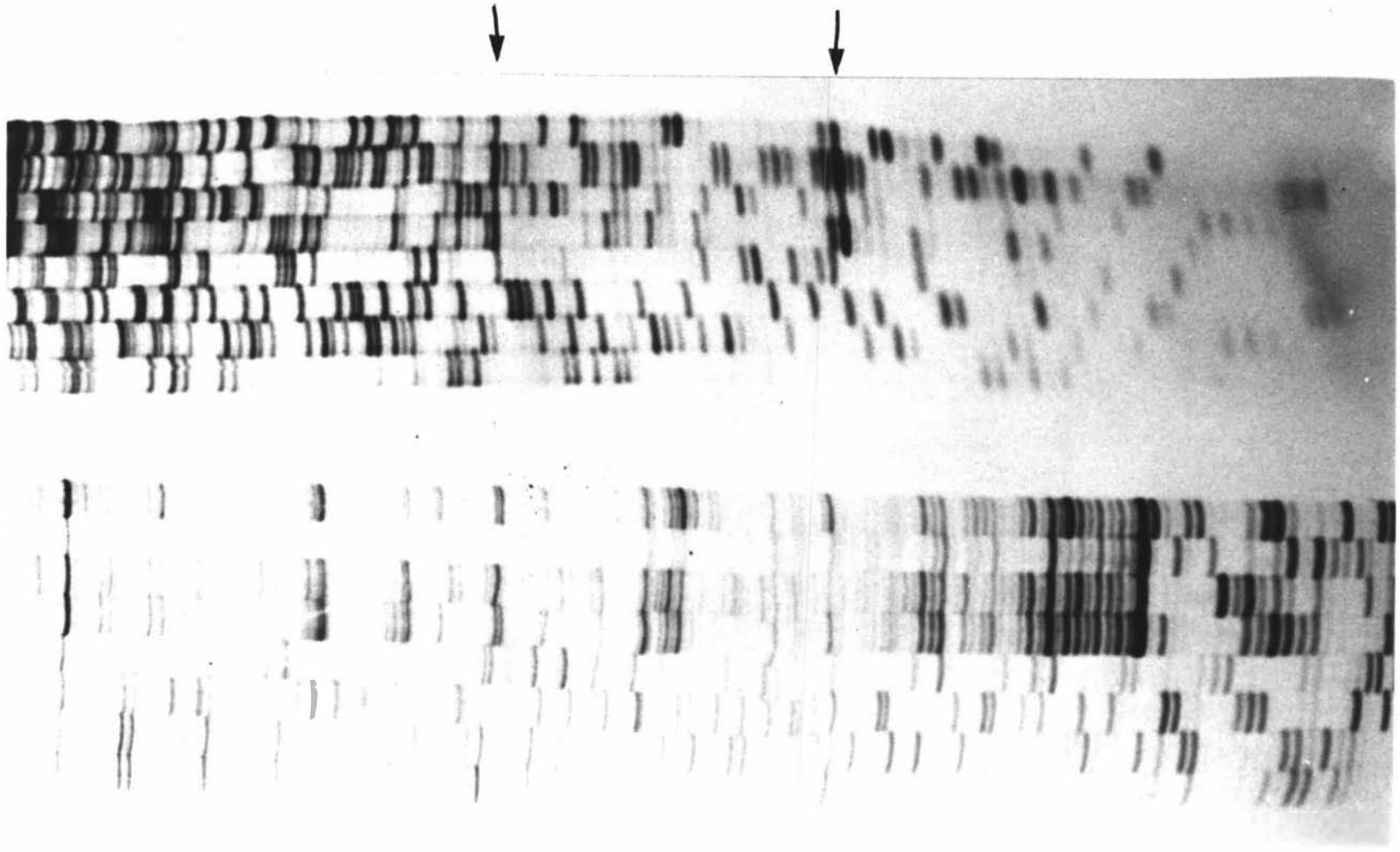
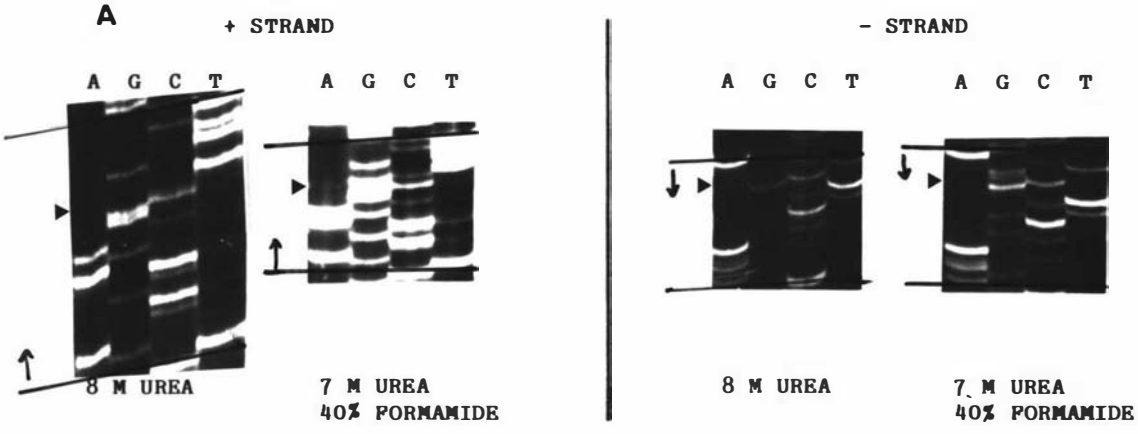


Figure 23:

Resolution of sequencing compressions using formamide gels.
Sections of sequencing gels (A) and the sequence read from them (B) show compressions in 8M urea gels and their resolution in 7M urea 40% formamide gels (arrowed). In the sequence read from these gels X's indicate regions of compression.

A



B

+ STRAND	8 M UREA	T G A C C G C C A G A C G G C X X T G T T C T
	7 M UREA 40% FORMAMIDE	T G A C C G C C A G A C G G C G C T G T T C T
- STRAND	8 M UREA	A C T X X C G G T C T G C C G C G A C A A G A
	7 M UREA 40% FORMAMIDE	A C T G G C G G T C T G C C G C G A C A A G A

Figure 24:

The consensus sequence of the *bar* gene region. The position of restriction enzyme sites for *Sal*I, *Xho*I, *Sph*I, *Alu*I and *Hae*III are indicated. The sequenced position of Tn5 insertions are shown by ▽(Fix$^+$) and ▼(Fix$^-$).

The Tn5 insertion numbers correspond to those in Section 3.5.1 and Figure 18.

The start and end of the putative open reading frame (ORF) for the *bar* gene is shown (bases 1487-2066).

```
   1  TCGAGTTCCT TCAGCACGTC CTTCATCGCT CGGTCCCCCA TGCGCGGTAC
      XhoI

  51  TTCTTGGCAG GCATTCGCCA GATATGGCAG AATAAGCCTA TGAAGAACAA

 101  CGTAGAAATA TCCGAGGATT TGAACCGCCG GATTGATATG CTCACGTCCC

 151  GCTCGACACT GACACGTGAT CAGATCATTG AGGATGCGCT GTCCCATGGA

 201  CGTTCGCTGG CGTGGCAAGA AAAATGGGTC GCAGGCCGTT CAGGCGGGAA
                                        HaeIII

 251  TAGAGGGCGC CGACCGTGGC CGACTTTGCC AATGAGGAAG AGATCGCGAC
                      HaeIII

 301  CGTTCTCAAC AAGTACAGTC AGGCATCAGC CTCTGTGTGA ATCGTCTGGA

 351  CCCGACACTA TCTGATCGAG CTAGCGTCAA TCGGCGACTA CACTGGCCAA
                       AluI                          HaeIII

 401  CATAAGCCGC GGGCCGCTGC CCGGATCGTC AGTGAGATCC ATTCAAAGAC
                HaeIII

 451  CCCCCGGCTG TTATCCGCCA ACCCGTTCAT CGGACGAGTT GGGGAGATCA

 501  AGGGCACGCG TGAAATGGTG ATTCCCGGCA CACCCGACAT CGTCGCCTAT

 551  CGGGTCACGG ACACTCAGAT AGAAATTCTG TTTGTGCAAC ATGGCGCCAG

 601  GAAGTGGCCG GATGAGGTGT GAGCGGACAT TTCGGACTGC TCTCTAGCCG
        HaeIII

 651  GCTTTTGCTG CGCTGCAACA TTACGCCCCT TGCATTCCGG GGCGAACTCA

 701  GCCATATGCA GCCCATAGGC GCTTGGGCCG GAGATTCCGG CCTTCTCGAC
                                HaeIII         HaeIII
```

```
 751  AAATGAGACT GGTTGCGTCT GTTTTCCCTC TTTCCGGCAA ATCGGCAAGG

 801  CGGCGAAAAA GCCCCGTGGC ATGATGCCTG CGGGCACGAC AGCGCAGCCG

 851  AGTGGACGCA CACGTCCGCC CGCCTTGTCG TTTCATATCA AATTTTTCGA

 901  CGGCAGGTTG CGCCCTGCCG CCATTGATGT TTGCGGCAAA GAAGCCGCGT

 951  GAAAGAAGAT TATTTTGACC AACGAAAACA CATTGCCCGG CAACGATACC

1001  GCGGAACTCT CCGGTTGTCG CAGCGCTGGG AATCACGGGT CGCGCTGCTC

1051  AAGGCCACCC ACAATGCCGG ATTCACCGAA CCGAAGCCGA TCCAGATGCA
        HaeIII

1101  GGCAATCCGC CCCAAGCTGG AAGGCCGTGA CAGTTTTCGG CATCGCGCAG
                    AluI      HaeIII           (56)

1151  ACGGGCTCCG GCAAGACCGC GGCCTTTGCG CTGCCTATCC TGTCGAAGAT
                            HaeIII

1201  CATCGCGCTC GGCACCAAGC GGCGCGCCAA GACGACCCGC GCGCTCATCC
                                    (41)

1251  TGGCGCCGAC CCGCGAGCTC GCCGTGCAGA TCGAGGACAC CATCAAGATC
                     AluI

1301  CTCGCCAAGG GCGCGCATGT CTCGACGGCA CTCGTGCTGG GCGGCGTCTC

1351  GCGCTTCAGC CAGGTGAAGA AGGTTGCTCC CGGCGTCGAC ATCCTCATTG
                                          SalI

1401  CCACGCCGGC CGATTGACCG ACCTGGTGCG CGAGGGCGAC CTCATCCTCT
            HaeIII
                                                /--- ORF ---
1451  CCGACACAAA TGGCTGGTGC TGGACGAGGG CGACCGCATG CTCGACATGG
                                          SphI

1501  GCTTCATCAA CGACGTCAAG CGCATCGCCA AGGCGACCGC GCCTGACCGC

1551  CAGACGGCGC TGTTCTCGGC CACCATGCCG GACGAGATCG CCGAACTGGC
                          HaeIII                        HaeIII

1601  CAAGGGTCTT CTGAAGAACC CGGTCCGTAT CGAAGTCTCG CCGCAAAGCA
                                    (239)

1651  CCGCGGCCGC CGAGATTGTC CAGGGCGTCG TCTTCGCCCG CACCAAGCAG
         HaeIII

1701  AAGCGCCAGG TTCTGTCGAC GATGCTCGCT GACGAGGCGA TGAAGTCCGT
                     SalI
```

```
1751  CATCATCTTT TCGCGCACCA AGCATGGCGC CGACCGGGTG ACCAAGGACC
                            (18)

1801  TCGAGCGCGA TGGCTTCAAG GCGGCCGTCA TCCACGGCAA CAAGTCGCAG
      XhoI                   HaeIII

1851  AATGCCCGTC AGAAGGCGCT GAACGATTTT CGCGATGGAT CGGTCCGCAT

1901  TCTGGTCGCG ACCGACATCG CGGCGCGCGG CATCGACGTC CCCGGCATCA

1951  GCCATGTCGT GAATTTCGAC CTGCCGGATG AGGCGGAAAG CTACGTCCAC
                                                  AluI

2001  CGCATCGGCC GCAGCCGGCC GCAACGGCAT GGACGGCATC GCGATCACGC
            HaeIII     HaeIII

              ---ORF---\
2051  TTTGCGATCC TTCTGAGAAC AGCAAGCTGC GCCAGGTCGA GCGCATCATC
                     (63)        AluI

2101  CGCACCAAAC TGCCGATCGT CGCCGACCAT CTCGGCAGCC CCGATCCGCA

2151  GCGCAATCCG GCTGAAAAGA ACGAGCGCTT CGAGCCCGCC AATGATCGCA

2201  ACGACCGCAA TGGCCGTCGT GACCGGCGCC GGCGGCGAAA ACAAGGCTCA
                  HaeIII

2251  ACGGCTTTGG CAGAGCGGTT CGGCGACACC TTTGTGAGCG ACCGAACGCA

2301  GCCTTCA
```

3.10 DETERMINATION OF THE POSITION OF Tn5 INSERTIONS WITHIN THE SEQUENCE

3.10.1 Subcloning Tn5::*Rhizobium* DNA Junction Fragments into M13mp9

Section 3.5.1 describes further Tn5 mutagenesis of the 7.5kb *Eco*RI fragment in pLAFR1 (pPN28). When the pPN28::Tn5 cosmids from strains PN1121 (#41), PN1123 (#18), and PN1135 (#239) were transferred and recombined into *Rhizobium* (Section 3.5.1), the resulting *Rhizobium* strains were Fix$^-$ on *Lotus pedunculatus*. The Tn5 insertions in strains PN1120 (#56) and PN1124 (#63) mapped adjacent to the above Tn5 insertions (Figure 18) but did not cause a Fix$^-$ phenotype when homogenotized into *Rhizobium*. Plasmid DNA was isolated (Section 2.7) from these five *E. coli* strains. The isolated plasmid was digested with *Hin*dIII and *Xho*I which produced *Xho*I -*Xho*I Tn5 junction fragments (Figure 25) which were cloned into *Sal*I digested M13mp9. Only the junction fragments had *Xho*I sites at both ends and so were able to ligate into the M13 vector. (The remainder of the pLAFR1 cosmid also had *Xho*I sites at each end but because of its size (27kb) did not ligate into the M13 vector.)

3.10.2 Sequencing the Tn5::*Rhizobium* DNA Junction Fragments

The junction fragments subcloned into M13mp9 were sequenced (Section 2.20) using a 15mer primer to the end of the Tn5 terminal repeat (Schofield and Watson, 1986; Egelhoff *et al.*, 1985). This produced sequence data reading from the Tn5 through the junction region into the *Rhizobium* DNA and allowed the Tn5 inserts to be precisely located in the sequence. An example is shown in Figure 26. The positions of the five Tn5 inserts discussed in Section 3.10.1 are shown in the consensus sequence (Figure 24).

At the point of insertion into the sequence Tn5 generates a 9 base repeat in the host DNA (Berg *et al.*, 1983; Egelhoff *et al.*, 1985). This was demonstrated for the Tn5 insertions sequenced in this project.

Figure 25:

Sequence mapping of Tn5 inserts in the *bar* locus.

Diagram illustrating the position of Tn5 - *Rhizobium* junction fragments (hatched bars) sequenced to determine the position of the Tn5 insertion (open bar) within the *Rhizobium* sequence. Enzyme sites are abbreviated as follows; E = *Eco*RI, H = *Hin*dIII, X = *Xho*I.

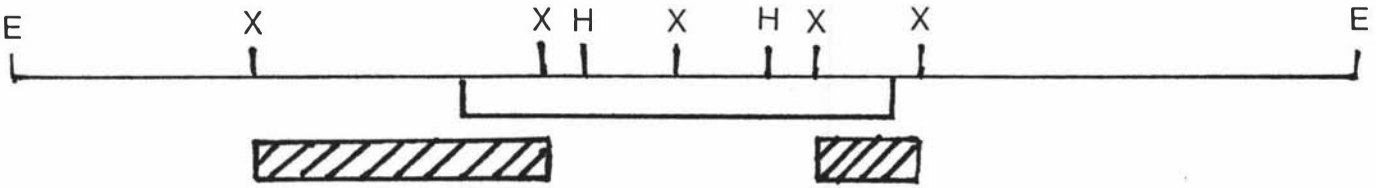
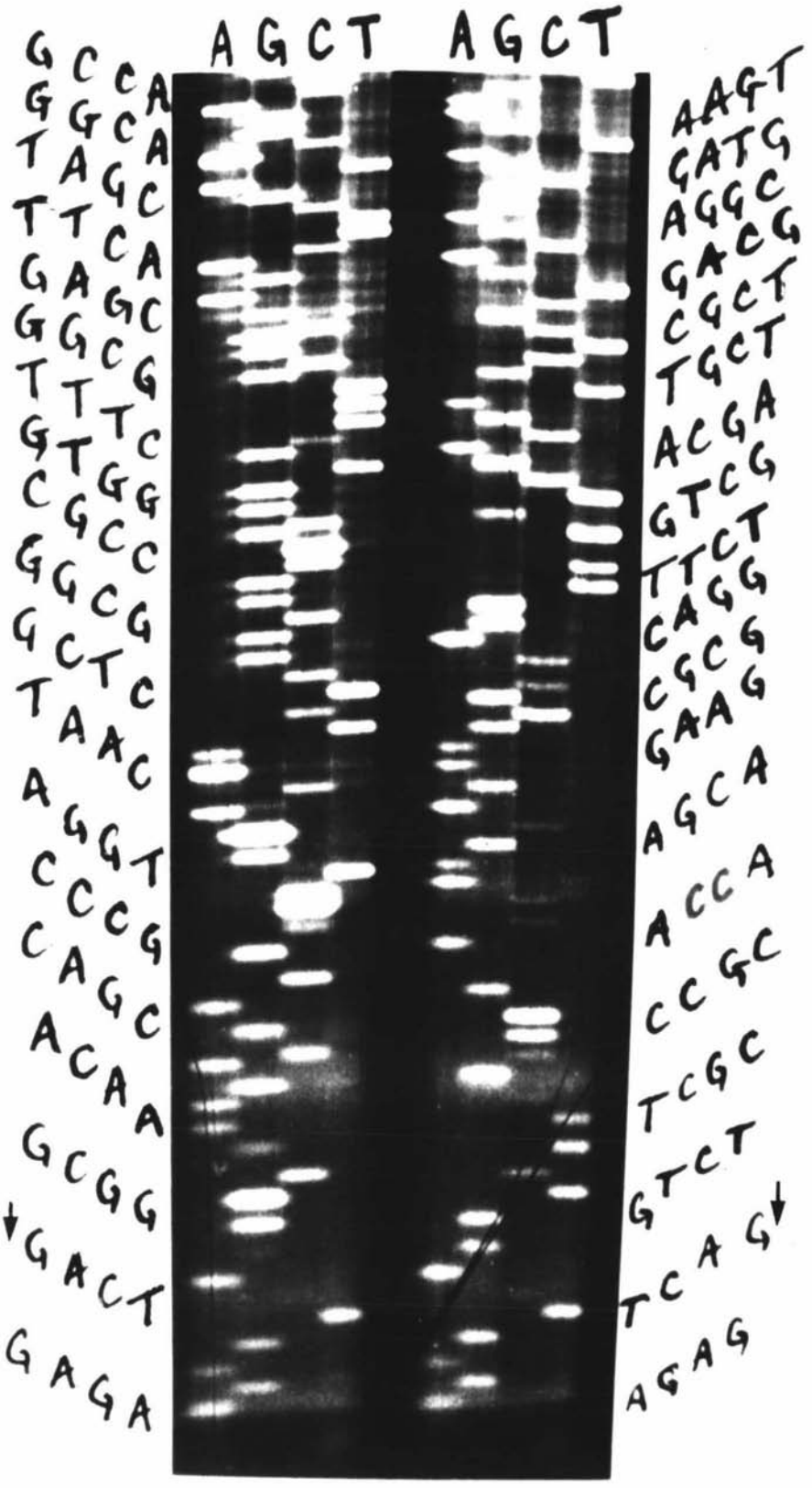

E
X
X
H
X
H
X
X
E

Figure 26:

DNA sequence of *Rhizobium*::Tn5 junction fragments.

Part of an autoradiograph from an 8M urea sequencing gel illustrating the sequence of two Tn5 - *Rhizobium* junctions. The sequence read from the gel is given. Arrows indicate the point of change in the sequence from Tn5 to *Rhizobium* DNA.


A G C T
A G C T
G C C A
G G C A
T A G C
T T C A
G A G C
G G C G
T T T C
G T G G
C G C C
G G C G
G C T C
T A A C
A G G T
C C C G
C A G C
A C A A
G C G G
G A C T
G A G A
A A G T
G A T G
A G G C
G A C G
C G C T
T G C T
A C G A
G T C G
T T C T
C A G G
C G C G
G A A G
A G C A
A C C A
C C G C
T C G C
G T C T
T C A G
A G A G

3.11 COMPUTER ANALYSIS OF THE SEQUENCE DATA

The consensus sequence obtained (Section 3.8) was analysed on a Vax 750 computer using the University of Wisconsin Genetics Computer Group (UWGCG) package (Deveraux *et al.*, 1984).

3.11.1 Open Reading Frames (ORF's)

The program FRAMES was used to identify all possible ORF's starting with an ATG initiation codon in both orientations of the consensus sequence. The output is shown in Figure 27. A number of ORF's of sufficient size to code for a gene exist in both orientations of the sequence.

3.11.2 TestCode Analysis (Fickett, 1982)

TestCode is a program which predicts whether a sequence codes for protein by examining the nucleic acid composition of every possible third base in the sequence and plotting a measure of the extent to which the nucleic acid composition of the third base is randomly selected. The testcode statistic is independent of reading frame. The output plot from this program is divided into three regions; the upper region predicts protein coding regions and the lower region predicts non-coding regions of the sequence. In the middle region the statistic is unable to make a significant prediction. The only variable in this program is the window size. The TestCode statistic is not claimed to make a significant prediction for windows of less than 200 base pairs. Figure 28 shows the TestCode plot for the 5' to 3' direction of the consensus sequence with a 200 base pair window. Two regions, from approximately 1100bp to 1400bp and from 1500bp to 2100bp were strongly predicted to be coding.

Figure 27:

Output plot from the UWGCG program FRAMES indicating the open reading frames (ORF's) in both orientations of the consensus sequence.

Note: This program only identifies ORF's which begin with an ATG initiation codon. ORF's labelled A and B in the 5' to 3' direction are discussed in further analysis of the sequence.

Figure 27

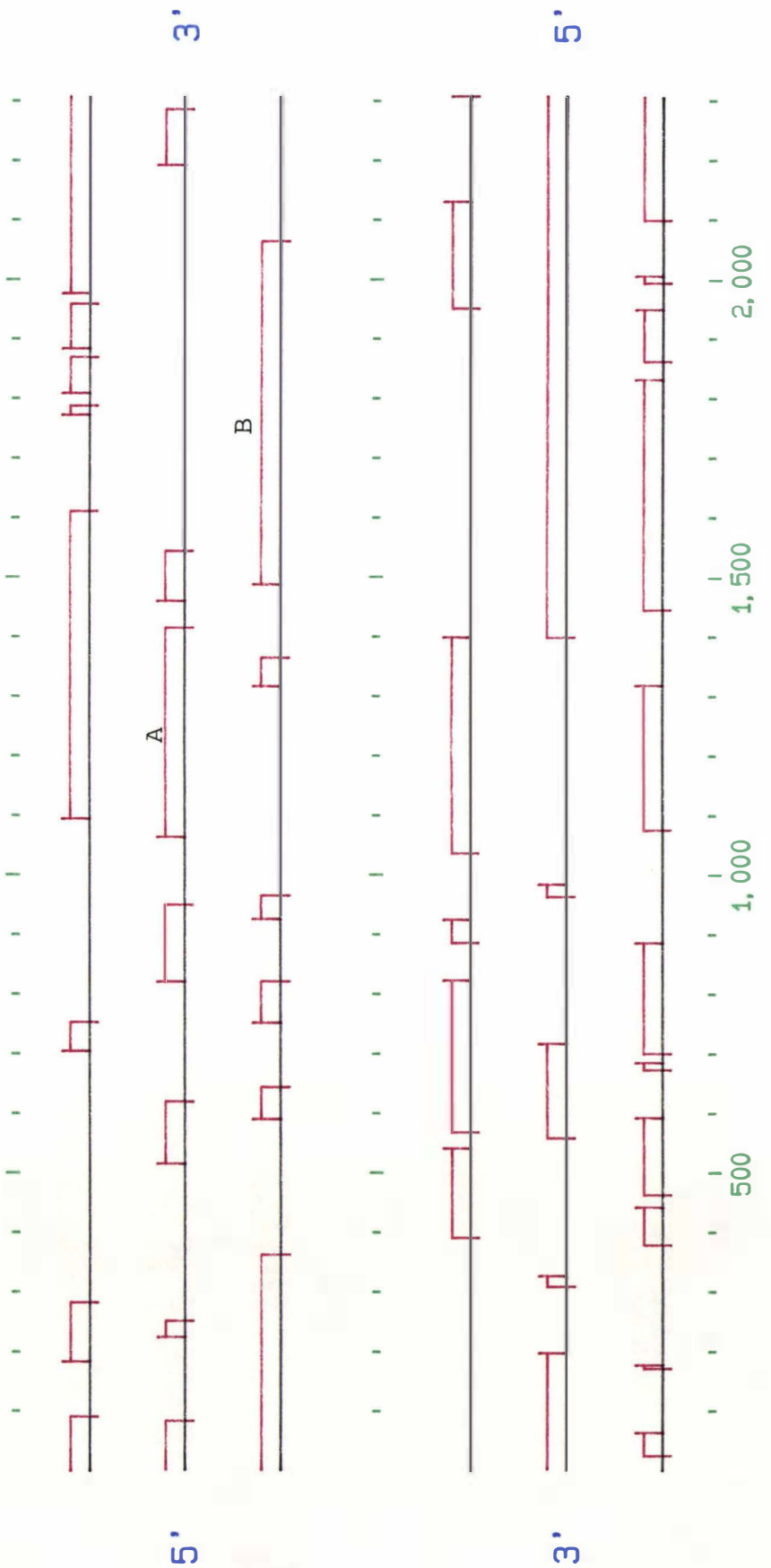
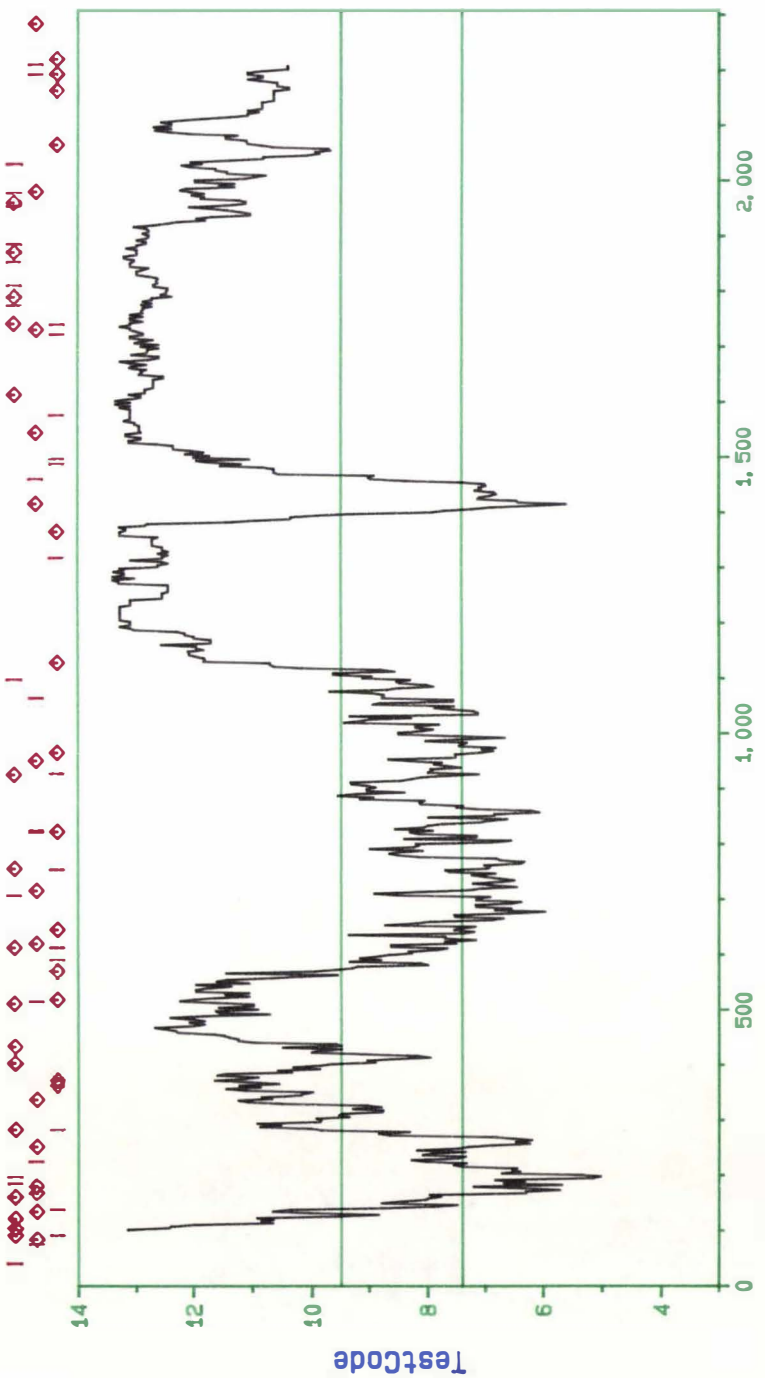


Figure 28:

Output plot from the UWGCG program TESTCODE for the consensus sequence in the 5' to 3' direction. The window size used for this plot was 200 base pairs. Above the plot are markings which identify the start (ATG) and stop (TGA) codons for each reading frame of the sequence. Starts are indicated by short vertical lines and stops by small diamonds. See Section 3.11.2 for discussion of the plot.

Figure 28



Above the plot are markings which identify the start (ATG) and stop (TGA) codons for each reading frame of the sequence. Starts are indicated by short vertical lines and stops by small diamonds. The two regions predicted to be coding correspond to open reading frames (A and B) identified by the FRAMES program (Section 3.11.1).

3.11.3 Third Position Compositional Bias (Bibb *et al.*, 1984)

This test uses the fact that 18 out of the 20 amino acids normally found in proteins allow a choice between G or C and A or U in the third codon position. Such a choice is possible for only 2 amino acids (leucine and arginine) in the first codon position and is not possible at all in the second codon position. Bibb *et al.*, (1984) demonstrated a strong positive correlation between the composition of the third base of the triplet and coding sequences. The output from this program (Figure 29) shows the compositional bias at the third codon position. Where the plots rise above the horizontal line the sequence is predicted to be coding. The window size determines the number of triplets considered at each step of the program. Reducing the window size (Figure 29B) may give a more precise indication of the position of the coding region however this also increases the "statistical noise" (Bibb *et al.*, 1984). Increasing the bias window size (Figure 29C) tends to smooth out the curves and make the presence of coding regions more obvious, however increasing the window size gives a less precise indication of the boundaries of the coding region. Figure 29A shows the plot for the third position GC bias in the 5' to 3' direction using a bias window of 25. Two regions, corresponding to open reading frames A and B, are predicted to be coding. These are from approximately bp 1100 to 1450 and bp 1480 to 2050 in the sequence.

Figure 29:

Third position GC bias plots of the consensus sequence (5' to 3' direction). A bias window of 25 base pairs was used for Figure 29A. The effect of reducing this window size to 10 base pairs is shown in Figure 29B, while the effect of increasing the window size to 50 base pairs is shown in Figure 29C. All possible open reading frames (ORF's) are indicated under each plot. ORF's labelled A and B in each plot correspond to those identified using the programs FRAMES and TestCode. For a discussion of these plots see Section 3.11.3.

Figure 29A

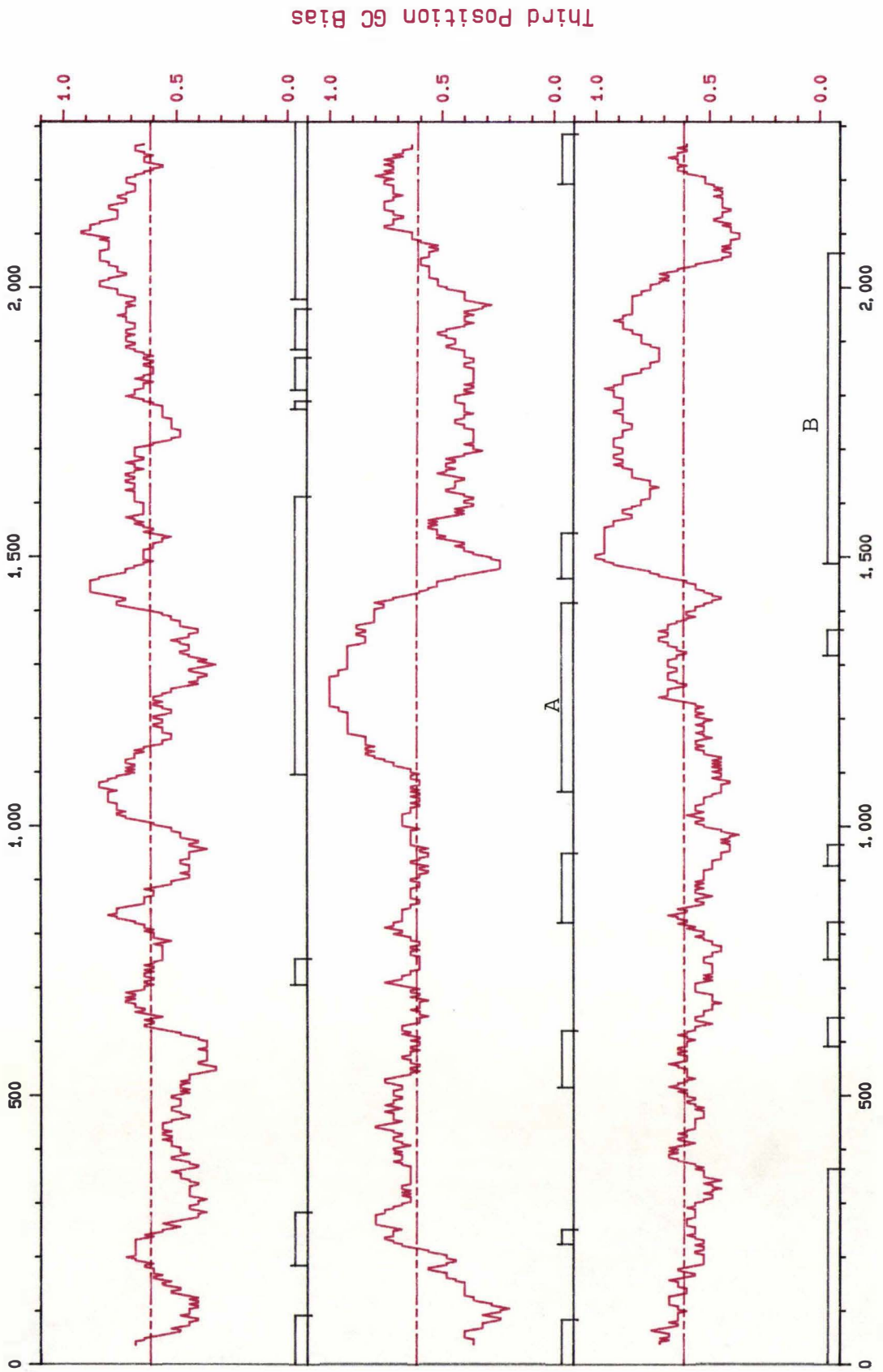

Third Position GC Bias
1.0
0.5
0.0
1.0
0.5
0.0
1.0
0.5
0.0
0
500
1,000
1,500
2,000
A
B

Figure 29B

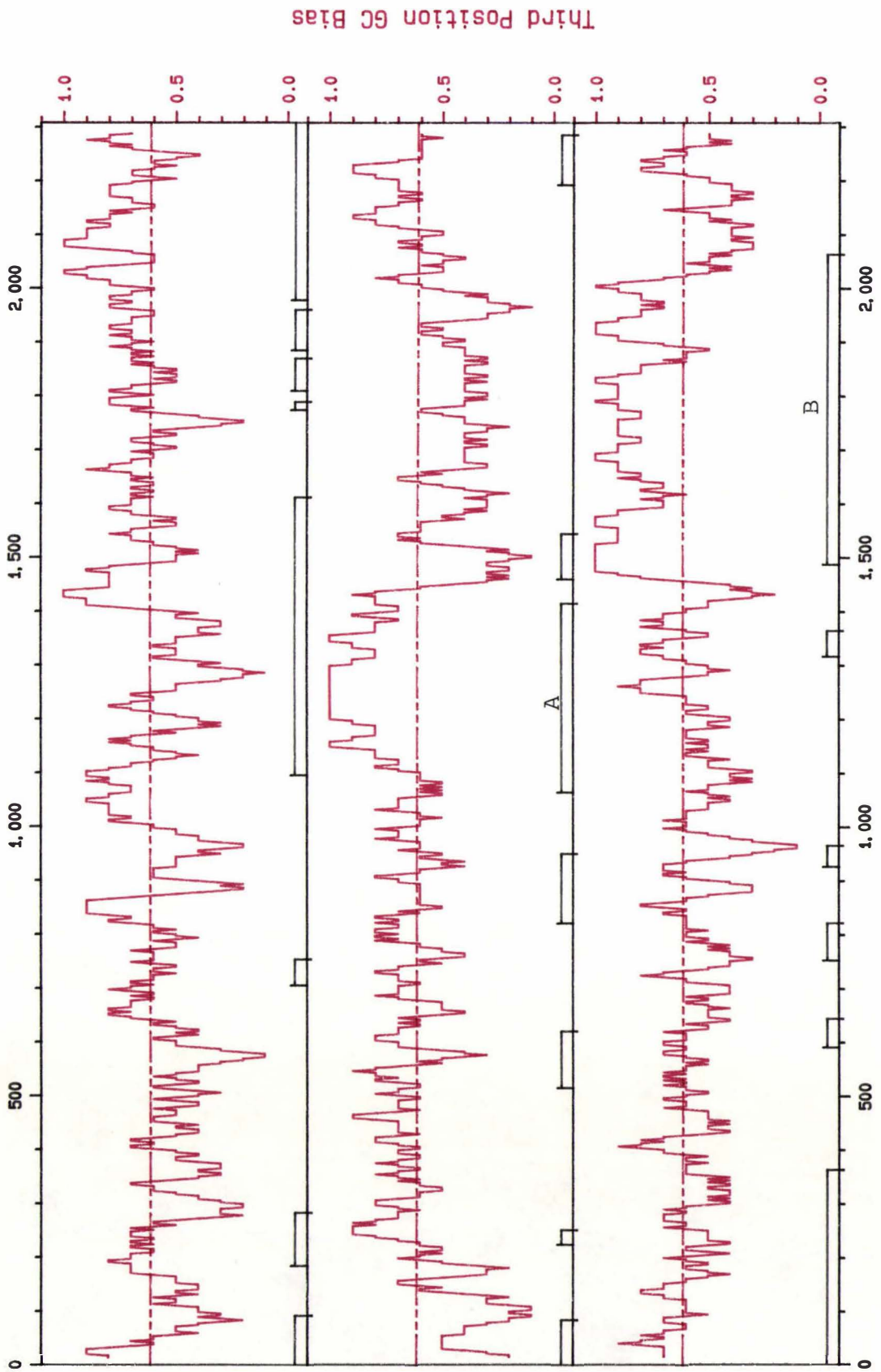
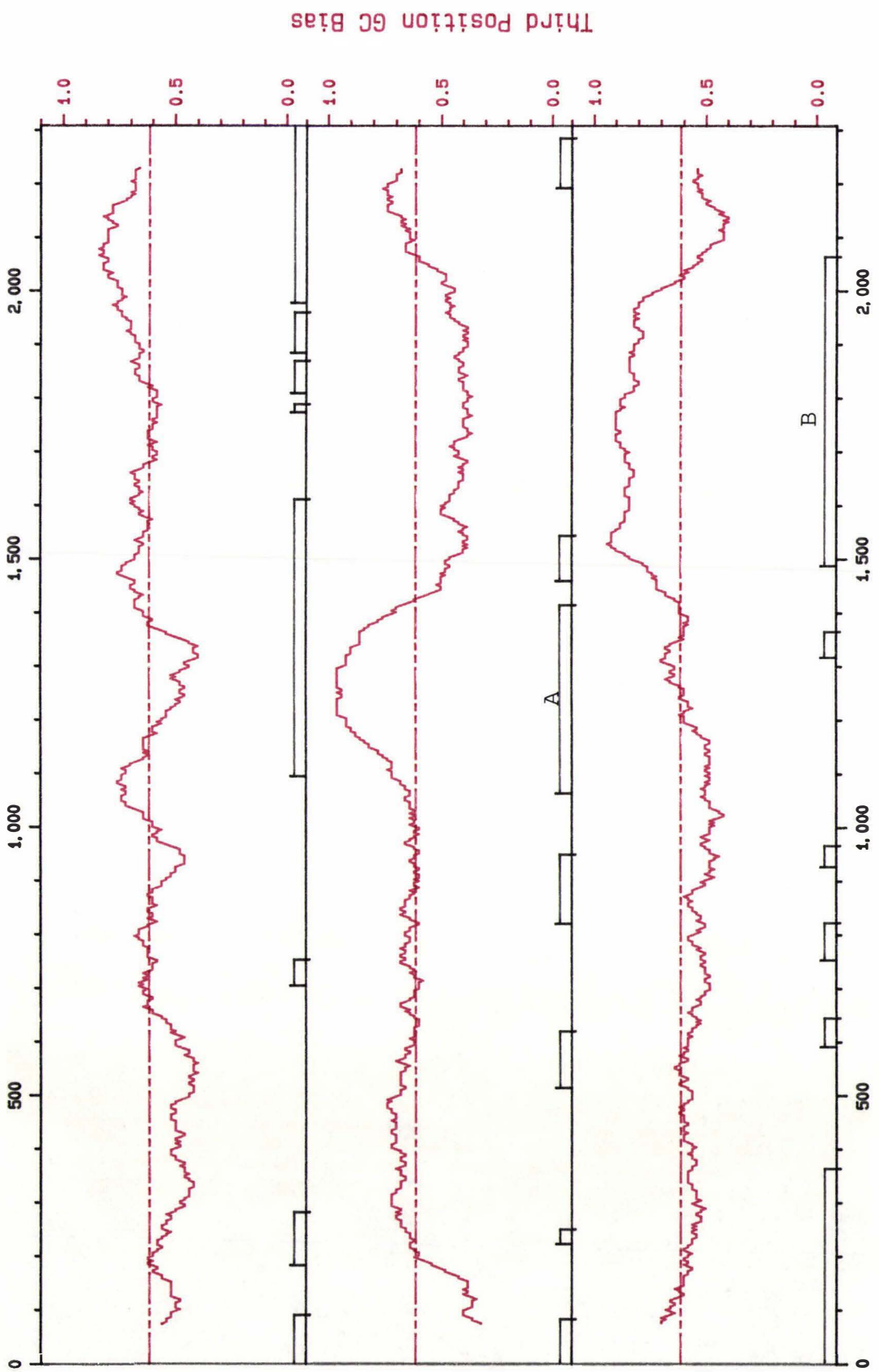


Figure 29C


Third Position GC Bias
1.0
0.5
0.0
1.0
0.5
0.0
1.0
0.5
0.0
0
500
1,000
1,500
2,000
A
B

3.11.4 Construction of a *Rhizobium* Codon Usage Table

In order to use the codon preference program (Gribskov *et al.*, 1984) it is necessary to have a codon usage table with which to compare the codon usage in the test sequence. Codon usage tables were constructed for *Rhizobium* and *Bradyrhizobium* sequences contained in our computer database (GenBank version 56, (Bilofsky and Burks, 1988); European Molecular Biology Laboratory (EMBL) version 15, (Cameron 1988); National Biomedical Research Foundation (NBRF) Nucleotide Sequence data library, (Sidman *et al.*, 1988)). The sequences used and the coding region of the sequence for which the codons were counted are listed in Table 11. The sequences of genes from *Bradyrhizobium* strains were used to construct a codon frequency table, Rhizslow.cod. The sequences of genes from fast growing *Rhizobium* strains were used to construct a codon frequency table, Rhizfast.cod. All *Rhizobium* and *Bradyrhizobium* sequences were combined into a single codon frequency table, Rhizall.cod. (Table 12), which was compared with the codon usage in the consensus sequence.

The relationship between the *Rhizobium* codon frequency tables and the codon frequency table constructed from *E. coli* genes with high levels of expression, Ecohigh.cod and the codon frequency table constructed from *E. coli* genes with low levels of expression, Ecolow.cod contained in the UWGCG package was examined using the program CORRESPOND. The D-squared value from this program is a measure of the differences between the codon frequency tables. The smaller the D-squared value the greater the similarity between the two tables. Identical tables have a D-squared value of zero. Table 13 shows the D-squared values obtained in applying the CORRESPOND program to the codon frequency tables discussed in this section. The results in Table 13 indicate a significant difference between *E. coli* codon usage and *Rhizobium* codon usage and between the Ecohigh.cod and Ecolow.cod codon usage tables. There is little difference between the *Rhizobium* codon usage tables, hence they were combined into one table, Rhizall.cod (Table 12).

Table 11: Sequences used in the construction of *Rhizobium* codon usage tables.

Database	Code	Sequence	Range(s) translated[1,2]	Reference
RHHNIFHA		*R.phaseoli* plasmid p42-D nitrogenase reductase gene (*nif*H), region a.	201 to 1094	Quinto *et al* (1985)
RHHNIFHB		*R.phaseoli* plasmid p42-D nitrogenase reductase gene (*nif*H), region b.	201 to 1094	Quinto *et al* (1985)
RHHNIFHC		*R.phaseoli* plasmid p42-D nitrogenase reductase gene (*nif*H), region c.	201 to 1094	Quinto *et al* (1985)
RHLFIXZ		*Rhizobium leguminosarum fix*Z gene	313 to 1392	Rossen *et al* (1984b)
RHLGNOD		*Rhizobium leguminosarum* nodulation genes *nod*A, *nod*B and *nod*C.	157 to 630 653 to 1303 1326 to 2600	Rossen *et al* (1984a)
RHLNODDFE		*R. leguminosarum* plasmid pRL1JI *nod*DFE region	354 to 1262 1930 to 2205 2209 to 3420	Shearman *et al* (1986)
RHLNODIJ		*R. leguminosarum* nodulation genes *nod*I and *nod*J	140 to 1075 1079 to 1358	Evans and Downie (1986)
RHMAGAS		*R.meliloti* delta-aminolevulinic acid synthetase gene	490 to 718	Leong *et al* (1985)
RHMHSNAD		*R.meliloti* megaplasmid pRme41b Sym *hsn* genes A, B, C, and D	R324 to 1058 R1549 to 2754 R2765 to 3037 3725 to 4468	Horvath *et al* (1986)
RHMNIFA		*Rhizobium meliloti nif*A gene	202 to 1824	Buikema *et al* (1985)
RHMNIFH		*R.meliloti* nitrogenase reductase (*nif*H) gene	159 to 1049	Torok & Kondorosi (1981)
RHMNOD		Sym megaplasmid (from *R. meliloti* 1021) *nod*D, *nod*A, *nod*B and *nod*C	R65 to 991 1257 to 1847 1844 to 2497 2512 to 3792	Egelhoff *et al* (1985) Jacobs *et al* (1985)
RHTNODG		*Rhizobium trifolii* Sym plasmid nodulation genes *nod*ABCDEF.	R1 to 336 R357 to 1004 R1001 to 1591 1826 to 2782 3266 to 3544 3544 to 3668	Schofield and Watson (1986)

Table 11: Cont'd

Database Code	Sequence	Range(s) translated[1,2]	Reference
RHBGSII	*Bradyrhizobium japonicum* gene *gln*II for glutamine synthetase II	1 to 990	Carlson & Chelm (1986)
RHBNIFB	*B.japonicum nif*B gene encoding nitrogenase	133 to 1629	Noti *et al* (1986)
RHJNIFDKO	*B.japonicum nif*DK operon coding for dinitrogenase alpha and beta subunits	376 to 1923 1942 to 2073	Kaluza & Hennecke (1984) Kaluza *et al* (1983)
RHJNIFH	*B.japonicum nif*H gene encoding nitrogenase Fe.	325 to 1209	Adams & Chelm (1984) Fuhrmann & Hennecke(1984)
RHBGLNA	*B.japonicum* glutamine synthetase I (*gln*A) gene	337 to 554	Carlson *et al* 1985
RHBHEMA	*B.japonicum hem*A gene encoding 5-aminolevulinic acid synthase	619 to 1848	McClung *et al* 1987
RHBNODDLA	*B.japonicum nod*D (3' end), *nod*L and *nod*A (5' end)	1 to 54 289 to 711 834 to 979	Nieuwkoop *et al* 1987
RHHPSI	*Rhizobium phaseoli* pRP2JI gene *psi* (polysaccharide inhibition)	80 to 340	Borthakner & Johnston 1987
RHINIFDK	Parasponia *Rhizobium nif*D and *nif*K genes coding for the alpha and beta subunits of the Mo-Fe protein of nitrogenase	176 to 1678 1767 to 3308	Weinman *et al* (1984)
RHINODD1	*Rhizobium* sp. MPIK3030 *nod*D1 gene	340 to 1308	Horvath *et al* 1987
RHLGLNA	*Rhizobium leguminosarum gln*A gene for glutamine synthetase I	251 to 586 665 to 2074	Colonna-Romano *et al* 1987
RHLNIFA	*Rhizobium leguminosarum* plasmid pRL6JI *nif*A gene, *fix*C gene and *nif*B gene	1 to 492 505 to 801 989 to 2548 2816 to 3301	Groenger *et al* 1987
RHMNODD1	*Rhizobium meliloti* 41 *nod*D1 gene	129 to 1055	Goettfert *et al* 1986
RHMNODD2	*Rhizobium meliloti* 41 *nod*D2 nodulation gene	119 to 1051	Goettfert *et al* 1986

Table 11: Cont'd

Database Code	Sequence	Range(s) translated[1,2]	Reference
RHMNODEFG	*Rhizobium meliloti* nodulation genes *nod*F, *nod*E and *nod*G	412 to 693 694 to 1902 2395 to 3132	Fisher *et al* 1987
RHMNODFEG	*Rhizobium meliloti* RCR2011 nodulation genes *nod*F, *nod*E and *nod*G	406 to 687 688 to 1896 2390 to 3127	Debelle & Sharma 1986
RHMNODH	*Rhizobium meliloti* RCR2011 nodulation gene *nod*H	284 to 1027	Debelle & Sharma 1986
RHTFIXX	*Rhizobium trifolii fix*C gene 3-terminus and *fix*X gene	1 to 172 185 to 481	Iismaa and Watson 1987

[1] Range translated - this indicates the coding portion of the sequence from which codons were counted. The numbers correspond to the numbering of the sequence in the database.

[2] R - denotes translation in reverse orientation to that in which the sequence is presented.

Table 12: Codon usage table constructed from the *Rhizobium* genes listed in Table 11.

Amino Acid	Codon	Total Number	Frequency /1000 codons	Fraction coding for amino acid
Gly	GGG	172.00	10.78	0.14
Gly	GGA	223.00	13.97	0.18
Gly	GGT	213.00	13.35	0.17
Gly	GGC	647.00	40.54	0.52
Glu	GAG	552.00	34.59	0.57
Glu	GAA	423.00	26.51	0.43
Asp	GAT	368.00	23.06	0.42
Asp	GAC	507.00	31.77	0.58
Val	GTG	360.00	22.56	0.33
Val	GTA	115.00	7.21	0.11
Val	GTT	176.00	11.03	0.16
Val	GTC	431.00	27.01	0.40
Ala	GCG	476.00	29.83	0.29
Ala	GCA	292.00	18.30	0.18
Ala	GCT	267.00	16.73	0.16
Ala	GCC	605.00	37.91	0.37
Arg	AGG	109.00	6.83	0.10
Arg	AGA	73.00	4.57	0.07
Ser	AGT	65.00	4.07	0.07
Ser	AGC	189.00	11.84	0.20
Lys	AAG	486.00	30.45	0.69
Lys	AAA	218.00	13.66	0.31
Asn	AAT	207.00	12.97	0.38
Asn	AAC	335.00	20.99	0.62
Met	ATG	459.00	28.76	1.00
Ile	ATA	116.00	7.27	0.12
Ile	ATT	223.00	13.97	0.23
Ile	ATC	636.00	39.85	0.65
Thr	ACG	230.00	14.41	0.28
Thr	ACA	111.00	6.96	0.14
Thr	ACT	136.00	8.52	0.17
Thr	ACC	343.00	21.49	0.42

Table 12: Cont'd

AmAcid	Codon	Number	Frequency	Fraction
Trp	TGG	177.00	11.09	1.00
End	TGA	31.00	1.94	0.61
Cys	TGT	60.00	3.76	0.22
Cys	TGC	215.00	13.47	0.78
End	TAG	7.00	0.44	0.14
End	TAA	13.00	0.81	0.25
Tyr	TAT	181.00	11.34	0.48
Tyr	TAC	194.00	12.16	0.52
Leu	TTG	235.00	14.73	0.15
Leu	TTA	50.00	3.13	0.03
Phe	TTT	176.00	11.03	0.31
Phe	TTC	388.00	24.31	0.69
Ser	TCG	246.00	15.41	0.27
Ser	TCA	97.00	6.08	0.10
Ser	TCT	85.00	5.33	0.09
Ser	TCC	245.00	15.35	0.26
Arg	CGG	176.00	11.03	0.17
Arg	CGA	118.00	7.39	0.11
Arg	CGT	180.00	11.28	0.17
Arg	CGC	409.00	25.63	0.38
Gln	CAG	326.00	20.43	0.65
Gln	CAA	174.00	10.90	0.35
His	CAT	173.00	10.84	0.43
His	CAC	228.00	14.29	0.57
Leu	CTG	420.00	26.32	0.28
Leu	CTA	122.00	7.64	0.08
Leu	CTT	276.00	17.29	0.18
Leu	CTC	417.00	26.13	0.27
Pro	CCG	335.00	20.99	0.43
Pro	CCA	172.00	10.78	0.22
Pro	CCT	115.00	7.21	0.15
Pro	CCC	155.00	9.71	0.20

Table 13: CORRESPOND analysis of Codon Frequency Tables

Between	and	D-squared
Ecohigh.cod	Ecolow.cod	3.389803
Ecohigh.cod	Rhizslow.cod	3.268904
Ecohigh.cod	Rhizfast.cod	3.277792
Ecohigh.cod	Rhizall.cod	3.214903
Ecolow.cod	Rhizslow.cod	2.078698
Ecolow.cod	Rhizfast.cod	1.063993
Ecolow.cod	Rhizall.cod	1.276757
Rhizslow.cod	Rhizfast.cod	0.492111
Rhizslow.cod	Rhizall.cod	0.306640
Rhizfast.cod	Rhizall.cod	0.051408

3.11.5 Codon Preference (Gribskov *et al.*, 1984)

The codon preference statistic for each reading frame shows the similarity of the codons in a window of that reading frame to a codon usage table. The codon preference plot of the correct reading frame for a gene may rise significantly above the background if the codon usage of the sequence is strongly biased and a suitable codon frequency table is used. Weakly expressed genes show much less bias in codon usage than do highly expressed genes; hence genes which are highly expressed are more likely to be detected by this program.

The consensus sequence in the 5' to 3' direction was compared with an *E. coli* highly expressed gene codon frequency table (Figure 30A). The two putative coding regions from previous programs (i.e. A, approx. 1050-1400 and B, 1480-2050) show an increased similarity of codon usage. This was also indicated when the sequence was compared with the *Rhizobium* codon usage table constructed as described in Section 3.11.4 (Figure

30C). In the reverse orientation (complement) (Figures 31A & B) there was a less marked codon preference within the possible open reading frames. Comparison with the codon usage table for weakly expressed *E. coli* genes to the sequence (Figure 30B) does not show any marked codon preference.

3.11.6 Rare Codon Usage

This program indicates codons used in each reading frame which occur with a low frequency in the codon usage table. Figure 32 shows the open reading frames and rare codons (indicated as short vertical lines) for each reading frame in the 5' to 3' direction. In Figure 32A the highly expressed *E. coli* gene codon frequency table was used. It is noticeable that there were fewer "rare" codons in the region approximately 1500 to 2050 bases corresponding to the putative open reading frame (B) in frame 3 of the plot. When the *Rhizobium* gene codon frequency table was used (Figure 32B) fewer rare codons were found throughout the sequence and there were no "rare" codons in the putative open reading frame (B).

Figure 30:

Output plot from the UWGCG program CODON PREFERENCE. Codon usage in the consensus sequence in the 5' to 3' direction was compared, using a comparison window of 25 codons, to the codon usage of: Figure 30A, highly expressed *E. coli* genes;
Figure 30B, weakly expressed *E. coli* genes;
and Figure 30C, *Rhizobium* genes.

Open reading frames indicated as A and B are those previously identified.

Figure 30A

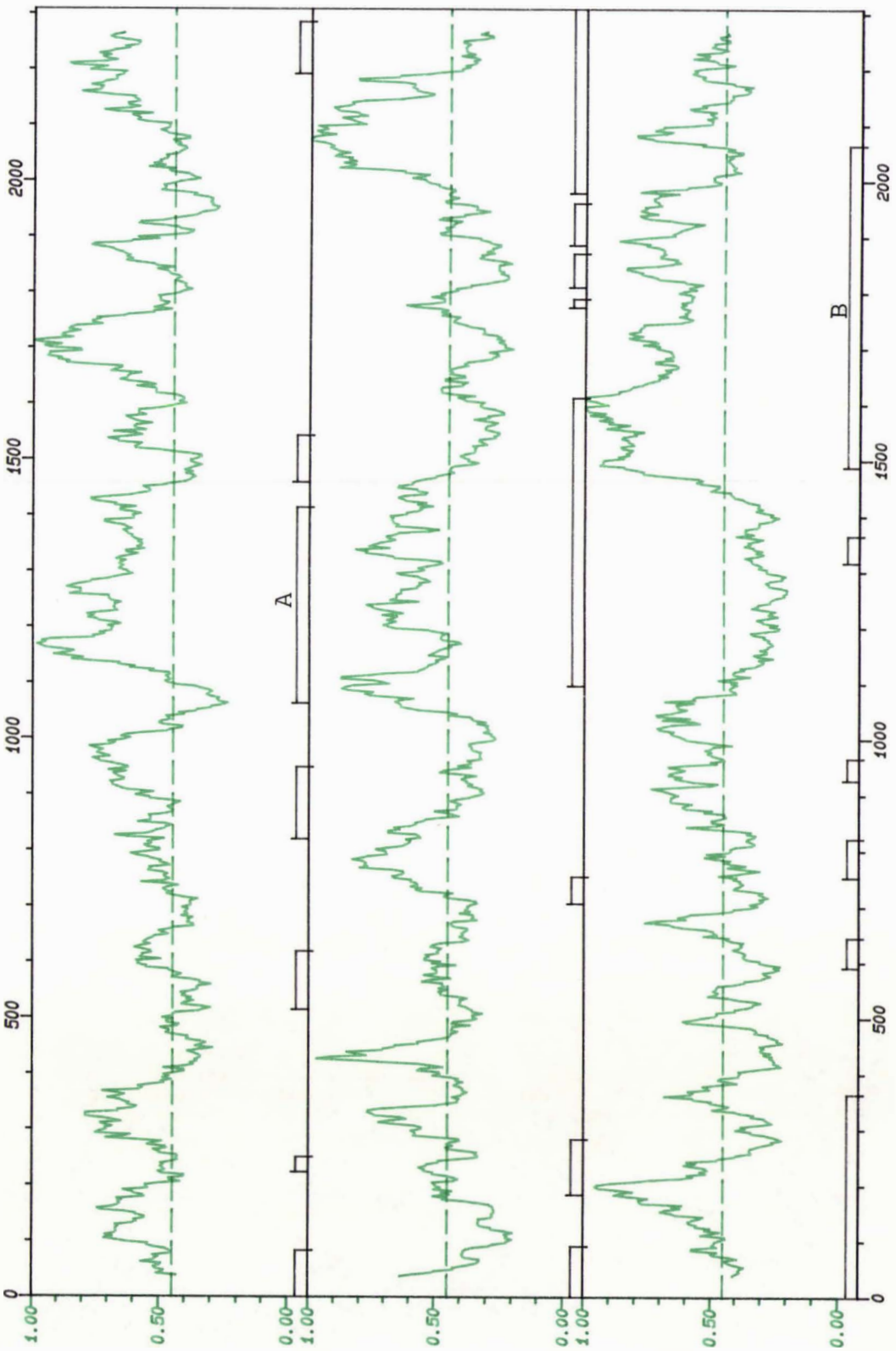


Figure 30B

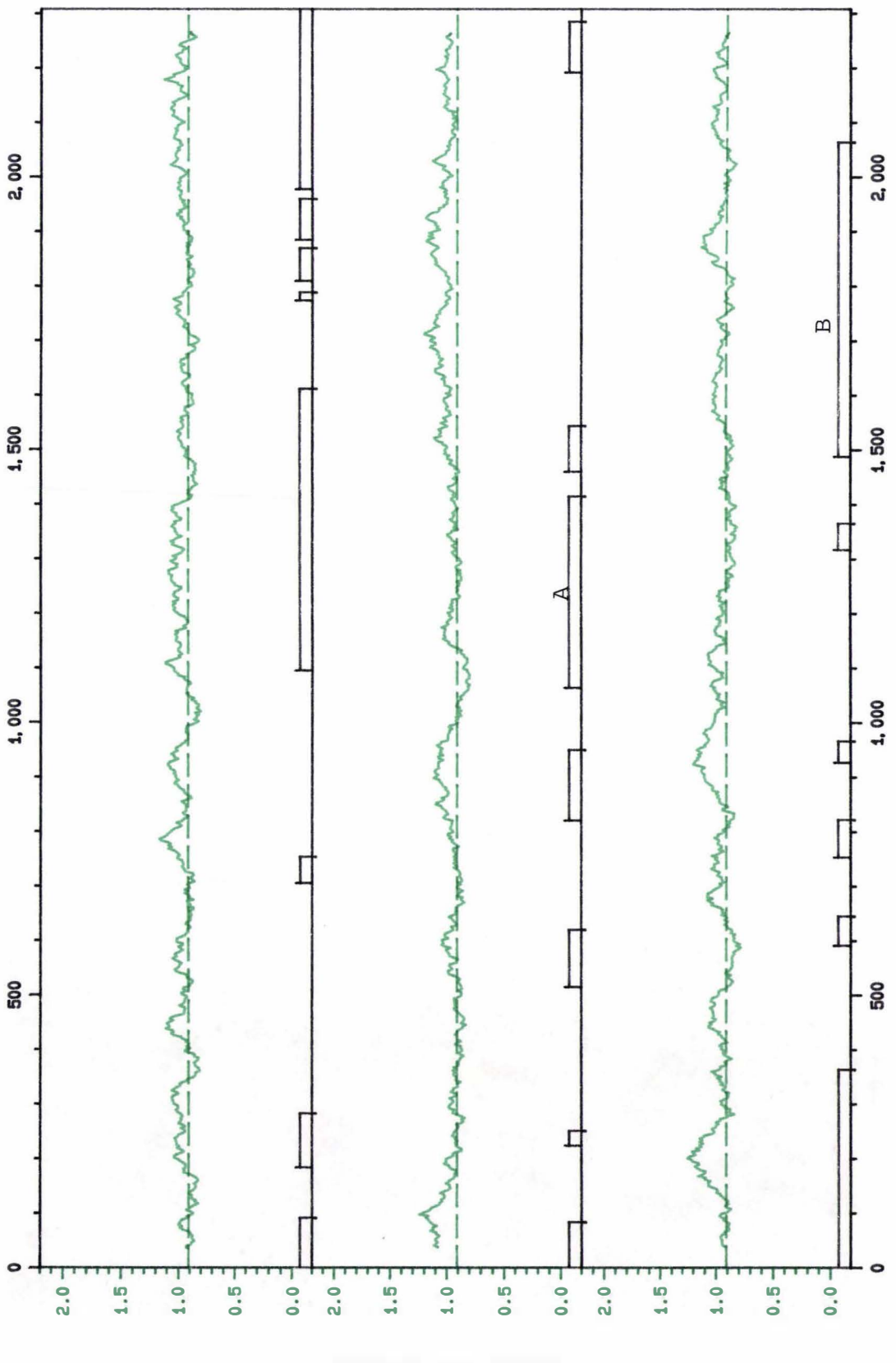


Figure 30C

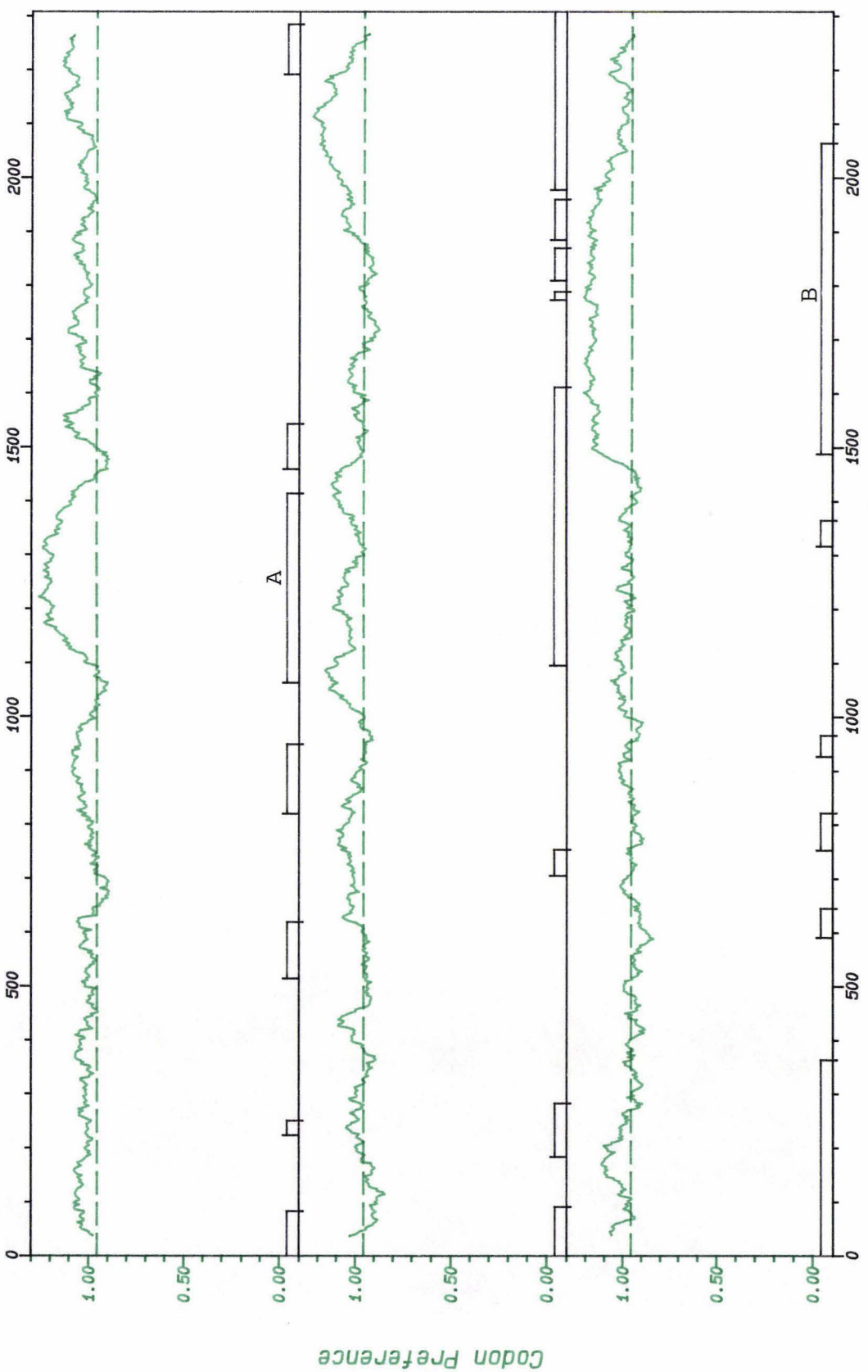


Figure 31:

Output plot from the UWGCG program CODON PREFERENCE. Codon usage in the consensus sequence in the 3' to 5' direction was compared using a comparison window of 25 codons, to the codon usage of: Figure 31A, highly expressed *E. coli* genes;
and Figure 31B, *Rhizobium* genes.

Figure 31A

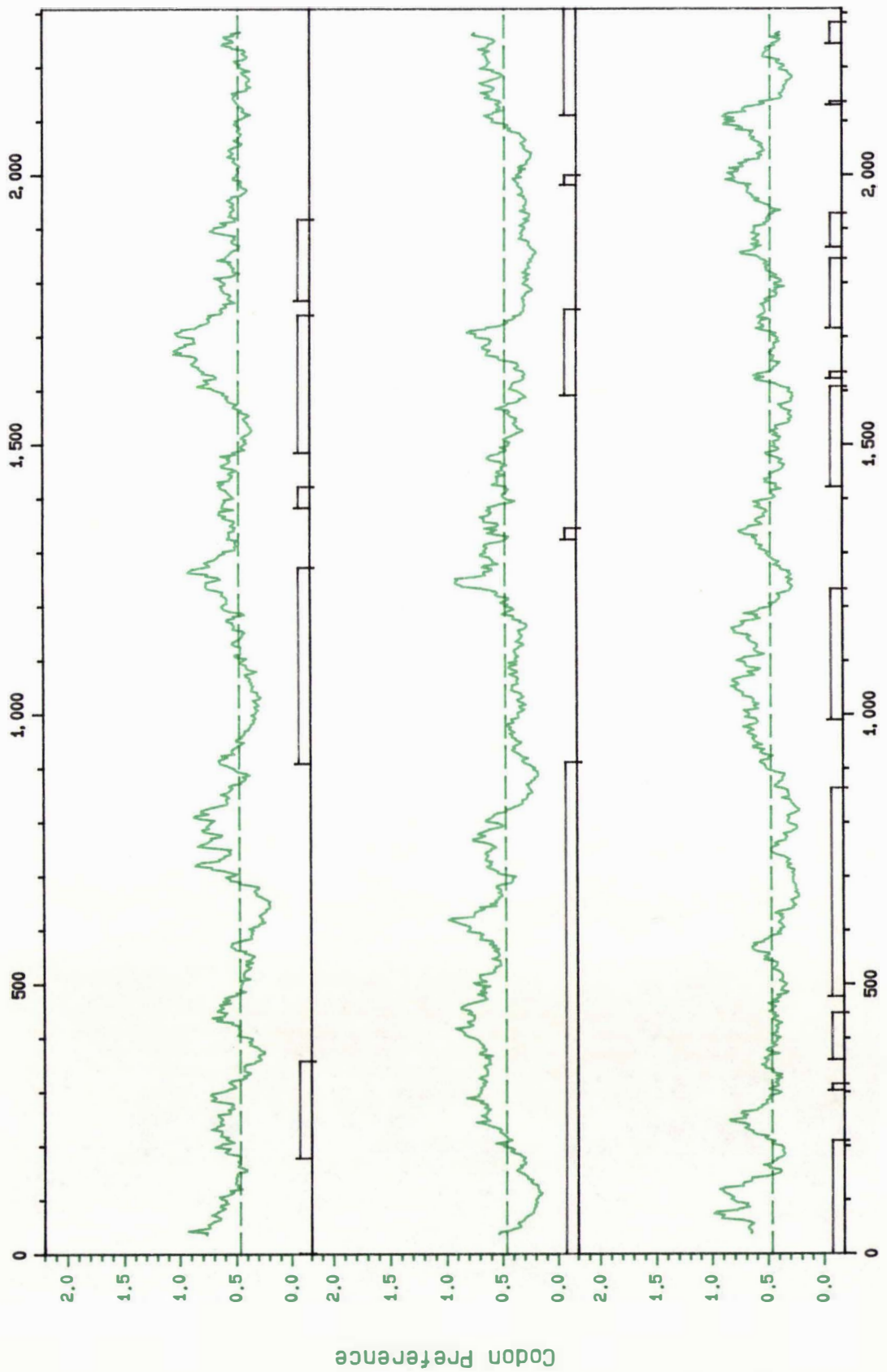


Figure 31B

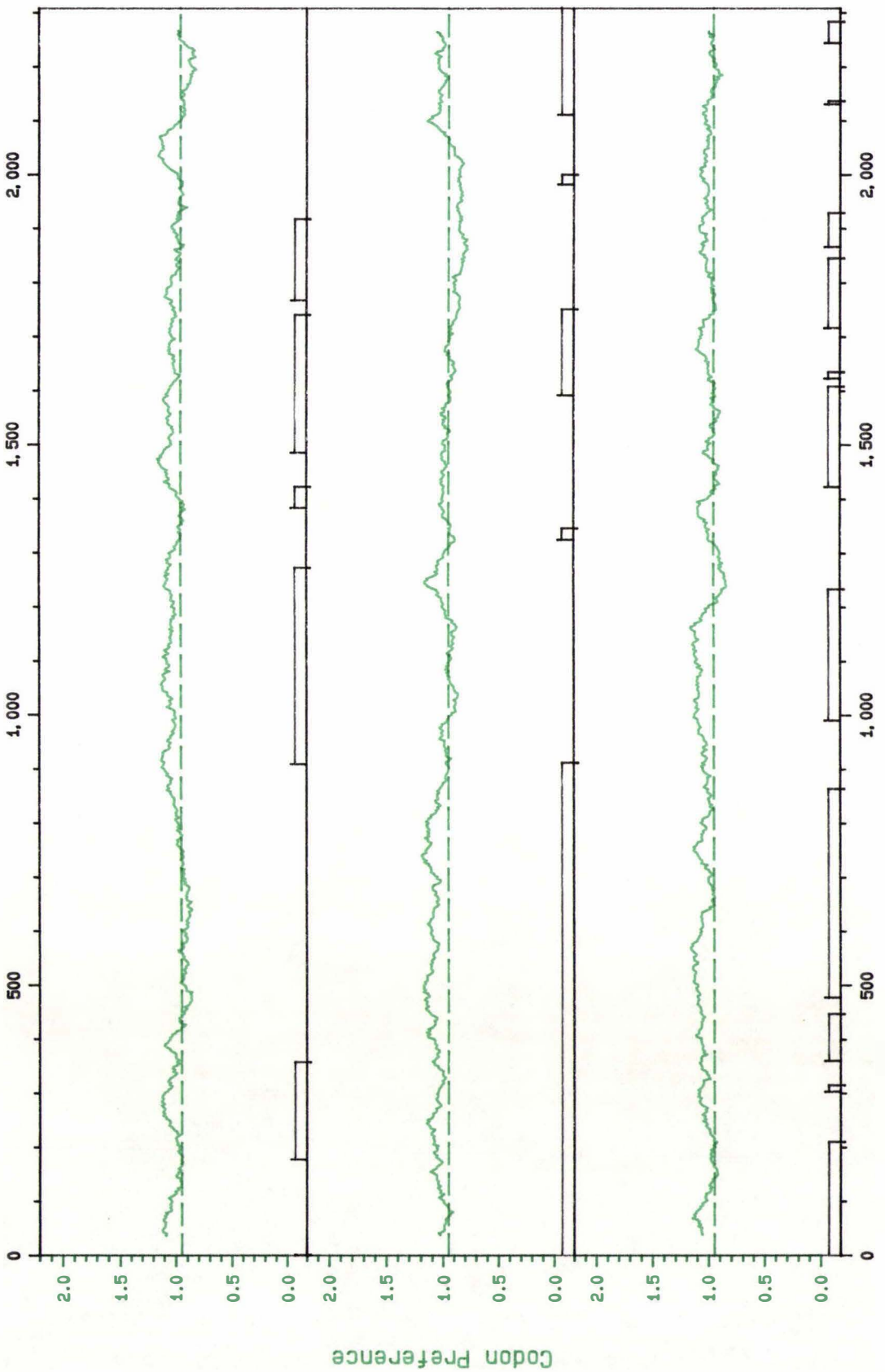


Figure 32:

Output from the UWGCG program CODON PREFERENCE indicating the use of rare codons throughout the consensus sequence in the 5' to 3' direction. Rare codon usage is indicated in each reading frame by a short vertical line. Open reading frames A and B correspond to those previously identified.

The codon usage is compared to codon usage tables constructed from highly expressed *E. coli* genes (Figure 32A) and from *Rhizobium* genes (Figure 32B). Section 3.11.6 discusses these plots.

Figure 32A

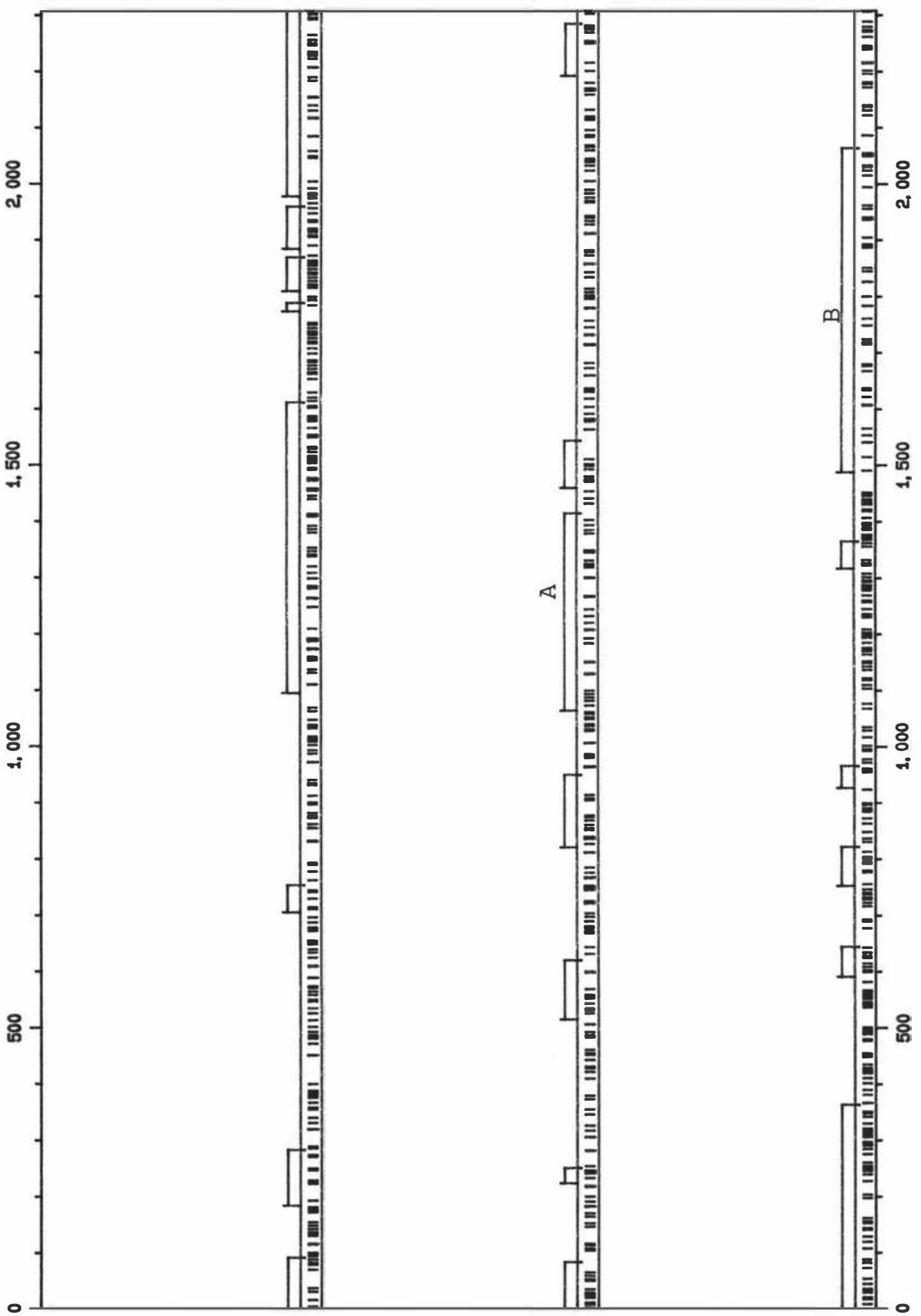
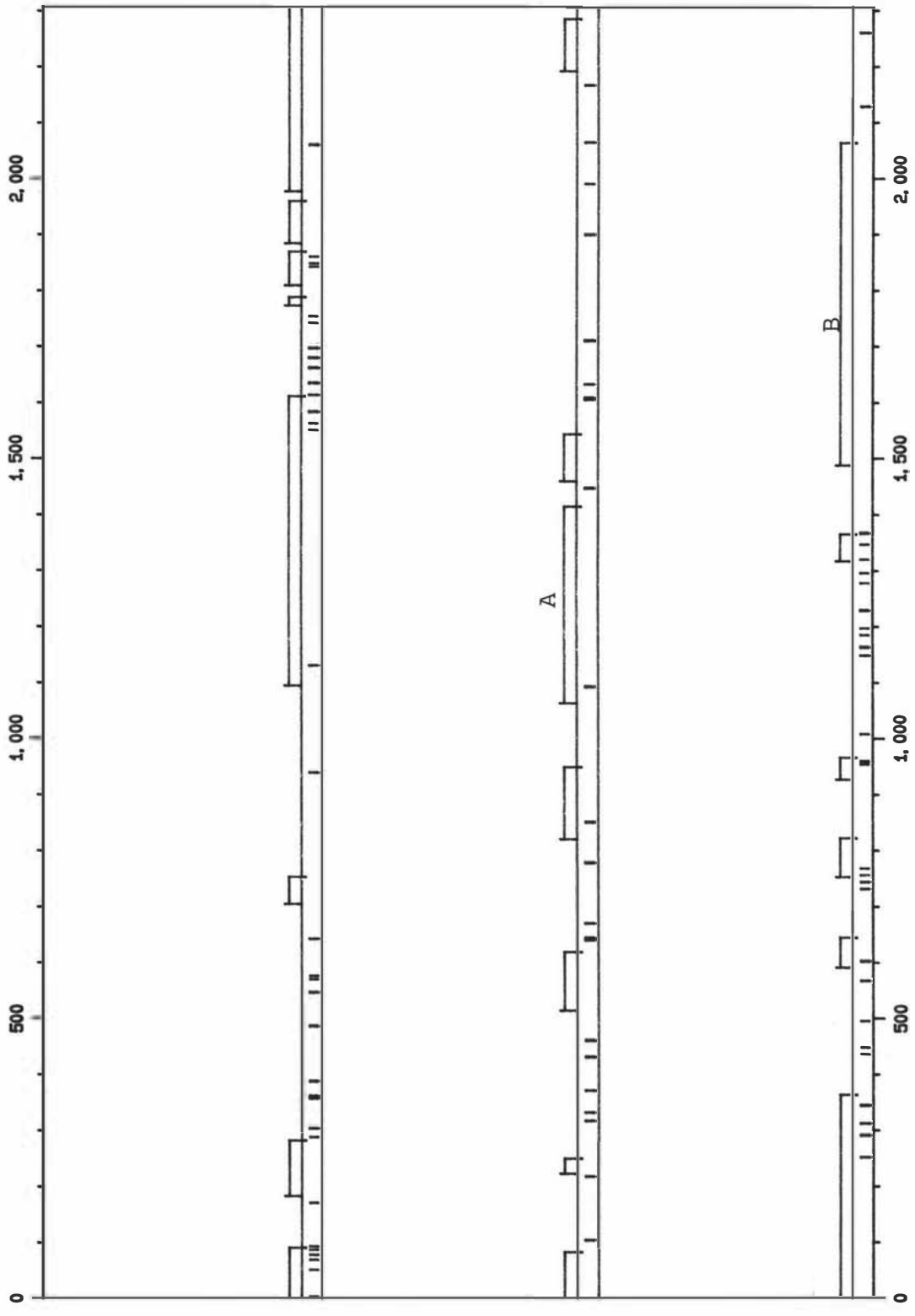


Figure 32B



3.12 FEATURES OF THE SEQUENCE

The information from TestCode (Section 3.11.2), compositional bias (Section 3.11.3), codon usage (Sections 3.11.5 and 3.11.6) and the position of the Tn5 insertions causing a Fix^- phenotype indicated that the open reading frame (B) from approximately 1500 to 2050 bases codes for a gene which is required for effective nodulation of *Lotus pedunculatus*. (This is more fully discussed in the Discussion (Section 4.9)).

3.12.1 A Potential Ribosome Binding Site

The sequence upstream of the putative gene was examined for a potential ribosome binding site (RBS) (Shine & Dalgarno sequence). Stormo *et al.*, (1982) proposed a set of "rules" for the identification of potential ribosome binding sites (Table 14).

A potential ribosome binding site (RBS) GGACGAGGG was identified 7 bases upstream from the presumed ATG initiation codon for the gene. This potential RBS either fits exactly or has a close match to 6 of the 7 RBS "rules" proposed by Stormo *et al.*, (1982) for *E. coli* ribosome binding sites. Only 39% of known *E. coli* genes have a RBS which fit rule 7. The "rules" proposed by Stormo *et al.*, and the matches with the putative RBS for the *bar* gene are shown in Table 14.

Table 14: Comparison of the ribosome binding site (RBS) "rules" proposed by Stormo *et al.*, 1982 and the putative RBS for the *bar* gene.

"Rule" No.	Description[1] Shine Dalgarno sequence - space - ATG	% of *E. coli* gene RB sites which fit "rule"	Matches with RBS for *bar* gene[2]
1.	AGG) GGA) - 4E5N - ATG GAG)	85	AGG - 8N - ATG GGA - 13N - ATG GAG - 9N - ATG
2.	AGG) GGA) - 3E6N - ATG GAG)	83	AGG - 8N - ATG GGA - 13N - ATG GAG - 9N - ATG
3.	AGG - 2E7N) GGA - 9N) - ATG GAG - 1E6N)	73	AGG - 8N - ATG GGA - 13N - ATG GAG - 9N - ATG
4.	AAGG) AGGA) GGAG)- 6E4N - ATG GAGG) AGGT)	77	GAGG - 8N - ATG
5.	AAGG) AGGA)- 4E5N - ATG GGAG) GAGG)	73	GAGG - 8N - ATG
6.	AGGA) GGAG)- 4E5N - ATG GAGG)	67	GAGG - 8N - ATG
7.	AAGGA) AGGAG)- 4E5N - ATG GGAGG)	39	no exact match - nearest matches ACGAG - 9N - ATG CGAGG - 8N - ATG

1. N is any base, E is any or no base (i.e.) 4E5N means 5 to 9 unspecified bases.

2. N is the number of bases between the indicated Shine Dalgarno sequence and the ATG initiation codon.

3.12.2 Possible Promoter Region for the Gene

Ronson and Astwood, (1985) proposed a "non-nitrogen" consensus promoter derived from possible promoters identified upstream of *Rhizobium* genes (Leong *et al.*, 1985, Buikema *et al.*, 1985).

This consensus promoter; TTPuANN-17b-PuAPuPuPuPu-4-5b-CA[1] was used to search for possible promoter regions, using the UWGCG programs BESTFIT (Smith and Waterman, 1981) and GAP (Needleman and Wunsch, 1970). Comparison of this consensus with the ALA synthetase sequence (Leong *et al.*, 1985) correctly identifies the promoter proposed in that paper.

Comparison of the consensus promoter with the sequence determined in this project identifies a number of regions which have good homology to the consensus promoter. These are 197 to 230bp, 1196-1229bp, 1797-1830bp, 2051-2100bp and 2143-2174bp. Of these regions, one (1196-1229bp) is in a position from - 272 to -293bp upstream of the proposed gene. Further, it is located in a small region (89 bases) between Tn5 insertions #56 and #41. Insertion of Tn5 at the position of #56 (-298bp) does not affect the function of the proposed gene so presumably the promoter is between this position and the initiation codon for the gene. Insertion of Tn5 at the position of #41 (-210bp) causes a Fix$^-$ phenotype, hence the promoter was either upstream of, or spanned this position. The promoter at 1196-1229bp, AAGATCATCGCGCTCGGCACCA was consistent with this data.

Significant homology was not found between this putative promoter sequence and any regions upstream of reported *Rhizobium* DNA sequences, including that of a gene involved in bacterial release reported by Morrison & Verma, (1987).

[1] N = any base; A, G, C, T Pu = purine base; A or G

3.12.3 Possible Termination Structures

The sequence downstream from the presumed stop (TGA) codon was examined for inverted repeats which had the potential to form stem-loop structures.

Two such regions were identified. The possible stem-loop structures are shown in Figures 33A and B.

Figure 33: Stem-loop structures downstream of the *bar* gene.

A.

```
      A  G
    C      C
  G          G
    C      C
      C  A
      T=A
      A=T
      G=C
      C  C
      C=G
      C=G
      G=C
      A=T
      C=G
(2136)    (2164)
```

B.

```
      C  G
    C      G
  A          C
    G      G
      T  C
      G=C
      C=G
      T  G
      G=C
      C=G
      C=G
      G=C
      C=G
      T=A
(2211)    (2238)
```

3.12.4 Direct and inverted repeats upstream from the presumed ATG initiation codon for the gene.

A number of direct and inverted repeats were identified immediately upstream from the presumed start of the gene. The longest of these was the longest repeat in the entire sequence and was an imperfect (21 out of 26 bases (81%)) 26 base repeat. The region from base 1424 to base 1449 was repeated from base 1468 to 1493. The presumed start of the gene was at base 1488.

```
1424  TGGTGCGCGAGGGCGACCTCATCCTC  1449
      || ||  |||||||||||  ||| |||
1468  TGCTGGACGAGGGCGACCGCATGCTC  1493
```

It was interesting to note that this included an imperfect repeat of the proposed Shine Dalgarno sequence for the gene (though not the ATG start codon).

A number of other short repeats existed within this region. These, and their direction are indicated by arrows on the following sequence:

```
GTCGACATCCTCATTGCCACGCCGGCCGATTGACCGA
                                    /
Sal I site

CCTGGTGCGCGAGGGCGACCTCATCCTCTCCGACACAA
 \          /    \     \

ATGGCTGGTGCTGGACGAGGGCGACCGCATG
           proposed       Start
         Shine Dalgarno   codon
            sequence     for gene
```

The significance (if any) of these repeated sequences is not known.

3.12.5 Putative Protein Translation

The region of the sequence between base 1488 and 2066 (578 base pairs) was translated into an amino acid sequence using the program TRANSLATE. This translation is given in Figure 34. The unmodified molecular weight is 21.24kD. Table 15 summarises the amino acid composition of the putative protein.

Table 15: Amino acid composition of the putative protein

Residue	Number	Mole %
Ala	23	11.917
Asx	0	0.000
Cys	0	0.000
Asp	15	7.772
Glu	8	4.145
Phe	8	4.145
Gly	11	5.699
His	7	3.627
Ile	13	6.736
Lys	12	6.218
Leu	12	6.218
Met	5	2.591
Asn	6	3.109
Pro	7	3.627
Gln	8	4.145
Arg	19	9.845
Ser	12	6.218
Thr	9	4.663
Val	16	8.290
Trp	0	0.000
Tyr	1	0.518
Glx	0	0.000

Summary for whole protein:

Molecular weight	=	21243.10
Average residue weight	=	110.068
Total residues	=	193
Charged residues	=	8

Figure 34:

The putative coding region of the *bar* gene sequence and the derived amino acid sequence (Section 3.12.5).

```
ATGCTCGACATGGGCTTCATCAACGACGTCAAGCGCATCGCCAAGGCGACCGCGCCTGAC
MetLeuAspMetGlyPheIleAsnAspValLysArgIleAlaLysAlaThrAlaProAsp

CGCCAGACGGCGCTGTTCTCGGCCACCATGCCGGACGAGATCGCCGAACTGGCCAAGGGT
ArgGlnThrAlaLeuPheSerAlaThrMetProAspGluIleAlaGluLeuAlaLysGly

CTTCTGAAGAACCCGGTCCGTATCGAAGTCTCGCCGCAAAGCACCGCGGCCGCCGAGATT
LeuLeuLysAsnProValArgIleGluValSerProGlnSerThrAlaAlaAlaGluIle

GTCCAGGGCGTCGTCTTCGCCCGCACCAAGCAGAAGCGCCAGGTTCTGTCGACGATGCTC
ValGlnGlyValValPheAlaArgThrLysGlnLysArgGlnValLeuSerThrMetLeu

GCTGACGAGGCGATGAAGTCCGTCATCATCTTTTCGCGCACCAAGCATGGCGCCGACCGG
AlaAspGluAlaMetLysSerValIleIlePheSerArgThrLysHisGlyAlaAspArg

GTGACCAAGGACCTCGAGCGCGATGGCTTCAAGGCGGCCGTCATCCACGGCAACAAGTCG
ValThrLysAspLeuGluArgAspGlyPheLysAlaAlaValIleHisGlyAsnLysSer

CAGAATGCCCGTCAGAAGGCGCTGAACGATTTTCGCGATGGATCGGTCCGCATTCTGGTC
GlnAsnAlaArgGlnLysAlaLeuAsnAspPheArgAspGlySerValArgIleLeuVal

GCGACCGACATCGCGGCGCGCGGCATCGACGTCCCCGGCATCAGCCATGTCGTGAATTTC
AlaThrAspIleAlaAlaArgGlyIleAspValProGlyIleSerHisValValAsnPhe

GACCTGCCGGATGAGGCGGAAAGCTACGTCCACCGCATCGGCCGCAGCCGGCCGCAACGG
AspLeuProAspGluAlaGluSerTyrValHisArgIleGlyArgSerArgProGlnArg

CATGGACGGCATCGCGATCACGCTTTGCGATCCTTCTGA
HisGlyArgHisArgAspHisAlaLeuArgSerPheEnd
```

3.12.6 Comparison of the Nucleic Acid and Protein Sequences to the Databases

The sequence data obtained in this project was compared with the sequences contained in our computer database. This comprises GenBank version 56 (Bilofsky and Burks, 1988); European Molecular Biology Laboratory (EMBL) version 15 (Cameron, 1988); National Biomedical Research Foundation (NBRF) Nucleotide Sequence Data Library version 33 (Sidman *et al.*, 1988) and the NBRF Protein Resource data base (version 16).

The data was compared using the program WORDSEARCH which uses the algorithm of Wilbur & Lipman, (1983) to find regions of possible similarity. The output from WORDSEARCH is used by the program SEGMENTS which uses the alignment algorithm of Smith and Waterman, (1981) to display the best segment of similarity in the region identified by WORDSEARCH. Lipman and Pearson, (1985) considered that the procedure of Smith and Waterman was the best way to search for segments of similarity, however the best parameters to use for the program SEGMENTS are not clear. The alignments may be different depending on the values assigned for match, mismatch, gap weight and gap length weight. For the searches below the default values assigned by the UWGCG package were used.

The data was searched in the following way:

(a) The entire consensus sequence.
(b) The open reading frame for the gene.
(c) The open reading frame for the gene with a word mask.
    The mask + + - + + - + + was used. This requires that the bases at position 1 and 2 of each codon match but does not require a match in the third codon position. This is considered to be a more sensitive method of finding matching regions.
(d) The protein translation of the open reading frame was compared with the NBRF protein data base.

The search finds the registers of comparison (diagonals) which have the largest number of short perfect matches. A score was given for each diagonal and a plot made showing the number of diagonals with each possible score. By examining the plot, it was possible to determine if any of the matches had a score which was significantly above those of the population of diagonals whose similarity to the test sequence was random.

No significant matches were found in any of the searches. The best match with the consensus sequence was that of the *Rhizobium meliloti* ***ntr*A** gene (Ronson *et al.*, 1987) which had an 83 base region with 69% similarity to the gene sequenced in this project. However at the amino acid level the similarity was only 36% (compared to 60.6% between the *Rhizobium meliloti* ***ntr*A** gene and the *Klebsiella pneumoniae* ***ntr*A** gene (Merrick and Gibbins, 1985)). Additionally, the *R. meliloti ntr*A gene is 523 amino acids long, the *K. pneumoniae ntr*A gene 477 amino acids long but the gene sequenced in this work coded for only 192 amino acids.

The use of a word mask in the search did not produce any more significant matches.

No significant similarity was observed between the translation of the gene sequenced in this study and any of the protein sequences in the NBRF database.

DISCUSSION

Chua *et al.*, (1985) successfully introduced Tn5 into the genome of *R. loti* strain NZP2037. Thirteen Tn5 induced symbiotic mutants of strain NZP2037 were isolated, one was Nod^- and twelve others produced ineffective (Nod^+ Fix^-) nodules on *Lotus pedunculatus*. Chua demonstrated that all of these mutants were derived from strain NZP2037 and that DNA from each mutant contained Tn5 inserted in a single *Eco*RI fragment which was not the *Eco*RI fragment which hybridized with a *nif*HDK probe. Chua *et al.*, (1985) characterised 9 Nod^+ Fix^- mutants by light and electron microscopy and divided them into 3 classes. Five mutants were defective in aspects of nodule initiation (Noi) (Table 1, Vincent, 1980); each contained a Tn5 insertion in a different *Eco*RI fragment, implying that there are at least five different loci on the *R. loti* genome involved in nodule initiation. A similar situation has also been demonstrated with widely scattered *fix* regions on the chromosome of *R. meliloti* (Forrai *et al.*, 1983).

The second class of Nod^+ Fix^- mutants showed an accumulation of poly-ß-hydroxybutyrate in the bacteroids, a large number of small vacuoles and an accumulation of starch in the infected plant cells. Chua, (1984) suggested that transport functions within the nodule were defective and carbon substrates were not being mobilized to the bacteroids. These mutants were classified as deficient in complementary functions (Cof).

4.1 THE BACTERIAL RELEASE (Bar) MUTANT

The final class of Nod$^+$ Fix$^-$ mutant isolated by Chua was designated a bacterial release (Bar) mutant. Nodules formed by this mutant contained enlarged infection threads with more bacteria within the infection thread than in the infection threads of a normal effective nodule. No evidence was found for the release of the bacteria from the infection threads into the plant cells.

A similar phenotype has been observed for auxotrophic and antibiotic resistant mutants Vincent, (1980). Ma *et al.*, (1982) isolated several Tn5-induced *R. leguminosarum* Nod$^+$ Fix$^-$ mutants which were partially blocked at the release stage. These mutants mapped at several different loci on the Sym plasmid indicating that there may be several genes involved in this stage of nodule development. Morrison and Verma, (1987) have recently cloned and sequenced a gene involved in bacterial release in *Bradyrhizobium japonicum.*

4.2 COMPLEMENTATION OF THE Bar MUTANT

Chua constructed a gene library to the wild type *R. loti* strain NZP2037 in the broad host range vector pLAFR1 (Friedman *et al.*, 1982). To isolate pLAFR1 cosmids which were capable of complementing the Tn5-induced symbiotic mutants of *R. loti*, the gene library was crossed *en masse* (Long *et al.*, 1982) with each of the mutants in a triparental mating in which *E. coli* HB101/pRK2013 was used to provide a helper plasmid to mobilize the cosmids (Ditta *et al.*, 1980). *Lotus pedunculatus* seedlings were inoculated with transconjugants from the triparental mating. In the cases where complementation of the Tn5 mutation occurred a further triparental mating was carried out to transfer the cosmid responsible for the complementation back into *E. coli* HB101 (Chua *et al.*, 1985). In this way three cosmids, pPN318, pPN319 and pPN320 were isolated which were able to complement the Bar mutant (strain PN239).

4.3 MAPPING THE COSMIDS COMPLEMENTING THE Bar MUTANT

Physical maps of these three cosmids were constructed using the restriction endonucleases *Eco*RI and *Hin*dIII (Figure 9). The size of the restriction fragments from pPN318 and pPN319 was identical and hybridization data confirmed that pPN318 and pPN319 contained the same region of the *R. loti* genome.

One feature of the physical map of the cosmid pPN320 is the presence of three adjacent 5.9kb *Eco*RI fragments. Whether the similar size of these fragments is coincidental or whether they represent a repeated portion of the genome is unknown but this may prove an interesting area for further investigation.

4.4 THE COMMON 7.5kb *Eco*RI FRAGMENT

A 7.5kb *Eco*RI fragment was present in each of the cosmids which were able to complement the mutation. Chua *et al.*, (1985) found that the *Eco*RI fragment containing the Tn5 insert was approximately 13.7kb in size, and given that Tn5 itself is 5.7kb, a wild type *Eco*RI fragment of approximately 8.0kb was postulated. The difference in size can be easily accounted for by the difficulty of accurately measuring the size of the hybridizing fragment in a total genomic digest.

The 7.5kb *Eco*RI fragment was subcloned (Section 3.2) into pBR328 (pPN27) and into the broad host range vector pLAFR1 (pPN28). These plasmids were transferred into *E. coli* HB101 to produce strains PN1000 and PN1001.

The introduction of pPN28 (from strain PN1001) into the Bar mutant PN239 resulted in the ability of PN239 to form normal nitrogen fixing nodules (Section 3.3). It was shown (Section 3.3) that in the complemented strain, bands were present which hybridized to the 7.5kb *Eco*RI fragment. One

band (13.2kb) was derived from PN239 (7.5kb + 5.7kb Tn5) the other, (7.5kb) from pPN28. Thus the *bar* mutation is complemented by pPN28 rather than the wild-type DNA being recombined into the PN239 genome.

It was concluded that the 7.5kb *Eco*RI fragment contained genetic material necessary for effective nitrogen fixation by *Rhizobium loti* on *Lotus pedunculatus*.

4.5 LOCATION OF Tn5 INSERTIONS

To facilitate the location of the Tn5 insertion within the 7.5kb *Eco*RI fragment a restriction map was constructed using the restriction endonucleases *Pst*I, *Xho*I and *Sal*I. The location of Tn5 in the 7.5kb *Eco*RI fragment was determined by taking advantage of the presence of *Hin*dIII sites within Tn5 (Jorgensen *et al.*, 1979). Total genomic DNA from strain PN239 was digested with *Hin*dIII and a Southern blot of this digest was hybridized with the 7.5kb *Eco*RI fragment. There were no *Hin*dIII sites within the wild type 7.5kb *Eco*RI fragment; the only *Hin*dIII sites are carried by the Tn5 insertion. From the cosmid map (Section 3.1) and hybridization data the size of the wild type *Hin*dIII fragment containing the 7.5kb *Eco*RI fragment and the relative positions of the *Hin*dIII and *Eco*RI sites were determined. In the genomic DNA from strain PN239 two fragments hybridized to the 7.5kb *Eco*RI probe; these were the 1.2kb of Tn5 DNA from the *Hin*dIII site in Tn5 to the end of Tn5 plus the *Rhizobium* DNA from the point of the Tn5 insert to the *Hin*dIII site within the *Rhizobium* DNA on each side of the Tn5 insert (Section 3.5). The limitation on determining the exact position of insertion by this method is the accuracy with which the distance migrated by hybridizing fragments can be measured on an autoradiograph. Nevertheless this indicated the position of the Tn5 insertion, and hence the region of the 7.5kb *Eco*RI fragment which was of most interest to us.

To define the extent of the gene, further Tn5 mutagenesis of the 7.5kb fragment was carried out using the pLAFR1 - 7.5kb plasmid (pPN28) (Section 3.5.1). The position of each of these insertions given in Figure 18 has the measurement constraints discussed above. Nevertheless, a region, presumed to contain the gene, was defined between insertions 56 and 63. Insertions 18, 41 and 78 caused a Fix$^-$ phenotype. These insertions were presumed to be in the gene itself or in regulatory sequences upstream of the gene which are required for expression of the gene. The position of these insertions indicated the region to be sequenced in order to obtain the sequence of the gene causing the Bar$^-$ phenotype. This region spanned two *Sal*I sites and a *Xho*I site (Figure 18).

4.6 DNA SEQUENCING

The initial sequencing strategy was to subclone defined restriction fragments so that sequence data could be obtained starting from each of these sites. The restriction endonucleases *Sal*I and *Xho*I both recognize the same internal tetranucleotide sequence (TCGA) and cleave this sequence to produce identical "sticky ends". Hence fragments with ends produced by *Xho*I digestion can be ligated into *Sal*I sites, in this case the *Sal*I site in the M13mp9 sequencing vector (Messing and Vieira, 1982). This allowed the 1.8kb *Xho*I fragment, the 0.3kb *Sal*I fragment, the 1.4kb *Xho*I - *Sal*I fragment and the 0.9kb *Sal*I - *Sal*I fragment (Figure 14 and Table 10) to be subcloned into M13mp9. Each of these fragments was cloned and sequenced in both orientations and hence sequence data was obtained from both ends of each fragment.

To obtain the sequence from the *Xho*I site to the right of the two *Sal*I sites the 2.7kb *Eco*RI - *Xho*I fragment was cloned into M13mp8 cut with *Eco*RI - *Sal*I. This vector has the cloning cassette in the opposite orientation to M13mp9 and allowing sequence data to be obtained from the *Xho*I end of the fragment rather than from the *Eco*RI end.

With this framework of sequence data based on restriction sites with known positions, the remainder of the sequence in the region of the gene was determined by "shotgun cloning" fragments obtained from *Hae*III and *Alu*I digestion (Table 10). Using this approach the sequence of the entire region was obtained in both directions. Alternative strategies now available involve the construction of a series of directed deletions of a longer fragment using *Bal*31 (Poncz *et al.*, 1982) or *Exo*III (Guo and Wu, 1982; Heinkoff, 1984) and sequencing a series of overlapping fragments. This would save a considerable amount of effort in screening or sequencing the large number of templates obtained by a "shotgun" approach.

The primer used for the sequencing reactions was usually the "universal" 17mer primer (TGACCGGCAGCAAAATG) which anneals 8 bases to the right of the *Sma*I cleavage point of M13mp9 (Figure 20). When this primer was used the sequence at the junction of the M13mp9 vector/*Rhizobium* DNA insert was sometimes difficult to determine. This problem was overcome by using a 15mer primer (TGCAGCACTGACCT) which anneals 19 bases further away from the cloning cassette in M13mp9 (Figure 20) thus giving a longer reading into the junction point and allowing this sequence to be more accurately determined.

4.7 SEQUENCE COMPRESSIONS

Both strands of DNA in the region of interest were sequenced. This was necessary to confirm the sequence since a number of compressions were observed in both orientations (Section 3.9). Compressions appear in a sequencing gel as irregular spacing between bands on the autoradiograph. They are caused by strong secondary structure in the labeled products of the sequencing reaction, so that fragment mobility in the gel is influenced by conformation as well as size (Martin, 1987). In addition to sequencing the opposite strand, two methods were used to overcome compressions in the sequenced strand. AMV reverse transcriptase gave a less ambiguous reading through the compressed areas (Section 3.9.1), however it was not suitable for primary sequencing as it was prone to premature termination, "stops", in reading the sequence. Stops appear as bands across all four lanes of the autoradiograph. They are believed to be caused by an area of strong secondary structure in the template which causes the polymerase to terminate synthesis (Martin, 1987). While this is less of a problem if the basic sequence is already known it makes determination of new sequence very difficult. The second method used to overcome compressions was the use of 7M urea 40% formamide gels (Section 3.9.2) instead of the normal 8M urea gels. This proved to be a cheap and effective method of resolving all the compressions observed, however the formamide gels were more difficult to handle and so this cannot be recommended as a primary sequencing method.

The strategy used was to first determine the sequence of both strands using Klenow fragment polymerase and 8M urea gels, then to resequence any compressed regions using Klenow fragment polymerase but separating the fragments on 7M urea 40% formamide gels.

Overlapping sequence data obtained from the templates shown in Figure 21 was assembled and edited to form a contiguous sequence of 2307 base pairs.

4.8 SEQUENCING THE POSITION OF Tn5 INSERTIONS

To define the region of interest within the sequence it was necessary to know the exact point of insertion of the Tn5's which caused a Fix$^-$ phenotype and the position of the Tn5 insertions which did not affect nitrogen fixation on the left and right of these insertions. The Tn5-*Rhizobium* junction fragments were subcloned into M13mp9 as described in Section 3.10 and sequenced using the 17mer Tn5 primer GTTCATCGCAGGACTTG (Schofield and Watson, 1986; Egelhoff *et al.*, 1985). This primer extends from base 15 to base 31 at the ends of Tn5. It was used to sequence from the Tn5 insertion through the junction into the *Rhizobium* DNA. In this way the position of the Tn5 inserts which do not cause a Fix$^-$ phenotype (56 and 63, Figure 18) and the Tn5 insets which cause a Fix$^-$ phenotype (41, 239 and 18, Figure 18) were determined. The culture for 78 was unfortunately lost and so the exact position of this insertion was not determined, however from fragment mapping (Section 3.5.1) its position of insertion was between insertions 18 and 41.

At the point of insertion of Tn5 into the host DNA a 9bp repeat of the host DNA is generated (Berg *et al.*, 1983; Egelhoff *et al.*, 1985). This was demonstrated for the Tn5 insertions sequenced in this study (Section 3.10.2).

4.9 ANALYSIS OF THE SEQUENCE

A consensus sequence was assembled from the data obtained by sequencing the overlapping templates using the Staden VTUTIL programs and the GELASSEMBLE program in the UWGCG package (Section 3.8 and Figure 24). The consensus sequence was analysed for the presence of a putative gene using a number of programs contained in the UWGCG package.

4.9.1 Open Reading Frames

An open reading frame (ORF) is a region of sequence containing triplet codons without any termination codons. The sequence in an open reading frame is potentially translatable into a protein. Section 3.11.1 and Figure 27 demonstrates that there are ORF's of sufficient size to code for a protein in both orientations of the sequence. Further information was therefore required in order to determine: (a) which orientation of the sequence was transcribed, and (b) which of the open reading frames in that orientation was the one coding for the gene of interest.

4.9.2 Testcode

Fickett, (1982) devised a statistic which (independently of reading frame) predicts whether a DNA sequence is coding or not. Fickett's statistic was calculated using the UWGCG program TESTCODE (Section 3.11.2). The two regions which TESTCODE predicted to be coding corresponded to open reading frames identified in Section 3.11.1.

4.9.3 Third Position Compositional Bias

An alternative method for the detection of possible coding regions in a sequence examines the compositional bias of the third position of each codon triplet (Bibb *et al.*, 1984). This method is independent of the codon usage of the organism but is dependent on reading frame. Section 3.11.3 indicates the results obtained using this program. A window size of 25 codons (Section 3.11.3) was found to give a satisfactory indication of the putative coding regions. Two regions in the 5' to 3' direction were predicted to be coding by this statistic; these corresponded to the same ORF's identified as coding by TESTCODE (Section 3.11.2).

4.9.4 Codon Preference

The codon preference statistic (Section 3.11.5) compares the similarity of codons in each reading frame to a codon usage table. Where the codon usage is biased toward the use of particular codons for an amino acid, and this bias is reflected in the codon usage table which is compared to the test sequence, this statistic can indicate the correct reading frame for a gene. Differences in codon usage between organisms are well documented (Grosjean & Fiers, 1982; Ikemura, 1982; Ikemura, 1985). It has been suggested that the highly biased codon usage which occurs in some genes is related to the expression level for that gene (Grantham *et al.*, 1981; Gouy and Gautier, 1982; Sharp and Li, 1986). A number of methods of comparing codon usage have been developed (Staden and McLachlan, 1982; Gribskov *et al.*, 1984; Sharp and Li, 1987). The UWGCG analysis package used in this project uses the codon preference plot method of Gribskov *et al.*, (1984) to examine the relationship between the test sequence and a selected codon usage table.

The *Rhizobium* sequence determined in this project was compared with several codon usage tables. The UWGCG package contains two codon usage tables based on *E. coli* sequences, the codon usage table Ecohigh.cod describes the codon usage of *E. coli* genes which are expressed at a high level while Ecolow.cod describes the codon usage of *E. coli* genes which have low levels of expression. Genes which are not strongly expressed in *E. coli* show less biased codon usage than genes which are strongly expressed (Sharp and Li, 1986). The *Rhizobium* and *Bradyrhizobium* gene sequences listed in Table 11 were used to construct a *Rhizobium* codon usage table (Section 3.11.4). Separate codon frequency tables were constructed for genes sequenced from slow-growing *Bradyrhizobium* strains (Rhizslow.cod) and fast-growing *Rhizobium* strains (Rhizfast.cod). All available complete gene sequences from both *Rhizobium* and *Bradyrhizobium* species (Table 11) were combined in Rhizall.cod; the codon frequency table which is presented in Table 12. As the data presented in Table 13 shows, the Rhizall.cod codon frequency table was significantly different from the *E. coli* codon frequency tables but not from either the Rhizslow.cod or Rhizfast.cod codon frequency tables. All *Rhizobium* codon frequency tables resembled the Ecolow.cod codon frequency table more closely than they resembled the Ecohigh.cod codon frequency table.

The *R. loti* sequence determined in this project showed two regions in which the codon usage showed greater similarity to the codon usage tables constructed from *Rhizobium* genes than to random codon usage. These two regions were the same as those identified using TESTCODE (Section 3.11.2) and third position compositional bias (Section 3.11.3).

4.9.5 Rare Codon Usage

The UWGCG codon preference program also indicates which codons in the test sequence occur at a low frequency in the codon usage table (rare codons). Using the *E. coli* based codon usage tables, there were fewer rare codons in the putative coding regions than in the non-coding regions of the sequence (Figure 32A). When the *Rhizobium* codon frequency table Rhizall.cod was used, no rare codons occurred in the putative coding regions (Figure 32B). This decrease in the use of rare codons in the putative coding regions was further evidence that these were coding regions.

4.10 ADDITIONAL EVIDENCE FOR THE *bar* GENE

Data obtained from transcriptional fusions using Tn3-HoHo1 (D. McSweeney and D.B. Scott, personal communication) indicated that promoters existed upstream of the gene identified in this work. Lac fusions in the same orientation as this gene were expressed whereas no lac fusions were expressed the opposite orientation. This was further evidence that the coding region identified in this study codes for a gene.

The putative ORF defined by the position of the Tn5 insertions and predicted by TestCode, compositional bias and codon usage analysis was translated into an amino acid sequence (Section 3.12.1 and Figure 34). The putative protein had an unmodified molecular weight of 21.2kDal. Maxicell expression of a cloned *Alu*I fragment containing part of this gene region produced a protein of approximately 18 kDal molecular weight (E. Rockman, unpublished results). Translation of the gene from the ATG codon to the *Alu*I site predicted a protein size of 18.3kDal.

4.11 A POTENTIAL RIBOSOME BINDING SITE

The sequence upstream of the putative gene was examined for a potential ribosome binding site (Section 3.12.2). The ribosome binding site indicated fits 6 of the 7 rules for *E. coli* translational initiation sites proposed by Stormo *et al.*, (1982) (Table 14), hence it was likely to be the ribosome binding site for the gene.

4.12 A POSSIBLE PROMOTER REGION FOR THE GENE

Determining the location of a possible promoter for the gene (Section 3.12.3) was more difficult since no promoter sequences have been reported for *R. loti*. Promoter sequences have been identified for *nif* and *nod* genes in *R. leguminosarum* and *R. meliloti* (Ausubel, 1984), however no significant homology between these and any upstream region of this gene was found. A consensus promoter sequence for *fix* genes regulated by the products of the *ntr*A gene (*ntr*A activatible genes) has been proposed (Gussin *et al.*, 1986). The region upstream from the *bar* gene did not show any homology to this consensus sequence; CTGGYAYR-N4-TTGCA.

Ronson and Astwood, (1985), proposed a consensus sequence; TTPuANN-17b-PuAPuPuPuPu-4-5b-CA for *Rhizobium* "non-nitrogen" promoters. Comparison of the upstream region of the *bar* gene sequence, with this consensus sequence using the UWGCG programs BESTFIT and GAP, indicated a possible promoter region. This region did not have perfect homology with the consensus promoter, however this would not be surprising as the sequences used to construct the consensus promoter do not have perfect homology with each other (Leong *et al.*, 1985; Buikema *et al.*, 1985), and they are from *R. meliloti* not *R. loti*. However the positional relationship between the putative promoter sequence and the Tn5 insertions was consistent with the identified sequence being a possible promoter.

4.13 DIRECT AND INVERTED REPEATS

The sequence immediately upstream of the presumed ATG start codon for the gene contained a number of direct and inverted repeats (Section 3.12.5). The function of these repeats (if any) is not known.

4.14 TERMINATION STRUCTURES

The sequence downstream of the gene contains two stem-loop structures (Section 3.12.4) which are possible *rho*-dependent transcription terminators (Adhya and Gottesman, 1978). Both structures involve 28 bases (including the loop and the base-paired stem) and are relatively G-C rich. Neither stem-loop structure had associated with it the run of U-residues normally found in *rho*-independent terminators, however both were good candidates for *rho*-dependent terminators, which have no particular consensus sequence.

4.15 AMINO ACID TRANSLATION OF THE SEQUENCE

The DNA sequence and its putative amino acid translation was compared with the various databases as described in Section 3.12.6. No significant homology was found between any previously sequenced genes and the gene sequenced in this study. This is not particularly surprising in view of the data in Section 3.6 which indicates that the DNA fragment containing this gene hybridizes only to DNA from *R. loti* strains.

4.16 POSSIBLE FUNCTION OF THE GENE

Strain PN239, the Tn5 mutant strain of strain NZP2037 is Fix$^-$ on *Lotus pedunculatus*, however this strain is Fix$^+$ on *Lotus corniculatus* (D.B. Scott, personal communication). This suggested that the gene was involved in host specificity rather than in a basic step of nodulation common to all rhizobia. Fix$^-$ mutants have been isolated which form empty nodules which have either no infection threads (Chua *et al.*, 1985; Finan *et al.*, 1985; Leigh *et al.*, 1985; Vandenbosch *et al.*, 1985; Dylan *et al.*, 1986) or contain aborted infection threads (Noel *et al.*, 1986). These Fix$^-$ mutants have been shown to have altered exopolysaccharide composition (Finan *et al.*, 1985; Leigh *et al.*, 1985), lipopolysaccharide (Noel *et al.*, 1986) or alterations in ß-glucan synthesis (Dylan *et al.*, 1986). While Chua, (1984) indicated that the gene locus studied in this investigation was involved in bacterial release from the infection thread, further studies (Ward *et al.*, 1989) have shown that the phenotype is more complex. Electron microscopy of nodules formed by *R. loti* strains with Tn5 mutations affecting this locus showed some released rhizobia which had failed to develop into mature bacteroids. In some cases there was extensive degradation of the peri-bacteroid membrane, a phenotype characteristic of a number of *nif* and *fix* mutations (Ma *et al.*, 1982; Hirsch and Smith, 1987).

This gene locus has been referred to as the *bar* (bacterial release) locus throughout this thesis. The original description of the mutation in strain PN239 was of a Bar$^-$ phenotype (Chua, 1984). Other Tn5 mutations in this locus (Section 3.5.1, Ward *et al.*, 1989) have had a more complex phenotype (see comments above), however all have been Nod$^+$ Fix$^-$ on *Lotus pedunculatus*. Electron microscopy of the Fix$^-$ nodules indicated that the mutant strains had a reduced ability to invade the host. This locus has therefore been renamed *inv* for invasiveness and the gene referred to as the *inv* gene (Ward *et al.*, 1989).

R. loti strain NZP2037 is one of a small group of *R. loti* strains which are able to effectively nodulate *L. pedunculatus*. It is possible that mutations in the gene sequenced in this project cause some change in the cell surface which restricts the host range effectiveness of this strain or the mutation may cause a biochemical change which reduces the ability of the rhizobia to induce the host response necessary to form an effective symbiosis. Morrison and Verma, (1987) have reported a *Bradyrhizobium japonicum* mutant in which the bacteria are not released into the host cells. That mutation affects the synthesis of peribacteroid membrane (PBM) nodulins suggesting that these may be required for the efficient release of rhizobia and subsequent development of mature bacteroids. Identification of the biochemical function of the product of the gene sequenced in this project will aid the understanding of the interaction of host and bacteria in forming an effective nitrogen fixing symbiosis.

CONCLUSIONS

A region of the *R. loti* genome which is involved in the establishment of an effective nitrogen fixing symbiosis between strain NZP2037 and *Lotus pedunculatus* has been mapped and sequenced. A gene coding for a putative protein of molecular weight 21.2kDal was identified from the sequenced position of Tn5 insertions and from computer analysis of the sequence data. The gene appears to be unique to *R. loti* and may have a role in the host specificity of *R. loti* on *Lotus pedunculatus*.

REFERENCES

Abe, M., Sherwood, J.E., Hollingsworth, R.I. and Dazzo, F.B. (1984). Stimulation of clover root hair infection by lectin-binding oligosaccharides from the capsular and extracellular polysaccharides of *Rhizobium trifolii.* J. Bacteriol. 160: 517-520.

Adams, T.H. and Chelm, B.K. (1984). The *nif*H and *nif*DK promoter regions from *Rhizobium japonicum* share structural homologies with each other and with nitrogen-regulated promoters from other organisms. J. Mol. App. Genet. 2: 392-405.

Adhya, S. and Gottesman, M. (1978). Control of transcription termination. Ann. Rev. Biochem. 47: 967-996.

Aguilar, O.M., Reilander, H., Arnold, W. and Puhler, A. (1987). *Rhizobium meliloti nif*N (*fix*F) gene is part of an operon regulated by a *nif*-A dependent promoter and codes for a polypeptide homologous to the *nif*K gene product. J. Bacteriol. 169: 5393-5400.

Ausubel, F.M. (1984). Regulation of nitrogen fixation genes. Cell. 37: 5-6.

Bachem, C., Banfalvi, Z., Kondorosi, E., Schell, J. and Kondorosi, A. (1986). Identification of host range determinants in the *Rhizobium* species MPIK3030. Mol. Gen. Genet. 203: 42-48.

Berg, D.E., Schmandt, M.A. and Lowe, J.B. (1983). Specificity of transposon Tn5 insertion. Genetics 105: 813-828.

Beringer, J.E. (1974). R-factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. 84: 188-198.

Beringer, J.E., Beynon, J.L., Buchanan-Wollaston, A.V. and Johnston, A.W.B. (1978). Transfer of the drug resistant transposon Tn5 to *Rhizobium*. Nature 276: 633-634.

Bibb, M.I., Findlay, P.R. and Johnson, M.W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30: 157-166.

Bilofsky, H.S. and Burks, C. (1988). The GenBank^{R} genetic sequence data bank. Nucl. Acids Res. 16: 1861-1863.

Bolivar, F., Rodriguez, R., Greene, P.J., Betlach, M., Heyneker, H.L., Boyer, H.W., Crosa, I. and Falkow, S. (1977). Construction and characterisation of new cloning vehicles. II. A multi-purpose cloning system. Gene 2: 95-113.

Borthakur, D. and Johnston, A.W.B. (1987). Sequence of *psi*, a gene on the symbiotic plasmid of *Rhizobium phaseoli* which inhibits exopolysaccharide synthesis and nodulation and demonstration that its transcription is inhibited by *psr*, another gene on the symbiotic plasmid. Mol. Gen. Genet. 207: 149-154.

Boyer, H.W. and Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. 41: 459-472.

Buchanan-Wollaston, A.V., Beringer, J.E., Brewin, N.J., Hirsch, P.R. and Johnston, A.W.B. (1980). Isolation of symbiotically defective mutants of *Rhizobium leguminosarum* by insertion of the transposon Tn5 into a transmissible plasmid. Mol. Gen. Genet. 178: 185-190.

Buikema, W.J., Szeto, W.W., Lemley, P.V., Orme-Johnson, W.H. and Ausubel, F.M. (1985). Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntr*C of *K. pneumoniae*. Nucl. Acids Res. 13: 4539-4555.

Burkardt, B. and Burkardt, H.J. (1984). Visualisation and exact molecular weight determination of a *Rhizobium meliloti* megaplasmid. J. Mol. Biol. 175: 213-218.

Callaham, D.A. and Torrey, J.G. (1981). The structural basis for infection of root hairs of *Trifolium repens* by *Rhizobium*. Can. J. Bot. 59: 1647-1664.

Calvert, H.E., Pence, M.K., Pierce, M., Malik, N.S.A. and Bauer, W.D. (1984). Anatomical analysis of the development and distribution of *Rhizobium* infections in soybean roots. Can. J. Bot. 62: 2375-2384.

Cameron, G.N. (1988). The EMBL data library. Nucl. Acids Res. 16: 1865-1867.

Casse, F., Boucher, C., Julliot, J.S., Michel, M. and Denarie, J. (1979). Identification and characterization of large plasmids in *Rhizobium meliloti* using agarose gel electrophoresis. J. Gen. Microbiol. 13: 229-242.

Carlson, R.W., Kalembasa, S., Turowski, D., Pachori, P. and Noel, K.D. (1987). Characterization of the lipopoly-saccharide from a *Rhizobium phaseoli* mutant that is defective in infection thread development. J. Bacteriol. 169: 4923-4928.

Carlson, T.A. and Chelm, B.K. (1986). Apparent eukaryotic origin of glutamine synthetase II from the bacterium *Bradyrhizobium japonicum*. Nature 322: 568-570.

Carlson, T.A., Guerinot, M.L. and Chelm, B.K. (1985). Characterization of the gene encoding glutamine synthetase I (*gln*A) from *Bradyrhizobium japonicum*. J. Bacteriol. 162: 698-703.

Chua, K.Y. (1984). A Molecular and Genetical Analysis of Symbiotic Genes in *Lotus* Rhizobia. PhD Thesis, Massey University.

Chua, K.Y., Pankhurst, C.E., MacDonald, P.E., Hopcroft, D.H., Jarvis, B.D.W. and Scott, D.B. (1985). Isolation and characterization of transposon Tn5-induced symbiotic mutants of *Rhizobium loti*. J. Bacteriol. 162: 335-343.

Clover, R.H., Kieber, J. and Signer, E.R. (1989). Lipopoly-saccharide mutants of *Rhizobium meliloti* are not defective in symbiosis. J. Bacteriol. 171: 3961-3967.

Colonna-Romano, S., Riccio, A., Guida, M., Defez, R., Lamberti, A., Iaccarino, M., Arnold, W., Priefer, U. and Puhler, A. (1987). Tight linkage of *gln*A and a putative regulatory gene in *Rhizobium leguminosarum*. Nucl. Acids Res. 15: 1951-1964.

Corbin, D., Ditta, G. and Helinski, D.R. (1982). Clustering of nitrogen fixation (*nif*) genes in *Rhizobium meliloti*. J. Bacteriol. 149: 221-228.

Crow, V.L., Jarvis, B.D.W. and Greenwood, R.M. (1981). Deoxyribonucleic acid homologies among acid-producing strains of *Rhizobium*. Int. J. Syst. Bacteriol. 31: 152-172.

Dart, P.J. (1977). Infection and development of leguminous nodules pp. 367-472 In R.W.F. Hardy and W.S. Silver (eds.) A Treatise on Dinitrogen Fixation, Section 3, Biology. Wiley (Interscience), New York.

Davis, E.O., Evans, I.J. and Johnston, A.W.B. (1988). Identification of *nod*X, a gene that allows *Rhizobium leguminosarum* biovar *viciae* strain TOM to nodulate Afghanistan peas. Mol. Gen. Genet. 212: 531-535.

Dazzo, F.B., Truchet, G.L., Sherwood, J.E., Hrabak, E.M., Abe, M. and Pankratz, S.H. (1984(a)). Specific phases of root hair attachment in the *Rhizobium trifoli* - clover symbiosis. Appl. Environ. Microbiol. 48: 1140-1150.

Dazzo, F.B., Truchet, G.L. and Hrabak, E.M. (1984(b)). Specific enhancement of clover root hair infection by trifolin A-binding lipopolysaccharide from *Rhizobium trifolii* p 413. In C. Veeger and W.E. Newton (eds.) Advances in Nitrogen Fixation Research. Nijhoff Publishers, The Hague.

Debelle, F. and Sharma, S.B. (1986). Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity of nodulation. Nucl. Acids Res. 14: 7453-7472.

De Bruijn, F.J. and Lupski, J.R. (1984). The use of transposon Tn5 mutagenesis in the rapid generation of correlated physical and genetic maps of DNA segments cloned into multicopy plasmids - A review. Gene 27: 131-149.

De Maagd, R.A., Rao, A.S., Mulders, I.H.M., Goosen-De Roo, L., Van Loosdrecht, M.C.M., Wijffelman, C.A. and Lugtenberg B.J.J. (1989). Isolation and characterization of mutants of *Rhizobium leguminosarum* bv. *viciae* 248 with altered lipopolysaccharides: Possible role of surface charge or hydrophobicity in bacterial release from the infection thread. J. Bacteriol. 171: 1143-1150.

De Maagd, R.A., Spaink, H.P., Pees, E., Mulders, I.H.M., Wijfjes, A., Wijffelman, C.A., Okker, R.J.H. and Lugtenberg, B.J.J. (1989). Localization and symbiotic function of a region on the *Rhizobium leguminosarum* sym plasmid pRL1JI responsible for a secreted, flavonoid-inducible 50-kilodalton protein. J. Bacteriol. 171: 1151-1157.

Denarie, J., Boistard, P., Casse-Delbart, F., Atherly, A.G., Berry, J.O. and Russell, P. (1981). Indigenous plasmids of *Rhizobium*. In K.L. Giles and A.G. Atherly (eds.) Int. Rev. Cytol. Suppl. 13: Academic Press, New York.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12: 387-395.

Diaz, C.L., Melchers, L.S., Hooykaas, P.J.J., Lugtenberg, B.J.J. and Kijne, J.W. (1989). Root lectin as a determinant of host-plant specificity in the *Rhizobium*-legume symbiosis. Nature 338: 579-581.

Ditta, G., Stanfield, S., Corbin, D. and Helinski, D.R. (1980). Broad host range DNA cloning system for Gram-negative bacteria. Construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. U.S.A. 77: 7347-7351.

Djordjevic, M.A., Redmond, J.W., Batley, M. and Rolfe, B.G. (1987). Clovers secrete specific phenolic compounds which either stimulate or repress *nod* gene expression in *Rhizobium trifolii*. EMBO J. 6: 1173-1179.

Djordjevic, M.A., Schofield, P.R., Ridge, R.W., Morrison, N.A., Bassam, B.J., Plazinski, J., Watson, J.M. and Rolfe, B.G. (1985). *Rhizobium* nodulation genes involved in root hair curling (Hac) are functionally conserved. Plant Mol. Biol. 4: 147-160.

Djordjevic, M.A., Zurkowski, W., Shine, J. and Rolfe, B.G. (1983). Sym plasmid transfer to symbiotic mutants of *Rhizobium trifolii, Rhizobium leguminosarum* and *Rhizobium meliloti.* J. Bacteriol. 156: 1035-1045.

Downie, J.A., Knight, C.D., Johnston, A.W.B. and Rossen, L. (1985). Identification of genes and gene products involved in the nodulation of peas by *Rhizobium leguminosarum*. Mol. Gen. Genet. 198: 255-262.

Dylan, T., Ielpi, L., Stanfield, S., Kashyap, L., Douglas, C., Yanofsky, M., Nester, E., Helinski, D.R. and Ditta, G. (1986). *Rhizobium meliloti* genes required for nodule development are related to chromosomal virulence genes in *Agrobacterium tumefaciens.* Proc. Natl. Acad. Sci. USA 83: 4403-4407.

Egelhoff, T.T., Fisher, R.F., Jacobs, T.W., Mulligan, J.T. and Long, S.R. (1985). Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nod*D is read divergently from *nod*ABC. DNA 4: 241-248.

Evans, I.J. and Downie, J.A. (1986). The *nod*I gene product of *Rhizobium leguminosarum* is closely related to ATP-binding bacterial transport proteins; nucleotide sequence analysis of the *nod*I and *nod*J genes. Gene. 43: 95-101.

Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. Nucl. Acids Res. 10: 5303-5318.

Finan, T.M., Hirsch, A.M., Leigh, J.A., Johansen, E., Kuldan, G.A., Deegan, S., Walker, G.C. and Signer, E.R. (1985). Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. Cell 40: 869-877.

Fisher, R.A., Swanson, J.A., Mulligan, J.T. and Long, S.R. (1987). Extended region of nodulation genes in *Rhizobium meliloti* 1021. II. Nucleotide sequence, transcription start sites and protein products. Genetics 117: 191-201.

Fisher, R.F., Tu, J.K. and Long, S.R. (1985). Conserved nodulation genes in *Rhizobium meliloti* and *Rhizobium trifolii*. App. Environ. Microbiol. 49: 1432-1435.

Forrai, T., Vincze, E., Banfalvi, Z., Kiss, G.B., Randhawa, G.S. and Kondorosi, A. (1983). Location of symbiotic mutations in *Rhizobium meliloti*. J. Bacteriol. 153: 635-643.

Friedman, A.M., Long, S.R., Brown, S.E., Buikema, W.J. and Ausubel, F.M. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene 18: 289-296.

Fuhrmann, M. and Hennecke, H. (1984). *Rhizobium japonicum* nitrogenase Fe protein gene (*nif*H). J. Bacteriol. 158: 1005-1011.

Goettfert, M., Horvath, B., Kondorosi, E., Putnoky, P., Rodriguez-Quinones, F. and Kondorosi, A. (1986). At least two *nod*D genes are necessary for efficient nodulation of alfalfa by *Rhizobium meliloti*. J. Mol. Biol. 191: 411-420.

Goodchild, D.J. (1978). The ultrastructure of root nodules in relation to nitrogen fixation. Int. Rev. Cytol. 41: 235-288.

Goodchild, D.J. and Bergersen, F.J. (1966). Electron microscopy of the infection and subsequent development of soybean nodule cells. J. Bacteriol. 92: 204-213.

Gotz, R., Evans, I.J., Downie, J.A., Johnston, A.W.B. (1985). Identification of the host range DNA which allows *Rhizobium legumosarum* strain TOM to nodulate cv. Afghanistan peas. Mol. Gen. Genet. 201: 296-300.

Gouy, M. and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. Nucl. Acids Res. 10: 7055-7074.

Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pave, A. (1980). Codon catalog usage and the genome hypothesis. Nucl. Acids Res. 8: r49-r62.

Grantham, R., Gautier, C. and Gouy, M. (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucl. Acids Res. 8: 1893-1912.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981). Codon catalog usage is a genome strategy modulated for gene expessivity. Nucl. Acids Res. 9: r43-r74.

Greenwood, R.M. and Pankhurst, C.E. (1977). The *Rhizobium* component of the nitrogen-fixing symbiosis. Proc. N.Z. Grassland Ass. 38: 167-174.

Gribskov, M., Devereux, I. and Burgess, R.R. (1984). The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucl. Acids Res. 12: 539-549.

Groenger, P., Manian, S.S., Reilaender, H., O'Connell, M., Priefer, U.B. and Puhler, A. (1987). Organisation and partial sequence of a DNA region of the *Rhizobium leguminosarum* symbiotic plasmid pRL6JI containing the genes *fix*ABC, *nif*A, *nif*B and a novel open reading frame. Nucl. Acids Res. 15: 31-49.

Grosjean, H. and Fiers, W. (1982). Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18: 199-209.

Guo, L-H. and Wu, R. (1982). New rapid methods for DNA sequencing based on exonuclease III digestion followed by repair synthesis. Nucl. Acids Res. 10: 2065-2084.

Gussin, G.N., Ronson, C.W. and Ausubel, F.M. (1986). Regulation of nitrogen fixation genes. Annu. Rev. Genet. 20: 567-591.

Halverson, L.J. and Stacey, G. (1985). Host recognition in the *Rhizobium*-soybean symbiosis. Evidence for the involvement of lectin in nodulation. Plant Physiology 77: 621-625.

Halverson, L.J. and Stacey, G. (1986(a)). Effect of lectin on nodulation by wild-type *Bradyrhizobium japonicum* and a nodulation-defective mutant. Appl. Environ. Microbiol. 51: 753-760.

Halverson, L.J. and Stacey, G. (1986(b)). Signal exchange in plant-microbe interactions. Microbiol. Rev. 50: 193-225.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351-359.

Higashi, S. (1967). Transfer of clover infectivity of *Rhizobium trifolii* to *Rhizobium phaseoli* as mediated by an episomic factor. J. Gen. Appl. Microbiol. 13: 391-403.

Higashi, S. and Abe, M. (1980). Promotion of infection thread formation by substances from *Rhizobium*. Appl. Environ. Microbiol. 39: 297-301.

Hirsch, A.M. and Smith, C.A. (1987). Effects of *Rhizobium meliloti nif* and *fix* mutants on alfalfa root nodule development. J. Bacteriol. 169: 1137-1146.

Holmes, D.S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114: 193-197.

Horvath, B., Bachem C.W.B., Schell, J. and Kondorosi, A. (1987). Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the *nod*D gene product. EMBO J. 6: 841-848.

Horvath, B., Kondorosi, E., John, M., Schmidt, J., Torok, I., Gyorgypal, Z., Barabas, I., Wieneke, U., Schell, J. and Kondorosi, A. (1986). Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for Alfalfa. Cell 46: 335-343.

Iismaa, S.E. and Watson, J.M. (1987). A gene upstream of the *Rhizobium trifolii nif*A gene encodes a ferridoxin-like protein. Nucl. Acids Res. 15: 3180.

Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. 158: 573-597.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2: 13-34.

Innes, R.W., Hirose, M.A. and Kuempel P.L. (1988). Induction of nitrogen-fixing nodules on clover requires only 32 kilobase pairs of DNA from the *Rhizobium trifolii* symbiosis plasmid. J. Bacteriol. 170: 3793-3802.

Innes, R.W., Kuempel, P.L., Plazinski, J., Cantercremers, H.C.J., Rolfe, B.G. and Djordjevic, M.A. (1985). Plant factors induce expression of nodulation and host-range genes in *R. trifolii*. Mol. Gen. Genet. 201: 426-432.

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

Jacobs, T.W., Egelhoff, T.T. and Long, S.R. (1985). Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nod*C. J. Bacteriol. 162: 469-476.

Jarvis, B.D.W., Pankhurst, C.E. and Patel, J.J. (1982). *Rhizobium loti*, a new species of legume root nodule bacteria. Int. J. Syst. Bacteriol. 32: 378-380.

Jordan, D.C. (1982). Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium japonicum* gen, nov., a genus of slow-growing, root nodule bacteria from leguminous plants. Int. J. Syst. Bacteriol. 32: 136-139.

Jorgensen, R.A., Rothstein, S.J. and Reznikoff, W.S. (1979). A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. Mol. Gen. Genet. 177: 65-72.

Kaluza, K. and Hennecke, H. (1984). Fine structure analysis of the *nif*DK operon encoding the alpha and beta subunits of dinitrogenase from *Rhizobium japonicum*. Mol. and Gen. Genet. 196: 35-42.

Kaluza, K., Fuhrmann, M., Hahn, M., Regensburger, B. and Hennecke, H. (1983). In *Rhizobium japonicum* the nitrogenase genes *nif*H and *nif*DK are separated. J. Bacteriol. 155: 915-918.

King, P.V. and Blakesley, R.W. (1986). Optimizing DNA ligations for Transformation. Focus (Bethesda Research Laboratories) 8: 1-3.

Kondorosi, E., Banfalvi, Z. and Kondorosi, A. (1984). Physical and genetic analysis of a symbiotic region of *Rhizobium meliloti*: Identification of nodulation genes. Mol. Gen. Genet. 193: 445-452.

Kondorosi, E. and Kondorosi, A. (1986). Nodule induction on plant roots by *Rhizobium*. TIBS 11: 296-298.

Leigh, J.A., Signer, E.R. and Walker, G.C. (1985). Exopoly-saccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. Proc. Natl. Acad. Sci. USA 82: 6231-6235.

Leong, S.A., Ditta, G.S. and Helinski, D.R. (1982). Heme biosynthesis in *Rhizobium*. J. Biol. Chem. 257: 8724-8730.

Leong, S.A., Williams, P.H. and Ditta, G.S. (1985). Analysis of the 5'-regulatory region of the gene for delta-aminolevulinic acid synthetase of *Rhizobium meliloti*. Nucl. Acids Res. 13: 5965-5976.

Lipman, D.J. and Pearson, W.R. (1985). Rapid and sensitive protein similarity searches. Science 227: 1435-1441.

Long, S.R., Buikema, W.J. and Ausubel, F.M. (1982). Cloning of *Rhizobium meliloti* nodulation genes by direct complementation of Nod$^-$ mutants. Nature 298: 485-488.

Ma, Q-S., Johnston, A.W.B., Hombrecher, G. and Downie, J.A. (1982). Molecular genetics of mutants of *Rhizobium leguminosarum* which fail to fix nitrogen. Mol. Gen. Genet. 187: 166-171.

Mackinnon, P.A., Robertson, J.G., Scott, D.J. and Hale, C.N. (1975). Legume inoculant usage in New Zealand. N.Z. J. Experi. Agri. 5: 35-39.

Martin, R. (1987). Overcoming DNA Sequencing Artifacts: Stops and Compressions. Focus (Bethesda Research Laboratories) 9: 8-10.

McClung, C.R., Somerville, J.E., Guerinot, M.L. and Chelm, B.K. (1987). Structure of the *Bradyrhizobium japonicum* gene *hem*A encoding 5-aminolevulinic acid synthase. Gene 54: 133-139.

Merrick, M.J. and Gibbins, J.R. (1985). The nucleotide sequence of the nitrogen-regulation gene *ntr*A of *Klebsiella pneumoniae* and comparison with conserved features in bacterial RNA polymerase sigma factors. Nucl. Acids Res. 13: 7607-7620.

Messing, J. and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand or double-digest restriction fragments. Gene 19: 269-276.

Miller, J.H. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Morrison, N. and Verma, D-P.S. (1987). A block in the endocytosis of *Rhizobium* allows cellular differentiation in nodules but affects the expression of some peribacteroid membrane nodulins. Plant Mol. Biol. 9: 185-196.

Mulligan, J.T. and Long, S.R. (1985). Induction of *Rhizobium meliloti nod*C expression by plant root exudate requires *nod*D. Proc. Natl. Acad. Sci. USA 82: 6609-6613.

Neddleman, S.B. and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453.

Newcomb, W. (1981). Nodule morphogenesis and differentiation. Int. Rev. of Cytology Suppl. 13: 247-298.

Nieuwkoop, A.J., Banfalvi, Z., Deshmante, N., Gerhold, D., Schell, M.G., Sirotkin, K.M. and Stacey, G. (1987). A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. J. Bacteriol. 169: 2631-2638.

Noel, K.D., Stacey, G., Tandon, S.R., Silver, L.E. and Brill, W.J. (1982). *Rhizobium japonicum* mutants defective in symbiotic nitrogen fixation. J. Bacteriol. 152: 485-494.

Noel, K.D., Vandenbosch, K.A. and Kulpaca, B. (1986). Mutations in *Rhizobium phaseoli* that lead to arrested development of infection threads. J. Bacteriol. 168: 1392-1401.

Noti, J.D., Folkerts, O., Turken, A.N. and Szalay, A.A. (1986). Organization and characterization of genes essential for symbiotic nitrogen fixation from *Bradyrhizobium japonicum* I110. J. Bacteriol. 167: 774-783.

Nuti, M.P., Ledeboer, A.M., Lepidi, A.A. and Schilperoort, R.A. (1977). Large plasmids in different *Rhizobium* species. J. Gen. Microbiol. 100: 241-248.

Nuti, M.P., Lepidi, A.A., Prakash, R.K., Schilperoort, R.A. and Cannon, F.C. (1979). Evidence for nitrogen fixation (*nif*) genes on indigenous *Rhizobium* plasmids. Nature 282: 533-535.

Pankhurst, C.E. (1974). Ineffective *Rhizobium trifolii* mutants examined by immune-diffusion, gel-electrophoresis and electron microscopy. J. Gen. Microbiol. 82: 405-413.

Pankhurst, C.E. (1977). Symbiotic effectiveness of antibiotic resistant mutants of fast- and slow- growing strains of *Rhizobium* nodulating *Lotus* species. Can. J. Microbiol. 23: 1026-1033.

Pankhurst, C.E., Broughton, W.J. and Wieneke, U. (1983). Transfer of an indigenous plasmid of *Rhizobium loti* to other rhizobia and *Agrobacterium tumefaciens.* J. Gen. Microbiol. 129: 2535-2543.

Pankhurst, C.E., Craig, A.S. and Jones, W.T. (1979). Effectiveness of *Lotus* root nodules. I. Morphology and flavolan content of nodules formed on *Lotus pedunculatus* by fast-growing *Lotus* rhizobia. J. Exp. Bot. 30: 1085-1093.

Pankhurst, C.E., Schwinghamer, E.A. and Bergersen, F.J. (1972). The structure and acetylene-reducing activity of root nodules formed by a riboflavin - requiring mutant of *Rhizobium trifolii.* J. Gen Microbiol. 70: 161-177.

Peters, N.K., Frost, J.W. and Long, S.R. (1986). A plant flavone, luteolin, induces expression of *Rhizobium meliloti* genes. Science 233: 977-980.

Philip-Hollingsworth, S., Hollingsworth, R.I., Dazzo, F.B., Djordjevic, M.A. and Rolfe, B.G. (1989). The effect of interspecies transfer of *Rhizobium* host-specific nodulation genes on acidic polysaccharide structure and *in situ* binding by host lectin. J. Biol. Chem. 264: 5710-5714.

Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. and Surrey, S. (1982). "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. (USA) 79: 4298-4302.

Prakash, R.K., Schilperoort, R.A. and Nuti, M.P. (1981). Large plasmids of fast-growing rhizobia: homology studies and location of structural nitrogen fixation (*nif*) genes. J. Bacteriol. 145: 1129-1136.

Prakash, R.K., Hooykaas, P.J.J., Ledeboer, A.M., Schilperoort, R.A., Nuti, M.P., Lepidi, A.A., Casse, F., Boucher, C., Julliot, J.S. and Denarie, J. (1978). Detection, isolation and characterisation of large plasmids in *Rhizobium*. pp 139-163. In W.E. Newton and W.H. Orme-Johnson (eds.) Proceedings of the III International Symposium on Nitrogen Fixation. University Park Press, Baltimore.

Quinto, C., De La Vega, H., Flores, M., Leemans, J., Cevallos, M.A., Pardo, M.A., Azpiroz, R., De Lourdes-Girard, M., Calva, E. and Palacios, R. (1985). Nitrogenase reducatse: A functional multigene family in *Rhizobium phaseoli*. Proc. Natl. Acad. Sci. (USA) 82: 1170-1174.

Redmond, J.W., Batley, M., Djordjevic, M.A., Innes, R.W., Kuempel, P.L. and Rolfe, B.G. (1986). Flavones induce expression of nodulation genes in *Rhizobium*. Nature 323: 632-635.

Renalier, M-H., Batut, J., Ghai, J., Terzaghi, B., Gherardi, M., David, M., Garnerone, A-M., Vasse, J., Truchet, G., Huguet, T. and Boistard, P. (1987). A new symbiotic cluster on the pSym megaplasmid of *Rhizobium meliloti* 2011 carries a functional *fix* gene repeat and a *nod* locus. J. Bacteriol. 169: 2231-2238.

Roberts, G.P. and Brill, W.J. (1981). Genetics and regulation of nitrogen fixation. Ann. Rev. Microbiol. 35: 207-235.

Robertson, J.G., Lyttleton, P., Bullivant, S. and Grayston, G.F. (1978). Membranes in lupin root nodules. I. The role of Golgi bodies in the biogenesis of infection threads and peribacteroid membranes. J. Cell. Sci. 30: 129-149.

Robertson, J.G. and Farnden, K.J.F. (1980). Ultrastructure and metabolism of the developing legume root nodule. pp 65-113. In Cohn. E. (ed.) The Biochemistry of Plants, Vol. 5. Academic Press, New York.

Ronson, C.W., Nixon, B.T. and Ausubel, F.M. (1987). Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell 49: 579-581.

Ronson, G.W. and Astwood, P.M. (1985). Genes involved in the carbon metabolism of bacteroids. pp.201-207. In H.J. Evans, P.J. Bottomley, and W.E. Newton, (eds.) Nitrogen Fixation Research Progress. Martinus Nijhoff, Boston.

Rossen, L., Davis, E.O. and Johnston, A.W.B. (1987). Plant-induced expression of *Rhizobium* genes involved in host specificity and early stages of nodulation. TIBS 12: 430-433.

Rossen, L., Ma, Q-S, Mudd, E.A., Johnston, A.W.B. and Downie, J.A. (1984a). Identification and DNA sequence of *fix*Z, a *nif*B-like gene from *Rhizobium leguminosarum*. Nucl. Acids Res. 12: 7123-7134.

Rossen, L., Johnston, A.W.B. and Downie, J.A. (1984b). DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nod*AB and C required for root hair curling. Nucl. Acids Res. 12: 9497-9508.

Rossen, L., Shearman, C.A., Johnston, A.W.B. and Downie, J.A. (1985). The *nod*D gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the nodA,B,C genes. EMBO J. 4: 3369-3373.

Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A. and Kondorosi, A. (1986). Conservation of extended promoter regions of nodulation genes in *Rhizobium*. Proc. Natl. Acad. Sci. (USA) 83: 1757-1761.

Ruvkun, G.B., Sundaresan, V. and Ausubel, F.M. (1982). Directed transposon Tn5 mutagenesis and complementation analysis of *Rhizobium meliloti* symbiotic nitrogen fixation genes. Cell 29: 551-559.

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

Schofield, P.R., Ridge, R.W., Djordjevic, M.A., Rolfe, B.G., Shine, J. and Watson, J.M. (1984). Host-specific nodulation is encoded on a 14kb DNA fragment in *Rhizobium trifolii*. Plant Mol. Biol. 3: 3-11.

Schofield, P.R. and Watson, J.M. (1986). DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nod*ABC and *nod*FE. Nucl. Acids Res. 14: 2891-2903.

Scott, D.B., Court, C.B., Ronson, C.W., Scott, K.F., Watson, J.M., Schofield, P.R. and Shine, J. (1984). Organization of nodulation and nitrogen fixation genes on a *Rhizobium trifolii* symbiotic plasmid. Arch. Microbiol. 139: 151-157.

Scott, D.B. and Ronson, C.W. (1982). Identification and mobilization by cointegrate formation of a nodulation plasmid in *Rhizobium trifolii*. J. Bacteriol. 151: 36-43.

Scott, D.B., Chua, K.Y., Jarvis, B.D.W. and Pankhurst, C.E. (1985). Molecular cloning of a nodulation gene from fast- and slow-growing strains of *Lotus* rhizobia. Mol. Gen. Genet. 201: 43-50.

Seaney, R.R. and Henson, P.R. (1970). Birdsfoot trefoil. Adv. Agron. 22: 119-157.

Sharp, P.M. and Li., W-H. (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucl. Acids Res. 14: 7737-7749.

Sharp, P.M. and Li., W-H. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. Nucl. Acids Res. 15: 1281-1295.

Shearman, C.A., Rossen, L., Johnston, A.W.B. and Downie, J.A. (1986). The *Rhizobium leguminosarum* nodulation gene *nod*F encodes a polypeptide similar to acyl-carrier protein and is regulated by *nod*D plus a factor in pea root exudate. EMBO J. 5: 647-652.

Shine, J. and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71: 1342-1346.

Sidman, K.E., George, D.G., Barker, W.C. and Hunt, L.T. (1988). The protein identification resource (PIR). Nucl. Acids Res. 16: 1869-1871.

Smetham, M.L. (1973). Pasture legume species and strains. pp 85-127. In Langer, R.H.M. (ed.) Pastures and Pasture Plants. A.H. and A.W. Reed, Wellington.

Smith, T.F. and Waterman, M.S. (1981). Comparison of biosequences. Adv. Appl. Math. 2: 482-489.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.

Spaink, H.P., Okker, R.J.H., Wijffelman, C.A., Pees, E. and Lugtenberg, B.J.J. (1977). Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Plant Mol. Biol. 9: 27-39.

Spaink, H.P., Okker, R.J.H., Wijffelman, C.A., Tak, T., Goosen-de Roo, L., Pees, E., Van Brussel, A.A.N., and Lugtenberg, B.J.J. (1989). Symbiotic properties of rhizobia containing a flavonoid-independent hybrid *nod*D product. J. Bacteriol. 171: 4045-4053.

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. Nucl. Acids Res. 8: 3673-3694.

Staden, R. and McLachlan, A.D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucl. Acids Res. 10: 141-156.

Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982). Characterization of translational initiation sites in *E. coli*. Nucl. Acids Res. 10: 2971-2996.

Szeto, W.W., Zimmerman, J.L., Sundaresan, V. and Ausubel, F.M. (1984). A *Rhizobium meliloti* symbiotic regulatory gene. Cell 36: 535-543.

Thornton, H.G. (1930). The early development of the root-nodule of lucerne. (*Medicago sativa* L.) Annals of Botany 44: 385-392.

Torok, I. and Kondorosi, A. (1981). Nucleotide sequence of the *Rhizobium meliloti* nitrogenase reductase (*nif*H) gene. Nucl. Acids Res. 9: 5711-5723.

Triplett, E.W., Roberts, G.P., Ludden, P.W. and Handelsman, J. (1989). What's new in nitrogen fixation. ASM News 55: 15-21.

Truchet, G., Michel, M. and Denarie, J. (1980). Sequential analysis of the organogenesis of lucerne (*Medicago sativa*) root nodules using symbiotically-defective mutants of *Rhizobium meliloti.* Differentiation 16: 163-172.

Vandenbosch, K.A., Noel, K.D., Kaneko, Y. and Newcomb, E.H. (1985). Nodule initiation elicited by noninfective mutants of *Rhizobium phaseoli.* J. Bacteriol. 162: 950-959.

Verma, D.P.S., Kazazian, V., Zogbi, V. and Bal, A, K. (1978a). Isolation and characterization of the membrane envelope enclosing the bacteroids in soybean root nodules. J. Cell Biol. 78: 919-936.

Verma, D.P.S., Zogbi, V. and Bal A.K. (1978b). A cooperative action of plant and *Rhizobium* in dissolving host cell wall during development of symbiosis. Plant Sci. Lett. 13: 137-142.

Vincent, J.M. (1980). Factors controlling the legume-*Rhizobium* symbiosis. pp 103-129. In W.E. Newton and W.H. Orme-Johnson (eds.)Nitrogen Fixation, Vol 2 University Park Press, Baltimore.

Ward, L.J.H., Rockman, E.S., Ball, P., Jarvis, B.D.W. and Scott, D.B. (1989). Isolation and characterization of a *Rhizobium loti* gene required for effective nodulation of *Lotus pedunculatus*. Mol. Plant - Microbe Interactions 2: 224-232.

Weinman, J.J., Fellows, F.F., Gresshoff, P.M., Shine, J. and Scott, K.F. (1984). Structural analysis of the genes encoding the molybdenum-iron protein of nitrogenase in the *Parasponia Rhizobium* strain ANU289. Nucl. Acids Res. 12: 8329-8344.

Whitfeld, P.L., Seeburg, P.H. and Shine, J. (1982). The human proopiomelanocortin gene: Organisation, sequence and interspersion with repetitive DNA. DNA 1: 133-143.

Wilbur, W.J. and Lipman, D.J. (1983). Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. (USA) 80: 726-730.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

Zaat, S.A.J., Wijffelman, C.A., Spaink, H.P., Van Brussel, A.A.N., Okker, R.J.H. and Lugtenberg, B.J.J. (1987). Induction of the *nod*A promoter of *Rhizobium leguminosarum* Sym plasmid pRL1JI by plant flavanones and flavones. J. Bacteriol. 169: 198-204.