Characterisation of a *Rhizobium loti* nodulation mutant.

A thesis presented in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy
in molecular genetics at
Massey University

Julie Marie Collins-Emerson

1991

ABSTRACT.

The aim of the project was to characterise the *Rhizobium loti* Nod$^-$ Tn*5* mutant strain, PN233. The Tn*5* insertion had been previously localised to a 7.1 kb *Eco* RI chromosomal fragment. This fragment was sub-cloned and a *Bam* HI/*Sal* I endonuclease restriction map for the region was determined. *Hind* III digests were utilised to identify the approximate location of the Tn*5* 233 insertion and those of four other Tn*5* insertions (4016, 4019, 4047 and 4053) in the 7.1 kb region. The 233 mutation was found to map to a 1.45 kb *Sal* I fragment and that of an overlapping 2.8 kb *Bam* HI fragment.

The 7.1 kb *Eco* RI fragment and a larger 22.7 kb fragment that encompassed this region, had been cloned into pLAFR1. The construct carrying the 22.7 kb fragment (pPN305) was crossed into four *R.l.* bv. *trifolii* strains, each mutant in one of the four common *nod* genes, A,B,C, and D. The construct was able to complement the *nod*C mutation indicating the presence of a *nod*C gene somewhere on the 22.7 kb region.

The mutations 4047 and 4053 had been found to map to either side of the 233 Tn*5* insertion. Both insertions affected nodule formation and were thus included in further plant complementation tests. These experiments involved crossing both the pPN305 and a construct bearing the smaller 7.1 kb *Eco* RI fragment (pPN25) into the *R. loti* and *R.l.* bv. *trifolii* Tn5 mutants. What was unusual about the results was that, while the 7.1 kb fragment was able to complement the mutations, the larger 22.7 kb fragment which encompasses that region could complement PN4047 and PN4053 but was unable to complement the PN233 mutant.

The 2.8 kb *Bam* HI and 1.45 kb *Sal* I fragments, to which the 233 insertion was mapped, and that of an adjacent 1.2 kb *Sal* I fragment, were sub-cloned and then *Bal* 31 digested in both orientations to create a series of overlapping fragments. These fragments were then sequenced. The data revealed that the 233 Tn*5* had inserted into the *R. loti nod*C gene. It was determined that the 4047 Tn*5* was also located in this gene, slightly upstream of 233, while 4053 had inserted into

the 5'-region of *nod*I which is downstream of *nod*C. *Nod*A was identified upstream of *nod*C indicating an arrangement of common *nod* genes different from the conventional *nod*ABCIJ found in other rhizobia. The promoter for these *nod* genes, the *nod* box, was located upstream of the *nod*A gene.

A particularly puzzling aspect of the results is that, while PN4047 is complemented by both pPN305 and pPN25, PN233, which has an insertion in the same gene, could only be complemented by the smaller fragment carried by the pPN25 construct. To explain this result, it is proposed that PN233 is producing a mutant NodC protein and that this, in combination with doubled copies of a gene or genes present elsewhere on the 22.7 kb fragment, is responsible for interfering with complementation in this mutant. Alternatively, it may be that the imbalance of doubled copies of downstream, co-transcribed genes in the presence of one copy of a functional *nod*C gene causes complementation failure.

ACKNOWLEDGEMENTS.

Sincere thanks goes to my chief supervisor Assoc. Prof. Eric Terzaghi for his guidance, patience, enthusiasm and friendship and to Prof. Barry Scott for his advice and assistance and for providing the research topic.

I wish to express my gratitude to the Department of Microbiology and Genetics for their support. Also, many staff members have been generous in their friendship and given of their time in either a technical or advisory capacity, in particular Robert C., Carolyn, Ron, Dawn, Shirley and Trish, while Laura Medhurst and Steven Thomas worked on the computer diagrams. This has been much appreciated.

The encouragement and friendship shown by Mike, Trish, Carolyn, David, Wil, Robert H., Grant, Mark and Chris and many members of the MGU past and present is most valued.

My parents are thanked for their encouragement and support throughout my education (and Dad for his proof reading).

Finally, sincere thanks go to my husband Stephen for his friendship, support and love - a vital ingredient in the success and completion of the project.

TABLE OF CONTENTS. page no.

LIST OF FIGURES.

LIST OF TABLES.

1.0 INTRODUCTION.

Microbes play an important role in the biological cycles of carbon, oxygen, nitrogen and sulphur which are fundamental to life on this planet. While nitrogen is abundant on Earth, most of it is not readily utilisable by plants. Consequently, one of the factors that most limits global agricultural productivity is the availability of fixed nitrogen. The dinitrogen bond is very strong. Consequently, industrial processes directed towards fixing nitrogen require specialised conditions and are energy-intensive. However, many microbes in either a free-living state and/or in a symbiotic relationship with a plant are capable of fixing atmospheric nitrogen by converting it to ammonia via the following chemical conversions.

$$N_2 + 6e^- + 6H^+ + 12ATP \xrightarrow[\text{NITROGENASE}]{Mg^{2+}} 2NH_3 + 12ADP + 12Pi$$

At present, a vast amount of money and effort is expended on applying industrial nitrogenous fertiliser, which is mainly a product of the Haber process, to agricultural crops. The impact of industrial fertilisers on the environment and the economic reality that nitrogen fertiliser is beyond the reach of many countries, is of worldwide concern. Research into biological nitrogen fixation is therefore of considerable interest, not only academically, but because of the potential applications it may have for crop improvement and productivity.

1.1 DINITROGEN-FIXING PLANT - MICROSYMBIONT ASSOCIATIONS.

1.1.1 NON-LEGUMINOUS ASSOCIATIONS.

A number of dicotyledonous, non-leguminous plants from phylogenetically unrelated families and genera, form nitrogen-fixing root nodules in symbiotic relationships with endophytes that mainly belong to the order Actinomycetales and to the genus *Rhizobium*. The actinomycete endophytes of non-legumes can be placed in one family Frankiaceae, with a single genus *Frankia* (Becking, 1975). Examples of

some plant genera that have members which bear non-leguminous nodules formed by actinomycete-host symbioses are *Casuarina*, *Myrica*, *Alnus*, *Cerocarpus*, *Coriaria*, *Comptonia*, and *Colletia* (Becking, 1975; Bowes *et al.*, 1977; Callaham *et al.*, 1979). Some non-leguminous, nodule-bearing dicotyledonous hosts capable of symbioses with *Rhizobium* or supposed *Rhizobium* species are *Trema*, *Parasponia*, *Zygophyllum*, *Fagonia*, *Viola* and *Opuntia* (Becking, 1975).

Two morphological types of actinomycetous nodules are known: 1) *Alnus*-type and 2) *Myrica-Casuarina*-type (Torrey and Callaham, 1978). *Alnus*-type nodules are found in representatives of Belutaceae, Elaegnaceae, Rhamnaceae, Coriariaceae and Rosaceae. The nodules comprise modified, often dicotomously-branched roots of arrested growth which usually have a coralloid appearance (Becking, 1975). *Myrica/Casuarina*-type nodule lobes give rise to a normal root which is negatively geotropic (Torrey and Callaham, 1978).

Actinomycetes enter the host plant via root hair infection. These hairs curl on invasion by the actinomycete (Torrey and Callaham, 1978; Callaham *et al.*, 1979; Lalonde, 1980). The hyphae perforate the root cortical cells by local degradation of the cell walls and penetrate the host cell cytoplasm where it is then surrounded by host plasma membrane and a thick polysaccharide material termed the capsule. The hyphae branch extensively in specific layers of the cortex, penetrating most of the host cytoplasm (Newcomb *et al.*, 1978; Lalonde, 1980). These nodules are highly modified lateral branches, both in their origin and in their development (Becking, 1975; Torrey and Callaham, 1978).

1.1.2 LEGUMINOSEAE AND *RHIZOBIUM*.

Much research effort is focused on nitrogen fixation in leguminous plants, many of which are of immense agricultural significance both as pasture and as food crops, examples being clover, lucerne (alfalfa), peas, beans, soya beans and peanuts. This symbiotic relationship enables them to grow in nitrogen-deficient soils and hence they are also of ecological importance, in that they can be used

to reclaim poor and nitrogen-deficient land. There are about 18,000 species in the family Leguminoseae and over 90 percent of plants in the sub-families Mimosodeae and Papilionoideae bear highly specialised root nodules which provide the appropriate microaerobic conditions necessary for the nitrogen-fixing bacterial symbiont, *Rhizobium* (Vincent, 1982).

Members of the genus *Rhizobium* characteristically invade the roots of leguminous plants and produce root nodules. Taxonomic classification of *Rhizobium* tends to be based on the plant affinity (cross-inoculation group) concept. Until recently, fast (generation time of under 6 hours) and slow (generation time greater than 6 hours) -growing rhizobia were grouped in the same genus *Rhizobium*. However studies of numerical taxonomy, RNA cistron similarities, DNA base ratio determination, nucleic acid hybridisation, immunology, composition of extracellular polysaccharides, carbohydrate utilisation and metabolism, bacteriophage and antibiotic susceptibilities, protein composition and types of intracellular inclusion bodies in bacteroids (Vincent, 1977; Elkan, 1981; Jordan, 1982; Trinick, 1980) have supported a major division of the genus into the fast-growing *Rhizobium* and the slow-growing *Bradyrhizobium* sp. (Buchanan-Wollaston *et al.*, 1980; Jordan, 1982).

Fast-growing *R. japonicum* isolated from Asian-type soya beans have physiological characteristics similar to other fast-growing rhizobia, but their symbiotic properties are similar to the cowpea miscellany (Stowers and Eaglesham, 1984). These types of rhizobia have been grouped into a new species, *R. fredii* (Scholla and Elkan, 1984; Sadowsky *et al.*, 1987). Likewise, the fast-growing strains that nodulate *Lotus* sp. show low DNA:DNA homology (10-15%) both with other *Rhizobium* sp. and with slow-growing strains able to form symbiotic relationships with *Lotus* (Crow *et al.*, 1981). These fast-growers have been grouped into a new *Rhizobium* species, *R. loti* (Jarvis *et al.*, 1982). *R. leguminosarum*, *R. trifolii* and *R. phaseoli* are now categorised as different *R. leguminosarum* biovars, i.e. *R.l.* bv. *viciae*, *R.l.* bv. *trifolii* and *R.l.* bv. *phaseoli*. Table 1.1 lists host plants and their corresponding microsymbionts.

Table 1.1: Microsymbionts and their respective host plants .

Bacterium	host plant
Rhizobium meliloti	lucerne (alfalfa)
R. leguminosarum bv. *viciae*	pea, vetch
bv. *trifolii*	clover
bv. *phaseoli*	bean
Rhizobium fredii	soya bean
Bradyrhizobium japonicum	soya bean
Rhizobium loti	*Lotus*
Azorhizobium caulinodans	*Sesbania* (stem nodules)
Rhizobium NG234	*Parasponia* (a non-legume)
Bradyrhizobium sp. Parasponia	*Parasponia* (a non-legume)

1.2 THE BIOLOGY OF RHIZOBIAL NODULATION.

1.2.1 NODULE ONTOGENY.

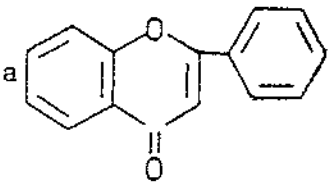
1.2.1.1 RHIZOBIAL COLONISATION OF ROOT HAIRS AND NODULE INITIATION.

The successful infection of the host plant by the microsymbiont is dependent upon a sequence of signals between plant and microbe. Plant root exudate will affect broth cultures of *Rhizobium* (Dazzo and Gardiol, 1984) while the culture filtrate of *Rhizobium* will induce root hair curling or branching (Yao and Vincent, 1976). This indicates that transfusable chemical signals play a role in early stages of colonisation and infection.

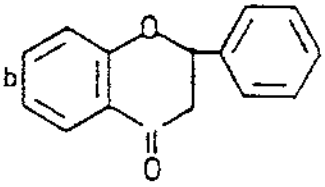
Gulash *et al.* (1984) observed that clouds of *R. meliloti* were attracted to localised sites on the surface of the infectible region of lucerne. Only *Rhizobium* strains formed these characteristic clouds suggesting that the microsymbionts were responding to a substance in the root exudate. Host plants exude flavonoid compounds that are derived from the phenylpropanoid biosynthetic pathway (reviewed by Rolfe and Gresshoff, 1988). These include flavones, flavànones, hydroxyflavanones and isoflavanones (Zaat *et al.*, 1989), (See Figure 1.1). These compounds can indirectly activate transcription of nodulation genes and play a role in host-specificity, as they can act as antagonists in non-homologous combinations. Khan *et al.* (1990) reported that isoflavones from a host plant can act as chemoattractants. It should be noted that flavonoids are closely related to some phytoalexins (see Figure 1.1), the plant-produced antibiotics synthesised upon invasion of the plant by micro-organisms. Parniske *et al.* (1990) found that *B. japonicum* was able to overcome the inhibiting effect of its host's phytoalexin, glyceollin, but not that of incompatible hosts. This area is of interest, as it is not known how the bacteria overcome the host defence system and get the plant to accept the invasion by the microsymbionts. Just as the plant communicates with the bacteria, via the production of an inducer which

Figure 1.1: The chemical structure of flavonoids and some phytoalexins.
a = flavones; b = flavonones; c = isoflavones (Firmin *et al.*, 1986); d = some bean phytoalexins (Deverall, 1982).

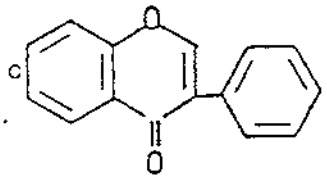
a

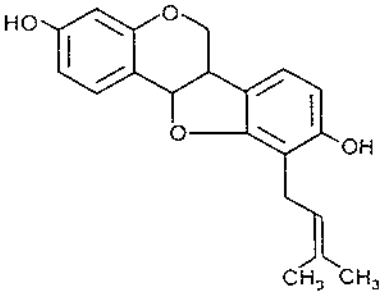


b

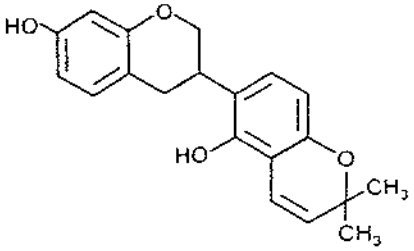


c



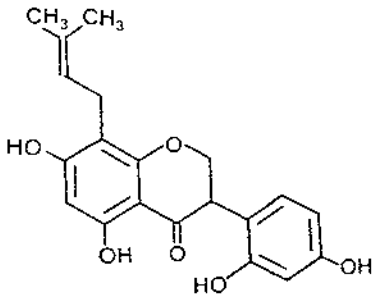


Phaseollidin



Phaseollinisoflavan

d



Kievitone

activates nodulation gene transcription, a signal molecule encoded by the *Rhizobium* early nodulation genes is important in establishing host specificity (Lerouge *et al.*, 1990). Colonisation of the root surface is the first step in the nodulation process.

One hypothesis concerning host-microsymbiont specificity is the "lectin recognition hypothesis", that is, that recognition between the legume and appropriate *Rhizobium* involves a binding of plant lectins to unique carbohydrates found exclusively on the *Rhizobium* cell surface (Dazzo and Gardiol, 1984). Lectins are found in many diverse eucaryotic and procaryotic organisms. They are divalent or multivalent carbohydrate-binding proteins that are grouped together because they agglutinate cells or other materials that display more than one saccharide of sufficient complementarity (Barondes, 1981). These molecules bind carbohydrate residues reversibly. Lectins can be glycosylated or non-glycosylated and are found in cells both as a membrane-bound and an unbound, freely-soluble form (Miller and Bowles, 1982). Though predominantly intracellular they are also detectable on cell surfaces or extracellularly and these molecules appear to interact with complementary saccharides on cell surfaces (Barondes, 1981).

The "lectin recognition hypothesis" suggests that the expression of host-specificity may involve preferential adsorption of rhizobia to root hairs (Bohlool and Schmidt, 1974; Dazzo *et al.*, 1976; Dazzo and Brill, 1978; Dazzo, 1980), by specific cross-bridging of antigenically related saccharide determinants on the surface of the bacterium and the cell wall of the host by multivalent host-encoded lectin (Dazzo *et al.*, 1979). However, soya beans that lack lectin can be nodulated by *R. japonicum* (Barondes, 1981). Also the lectin levels in the soya bean roots studied had fallen to undetectable levels about 2 weeks after planting but the highest rate of nodulation was 2-4 weeks after sowing (Barondes, 1981). It had been reported that some strains of rhizobia do not bind to lectins from the host that they nodulate (Dazzo and Hubbell, 1975b). This information tends not to support the lectin

hypothesis. However, some studies that suggest lectin is not involved with host-specifity have involved seed lectin and this can differ from root lectins (Law and Strijdom, 1977; Barondes, 1981; Dazzo and Gardiol, 1984).

Work by MacGregor and Alexander (1972) and Dazzo and Hubbell (1975a; 1975b) showed that there were differences in the extracellular polysaccharide (which includes capsular polysaccharides, CPS; exopolysaccharides, EPS; lipopolysaccharides, LPS and cyclic β-1,2-glucans [reviewed by Rolfe and Gresshoff, 1988]) and surface antigen composition between infective and non-infective strains of *Rhizobium*. This included an experiment that compared a non-nodulating mutant *Rhizobium* with its nodulating parental stain (MacGregor and Alexander, 1972). The complete lipopolysaccharide structure is required by *R.l.* bv. *phaseoli* strain CE3 to nodulate bean properly (Brink *et al.*, 1988) and mutants lacking the O-antigen lipopolysaccharide are defective in nodule development (Carlson *et al.*, 1988).

One lectin that has been investigated is trifolin, a clover lectin. This glycoprotein accumulates at clover root hair tips (Dazzo *et al.*, 1979). Fixed nitrogen (eg. NO_3^-, NH_4^-) in the rooting medium regulates the levels of trifolin on clover root hairs and hence the ability of rhizobia to adhere to these surfaces (Dazzo *et al.*, 1979). It is interesting to note that nodulation is regulated by the amount of fixed nitrogen available.

The growth phase of the rhizobia is another important factor affecting binding of bacteria to the root (Dazzo and Gardiol, 1984). Various groups have found that the capsular polysaccharides of rhizobia change with the age of the culture (Bhuvaneswari *et al.*, 1977; Dazzo *et al.*, 1979; Dazzo and Gardiol, 1984; Mikiko *et al.*, 1984) and that change affects their binding ability. White clover had more infected root hairs when inoculated with cells in early stationary phase than with cells in mid-exponential phase (Dazzo and Gardiol, 1984). *R.l.* bv. *trifolii* 0403, grown on agar plates for 5 days attached in greater numbers to clover root hairs than those grown

on plates for 3 or 7 days (Dazzo *et al.*, 1979). The above suggests that the lectin binding sites on rhizobia may be transient rather than constitutive components of the cell surface (Bhuvaneswari *et al.*, 1977; Dazzo and Truchet, 1984).

Host plants also appear to regulate infection. Clover root exudate has enzymatic activity that alters and erodes the bacterial capsule and this favours polar attachment via polar-lectin receptors (Dazzo *et al.*, 1982; Halverson and Stacey, 1986). Another example of plant products influencing infection is where flavones from the root washings of clover induce expression of genes required for nodulation by *R.l.* bv. *trifolii* (Redmond *et al.*, 1986). Likewise, luteolin, a flavone from lucerne, induces expression of *R. meliloti* nodulation genes (Peters *et al.*, 1986).

During the early stages of infection, which can be in a matter of minutes after inoculation (Turgeon and Bauer, 1982), the bacteria attach to the root hairs via a hapten-reversible interaction and then later become irreversibly anchored to the host cell. These steps are an early expression of host-specificity in the legume-*Rhizobium* symbiosis (Dazzo and Gardiol, 1984). There is some evidence that there are multiple mechanisms for rhizobial attachment to the host - some specific and some non-specific (Dazzo and Gardiol, 1984).

Bacteria can enter the host plant in one of three ways:

1) infection via infection thread and root hairs,

2) crack entry with no observable infection thread formation, and

3) crack entry with infection thread formation (Rolfe and Shine, 1984).

The first mode of entry is common and hence reasonably well investigated. The earliest visible sign that a successful infection may be proceeding is when the root hairs start to curl. Root hair curling can be caused by non-compatible host-microsymbiont associations (Li and Hubbell, 1969) indicating that this response is not host-specific. Also, the number of root hair infections is generally less than the number of curled root hairs for clover (Purchase, 1958). In most plants the area of infection is localised to

the cells above the zone of elongation and just below the position of the smallest root hairs, (Bhuvaneswari *et al.*, 1980; Pueppke, 1983), except in the mature root zone where infections can be initiated during lateral branching near the base of an older root hair (Nutman, 1959; Callaham and Torrey, 1981). The window of infectability of given host cells is therefore generally restricted to a few hours (Bhuvaneswari *et al.*, 1980).

The root hairs grow around the polarly-attached rhizobia (Vincent, 1980), and an infection thread develops from the most acutely curled region of the root hair (Bauer, 1981). In soya bean, root hair curling occurs within 12 hours and the infection thread is visible, under light microscopy, within 24 hours of inoculation (Turgeon and Bauer, 1982). Within 2 to 4 hours, infection threads that will give rise to nitrogen-fixing nodules apparently block nodule formation at adjacent sites on the root (Bauer, 1981), implying that there is a signal other than nitrogen production which controls nodule formation. Consequently, effective nodulation is a negatively self-regulating phenomenon (Nutman, 1981) by which a plant controls the number of nodules and avoids excessive energy expenditure on unnecessary nodule development (Verma and Nadler, 1984). The fact that ineffective strains of rhizobia will produce more nodules than effective strains, supports the presence of a feed-back inhibition system (Maier and Brill, 1976).

The curled root hair forms a "pocket" that traps the bacteria and creates a microenvironment conducive to production of wall "loosening" enzymes necessary for invasion (Callaham and Torrey, 1981; Verma and Nadler, 1984). These rhizobial enzymes dissolve the cell wall matrix leaving a layer of depolymerised wall microfibrils. Growth pressure from the bacterial colony "stretches" the weakened wall forming a bulge into an interfacial zone between the wall and plasmalemma (Ridge and Rolfe, 1985). The infection thread is initiated by an invagination of the root hair wall (Napoli and Hubbell, 1975). The apparent continuity of the cell and infection thread walls is due to a new wall

layer, which is deposited by the actively-streaming host cytoplasm, internal to the existing wall (Callaham and Torrey, 1981). This new wall is thicker and more fibrillar in appearance with a thickness of 0.5 um in clover (Callaham and Torrey, 1981).

Turgeon and Bauer (1982) observed that infected root hairs have increased cytoplasm, prominent nuclei and smaller vacuoles than uninfected root hairs. Prominent features of the cytoplasm surrounding the infection site are endoplasmic reticulum, occasional dictyosomes and numerous mitochondria (Callaham and Torrey, 1982). Vesicular bodies, similar to plasmalemmasomes, accumulate at the penetration site in a manner that parallels host-pathogen systems (Ridge and Rolfe, 1985). It has been suggested that the coated vesicles of the plant, which appear to arise from Golgi bodies, are involved in infection thread biosynthesis (Robertson *et al.*, 1978). In soya bean, it is not uncommon to find two infection threads in one root hair (Ranga Rao and Keister, 1978). Successful nodulation is generally achieved by one strain of *Rhizobium* (Dart, 1974), although nodules can be occupied by more than one strain. Non-nodulating mutants, when mixed with efficient strains, can nodulate plants by "sneaking" into the root along with the invading partner (Trinick, 1982). The infection thread grows down the length of the root hair, through the epidermis and ramifies through the cortex. When the infection thread crosses from one cell to another in the cortex, there appears to be a repetition of the infection site invagination process (Napoli and Hubbell, 1975).

The peanut, *Arachis hypogaea*, is an example of the crack entry mode of invasion where no infection thread is observed. Nodules on these plants arise solely at root junctions (Dart, 1975), the only place where root hairs exist on their hosts (Chandler, 1978). Rhizobia enter the plant where the root hairs emerge and occupy the place between the root hair wall and the adjoining epidermal and cortical cells, hence the term "crack entry". The bacteria multiply and then invade the plant cells. The invaded basal cells divide and, with repeated divisions, become smaller and become incorporated into the nodule tissue (Chandler, 1978).

While the term "*Rhizobium*-legume symbiosis" has become synonymous with the term "root nodule symbiosis", it is important to note that nodulation and symbiotic nitrogen-fixation also occurs on the stem of some legumes. Three genera of the family Leguminoseae have been reported to form stem nodules: *Aeschynomene* and *Sesbania* of the subfamily Papilionoideae, and *Neptunia*, of the Mimisoideae subfamily (Legocki and Szalay, 1984). The initial mode of infection in these cases remains unknown. Tsien *et al.* (1983) showed the presence of branched intercellular infection threads in *Sesbania rostrata;* however no infection threads have been found in *Aeschynomene* plants to date (Legocki and Szalay, 1984).

1.2.1.2 FURTHER NODULE DEVELOPMENT.

In many species studied it is observed that increases in cortical cytoplasm and in mitotic division occur in advance of the penetrating infection thread (Dart, 1975; Newcomb *et al.*, 1979; Turgeon and Bauer, 1982; Verma and Nadler, 1984). This suggests that a diffusible substance produced by the infection thread may stimulate the cortical cells to divide. Pure cultures of *Rhizobium* have been reported to produce cytokinins and indole acetic acid (IAA) (Dart, 1975; Phillips and Torrey, 1972; Verma and Nadler, 1984) which stimulate cell division. It is therefore possible that alterations in the balance of these compounds may be responsible for the increased cortical division. Also, root nodules have been found to be rich in auxins (Lie, 1964). The presence of a diffusible molecule is also supported by experiments where the early stages of cortical division are initiated by rhizobial mutants which fail to cause root hair curling and infection thread formation (Knight *et al.*, 1986). Recently an alfalfa-specific nodulation signal molecule designated NodRm-1 has been identified and has been shown to stimulate root hair curling and cell division (Lerouge *et al.*, 1990).

Figure 1.2: The structure of determinate and indeterminate nodules. (Sutton, 1983).

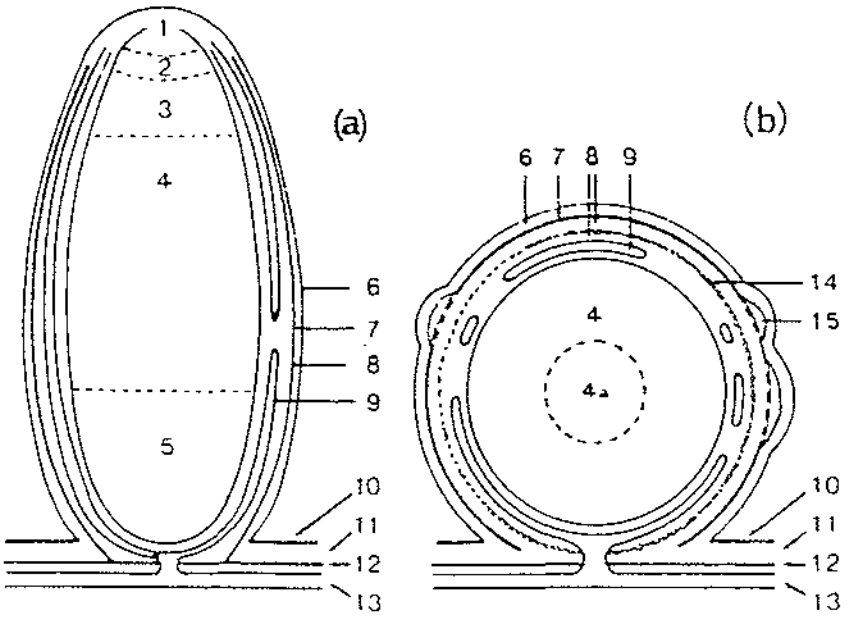
(a) Indeterminate (apical) nodule

(b) Determinate (spherical) nodule

1. nodule meristem
2. zone of infection thread growth and cell penetration
3. zone of expanding infected cells
4. zone of mature bacteroid-containing cells

4a. zone where senescence usually commences

5. senescent bacteroid-containing tissue
6. outer cortex
7. nodule endodermis, continuous with root endodermis at the nodule base in (a); interrupted over the vascular bundles and at the nodule base in (b)
8. inner cortex
9. vascular bundle (xylem, phloem, pericycle and bundle endodermis) continuous with root vascular tissue and endodermis at nodule base
10. root epidermis
11. root cortex
12. root endodermis
13. root xylem and phloem
14. sclerenchyma
15. phellogen tissue overlaying vascular traces.


(a)
1
2
3
4
5
6
7
8
9
10
11
12
13
(b)
6 7 8 9
14
15
4
4a
10
11
12
13

An alternative hypothesis to that of the diffusible division-inducing signal is that the invading bacteria alter pre-existing positional gradients that control the placings of meristematic activity (as proposed by Warren Wilson, 1978).

The exact nature of the cells which respond to the division stimulus is unknown. Wipf and Cooper (1940) reported tetraploid cells in the cortex of a number of leguminous plants. Nuclei containing twice the normal DNA content have been found in white clover root cortical cells (Gresshoff and Mohapatra, 1981) and polyploid mitoses in the root cortex are characteristic of developing pea nodules (Phillips and Torrey, 1970). Explants of pea cortical tissue have shown that tetraploid cells are stimulated to divide at lower cytokinin levels than diploid cells (Nagl and Rucker, 1974) and this lends support to the hypothesis that the tetraploid cells may act as "target" tissues for nodule initiation. That is, cytokinin and auxins which the rhizobia produce stimulate these cells to divide and hence initiate meristematic activity (Wipf and Cooper, 1940). Alternatively, the invading rhizobia may stimulate diploid cells to become polyploid (Libbenga and Bogers, 1974; Truchet, 1978). Neither of these hypotheses have been conclusively disproved.

Leguminous root nodules tend to be initiated in the cortex (Libbenga and Bogers, 1974; Dart, 1975; Chandler, 1978; Newcomb, *et al.*, 1979; Newcomb, 1980; Bauer, 1981) or possibly in the pericycle in some species (Dart, 1975; Collins, 1983). Cell divisions, due to the presence of an infection thread, are usually opposite the protoxylem points in the root stele (Phillips, 1971; Sprent, 1979; Collins, 1983). Laterals arise from the same position lending support to the hypothesis that both nodules and laterals arise from the same "Anlage" and then bifurcate early in development. This would indicate that the sum total of nodule and lateral initials for a plant would be relatively constant, a feature noticed by Nutman (1948) in red clover. However, laterals always arise opposite protoxylem poles whereas nodules need not (Sprent, 1979), so the exact relationship between nodules and laterals, if any, remains debatable.

The overall direction of the infection thread is towards the stele. In the case of the pea plant, the infection thread passes through cortical cell walls and the intercellular spaces (Newcomb, 1980). Cell division is stimulated and this results in the formation of a nodule meristem. The meristem is characterised by isodiametric cells having densely staining cytoplasm, many small vacuoles, a prominent nucleus (Newcomb, 1976), which often has several nucleoli, a few mitochondria, numerous plastids and free ribosomes and relatively little rough endoplasmic reticulum. Dependent upon the arrangement of the meristem, the nodule formed is either indeterminate (apical meristem) as in pea and clover, or determinate (peripheral meristem), examples being soya bean and siratro nodules (see Figure 1.2). Slow-growing rhizobia tend to infect plants that produce determinate nodules; temperate legumes that develop indeterminate nodules tend to be invaded by fast-growing rhizobia (reviewed by Ludwig and de Vries, 1986). The size and shape of the nodule is determined by the host (Newcomb *et al.*, 1979; Sutton, 1983). In fact, some alfalfa plants can form non-nitrogen fixing structures that possess all the features characteristic of indeterminate nodules (Truchet *et al.*, 1989), indicating that the host controls nodule structure. The nodule primordium continues to grow as more cortical layers are incorporated into the nodule. The peripheral cortex of the nodule, the endodermis and the pericycle are not invaded by the bacteria (Libbenga and Bogers, 1974). The infection thread ramifies through the meristematic zone. At no time during the course of a normal infection are the bacteria "naked" in the root hair cytoplasm (Vincent, 1980). They are surrounded by a large amount of matrix (Vincent, 1980), possibly polysaccharide in nature, in which they continue to divide.

In spherical nodules the meristematic activity is restricted to a cambium-like layer 2-3 cell layers wide surrounding the bacteroid zone. Other meristematic regions appear in the cortex and these give rise to the vascular traces (Dart, 1977). In plants that have no infection thread, rhizobia appear to be dispersed entirely by cell division. In soya bean, (*Glycine max*), initial cortical divisions produce an outer globular mass of cytoplasm-rich cells and an inner

region of vacuolated cells adjacent to the stele. The infection threads invade the outer mass and some of them penetrate actively dividing cells. By 12-18 days mitotic activity is mostly finished and any subsequent increase in the bacteroid zone is due to cell enlargement (Sutton, 1983).

In indeterminate nodules the infected cells do not divide, the bacteria being spread by the ramification of the infection thread into new cells formed by meristematic activity (Sutton, 1983). As the initial meristematic centre is lost to bacteroid formation, new meristematic activity is initiated in neighbouring cortical tissue (Libbenga and Harkes, 1973). As the development in indeterminate nodules proceeds, more cortical layers contribute to the nodule. The peripheral layer and apical meristem of the nodule are not invaded (Libbenga and Harkes, 1973).

Not all cells in the bacteroid zone are invaded or contain rhizobia. These cells link up to form channels from the cortex to the bacteroid tissue. In determinate nodules, these cells form radial files whereas in indeterminate nodules they tend to form longitudinal files. There are few uninvaded cells in nodules in which rhizobia are dispersed solely through plant cell division (Dart, 1977).

The mature root nodule has an uninvaded cortex of very large, vacuolated, thick-walled cells. Within it lies a nodule endodermis connected to the root endodermis and, internal to this, there is a layer of scleroids and several vascular traces. A periderm layer may be present outside the endodermis. The nodule endodermis surrounds the developing nodule and merges with the meristem at the apex of cylindrical nodules and usually has plant cells with thickened suberised walls (Dart, 1975).

Vascular traces differentiate in the root cortex just after the bacteria have dispersed through the nodule initial (Dart, 1977). The vascular bundles are initiated by division of cortical cells (Dart, 1977) or endodermal and pericyclic cells (Sutton, 1983) adjacent to the middle of the nodule and in a direction mainly parallel to the root radius, forming a procambial strand that develops toward the root

stele. Protoxylem and protophloem cells elongate radially in rows and the xylem cells form a secondary thickening (Sutton, 1983). Differentiation of the cells in the vascular bundles is generally acropetal beginning at the root vascular connection. Nodules may have one to five vascular traces, the number being dependent on the plant species (Dart, 1975). They are mostly connected to one protoxylem point in the root. The vascular traces branch dicotomously at the base of the nodule. Further branching occurs in the nodule cortex. In determinate nodules the vascular bundles may anastomose at the distal end of the nodule but not in indeterminate nodules (Dart, 1977; Sutton, 1983).

Each trace has its own endodermis, with Casparian thickening, that is joined to the root endodermis. Surrounding the phloem and xylem are pericycle cells, some of which have been reported to develop into transfer cells (Dart, 1975). Newcomb and Peterson (1979), however, reported that it is the xylem parenchyma cells of the root tissue adjacent to the nodules that develop into the transfer cells. These specialised cells are characterised by ingrowths of the cell wall. The ingrowths are located adjacent to the pits in the xylem elements. The ontogeny and time of development of transfer cells vary in different species. These cells export amino acids from the nodule to the vascular bundle (Newcomb and Peterson, 1979) and mediate the symplastic flow of sugars from the phloem to the bacteroid tissues (Dart, 1975).

The release of bacteria from the infection thread usually commences with the formation of a vesicle at the terminus of the infection thread (Dart, 1977). Rhizobia are released into the cell by a process which resembles endocytosis. Individual bacteria embed in the infection thread membrane and are "budded-off" into individual membrane vesicles - the peribacteroid membranes (Sutton, 1983). The bacteria are never naked within the plant cell cytoplasm as they are surrounded by the infection thread or the peribacteroid membrane; both of host cell origin. It is possible that this represents a modified defence response. In some nodules, such as those of cowpea and soya

bean, the host cells enlarge and this is accompanied by divisions of the bacteroids within the peribacteroid membrane. In others, such as clover and alfalfa, each bacterium enlarges and remains enclosed singly in the membrane. Bacteroids have decreased DNA content, fewer ribosomes and changed cytochromes (which allows for respiration at the lower O_2 concentrations within nodules) and cell wall composition (Bergersen, 1974).

The first outward sign of senescence is when the nodule, which is usually red or pink due to the pigment leghaemoglobin, turns to a brown, grey or green colour, representing a breakdown of the nodule pigment. In indeterminate nodules senescent tissue is found at the base of all but the youngest nodules. This zone advances into the younger symbiotic tissue laid down by the apical meristem. Under favourable conditions, the senescent tissue may regenerate providing the meristem has not yet degenerated (Sutton, 1983). In determinate nodules, senescence commences at the centre of the organ and continues to grow out towards the periphery of the nodule (Dart, 1975). In *A. hypogaea* a layer of suberised cells develops at the base of the nodule and proceeds to sever the vascular traces to the nodule resulting in senescence. The suberised layer also forms a seal on the root surface after nodule degeneration is complete (Sutton, 1983).

During senescence the bacteroid tissue breaks down to form zoogloeal masses around the plant nuclei. The bacteroids decompose to granules and finally disappear while rod-shaped bacteria proliferate in the intercellular spaces in *M. sativa*, *T. repens* and *P. sativum* (Sutton, 1983). Electron microscopic observations of *P. sativum* showed that multiple vacuoles appeared in senescing bacteroids followed by the disintegration of both host cell and bacteroid structures (Sutton, 1983). Early signs of degeneration are irregular contraction of the nucleus and clumping of nucleoplasm, fragmentation and vesicularisation of the endoplasmic reticulum, reduced numbers of ribosomes, disappearance of starch granules and swelling of some mitochondria. The tonoplasts disappear followed by the disintegration of the entire contents of the host cell (Sutton, 1983).

As can been seen from the information outlined above, the development of a nitrogen-fixing nodule requires close communication and co-ordination between the plant and bacterium. The genetics of nodule formation and nitrogen fixation are discussed in the following sections.

1.3 THE GENETICS OF NODULATION.

With the exception of *R. loti* (Pankhurst *et al.*, 1983; Chua *et al.*, 1985), *B. japonicum* (Long, 1989) and *R. fredii* (Stanley *et al.*, 1986), whose nodulation (*nod*) and nitrogenase structural genes (*nif*) are, or appear to be chromosomally borne, the *nod*, *nif* and symbiotic nitrogen-fixing (*fix*) genes in most rhizobia are located on large indigenous plasmids (Denarie *et al.*, 1981; Hooykaas *et al.*, 1981; Sadowsky and Bohlool, 1983). These are commonly referred to as the Sym (symbiotic) plasmids (Zurkowski and Lorkiewicz, 1979; Wijffelman *et al.*, 1985).

1.3.1 NODULATION GENES.

1.3.1.1 COMMON *NOD* GENES.

There are a number of highly conserved *nod* genes that appear to be functionally interchangeable among rhizobia (Fisher *et al.*, 1985; Long, 1989). These are called the "common" *nod* genes and include *nod*ABC and D. *Nod*ABC are involved in root hair curling (Debelle *et al.*, 1986) and stimulation of cell division in the plant root (Dudley *et al.*, 1987).

The NodA and NodB products, which are predicted to be M_r 21,778 (196 aa) and M_r 23,798 (217 aa) in *R. meliloti* (Torok *et al.*, 1984; Egelhoff *et al.*, 1985) respectively, are thought to be located within the cytosol (Schmidt *et al.*, 1986). Experiments on NodC, which is calculated to be M_r 46,768 (426 aa) in *R. meliloti* (Torok *et al.*, 1984), as well as deductions from the amino acid sequence, suggest that it is a transmembrane protein (John *et al.*, 1988). In *R.*

meliloti, *nod*C is known to be involved with the production of a signal molecule (Lerouge *et al.*, 1990). As the *nod*ABC genes are all highly conserved and readily interchangeable between rhizobia, they must perform a fundamental function in the nodulation process.

*Nod*IJ are located downstream of *nod*ABC. *Nod*IJ mutations in *R.l.* bv. *viciae* cause a slight delay in nodulation followed by an unusually large number of curled root hairs and infection threads (Debelle *et al.*, 1986) and exaggerated hair curling (Hac^{++}) (Innes *et al.*, 1985). Mutations in *nod*IJ in *B. japonicum* have also been reported to almost abolish chemotaxis expressed in the presence of daidzen and genistein, the host plant flavonoids (Khan *et al.*, 1990). Based on homology studies, NodIJ proteins may be involved in a membrane transport system (Evans and Downie, 1986). Given the close proximity of *nod*ABCIJ, and the polar effect of Tn*5* mutations, these genes are thought to comprise an operon (Marx, 1985; Schofield and Watson, 1986). The stop codon for *nod*A and the start codon for *nod*B, overlap in *R.l.* bv. *trifolii* and *R. meliloti*, i.e. ATGA, suggesting that translational coupling may occur in these organisms (Schofield and Watson, 1986; Little *et al.*, 1989). Genes which exhibit such overlapping of stop and start codons often code for polypeptides that function as a complex, or otherwise, are those that need to be closely co-ordinated. It is possible that translation coupling ensures the correct concentrations of such peptides in the cell (Normark *et al.*, 1983).

With the exception of *nod*D, *nod* genes require the NodD protein and the presence of plant inducers (flavonoids) to be expressed (Mulligan and Long, 1985; Lugtenberg *et al.*, 1988). *Nod*D is constitutively expressed in free-living rhizobia (reviewed by Windsor, 1989) and also autoregulates its own expression in *R.l.* bv. *viciae* (Rossen *et al.*, 1985). Like *nod*IJ, *nod*D in *B. japonicum* has been implicated in chemotactic responses to flavonoids (Khan *et al.*, 1990). Three copies of *nod* D (D_1, D_2 and D_3) have been identified in *R. meliloti* (Honma and Ausubel, 1987), however, these do not appear to be autoregulated (Hong *et al.*, 1987) and it is unknown whether *nod*D_2 and *nod*D_3 are constitutively expressed (Windsor, 1989). Nod D_1, D_2 and D_3

differ in their response to pure flavonoids. They will all induce *nod*ABC and host-specificity genes, but activate expression in conjunction with different flavonoid specificity (Gyorgypal *et al.*, 1988). Applebaum *et al.* (1988) reported that *B. japonicum nodD*$_1$ could complement a *R.l.* bv. *trifolii nod*D mutation but that *B. japonicum nodD*$_2$ could not. In another example, a *R. meliloti nod*D gene was capable of complementing a *R.l.* bv. *trifolii nod*D mutant enabling nodulation of *T. repens* but not *T. pratense* (Spaink *et al.*, 1987b). So, while *nod*D has been designated a "common" *nod* gene, it is important to note that the different *nod*Ds play a role in host-specificity.

*Nod*D is read divergently from *nod*ABCIJ in *R. meliloti* and *R.l.* bv. *viciae*; the open reading frame (ORF) codes for an unmodified protein of M_r 34,879 (308 aa) in *R. meliloti* (Egelhoff *et al.*, 1985). Sequence analysis suggests that this protein belongs to a family of transcription-activating proteins (Henikoff *et al.*, 1988). This is consistent with the finding that NodD binds to the *nod* box (see section 1.3.2) and, together with the plant inducers, regulates the expression of *nod* genes read downstream from the *nod* box (Hong *et al.*, 1987; Kondorosi *et al.*, 1989).

1.3.1.2 HOST-SPECIFICITY GENES.

*Nod*EFGHQP are the host-specificity genes identified in *R. meliloti*, while *nod*EFLMNT (and *nod*X in some strains) are those found in *R.l.* bv. *viciae* to date (Long and Ehrhardt, 1989). Although there is a high degree of similarity between *nod*EF in *R.l.* bv. *viciae* and *R. meliloti*, they cannot complement each other and therefore are not considered "common" *nod* genes (Debelle and Sharma, 1986). *Nod*EF, and possibly *nod*G, belong to the same operon in *R. meliloti*. As the stop and start codons for *nod*E and *nod*F overlap, it is thought that they, like *nod*A and *nod*B, are translationally coupled. The *nod*EFG genes are involved with infection thread development (Debelle *et al.*, 1986). However, Tn*5* mutations in *nod*EF have been reported to be "leaky" in expression, so it is possible that they may not be totally

essential for nodulation (Swanson *et al.*, 1987). *Nod*EFG are predicted to be M_r 41,826 (402 aa), M_r 9,770 (93 aa) and M_r 26,087 (245 aa) respectively (Fisher *et al.*, 1987b). Sequence homology studies have highlighted the similarity of the *nod*G product to the ribitol dehydrogenase protein in *Klebsiella aerogenes* (Debelle and Sharma, 1986), while NodE shares similarity with a β-ketoacyl synthase (Downie, 1989; Sherman *et al.*, 1989) and NodF with acyl-carrier proteins (Shearman *et al.*, 1986).

*Nod*H is a host-specificity gene apparently unique to *R. meliloti* (Long and Ehrhardt, 1989). It is predicted to be M_r 28,854 (247 aa) (Fisher *et al.*, 1987b) and inhibits nodulation in heterologous hosts (Debelle *et al.*, 1986). It appears that the NodH product could be enzymatically involved in modifying the *nod*ABC product (Faucher *et al.*, 1988; Banfalvi and Kondorosi, 1989). Two newly identified *nod* genes, Q and P are also associated with host-specificity functions in *R. meliloti*. They are reported to have ATP sulphurylase activity and it is proposed that the genes synthesise an activated sulphate that is an intermediate in the formation of the alfalfa-specific sulphated NodRm-1 factor (Schwedock and Long, 1990).

*Nod*LMN are the best characterised host-specificity genes in *R.l.* bv. *viciae* (Long, 1989). *Nod*MN, together with *nod*T, form an operon which is preceded by a *nod* box. *Nod*L is in the same operon as the *nod*FE genes (Surin and Downie, 1988). Based on sequence comparisons, it is thought that the NodL protein may function as an acetyl-transferase (Downie, 1989). While a mutation in *nod*L will strongly reduce nodulation in peas, *Lens* and *Lathyrus*, Surin and Downie (1988) reported that it had little effect on nodulation of *Vicia* species. *Nod*M has been reported to bear a resemblance to an amido-phosphoribosyl transferase and may therefore modify EPS sugar residues with amidos (Weinman *et al.*, 1988). Another gene involved in host-specificity is *nodX*, found in some strains of *R.l.* bv. *viciae* and required for the nodulation of Afghanistan peas (Gotz *et al.*, 1985).

The function of the newly-identified *nod*O (de Maagd *et al.*, 1989) and *nod*T (Canter Cremers *et al.*, 1989) genes in *R.l.* bv. *viciae* is

unclear. It is known however, that the NodO protein does share homology in part with the haemolysin protein (HlyA) of *E. coli* and is thought to be secreted. This gene is preceded by a poorly conserved *nod* box. NodT has a signal peptide similar to those of proteins that are destined for the outer membrane (Economou *et al.*, 1990). Two different open reading frames are found between the *nod* box and *nodA* in *Bradyrhizobium* sp. (*Parasponia*) (Scott, 1986) and *B. japonicum* (Nieuwkoop *et al.*, 1987) and designated *nodK* and *nodY* respectively. It is not known whether these ORFs are translated, and if so, what roles the genes play. The arrangement of *nod* genes in *R. meliloti* and *R.l.* bv. *viciae* are shown in Figure 1.3.

1.3.2 THE *NOD* BOX.

A highly conserved, reiterated sequence is located upstream of the inducible *nod*ABCIJ gene cluster and those of the host-specificity genes. This 49 bp sequence is referred to as the *nod* box and is involved with co-ordination of *nod* gene expression (Schofield and Watson, 1986). The conserved region comprises subsequences of 7 bp, 5 bp and 25 bp (Rostas *et al.*, 1986). Spaink *et al.* (1987a) compared a number of *nod* boxes from different rhizobia and, based on that analysis, arrived at the following consensus sequence:

YATCCAY..YUYUGATG....Y.ATC.AAACAATCUATTTTACCAATCY

(Y = pyrimidine; U = purine; . = any base or base pair deletion)

Futhermore, Wang and Stacey (1990) suggest that *nod* boxes appear to comprise a repetitive sequence structure that is outlined below.

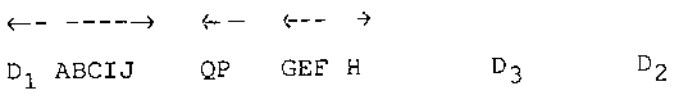
```
   GC AA A    GC AA A       GC AA A    GC AA A
-AT  A  Y -AT  A  Y -N4-AT  A  Y -AT  A  Y -
   CG TT T    CG TT T       CG TT T    CG TT T
```

Y = pyrimidine; N = any base; - separates the repeated units.

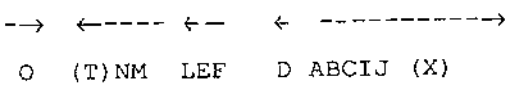
Spaink *et al.* (1987a) reported that the inducible promoter of *nod*A in *R.l.* bv. *viciae*, was contained within a fragment 72 bp long that carried the *nod* box and the loosely conserved sequence, AT(T)AG, approximately 13 bp downstream of the consensus sequence. The *nod* box is a *cis*-regulatory element (reviewed by Windsor, 1989). As mentioned previously, the *trans*-acting NodD product binds to the *nod* box and induces plant flavonoid-dependent transcription of the other *nod* loci

Figure 1.3: The arrangement of *nod* genes in *R. meliloti* and *R.l.* bv. *viciae* (de Maagd *et al.*, 1989; reviewed by Long, 1989). ----→ represents known direction of transcription and co-transcribed genes; (T) is an ORF; (X) is found in *R.l.* bv. *viciae* strain TOM.

R. meliloti



R.l. bv. *viciae*



(Horvath *et al.*, 1987; reviewed by Windsor, 1989). In *R. meliloti*, *nod* gene expression has been found also to be under the control of a repressor protein. This protein was found to bind to the DNA downstream of the *nod* box in the region containing the *nodD* and *nodA* transcriptional start sites (Kondorosi *et al.*, 1989).

1.3.3 THE ROLE OF EXTRACELLULAR POLYSACCHARIDES IN NODULATION.

Wild-type rhizobia often produce large quantities of extracellular polysaccharides. These comprise the neutral β-1,2-glucans and acidic exopolysaccharides (EPSs), which are attached to the outer membrane surface, and the lipopolysaccharides (LPSs) which form part of the outer membrane and contain the O-somatic antigens (Gray and Rolfe, 1990). The various rhizobial species produce differently structured EPSs which comprise oligosaccharide repeat units. The synthesis of these units is thought to take place at the inner membrane surface (Sutherland, 1985).

EPS, LPS and β-1,2-glucan are involved in the formation or maintenance of infection threads in a number of nodule-forming symbioses (reviewed by Martinez *et al.*, 1990). It has been reported that EPS is required for nodulation of indeterminate type nodules, though unnecessary for determinate type ontogeny. For example, a *R.l.* bv. *phaseoli* Exo^- mutant could produce fully effective determinate nodules on beans, however, the same mutations in *R.l.* bv. *viciae* and *R.l.* bv. *trifolii* formed defective indeterminate nodules on peas and clovers respectively (Gray and Rolfe, 1990). These observations suggest that EPSs play an important role in the development of indeterminate nodules but are not required for determinate nodule ontogeny. About 16 exopolysaccharide-related genes (exo) have been identified in *R. meliloti* and many have yet to be extensively characterised (reviewed by Martinez *et al.*, 1990).

1.3.4 NITROGEN FIXATION GENES.

Nodule development and the maintenance of symbiotic nitrogen fixation involves a complex interplay of signals between host and

rhizobia. Genes involved with nitrogen fixation in rhizobia are usually divided into two groups; *nif* and *fix* genes (Long, 1989). Those *Rhizobium* genes which have homologues in free-living nitrogen-fixing organisms like *Klebsiella pneumoniae* are referred to as *nif* genes and, by convention, these genes share the same name as their homologues in *K. pneumoniae*. Those genes whose role is not known to be analogous to a free-living function are designated as *fix* genes (Kondorosi, 1986). A map of the *nif* region in *K. pneumoniae* M5a1 and the functions of the genes are shown in Figure 1.4 and Table 1.2 respectively while Figure 1.5 illustrates the arrangement of *fix* and *nif* genes in *R. meliloti* and *B. japonicum*.

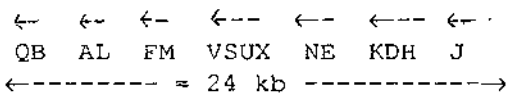
Fix^- rhizobia have been identified with mutants defective in haem biosynthesis, ammonia assimilation, carbohydrate metabolism, nitrate reduction, glutamine synthetase (Kondorosi, 1986) and cell surface components required for the production of signals that induce late nodulin expression (Kondorosi *et al.*, 1988a). Two *R. meliloti fix* genes whose function has been determined are *fix*LJ, which regulate the expression of other *nif* and *fix* genes (Boistard *et al.*, 1988). In *B. japonicum* and *R. meliloti* the regulation of *nif* and *fix* genes occurs by a cascade of two consecutive gene activation steps (Hennecke *et al.*, 1988a; Kahn *et al.*, 1988). (See Figure 1.6).

1.3.5 NODULINS.

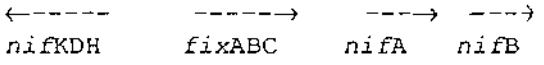
Nodulins are plant nodule-specific gene products, mostly expressed after the initiation of nodule development (reviewed by Rolfe and Gresshoff, 1988). Early nodulins are involved in the infection process and nodule morphogenesis (Schell *et al.*, 1988; Scheres *et al.*, 1990) and those studied so far appear to be (hydroxy)proline rich proteins that may be cell wall constituents (Gloudemans *et al.*, 1988). In more mature nodules, late nodulins that are activated at various stages such as, the release of bacteria into the cell and the induction of metabolites, play a role in nitrogen fixation and assimilation. Examples of the latter category being leghaemoglobin, uricase, glutamine synthetase and sucrose synthetase production (Mauro *et al.*, 1985; Lara *et al.*, 1988).

Figure 1.4: A map of the *nif* region of *K. pneumoniae* M5al (Postgate, 1982).
←-- represents the direction of transcription and co-transcribed genes.

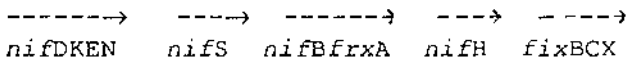
Figure 1.5: The arrangement of *fix* and *nif* genes in *R. meliloti* (Ausubel *et al.*, 1985) and *B. japonicum* (Hennecke *et al.*, 1988a).



R. meliloti



B. japonicum



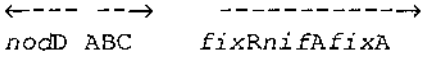
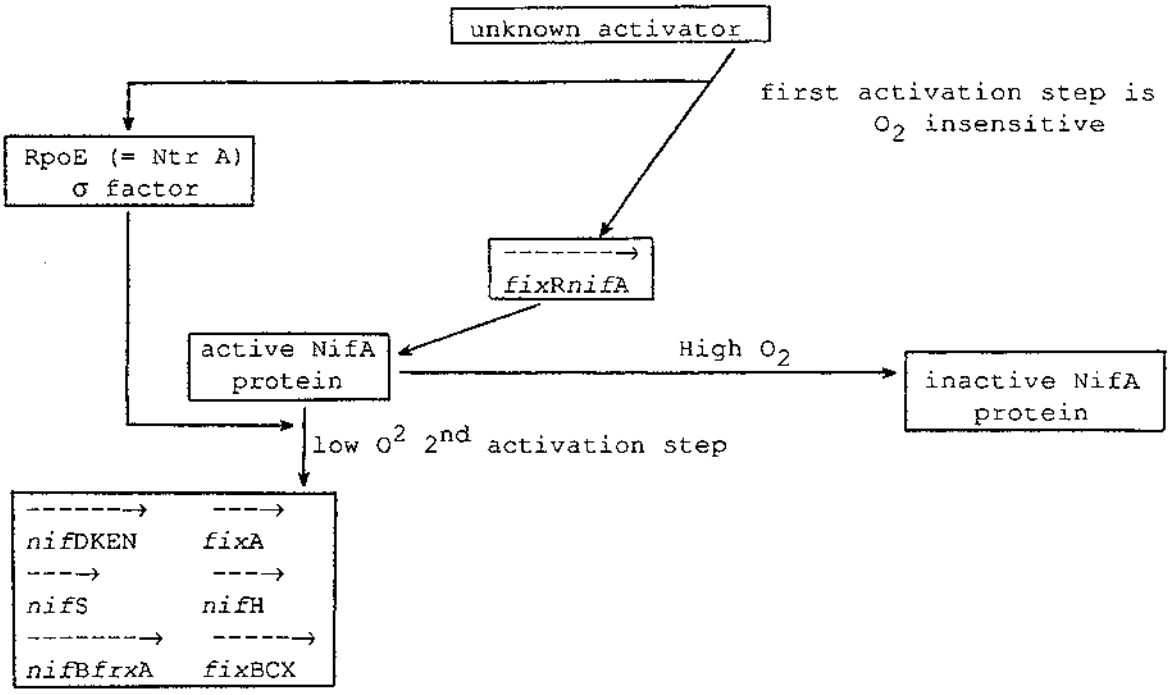


Table 1.2: The functions of *K. pneumoniae nif* genes. (Postgate, 1982; reviewed by Martinez *et al.*, 1990).

Q	unknown	V	processes NifKD protein
B	involved in processing the nitrogenase protein product of *nif*KD	U	unknown
		X	unknown
		N	processing NifKD
A	regulates expression of many of the *nif* and *fix* genes	E	processes NifKD
		Y	unknown
L	regulatory	K	codes for β-subunit of NifKD
M	processing the protein product of *nif*H	D	codes for α-subunit of NifKD
		H	codes for subunit of NifH
S	synthesises iron molybdenum co-factor	J	electron input into nitrogenase

Figure 1.6: The regulation of *nif* and *fix* genes in *B. japonicum* and *R. meliloti* (Hennecke et *al.*, 1988a).


unknown activator
first activation step is
O_2 insensitive
RpoE (= Ntr A)
σ factor
fixRnifA
active NifA
protein
High O_2
inactive NifA
protein
low O^2 2nd activation step
nifDKEN
fixA
nifS
nifH
nifBfrxA
fixBCX

1.4 NITROGEN FIXATION - THE PHYSICAL AND BIOCHEMICAL ENVIRONMENT.

Nitrogen fixation resulting from the symbiotic relationship between legumes and rhizobia is the greatest net contributor of fixed nitrogen to the biosphere (Ludwig and de Vries, 1986). Because dinitrogen is a very stable molecule, a great deal of energy is required for its conversion to ammonium (NH_4^+), and in addition, the rhizobial symbiont requires that certain conditions be met. Some of the requirements of N_2 fixation include a low reducing potential, several ATPs for each NH_4^+ molecule formed, a mechanism for protecting the oxygen-sensitive nitrogenase from damage, and a complex of metalloproteins (Brill, 1979). Due to the high energy cost of this reaction, N_2-fixation is tightly controlled and there is cessation of production when NH_4^+ is no longer needed. The reaction is subject to feedback inhibition (Brill, 1979).

1.4.1 THE PHYSICAL ENVIRONMENT OF NITROGEN FIXATION.

Nitrogen fixation is carried out within the bacteroid, a differentiated form of the bacterium, which is enveloped by the peribacteroid membrane. Apart from providing nutrients and protection for the bacterium, the structure of the nodule helps to create the microaerobic environment necessary for nitrogen fixation. Present within the mature nodules are networks of air passages (Bergersen and Goodchild, 1973) that provide sufficient nitrogen and oxygen for the reaction; this is limited by the permeability of the cortical cells (Tjepkema, 1981; Postgate, 1982). This compartmentation of the nitrogen-fixing process, together with the presence of the oxygen-carrying leghaemoglobin, regulates oxygen pressure within the nodules.

1.4.2 LEGHAEMOGLOBIN.

Haem and globin can combine spontaneously to form leghaemoglobin (Dilworth and Appleby, 1979) and its assembly takes place external to the bacteroids (Appleby, 1974). This is the dominant haemoprotein in effective nodules and performs a critical role in the supply of oxygen to the bacteroid for the respiratory activity related to nitrogen

fixation. The considerable storage and buffering capabilities of leghaemoglobin enable the protein to perform the dual role of supplying oxygen to the bacteroids while, at the same time, keeping the level of free oxygen low (Postgate, 1982). Leghaemoglobin also facilitates the delivery of oxygen to the specific oxidase systems of the bacteroids at precisely defined tensions favourable for ATP formation (Dilworth and Appleby, 1979). The oxidase systems of bacteroids are specifically adapted for maintaining respiration in the low O_2 environment in root nodules (Bergersen, 1980).

The plant contains the genetic information that specifies the type of leghaemoglobin produced. While haem is produced by both symbiotic partners - as each needs haem for cytochromes, catalase and peroxidase - the bacterium provides haem for the creation of leghaemoglobin. The onset of leghaemoglobin production is closely linked to the start of nitrogen fixation (Dilworth and Appleby, 1979). This protein is reported to be localised in the host cell cytoplasm external to the peribacteroid membrane (Bisseling *et al.*, 1986; Werner *et al.*, 1988).

1.4.3 NITROGENASE.

Nitrogenase comprises two proteins, component I, that contains molybdenum, iron and acid-labile sulphur and component II, a smaller iron-sulphur protein (Eady and Smith, 1979). During nitrogen fixation, component II binds two ATPs and is reduced by one electron equivalent which is derived from an electron transport system. The reduced component II + 2ATP complex in turn transfers an electron to a partially-reduced component I and concomitantly hydrolyses 2ATP to 2ADP + 2Pi (reviewed by Ludwig and de Vries, 1986). The oxidised component II is released from the reduced component I and the cycle is repeated five more times to provide the six electron equivalents needed to convert dinitrogen to ammonium (Burns, 1979; Postgate, 1982), which is subsequently exported from the bacteroid (reviewed by Ludwig and de Vries, 1986). A portion of the electron flux through nitrogenase is transferred to protons to form H_2 (Burris *et al.*, 1981;

Evans *et al.*, 1981). Some rhizobial strains possess an uptake hydrogenase which can couple re-oxidation of the hydrogen with electron transport to oxygen via oxidative-phosphorylation and hence, produce ATP (reviewed by Ludwig and de Vries, 1986). Nitrogenase turnover is illustrated in Figure 1.7.

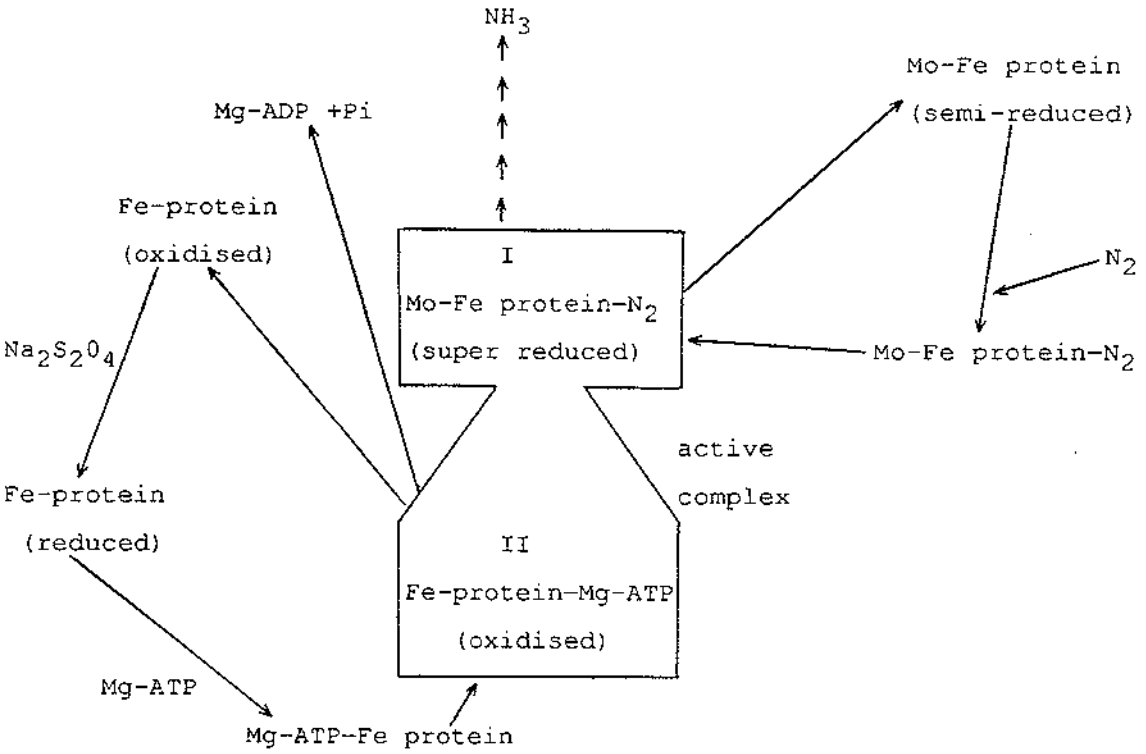
1.5 THE PROJECT.

1.5.1 BACKGROUND.

Lotus is a hardy leguminous plant that grows well in high country conditions and hence is an important pasture crop in southern New Zealand. *Lotus* is nodulated by both the fast-growing bacterium, *Rhizobium loti* and the slow-growing *Bradyrhizobium* sp. *Lotus*. The wild-type strain, NZP2037, forms effective nodules on *L. pedunculatus*, *L. corniculatus* and *L. tenuis* (Pankhurst *et al.*, 1979), while another strain, NZP2213, forms ineffective root nodules on *L. pedunculatus*. The difference between the host ranges of these two strains is one reason that prompted the study of this system.

Transfer of NZP2037's large indigenous plasmid failed to confer the ability to nodulate to other strains while plasmid-cured derivatives could still nodulate effectively, indicating that its nodulation genes, as opposed to those of many of the other rhizobia studied, are chromosomally borne (Chua *et al.*, 1985). This is also the case with the nitrogen fixation and flavolan resistance genes (Pankhurst *et al.*, 1983). The chromosomal location of these genetic elements made the study of *R. loti* nodulation genetics of interest. Tn*5* mutagenesis of *R. loti* NZP2037 was performed and several mutants isolated. Among these was a Nod^- mutant which blocked root hair curling and which was designated PN233 (Chua *et al.*, 1985). A pLAFR1 cosmid gene library was constructed for NZP2037 and complementary wild-type DNA was identified by the technique of *in planta* complementation (Chua *et al.*, 1985). The PN233::Tn*5* insertion was localised to a 7.1 kb *Eco* RI chromosomal fragment.

Figure 1.7: Nitrogenase turnover during conversion of N_2 to NH_3 (Eady and Smith, 1979).


NH_3
Mg-ADP +Pi
Mo-Fe protein
(semi-reduced)
Fe-protein
(oxidised)
I
Mo-Fe protein-N_2
(super reduced)
N_2
$Na_2S_2O_4$
Mo-Fe protein-N_2
active
complex
Fe-protein
(reduced)
II
Fe-protein-Mg-ATP
(oxidised)
Mg-ATP
Mg-ATP-Fe protein

1.5.2 THE PROJECT AIM.

The aim of the project was to characterise the PN233::Tn*5* mutant through plant complementation tests and by sequencing the region of the insert as outlined below. In order to determine whether the Tn*5* had insertionally inactivated one of the common *nod* genes, it was decided to conduct plant complementation tests. That is, wild-type NZP2037 DNA from the same region as that into which the Tn*5* had inserted, was to be crossed into common *nod* gene mutants of *R.l.* bv. *trifolii* to test whether the mutations could be complemented. It was hoped that these experiments would genetically identify the PN233::Tn*5* mutation (henceforth referred to as the 233 insertion or mutation).

Secondly, a restriction enzyme map for the 7.1 kb *Eco* RI region would be constructed. This, together with the restriction map of the same *Eco* RI fragment bearing the Tn*5*, would be used to locate the site of insertion more accurately and to identify smaller fragments useful for sequencing purposes.

The final step was to be the sequencing of the region into which the Tn*5* had inserted. It was hoped that this would either confirm plant complementation results, or if the mutation did not appear to be in a common *nod* gene, elucidate its function. The materials and techniques employed in the procedures described above, are outlined in the following chapter.

2.0 MATERIALS AND METHODS.

2.1 MEDIA.

Note:

1) Solutions and media were brought to the correct pH with either HCl or NaOH unless otherwise stated.
2) Sterilisation was achieved by autoclaving at 121°C for 15 minutes.

2.1.1 LURIA-BERTANI BROTH. (LB) (Miller, 1972).

(An *E. coli* growth medium).	(per litre)
Tryptone (BBL)	10 g
Yeast Extract (BBL)	5 g
NaCl	5 g
For top agar (Davis)	4 g
For bottom agar (Davis)	12 g

Ajust pH to 7.0

2.1.2 TRYPTONE YEAST EXTRACT BROTH. (TY) (Beringer, 1974).

(A *Rhizobium* and *E. coli* growth medium).	(per litre)
Tryptone (BBL)	5 g
Yeast Extract (BBL)	3 g

Before use, add 0.6 ml sterile 1 M $CaCl_2$/100 ml.

2.1.3 YT MEDIUM. (from Massey laboratory protocols).

(A rich, *E. coli* growth medium).	(per litre)
Tryptone (BBL)	16 g
Yeast Extract (BBL)	10 g
NaCl	5 g

Adjust pH with 0.5 ml 5 M NaOH.

2.1.4 YEAST EXTRACT MANNITOL BROTH. (YEM) (Vincent, 1970)

(A medium for purifying fungus-contaminated *E. coli* or *Rhizobium* stocks).

	(per litre)
K_2HPO_4	0.5 g
$MgSO_4.7H_2O$	0.2 g
NaCl	0.1 g
Mannitol (BDH)	10.0 g
Yeast Extract (BBL)	0.4 g
Agar	15.0 g

Add 3 g/l $CaCO_3$ if pH is below 6.8.

For YEM plus Actidione, (a cycloheximide fungicide), add Actidione (Upjohn Corp.) to 0.002%.

2.1.5 S10-DEFINED MEDIUM. (Scott and Ronson, 1982).

(A *Rhizobium* growth medium).

Stock Solutions.

(A) Salts/litre.

$MgSO_4.7H_2O$	25.0 g
$CaCl_2.2H_2O$	2.0 g
Fe EDTA (or 0.66 gm $FeCl_3$ + 1.5 gm Na_2 EDTA)	1.5 g
NaCl	20.0 g

(B) Trace Elements/litre.

$ZnSO_4.7H_2O$	3 mg
$Na_2MoO_4.2H_2O$	40 mg
H_3BO_3	50 mg
$MnSO_4.4H_2O$	40 mg
$CoSO_4.5H_2O$	4 mg
$CoCl_2.6H_2O$ (0.2 gm/l)	2 ml

(C) Vitamins/50 ml.

Thiamine HCl	50 mg
Ca pantothenate	100 mg
Biotin	1 mg

Dissolve by heating gently on a hot plate.

(D) NH_4Cl/litre. 18 g

(E) Phosphates/litre.	
K_2HPO_4	100 g
KH_2PO_4	100 g
(F) Bromothymol Blue/100 ml. (BDH # 20021)	0.2 g
(G) Na succinate/litre. (0.5 M)	33.8 g

To make media:-

	(per 250 ml)
1. Salts (A)	2.5 ml
2. Trace Elements (B)	0.25 ml
3. NH_4Cl (D)	1.5 ml
4. MES [2-(N-Morpholino)ethane-sulphonic acid] (BDH)	2.5 g
Adjust pH to 6.2.	
5. Histidine	25.0 mg
6. Bromothymol Blue (F), optional.	2.5 ml
After autoclaving, add aseptically:-	
7. Vitamins (C)	0.25 ml
8. Phosphates (E)	1.25 ml
9. Carbon source (G)	5.0 ml
For plates add 15 g/litre of agar.	

2.1.6 **M9 MEDIUM.** (+ glucose, + thiamine) (Maniatis *et al.*, 1982)

(A minimal medium used for transformations).	(per litre)
Na_2HPO_4	6.0 g
KH_2PO_4	3.0 g
NaCl	0.5 g
NH_4Cl	1.0 g
Adjust pH to 7.4.	
Autoclave, then add aseptically,	
10% $MgSO_4$ stock	2.5 ml
1% $CaCl_2$ stock	1.5 ml
20% Glucose stock	12.0 ml
0.337% Thiamine stock	5.0 ml
Agar (for plates)	12.0 g

2.1.7 **FAHRAEUS MEDIUM. (FM)** (Carroll and Gresshoff, 1983).

(A nitrogen-deficient medium for culturing seedlings).

Stock Solutions. (40x Stock)	(per litre)
KH_2PO_4	4.0 g
$Na_2HPO_4.12H_2O$	6.0 g
$MgSO_4.7H_2O$	4.8 g
$CaCl_2.2H_2O$	4.0 g

Make 500 ml of each stock separately and store in the cold.

Iron Chelate + EDTA. (200x Stock)	(per 100 ml)
$FeSO_4.7H_2O$	557 mg
Na_2EDTA	745 mg

Make solutions separately, heating gently to dissolve, then combine and store in the cold.

Trace Elements. (1000x Stock) (Gresshoff and Doy, 1974).

	(per 100 ml)
$MnSO_4.4H_2O$	100.0 mg
H_3BO_3	30.0 mg
$ZnSO_4.7H_2O$	30.0 mg
$Na_2MoO_4.2H_2O$	2.5 mg
$CuSO_4.5H_2O$	2.5 mg
$CoCl_2.6H_2O$	2.5 mg

To make 1 litre of FM:-

To 500 ml of distilled H_2O, add in order, 25 ml of concentrated stocks of: KH_2PO_4, Na_2HPO_4, $MgSO_4$, and $CaCl_2$. Then add 5 ml of Iron Chelate and 3 ml of Trace Element stock solutions. Add 392 ml H_2O.

Adjust pH to 6.5. For plates add 12 g/l agar. Sterilise.

2.2 ANTIBIOTIC STOCK SOLUTIONS.

Neomycin.	100 mg/ml in H_2O.
Tetracycline.	1 mg/ml in methanol.
Chloramphenicol.	1 mg/ml in H_2O.
Ampicillin.	5 mg/ml in H_2O.
Kanamycin.	10 mg/ml in H_2O.
Gentamycin.	400 mg/ml in H_2O.
Streptomycin.	100 mg/ml in H_2O.

2.3 BACTERIAL STRAINS.

2.3.1 MAINTENANCE OF BACTERIAL CULTURES.

2.3.1.1 SHORT-TERM STORAGE ON PLATES.

Overnight cultures of the various strains of bacteria were grown and then a loopful of each culture spread onto suitable selective agar plates. The *E. coli* strains were left to grow overnight at 37°C, while the *Rhizobium* was grown for a few days at 30°C, and then both were sealed with Parafilm. *E. coli* were stored at room temperature, while *Rhizobium* cultures were kept at 4°C. Plates were left in an inverted position. These should be replated every few months.

2.3.1.2 MEDIUM-TERM STORAGE ON AGAR SLOPES.

The cultures were also stored on slopes. 10 ml of molten agar, containing appropriate antibiotics, was dispensed into sterile universal bottles and left to set on an angle in such a way as to create a steep agar slope. A loopful of overnight culture was streaked onto the agar surface, the tubes capped, sealed with Parafilm, and stored at the temperatures indicated above for the various genera. Cultures can be stored in this manner for about a year.

2.3.1.3 LONG-TERM STORAGE IN GLYCEROL.

1) Fill 1.5 ml Eppendorf tubes with 200 ul glycerol and a few anti-bumping beads. Cap tubes and then autoclave.
2) Grow overnight cultures of the cells to be stored.
3) Label Eppendorf tubes bearing in mind that the labelling must survive freezing procedures.
4) Add 0.8 ml of overnight culture to the appropriate tube aseptically.
5) Place in either the ¯20°C or ¯70°C freezer.
6) When removing from freezer for use, either scratch the frozen solution with a sterile needle and streak it onto a plate or spot it into a broth.

2.4 TRIPARENTAL CROSSES.

In this procedure (Ditta *et al.*, 1980) a helper plasmid is used to transfer a plasmid that is not self-transmissible, from a donor to a recipient bacterium by employing a "helper" plasmid from a third strain to mobilise the donor's plasmid.

Overnight cultures of the three strains were grown. 100 ul of the recipient and 50 ul of the donor and helper strains were pipetted together to form a large drop on a non-selective plate and then dried-down in a Laminar Flow cabinet (Airpure). The bacteria were left to grow up overnight. A loopful of the mixed culture was then streaked out onto a plate that selected for recipients possessing the required plasmid.

2.5 *TRIFOLII REPENS* (WHITE CLOVER) AND *LOTUS PEDUNCULATUS* CULTURE PROCEDURES. (Collins, 1983).

2.5.1 STERILISATION OF CLOVER AND *LOTUS* SEEDS.

Nodulation tests were carried out using *Lotus pedunculatus* cv. "Grasslands Maku", while the white clover (*T. repens* cv. Huia) seeds were obtained from Hodder and Tolley Ltd., Palmerston North.

Table 2.1: Bacterial strains and bacteriophage .

Strain Designation.	Characteristics and Selection.	Source or Reference.
E. coli		
HB101	F⁻ *pro leu thi lac* Y St^R r- m⁻ Endo I⁻ *recA*⁻	Boyer and Roulland-Dussoix, (1969)
JM101	(*lac proAB*) *sup*E *thi* [F' *proAB lacI*^{q} Z/\M15]	Yanisch-Perron *et al.*, (1985)
JM109	(*lac proAB*) *sup*E *thi recA rel*A *gyr*A *hsd*R [F'*proAB lacI*^{q} Z/\M15]	Yanisch-Perron *et al.*, (1985)
XL-1	*rec*A1 *end*A1 *gyr*A96 *thi hsd*R17 *sup*E44 *rel*A1 lambda⁻ *lac*⁻ [F' *proAB lacI*^{q} Z/\M15 Tn*10*]; T^R	B. Mansfield, Massey, P.N., N.Z.
PN231	HB101/pPN305; T^R	Scott *et al.*, (1985)
PN457	HB101/pPN25; T^R	Scott *et al.*, (1985)
PN445	HB101/pPN354; A^R, T^R	Scott *et al.*, (1985)
PN226	HB101/pPN301; A^R, K^R, T^R	Scott *et al.*, (1985)
PN384	HB101/pPN330; T^R	Scott *et al.*, (1985)
PN342	HB101/pRK2073	Scott *et al.*, (1985)
PN333	HB101/pPBR328; A^R, C^R, T^R	D.B. Scott, Massey, P.N., N.Z.
PN600	HB101/pPN26; T^R	Scott *et al.*, (1985)
PN1053	HB101/pPN306::Tn*5* in the *Eco* RI fragment; K^R, T^R	C. Pankhurst, DSIR, P.N., N.Z.
PN1054	HB101/pPN306::Tn*5* in the 7.1 *Eco* RI fragment; K^R, T^R	C.Pankhurst, DSIR, P.N., N.Z.
PN1055	HB101/pPN306::Tn*5* in the 7.1 *Eco* RI fragment; K^R, T^R	C. Pankhurst, DSIR, P.N., N.Z.
PN1056	HB101/pPN306::Tn*5* in the 7.1 *Eco* RI fragment; K^R, T^R	C. Pankhurst, DSIR, P.N., N.Z.
PNJ019	HB101/pBRJ12; A^R	This study
PNJ020	HB101/pBRJ14; A^R	This study
PNJ024	HB101/pGEM-2; A^R	This study
PNJ025	HB101/pJ14012; A^R	This study
PNJ026	HB101/pJ14011; A^R	This study
PNJ027	HB101/pJ28003; A^R	This study
PNJ028	HB101/pJ28006; A^R	This study
PNJ029	XL1/pJ12009; A^R, T^R	This study
PNJ030	XL1/pJ12010; A^R, T^R	This study
Bacteriophage.		
M13mp18		Messing and Vieira, (1982)

TABLE 2.1: Bacterial strains and bacteriophage. (continued).

Strain Designation.	Characteristics and Selection.	Source or Reference.
R. loti		
NZP2037	Nod^+ Fix^+ (*L. pedunculatus*, *L. tenuis*)	DSIR culture collection, P.N., N.Z.
PN233	NZP2037 *str-1*, *nod-1*::Tn*5*; N^R	Chua *et al.*, (1985)
PN4016	NZP2037 St^R derivative/pPN306::Tn*5*; N^R, G^R, St^R	C. Pankhurst, DSIR, P.N., N.Z.
PN4019	NZP2037 St^R derivative/pPN306::Tn*5*; N^R, G^R, St^R	C. Pankhurst, DSIR P.N., N.Z.
PN4047	NZP2037 St^R derivative/pPN306::Tn*5*; N^R, G^R, St^R	C. Pankhurst, DSIR P.N., N.Z.
PN4053	NZP2037 St^R derivative/pPN306::Tn*5*; N^R, G^R, St^R	C. Pankhurst, DSIR P.N., N.Z.
PNJ009	PN233/pPN305; N^R, T^R	This study
PNJ010	PN4047/pPN305; N^R, T^R	This study
PNJ011	PN4053/pPN305; N^R, T^R	This study
PNJ012	PN233/pPN26; N^R, T^R	This study
PNJ013	PN4047/pPN26; N^R, T^R	This study
PNJ014	PN4053/pPN26; N^R, T^R	This study
PNJ015	PN233/pPN25; N^R, T^R	This study
PNJ016	PN4047/pPN25; N^R, T^R	This study
PNJ017	PN4053/pPN25; N^R, T^R	This study
PNJ018	NZP2037/pPN305; T^R	This study

Table 2.1: Bacterial strains and bacteriophage. (continued)

Strain Designation.	Characteristics and Selection.	Source or Reference.
R.l. bv. *trifolii*		
PN100	Nod$^+$ Fix$^+$ *str-1 rif-1*	Scott and Ronson, (1982)
ANU851	*nodD*$^-$ Tn*5*-induced mutant of ANU843; N^R	Schofield *et al.*, (1983)
ANU277	*nodC*$^-$ Tn*5*-induced mutant of ANU843; N^R	Schofield and Watson (1986)
ANU249	*nodB*$^-$ Tn*5*-induced mutant of ANU843; N^R	Schofield and Watson (1986)
ANU252	*nodA*$^-$ Tn*5*-induced mutant of ANU843; N^R	Schofield and Watson (1986)
PNJ001	ANU851/pPN305; N^R, T^R	This study
PNJ002	ANU277/pPN305; N^R, T^R	This study
PNJ003	ANU249/pPN305; N^R, T^R	This study
PNJ004	ANU252/pPN305; N^R, T^R	This study
PNJ005	ANU851/pPN25; N^R, T^R	This study
PNJ006	ANU277/pPN25; N^R, T^R	This study
PNJ007	ANU249/pPN25; N^R, T^R	This study
PNJ008	ANU252/pPN25; N^R, T^R	This study
PNJ031	ANU851/pPN26; N^R, T^R	This study
PNJ032	ANU252/pPN26; N^R, T^R	This study
PNJ033	ANU249/pPN26; N^R, T^R	This study
PNJ034	ANU277/pPN26; N^R, T^R	This study

KEY: A = Ampicillin (100 ug/ml); K = Kanamycin (50 ug/ml); N = Neomycin (500ug/ml for *R. loti*, exceptions being PN4016, PN4019, PN4047 and PN4053 where 400 ug/ml was used and, for *R.l.* bv. *trifolii*, where 100 ug/ml was used); T = Tetracyline (15 ug/ml for *E. coli*, exceptions being XL1, PNJ029 and PNJ030 where 10 ug/ml was used and, for *Rhizobium*, where 2 ug/ml was used); C = Chloramphenicol (25 ug/ml); St = Streptomycin (250 ug/ml); G = Gentamycin (200 ug/ml). DSIR = Department of Scientific and Industrial Research; P.N., N.Z. = Palmerston North, New Zealand.

Table 2.2: Plasmids.

Designation.	Characteristics and Antibiotic Markers.	Source or Reference.
pLAFR1	21.6 kb [a cosmid derivative of the low copy number, broad host range vector pRK290]; T^R.	Friedman *et al.*, (1982)
pBR328	4.9 kb [a pBR322 derivative]; A^R, C^R, T^R.	Bolivar *et al.*, (1977)
pPN305	NZP2037 *nod* pLAFR1 cosmid ; T^R.	Scott *et al.*, (1985)
pPN25	pLAFR1 cosmid + NZP2037 7.1 kb *Eco* RI fragment; T^R.	Scott *et al.*, (1985)
pPN354	pBR328 + NZP2037 7.1 kb *Eco* RI fragment; A^R, T^R.	Scott *et al.*, (1985)
pPN301	pBR328 + NZP2037 7.1 kb *Eco* RI fragment::Tn*5*; A^R, T^R, K^R, N^R.	Scott *et al.*, (1985)
pRK2073	a K^S derivative of pPK2013; used as a mobilising plasmid.	Leong *et al.*, (1982)
pPN26	*R.l.* bv. *trifolii* PN100 *nod* pLAFR1 cosmid; T^R.	Scott *et al.*, (1985)
pPN306	NZP2037 *nod* pLAFR1 cosmid containing 30 kb *Eco* RI fragment; G^R, T^R.	Scott *et al.*, (1985)
pGEM-2	2.9 kb cloning vector; A^R.	Melton *et al.*, (1984)
pJ14011	pGEM-2 + 1.4 kb *Sal* I NZP2037 fragment, orientation #11; A^R.	This study
pJ14012	pGEM-2 + 1.4 kb *Sal* I NZP2037 fragment, orientation #12; A^R.	This study
pJ28003	pGEM-2 + 2.8 kb *Bam* HI NZP2037 fragment, orientation #3; A^R.	This study
pJ28006	pGEM-2 + 2.8 kb *Bam* HI NZP2037 fragment, orientation #6; A^R.	This study
pJ12009	pGEM-2 + 1.2 kb *Sal* I NZP2037 fragment, orientation #9; A^R.	This study
pJ12010	pGEM-2 + 1.2 kb *Sal* I NZP2037 fragment, orientation #10; A^R.	This study
pBRJ12	pBR328 (with a spontaneous deletion of ≈ 300 bp in T^R gene) + 1.2 kb *Sal* I fragment; A^R.	This study
pBRJ14	pBR328 (with a spontaneous deletion of ≈ 300 bp in T^R gene) + 1.4 kb *Sal* I fragment; A^R.	This study

KEY: A = Ampicillin; C = Chloramphenicol; G = Gentamycin; N = Neomycin for rhizobia or Kanamycin for *E. coli*; St = Streptomycin; T = Tetracycline.

1) Approximately 200 clover seeds were rinsed in a plastic petri-dish with 20 ml distilled H_2O. No more than 200 were placed in a dish because the concentration of a germination inhibitor released collectively by the seeds would have been sufficiently high to affect germination.
2) One drop of detergent, Triton X-100, was placed into each petri-dish and the suspension swirled a few times. The detergent breaks the surface tension and facilitates wetting of the seeds resulting in the seeds sinking to the bottom of the dish.
3) A saturated solution of calcium hypochlorite (bleach), was made by placing approximately 2 g of the bleach in 50 ml of distilled H_2O. The solution was stirred and left to stand for 10-15 minutes.
4) To sterilise seeds, 2.2 ml of the saturated bleach solution was added resulting in a final concentration of 10% calcium hypochlorite.
5) The mixture was swirled and left to stand for 10 minutes.
6) The petri-dish was drained completely and approximately 20 ml of sterile, distilled H_2O was added to wash the calcium hypochlorite from the seeds. The plate was drained again and the procedure repeated.
7) After the final rinse, 20 ml of sterile, distilled H_2O was added. The petri-dish was sealed with a Parafilm strip (approximately 2 cm wide) to prevent accidental opening and spillage.
8) The petri-dish was left in the dark overnight at room temperature.
9) The sterile, imbibed white clover seeds were transferred onto sterile FM plates, about 50 seeds per plate, and allowed to germinate. The plates were sealed with Parafilm and incubated under lights in an upright, vertical position in a growth cabinet/room set at 22°C. The germinated seedlings were transferred into growth tubes containing FM the next day.

2.5.2 INOCULATION OF SEEDLINGS.

Two methods were used during complementation trials. The first involved pipetting a few drops of the appropriate overnight culture, grown up in S10 medium, onto each FM plate and distributing evenly across the plate with a glass spreader. Two germinated seedlings were placed 2/3 of the way up each of the plates and left to lie in a horizontal position for about 60 minutes. This facilitated adhesion of the seedlings to the surface of the agar. Plates were then sealed with Parafilm. A 2 cm slit was made in the Parafilm at the top of each plate to enable gas exchange.

The second method involved pouring 10 ml of molten FM agar into test tubes and allowing these to set on a steep slope; approximately at a 60°angle. A germinated seedling was then placed near the top of an agar slope using either tweezers or a loop, under aseptic conditions. A few drops of inoculum was placed on the seedling, then the tube capped lightly.

2.6 MICROSCOPY OF NODULE SECTIONS.

Embedding of the nodules for sectioning and some of the electron microscopy work was very kindly performed by Doug Hopcroft of DSIR, Palmerston North.

2.6.1 EMBEDDING PROCEDURE.

Phosphate Buffer.

Na_2HPO_4 (pH 7.2)	0.1 M

Primary Fixative.

glutaraldehyde	3% (v/v)
formaldehyde (in 0.1 M Na_2HPO_4, pH 7.2)	2% (v/v)

Osmium Tetroxide Buffer.

OsO_4 in Na_2HPO_4 (pH 7.2)	1% (w/v)

Acetone/Resin.

50:50 mixture of acetone (AnalaR) and Polarbed S12 resin.

Method:

Nodules were excised from the root and the tissue then sliced under primary fixative in a plastic petri-dish. The samples were then transferred to glass vials and fresh fixative added. The tissue was left to fix at room temperature for two hours then washed 3 times in phosphate buffer. This was followed by staining for 30 minutes in osmium tetroxide buffer. After a further 3 washes in phosphate buffer, the samples were dehydrated through a series of acetone/water washes, the final two steps being in 100% acetone. The infiltration step involved stirring the samples overnight on a magnetic stirrer, then transferring them to a 100% resin solution for a further 8 hours with continued stirring. Specimens were embedded in fresh resin using silicone rubber moulds and left to cure at 60°C for 48 hours (Craig and Williamson, 1972).

2.6.2 SECTIONING AND MICROSCOPY.

A Reichert Ultracut E microtome, employing a diamond knife, was used for sectioning. Sections were either viewed with a Ziess microscope and photographed with Ilford FP4 35 mm film or, picked up onto copper grids, double stained with ethanolic uranyl acetate followed by lead citrate (Craig and Williamson, 1972). Grids for electron microscopy were examined with a Phillips ZOIC transmission electron microscope. Kodak Fine Grain positive film was used for the photography.

2.7 BUFFERS AND SOLUTIONS USED IN DNA PREPARATION.

2.7.1 *HAE* III BUFFER - A UNIVERSAL BUFFER. (from Massey laboratory protocols).

(10x)

Tris base (pH 7.6)	60 mM
$MgCl_2$	100 mM
β-mercaptoethanol	100 mM

2.7.2 TRIS BORATE EDTA BUFFER. (TBE) (Maniatis *et al.*, 1982).

(For agarose gels).

(10x Stock)	(per litre)
Tris base [Tris(hydroxymethyl)aminomethane] (USB)	108.0 g
EDTA (or 11.8 g of $Na_2EDTA.2H_2O$)	9.2 g
Boric acid	55.0 g

Adjust pH to 8.2.

2.7.3 SEQUENCING TRIS BORATE EDTA BUFFER. (D.B. Scott, personal communication).

(For sequencing gels).

(10x stock)	(per litre)
Tris base	162.0 g
$Na_2EDTA.2H_2O$	9.5 g
Boric acid	27.5 g

Adjust pH to 8.8.

2.7.4 TE BUFFER. (10:0.1) (Maniatis *et al.*, 1982).

(Tris 10 mM:EDTA 0.1 mM)	(per 100 ml)
1 M Tris base (pH 8.0)	1 ml
250 mM Na_2EDTA (pH 8.0)	40 ul

Note: Adjust accordingly for other Tris:EDTA ratios.

2.7.5 **10% SODIUM DODECYL SULPHATE.** (SDS) (Maniatis *et al.*, 1982). (Also called sodium lauryl sulphate)

Dissolve 100 g of electrophoresis-grade SDS in 900 ml of H_2O. Heat to 68°C. Adjust the pH to 7.2 by adding a few drops HCl (conc.). Adjust volume to 1 litre. Note: Weigh SDS whilst wearing a face mask.

2.7.6 **LOADING BUFFER.** (from Massey laboratory protocols).

(For loading DNA samples onto agarose gels)	(per 20 ml)
sucrose	4.0 g
EDTA (250 mM)	0.4 ml
SDS	0.2 g
bromophenol blue	40.0 mg

Note: Wear a face mask while weighing SDS.

2.7.7 **STANDARD SALINE CITRATE.** (SSC) (Maniatis *et al.*, 1982).

	(per litre)
NaCl	8.8 g
Na_3 citrate.$2H_2O$	4.4 g

2.8 **PREPARATION OF DIALYSIS TUBING FOR DNA PURIFICATION.** (Maniatis *et al.*, 1982).

1) Boil 1 packet of tubing in 1 litre of 5% (w/v) sodium carbonate for 15 minutes. Change solution and repeat the process until there is no detectable colour or odour in the solution.
2) Pour off the final amount and boil for a further 15 minutes in 1 litre of distilled H_2O.
3) Boil for 15 minutes in 1 litre of 1 mM EDTA.
4) Wash tubing for 5 minutes in distilled H_2O.
5) Boil for 15 minutes in 1 litre of 1 mM EDTA. Allow this solution and the tubing to cool. Cover beaker with foil and store at 4°C.

2.9 DNA PREPARATION PROCEDURES.

2.9.1 DNA EXTRACTIONS.

2.9.1.1 PHENOL AND CHLOROFORM PREPARATION FOR DNA EXTRACTIONS.

Chloroform in phenol/chloroform extractions of DNA was chloroform:iso-amyl alcohol (24:1) but will henceforth be referred to as chloroform in phenol extractions. The phenol used was Tris-equilibrated to pH ≈ 8.0 by the following procedure as outlined by M.J. O'Hara, (1989).

Approximately 750 ml of phenol was melted and poured into a 1 litre round-bottom flask. Anti-boiling chips were added and then the distillation equipment assembled. The phenol was heated to 180°C and maintained at this temperature till about 100 ml of the original volume remained. Heat was removed and the equipment was left to cool. All glassware was washed in hot water, then 95% ethanol followed by hot water again. The re-distilled phenol was stored at ⁻20°C until required.

Re-distilled phenol or AR grade phenol was thawed and a sufficient amount removed for extraction. The melted phenol was then transferred to a separating funnel and 0.1% (w/v) 8-hydroxyquinoline (Sigma) was added and dissolved. The phenol was then washed twice with 1 M Tris buffer, pH 8.0, containing 0.2% β-mercaptoethanol (1 M). This was usually sufficient to raise the pH of the phenol to 8.0. The phenol was transferred to a brown glass bottle and an equivalent volume of 0.1 M Tris pH 8.0, containing 0.2% β-mercaptoethanol, was added. The phenol was stored at 4°C.

2.9.1.2 PHENOL/CHLOROFORM EXTRACTIONS OF DNA.

For extractions, the usual procedure was to wash the DNA solution once with an equal volume of phenol. The mixture was then centrifuged to separate the phases and the top aqueous layer drawn off with a pipette. This phase was then washed with an equal volume of phenol:chloroform (1:1), spun, separated and finally washed with an equal volume of chloroform.

2.9.2 DNA PRECIPITATION.

DNA was precipitated with 1/20 volume of 5 M NaCl plus 2x volume of ethanol at ⁻20°C for 30 minutes, unless otherwise stated. The DNA was then pelleted in an Eppendorf centrifuge for 30 minutes.

2.10 PLASMID PREPARATION PROCEDURES.

2.10.1 LARGE SCALE PLASMID PREPARATION: ALKALINE LYSIS METHOD.

(based on Maniatis *et al.*, 1982).

Amplification in a rich medium.

1) Inoculate 10 ml of LB containing a suitable antibiotic with a single bacterial colony and grow overnight at 37°C with vigorous shaking.
2) The next morning, inoculate 25 ml of LB medium in a 100 ml flask containing the appropriate antibiotic with 100 ul of the overnight culture. Incubate at 37°C with vigorous shaking until the culture has reached late log phase (OD_{600} = 0.6).
3) Inoculate 25 ml of the late log phase culture into 500 ml of LB medium containing the appropriate antibiotic, pre-warmed to 37°C in a 2 litre flask. Incubate for 2.5 hours at 37°C while shaking vigorously. The OD_{600} of the culture should be approximately 0.4.
4) If the plasmid carries the ColE1 replicon add 2.5 ml of chloramphenicol (34 mg/ml in ethanol) for amplification purposes. The final concentration of chloramphenicol will be 170 ug/ml.
5) Incubate with vigorous shaking for a further 12-16 hours.
6) Harvest the bacterial cells by centrifugation at 4000*g* for 10 minutes at 4°C. Discard the supernatant.
7) Wash the cells in 100 ml of ice-cold STE and centrifuge again.

STE Buffer.

NaCl	0.1 M
Tris.HCl (pH 7.8)	10.0 mM
EDTA	1.0 mM

Lysis by alkali treatment.

8) Resuspend the bacterial pellet in 10 ml of solution 1 containing 5 mg/ml lysozyme.

Solution 1.

glucose	50 mM
Tris.HCl (pH 8.0)	25 mM
EDTA	10 mM

Note: Solution 1 can be autoclaved and stored at 4°C. Powdered lyzozyme should be dissolved in the solution just before use.

9) Transfer to a Beckman SW27 polyallomer tube (or its equivalent) and stand at room temperature for 5 minutes.

10) Add 20 ml of freshly made solution 2. Cover the top of the tube with Parafilm and mix gently by inverting the tube several times. Stand on ice for 10 minutes.

Solution 2.

NaOH	0.2 N
SDS	1.0 %

Note: This solution should be made from stock solutions of 10 N NaOH and 20% SDS.

11) Add 15 ml of an ice-cold solution of 5 M potassium acetate (pH 4.8) prepared as follows: to 60 ml of 5 M potassium acetate, add 11.5 ml of glacial acetic acid and 28.5 ml of H_2O. The resulting solution will be 3 M with respect to potassium and 5 M with respect to acetate. Cover the top of the tube with Parafilm and mix by inverting the tube sharply a few times. Stand for 10 minutes on ice.

12) Centrifuge in a Beckman SW27 at 20,000 rpm (≈70,000*g*) for 20 minutes at 4°C. The cell DNA and bacterial debris should form a tight pellet at the bottom of the tube.

13) Transfer equal quantities (≈ 18 ml) of the supernatant into each of two 30 ml Corex tubes.

14) Add 0.6 volumes (≈ 12 ml) of isopropanol to each tube to precipitate the DNA. Mix well and let stand at room temperature for 15 minutes.

15) Spin down in a Sorvall SS-34 rotor for 15 minutes at 15,000 rpm (≈ 27,000*g*).

16) Resuspend pellet in 400-600 ul TE (10:0.1), pH 8.0.

Purification of closed circular DNA by centrifugation on a caesium chloride-ethidium bromide gradient.

17) Measure the volume of DNA solution. For every millilitre add 1.08 grams of solid caesium chloride. Mix gently until all of the salt is dissolved.

18) Add 0.8 ml of a solution of ethidium bromide (10 mg/ml in H_2O) for every 10 ml of caesium chloride/DNA solution. Mix well. The final refractive index should be 1.3860-1.3920 and the concentration of ethidium bromide should be approximately 600 ug/ml.

19) Transfer the solution to a tube suitable for ultracentrifugation, filling the remainder of the tube with a solution of caesium chloride-ethidium bromide of the same density and ethidium bromide concentration as described above. Leave a 2 mm space from the rim of the tube to allow room for the cap.

20) Centrifuge at 45,000 rpm (≈120,000*g*) for 16 hours at 20°C.

21) Two bands of DNA should be visible in ordinary light. A hand-held UV lamp, set on a long wave frequency, can be used to help visualise bands if they are faint. The upper band comprises linear bacterial DNA and nicked circular plasmid DNA, while the lower band consists of closed circular plasmid DNA.

22) Remove the cap from the tube. The closed circular band can be isolated by careful collection with a pipette. Alternatively, a hypodermic needle (# 18 gauge) can be inserted through the side of the tube, just below the level of DNA, and the band drawn off.

23) The ethidium bromide is removed by adding an equal volume of isoamyl alcohol or, water-saturated 1-butanol.

24) Mix the phases by pipetting vigorously.

25) Centrifuge at 1,500*g* for 3 minutes at room temperature.

26) Transfer the lower aqueous phase to a clean glass tube.

27) Repeat the extraction 4-6 times or until all trace of the pink colour has been removed from the aqueous solution.

28) Dialyse the aqueous phase against sterile TE (10:0.1), pH 8.0, in the following manner. Seal the aqueous phase in prepared dialysis tubing (See 2.8). Fill a flask with 1 litre of TE and place in the cold on a stirrer. Dialyse the aqueous phase, changing the buffer twice the first day and once the day after.

29) DNA concentration and purity can be determined in the following manner. An optical density reading of "1" at OD_{260} represents a concentration of 50 ug/ml of DNA. Therefore, DNA concentration can be deduced by the following equation:

OD_{260} value x 50 x dilution factor = ug/ml of DNA.

A OD_{260}/OD_{280} value equal to about 1.8 or more indicates that the DNA solution is acceptably clean. Values significantly below 1.8 represent protein contamination and in this case, the DNA requires further purification.

2.10.2 MEDIUM SCALE PLASMID PREPARATION.

(based on Maniatis *et al.*, 1982).

1) Inoculate 5 ml of LB containing an appropriate antibiotic with a single colony and grow overnight with vigorous shaking.
2) Inoculate 600 ml LB plus antibiotic and grow in a shaker overnight.
3) The following morning, spin down the cells in a pre-cooled GSA rotor at 4,500 rpm (≈3,300*g*) using large screw-top containers no more than 2/3 full. Spin for 10 minutes.
4) Resuspend pellet in 10 ml of solution 1 plus lysozyme.

   Solution 1 + lysozyme.

glucose	50 mM
Tris.HCl (pH 8.0)	25 mM
EDTA	10 mM

   a "spatula-tip" quantity of lysozyme
5) Transfer to sterile 30 ml corex tubes. (The volume should be split into 2 x 6 ml lots).
6) For each 6 ml of pellet + solution 1 add 12 ml of freshly prepared NaOH/SDS (0.2 M NaOH, 1% SDS). Leave on ice for 10 minutes.

7) To each tube add 7.5 ml ice-cold 5 M KOAc, pH 4.8.

To make the KOAc solution:

5 M potassium acetate	60.0 ml
glacial acetic acid	11.5 ml
H_2O	28.5 ml

The resulting solution is 3 M with respect to K^+ and 5 M with respect to the acetate. Cover with Parafilm and mix thoroughly by inverting several times. Leave on ice for 10 minutes.

8) Centrifuge at 4°C in a SS-34 rotor for 5 minutes at 10,000 rpm (≈12,000*g*).

9) Divide the supernatant between two fresh Corex tubes and add an equal volume of phenol/chloroform. Cover with Parafilm and mix carefully by inverting twice.

10) Centrifuge for 10 minutes at 5,700 rpm (≈6,300*g*) in an HS-4 swinging bucket rotor.

11) Transfer the supernatant to fresh Corex tubes and precipitate with an equal volume of isopropanol. Leave at ⁻20°C for 60-90 minutes.

12) Spin down in SS-34 at 15,000 rpm (≈27,000*g*) for 15 minutes. Dry pellet under vacuum then resuspend in 400-600 ul TE (10:0.1), pH 8.0.

Note: The DNA can be further purified by running through a G-50 column with TE, pH 8.0.

2.10.3 SMALL SCALE PLASMID PREPARATION FOR *E. COLI*. (RAPID BOIL METHOD).

(based on Holmes and Quigley, 1981).

1) Inoculate 5 ml of medium containing an appropriate antibiotic. Incubate overnight at 37°C shaking vigorously.

2) Centrifuge 1.5 ml of culture in an Eppendorf tube.

3) Remove the supernatant by aspiration leaving the bacterial pellet as dry as possible.

4) Resuspend pellet in 350 ul of HQ-STET.

HQ-STET

sucrose	8.0% }	
Triton X-100	0.5% }	can be
EDTA (pH 8.0)	50 mM }	sterilised
Tris.HCl (pH 8.0)	10 mM }	

5) Add 25 ul of a freshly prepared solution of lysozyme (10 mg/ml in Tris.HCl, pH 8.0). Mix by vortexing for 3 seconds.
6) Place the tube in a boiling-water bath for 40 seconds.
7) Centrifuge immediately for 10 minutes at room temperature in an Eppendorf centrifuge.
8) Remove the resulting pellet from the tube with a sterile toothpick.
9) Ethanol precipitate the DNA and leave at $^{-}$20°C for a minimum of 1/2 hour.
10) Centrifuge for 1/2 hour at 4°C.
11) Dry the pellet and suspend it in 50 ul of TE (10:0.1), pH 8.0.

2.11 DNA RESTRICTION ENDONUCLEASE DIGESTS.

A typical digest.

DNA	1.0 ug
REact buffer 10x	2.5 ul
restriction enzyme	≈ 1.0 U
H_2O	to 25.0 ul

Digests were incubated at the appropriate temperature, for 1-2 hours. However, scaled-up reactions employing relatively stable enzymes were often left overnight. Small samples of the digests were generally run on mini-gels, with loading buffer (see 2.7.6), to check for reaction completion. A lambda standard (generally *Hind* III) was run along side the digests to determine the approximate size of the resulting linear fragments.

2.12 PREPARATION OF LAMBDA STANDARDS.

Standards that contained DNA fragments of known size were run against fragments of unknown size. The mobilities of the "standard" fragments were measured and then plotted against $\log_{10}$ of their known molecular weights. The sizes of the unknown fragments could thus be read off the graph by their relative mobilities. Lambda DNA, restricted with a suitable enzyme, was used for making the standards (Sanger *et al.*, 1982).

1) Cut 5 ug of DNA in a total of 50 ul with 2 ul of the appropriate enzyme.
2) Check that the digestion has gone to completion by running a small sample on an agarose gel.
3) Heat kill at 65°C for 10 minutes then rapidly cool on ice.
4) Dilute to 500 ul with:-

    50 ul sample
    325 ul H_2O
    125 ul SDS dye (see 2.7.6)

2.13 DNA AGAROSE GEL ELECTROPHORESIS.

Agarose gel electrophoresis is the standard manner for separating, identifying and isolating DNA fragments. Generally, most agarose gel electrophoresis is performed on horizontal slab gels. Powdered agarose is added to electrophoresis buffer (TBE; see 2.7.2) to the correct concentration and the slurry heated until the agarose is dissolved. The solution is then cooled to about 50°C and poured into the gel mold. A comb is placed at one end of the gel, the teeth forming the sample wells when extricated. After the agarose is set, the comb is carefully removed and sufficient buffer added to cover the gel to the depth of about 1 mm.

DNA samples are mixed with a loading buffer (see 2.7.6) and loaded into the pre-formed wells with an autopipette. Movement of the samples is monitored with the aid of the dye incorporated in the loading buffer. At completion of the run, the gel is carefully removed and placed in an ethidium bromide bath (0.5 ug/ml in H_20) and sufficient time allowed to stain the gel; this is dependent on gel size and thickness. Ethidium bromide is a fluorescent dye containing a planar group that intercalates between the stacked bases of nucleic acids. UV-irradiation of the gel is used to visualise nucleic acids as they fluoresce more strongly than the background dye.

Table 2.3: Restriction endonucleases.

Restriction endonucleases.	Reaction buffer.	Reaction temperature.	Source.
Alu I	REact 1	37°C	BRL
Ava I	REact 2	37°C	BRL
Bam HI	REact 3	37°C	BRL
Bgl II	REact 3	37°C	BRL
Eco RI	REact 3	37°C	BRL
Hae III	REact 2	37°C	BRL
Hind III	REact 2	37°C	BRL
Kpn I	REact 4	37°C	BRL
Pst I	REact 2	37°C	BRL
Pvu I	REact 7	37°C	BRL
Pvu II	REact 6	37°C	BRL
Sal I	REact 3	37°C	BRL
Sma I	REact 4	30°C	BRL
Sph I	REact 6	37°C	BRL
Taq I	REact 2	65°C	BRL

Mini-gels which are miniature versions of the above apparatus, are ideal for rapid analysis of small quantities of DNA. A more complete description of agarose gel electrophoresis can be found in Maniatis *et al.*, (1982).

Note: mini-gel dimensions - 9.5 cm x 7 cm; ≈ 20 ml agarose.
mini-sub dimensions - 12 cm x 8 cm; 50 ml agarose.
large gel dimensions - 12 cm x 17 cm; 100 ml agarose.

2.14 RECOVERY OF DNA FROM SEAPLAQUE (LOW-MELTING POINT) AGAROSE GELS.

2.14.1 ELECTROLUTION. (Maniatis *et al.*, 1982).

1) Run a SeaPlaque gel and then localise the desired band with a long-wave-length UV lamp (300-360 nm).
2) Using a clean scalpel blade, cut out a slice of the agarose containing the band.
3) Cut a short section of prepared dialysis tubing and tie or secure one end with dialysis clips to create a bag.
4) Fill the bag with 0.5x TBE. Transfer the excised slice to the bag using forceps and allow the agarose slice to sink to the bottom. Tie or clip the remaining end of the bag just above the gel slice, trying to avoid forming air bubbles in the process.
5) Immerse bag in a shallow layer of 0.5x TBE in an electrophoresis tank. Pass electric current through the bag (about 100V) for 2 or 3 hours. The DNA will be electroluted out of the agarose and onto the inner wall of the dialysis tube during this time.
6) Reverse the current for 2 minutes to release the DNA from the wall of the bag.
7) Open the dialysis bag and carefully recover the buffer surrounding the agarose slice. Using an Eppendorf pipette wash out the bag with 200 ul of 0.5x TBE.
8) The DNA can then be purified by phenol/chloroform extraction followed by ethanol precipitation.

Note: The agarose slice can be examined under UV light to determine if all the DNA has been electroluted.

2.14.2 **PHENOL/FREEZE TECHNIQUE.** (based on Thuring *et al.*, 1975).

1) Excise the DNA band from a SeaPlaque gel using a clean scalpel blade.
2) Place in an Eppendorf tube and add an equivalent volume of phenol.
3) Place the tube at 65°C until the agarose melts. Vortex then place at $^{-}$20°C until the solution freezes.
4) Spin for ten minutes in a bench centrifuge. The solution should separate into powdered agarose at the bottom and a layer of phenol topped by the aqueous (TBE) phase containing the DNA. If the gel has not disintegrated completely, refreeze and spin again.
5) Draw off the aqueous phase and phenol/chloroform extract.
6) Ethanol precipitate.

2.15 **LIGATIONS.**

LIGATION BUFFER. (from Bethesda Research Laboratories)

(5x reaction buffer)

Tris.HCl	250 mM
$MgCl_2$	50 mM
ATP	5 mM
DTT (dithiothreitol)	5 mM
PEG (polyethylene glycol) (8000)	25% (w/v)

2x T4 LIGATION COCKTAIL.

5x ligation buffer	10.0 ul
H_2O	15.0 ul
T4 DNA ligase (at 1U/ul)	0.5 ul

TCM BUFFER.

Tris.HCl (pH 7.5)	10 mM
$MgCl_2$	10 mM
$CaCl_2$	10 mM

Note: With ligations, it is generally advisable to CAP-treat vectors that have been cut with only one restriction enzyme to prevent self-ligation of the vector.

2.15.1 **CAP-TREATING VECTORS FOR LIGATION.** (B. Mansfield, Massey; personal communication).

1) After restriction-cutting approximately 1 ug of vector, add 1 ul CAP (calf alkaline phosphatase; Boehringer Mannheim) and incubate at 37°C for 1/2 hour.
2) Add an additional 1 ul CAP and leave at 37°C for another 1/2 hour.
3) Add sufficient SDS to make the solution 1% with respect to SDS. Add 0.5 mg/ml Proteinase K and incubate at 37°C for 1/2 hour.
4) Phenol/chloroform, chloroform, then ethanol precipitate.
5) Resuspend in TE.

2.15.2 **DNA LIGATIONS.**

2.15.2.1 **LIGATION: METHOD 1.** (from Massey laboratory protocols).

1) Prepare a reaction mixture cocktail for 5 ligations as follows:

distilled H_2O	23 ul
10x universal buffer (*Hae* III buffer; see 2.7.1)	5 ul
10 mM ATP	5 ul
suitably cut vector (≈ 20 ng/ul)	5 ul
T4- DNA ligase (New England Biolabs)	2 ul

2) Set up ligations in 750 ul Eppendorf tubes as follows:
   a) For the "control" add 2 ul of H_2O to 8 ul of reaction cocktail.
   b) Add 2 ul of insert DNA (≈ 100 ng/ml) to 8 ul of reaction cocktail.
3) Mix by gently vortexing and bring any liquid off the wall of the tube by spinning in an Eppendorf for a few seconds.
4) Leave in refrigerator overnight.
5) Ligations may be checked on a mini-gel.

2.15.2.2 **LIGATION: METHOD 2. (in-gel ligations)** (modified version of Struhl, 1985).

1) Excise the required fragments from a SeaPlaque agarose gel with a clean scalpel blade.

2) Melt the agarose by heating to 70°C for 10 minutes.
3) Mix approximately 5 ul of each of the vector and insert DNAs. (This may have to be adjusted slightly according to the concentrations of each, equal quantities of the DNAs being the desired effect. Concentrations may be checked on a mini-gel.) Equilibrate at 37°C. Add 10 ul of 2x T4 buffer and incubate at room temperate for 4 - 12 hours.
4) Melt ligation mixture to 70°C for 10 minutes. Check the volume and then dilute 10 fold with TCM buffer. Use this mixture for *E. coli* transformations.
5) Ligations may be checked on a mini-gel.

2.16 *BAL* 31 DNA DIGESTION. (based on Maniatis *et al.*, 1982).

Bal 31 nuclease acts as a highly specific, single-stranded endodeoxyribonuclease and exonuclease that catalyses the removal of small oligonucleotides or mononucleotides from the 5'- and 3'-termini of double- and single-stranded DNA (Maniatis *et al.*, 1982).

Under suitable conditions, the enzyme will remove nucleotides from both the 5'- and 3'-termini of a linear DNA molecule in a controlled manner. This attribute makes *Bal* 31 a useful tool for sequencing as a segment of DNA can be progressively shortened, forming a family of overlapping fragments through the region of interest, thereby eliminating the need to sub-clone smaller restriction fragments of a size suitable for sequencing in one step.

5x *Bal* 31 buffer.

		(per 100 ml)
$CaCl_2.2H_2O$	(60 mM)	0.66 g
NaCl	(3 M)	17.56 g
EDTA	(5 mM)	0.17 g
Tris	(100 mM)	1.21 g

Adjust pH to 8.1

After sterilising add $MgCl_2$ to a concentration of 60 mM (0.57 g).

1) Approximately 30 ug of the DNA is digested with the first enzyme to linearise the plasmid.

2) Check the digest on a mini-gel.

3) Extract with an equal volume of phenol:chloroform (1:1).
Extract with an equal volume of chloroform.
Ethanol precipitate and leave in $^{-}$20°C freezer for two hours.

4) Pellet DNA by centrifugation for 30 minutes at 12,000*g* in a bench-top Eppendorf centrifuge at 4°C. Resuspend in ≈ 30 ul sterile H_2O.

Trial Digest.

5) Take a small sample of the cut DNA and set up a trial digest.

linearised DNA	4.0 ul
5x *Bal* 31 buffer	50.0 ul
100 mM $MgCl_2$	31.5 ul
H_2O	160.5 ul

6) Equilibrate to 30°C.

7) Add 4 ul of *Bal* 31 (BRL;1 unit/ul) and mix quickly. Incubate at 30°C.

8) Take 20 ul samples at 2 minute time intervals. Stop each reaction by adding an equal volume of phenol and mix thoroughly.

9) Ethanol precipitate the samples.

10) Pellet DNA by centrifugation at $^{-}$4°C for 1/2 hour then vacuum-dry. Resuspend in 30 ul TE (10:0.1).

11) Load the samples on a 0.7% mini-gel along with a lambda *Hin*d III standard and determine the average number of nucleotides lost per minute. From this information the required sampling times for the fragment in question are deduced and the main *Bal* 31 digestion can be scaled-up from the trial one. (N.B. enzyme concentration and rate of digest are approximately inversely proportional).

12) Carry out the large scale digest as before and check a small amount of each sample time on a mini-gel against a standard.

13) Cut the *Bal* 31-digested DNA with the second enzyme.

14) Check on a mini-gel that the digestion has gone to completion.

15) Add 0.2 volume of SDS dye mix.

16) Load the samples onto a low-melting point [SeaPlaque; (FMC)] agarose gel.

17) Stain gel with ethidium bromide.
18) Visualise the DNA using a long-wave UV lamp.
19) Cut out the desired fragments.
20) Isolate the fragments by employing the phenol/freeze procedure (2.14.2) followed by ligations using method 1 (2.15.2.1), or proceed directly with in-gel ligations (2.15.2.2).

2.17 **PREPARATION OF COMPETENT CELLS FOR TRANSFORMATIONS.** (based on Cohen *et al.*, 1972).

Note: JM101, JM109 or XL1 were the cultures used in transformations.

1) Prepare an overnight culture of cells at 30°C in YT medium starting from a single colony. Retain this culture if transforming later in the day.
2) Sub the overnight culture 1/100 in 25 ml of YT medium and shake vigorously (300 rpm) at 37°C, 2 hours for JM101 and JM109 or 4 hours for XL1. (500 ul 20% glucose and 250 ul $CaCl_2$ were added in later experiments to increase the efficiency of transformation). (Note: In transformations using XL1, the medium for overnight growth included tetracycline at 10 ug/ml).
3) Harvest the cells in 30 ml Corex tubes by centrifugation at 5000 rpm for 10 minutes in a Sorvall SS-34 at 4°C.
4) Pour off the supernatant and resuspend in 10 ml of ice-cold 50 mM $CaCl_2$. Resuspend by pipetting gently up and down, preferably with cold pipettes. Do not vortex.
5) Leave cells on ice for 20 minutes.
6) Harvest cells at 5000 rpm for 10 minutes in a SS-34 rotor at 4°C.
7) Pour off supernatant and resuspend cells gently in 2.5 ml of ice-cold 50 mM $CaCl_2$.
8) Incubate cells on ice until ready to do transformations. Cells can be stored for a few days in this manner.

2.18 TRANSFORMATION OF *E. COLI*.

2.18.1 TRANSFORMATION OF *E. COLI* WITH PLASMID DNA. (based on Maniatis *et al.*, 1982).

1) Add 200 ul of competent cells to the ligation mixture and incubate on ice for 40 minutes.
2) To heat shock, transfer to a water bath, that has been preheated to 42°C, for 2 minutes.
3) Transfer the heat shocked cells to shaking water bath, set at 32°C, for 1 1/2 hours.
4) Add 1 ml of LB soft agar and pour onto selective plates.
5) Allow plates to set for 15 minutes then invert and incubate overnight at 37°C.

2.18.2 TRANSFORMATION OF *E. COLI* WITH M13 REPLICATIVE FORM (RF) DNA. (based on Maniatis *et al.*, 1982).

The filamentous coliphage M13 contains a single-stranded, 6.4 kb DNA molecule (Old and Primrose, 1981). This phage will only infect enteric bacteria that harbour F pili: absorption of M13 appears to be at the terminus of the F pilus. The infected cells do not lyse, but continue to grow and divide, however, at a slower rate than uninfected cells.

The single-stranded DNA enters the bacterium and is then converted to a double-stranded replicative form (RF), that multiplies until there are about 100 molecules in the cell. At this stage, the replication becomes asymmetrical due to the accumulation of a viral-encoded, single-stranded specific DNA binding protein (SSB) that prevents synthesis of the complementary strand (Lewin, 1983). After this, viral particles containing single stranded DNA are produced and then extruded from the bacterial cell.

Single-stranded filamentous phages can function as convenient vectors for DNA sequencing. In the RF state, the vector can be treated like a plasmid and both the RF and single-stranded DNA can be used to transfect *E. coli*. There is a 507 bp intergenic region in M13 that

contains the origin of replication site, (Ori) (Old and Primrose, 1981). The *E. coli lac* regulatory region and the sequence for the α-peptide of β-galactosidase, containing a versatile multiple cloning site polylinker, has been ligated into the M13 intergenic region (Old and Primrose, 1981; Yanisch-Perron *et al.*, 1985). An *E. coli* with a deletion in the β-galactosidase α-peptide region, is employed as the host strain. IPTG, which is added to the medium, functions as an inducer for β-galactosidase, while X-gal, also added to the medium, acts as a chromogenic substrate analog. When the host is transformed with M13, the α-peptide carried on the vector complements the host's *lac* deletion. X-gal is processed and a blue dye is liberated that precipitates within the cells resulting in blue plaques. Inserts in the polylinker region of the vector disrupt the α-peptide production, and result in failure of complementation and production of uncoloured plaques. There is a family of M13-based vectors available of which M13mp18 was selected for use in this project (Messing and Vieira, 1982).

1) Add 200 ul of competent cells to a ligation mixture and incubate on ice for 40 minutes.
2) Place freshly prepared M9 (+ glucose, + thiamine) plates in a 37°C incubator with the lids slightly ajar to allow the agar surface to dry.
3) Just prior to heat shocking the cells, set up the glass culture tubes in a heat block at 42°C and add to each tube in order:
   * molten LB top agar (warm to the touch) 3 ml
   * BCIG (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, or also known as X-gal; Sigma @ 20 mg/ml in dimethylformamide; stored at ⁻20°C) 20 ul
   * IPTG (isopropyl-β-D-thiogalactopyranoside; Sigma @ 24 mg/ml in H_2O; stored at ⁻20°C. 20 ul
   * Overnight culture from which the competent cells were made. 200 ml
4) Heat-shock the competent cells containing the ligation mix at 42°C for 2 minutes. Then place back on ice.

5) Vortex the top agar mixture and the heat-shocked cells together briefly and pour quickly onto the M9 plates. Allow to set for 15 minutes, invert and incubate overnight at 37°C.

2.19 PROCEDURE FOR MAKING [^{32}P]-LABELLED PROBES. (Taylor *et al.*, 1976; Whitfield *et al.*, 1982).

Reagents.

1) Deoxycytidine 5'[^{32}P] triphosphate (3000 Ci/mM, 10 mCi/ml).
2) Universal buffer 10x (*Hae* III buffer).
3) Random primers (prepared by D.B. Scott [Massey University, N.Z.] from herring sperm DNA). (25 mg/ml).
4) DNA polymerase I (Klenow fragment from Boehringer Mannheim).
5) Deoxyribonucleoside triphosphates (Sigma): dATP, (20 mM); dGTP, (20 mM); dTTP, (20 mM).
6) *Hae* III restriction enzyme (New England Biolabs).
7) Sephadex G-50 (Pharmacia, fine grade) equilibrated with TES (10 mM Tris.HCl, 1 mM Na_2EDTA, 100 mM NaCl) buffer, (pH 8.0).
8) 0.25 M Na_2EDTA, pH 8.0.
9) Phenol (equilibrated with TE (50/20) buffer, pH 8.0).
10) Chloroform.

Method.

1) Digest 0.25-1.0 ug of the DNA to be labelled with *Hae* III restriction enzyme in a 25 ul reaction mixture for 30 minutes at 37°C.
2) Add 100 ug (i.e. 4 ul) of random primers to the mix, heat to 100°C for 2 minutes and then cool on ice rapidly.
3) Add in order:-

sterile, distilled H_2O	2.5 ul
10x universal (*Hae* III) buffer	1.5 ul
dTTP	1.0 ul
dATP	1.0 ul
dGTP	1.0 ul
[^{32}P]-dCTP	3.0 ul
DNA polymerase I	1.0 ul

Mix then incubate at 37°C for 30 minutes.

4) Stop the reaction by adding 2 ul of 0.25 M Na_2EDTA. The reaction is extracted with phenol/chloroform and the aqueous phase loaded onto the Sephadex G-50 column. To prepare the column, place a small amount of glass fibre in the base of a 5 ml syringe and pack it firmly with a pair of tweezers. Add Sephadex, in TES, to the top of the syringe. Make a hole in the lid of an Eppendorf tube and insert the end of the syringe. Spin for 2 minutes at 1,085*g*. Add more Sephadex solution and spin again. The final packed column should be approximately level with the "1 ml " mark. The column is then ready to load with the aqueous phase. Spin again, collecting the DNA peak in a fresh Eppendorf tube. Store at $^{-}$20°C. DNA probes were usually labelled to a specific activity of 1-5 x 10^8 cpm/ ug of DNA.

2.20 SCREENING OF RECOMBINANT CLONES BY HYBRIDISATION OF [^{32}P]-LABELLED PROBES TO SINGLE M13 PLAQUES. (Benton and Davis, 1977).

1) Handling the 0.45 um, 8 cm diameter millipore filters (Millipore Corporation, Bedford, M.A. U.S.A.) with tweezers, lay them on top of the agar surfaces of the plates and leave for 5 minutes to allow for plaque adsorption.
2) Mark the filters in such a manner as to allow positive identification of hybridising plaques.
3) Carefully lift the filters off the plates and air-dry.
4) Bake *in vacuo* at 80°C for 2 hours.
5) Filters are pre-hybridised by placing them in a small plastic container (Tupperware). Sufficient 10x Denhardt's solution is added to cover the filters, and the container sealed with the lid. The filters are left to pre-hybridise, while gently shaking at 65°C for 2 hours.

10x Denhardt's solution. (Southern, 1975).

Hepes buffer (1 M; pH 7.0)	25.00 ml
20x SSC	75.00 ml
herring sperm DNA (28 mg/ml)	0.32 ml
E. coli tRNA (10 mg/ml)	1.00 ml
20% (w/v) SDS	2.50 ml
Ficoll (Sigma 70)	1.00 g
BSA (bovine serum albumin)	1.00 g
PVP (polyvinylpyrrolidone)	5.00 g
distilled H_2O	397.00 ml

20x SSC. (per litre) (Maniatis *et al.*, 1982).

NaCl	173.0 g
sodium citrate	88.2 g
Adjust pH to 7.0	

6) Heat probe to 100°C for 2 minutes then add to filters. Leave overnight, at 65°C, shaking gently.
7) Wash filters in 2x SSC for 15 minutes. Pour off solution and repeat two more times.
8) Dry filters between two sheets of 3MM paper at room temperature. Then cover with "Gladwrap" and expose to Ilford Curex X-ray film with Cronex intensifying screens for 1-5 days at $^{-}$70°C.

2.21 PREPARATION OF TEMPLATE DNA AND REPLICATIVE FORM (RF) DNA.

(based on Sanger *et al.*, 1977).

PEG/NaCl solution. (20% PEG in 2.5 M NaCl)

	(per 100 ml)
PEG (polyethylene glycol 6000)	20.0 g
NaCl	14.6 g

1) Set up an overnight culture of JM101 in YT medium at 37°C starting from a single colony.
2) Add 250 ul of the overnight culture to 25 ml YT medium and dispense 1 ml aliquots into culture tubes.

3) Shake vigorously (300 rpm) at 37°C for 5 1/2 to 6 hours.
4) Harvest cells by centrifugation for 5 minutes in an Eppendorf centrifuge.
5) Pour supernatant (containing the M13 bacteriophage) into another Eppendorf tube. Aspirate off the liquid remaining in the tube and retain the pelleted cells for preparation of double-stranded, replicative form (RF) DNA by the rapid boiling method (See 2.10.3).
6) Add 200 ul of PEG/NaCl solution to the supernatant. Mix then leave at room temperature for 15 minutes.
7) Centrifuge for 5 minutes in an Eppendorf centrifuge.
8) Pour or aspirate off the supernatant then respin a few seconds to bring the liquid off the wall. Carefully aspirate off the remaining supernatant. The PEG pellet of phage should be visible.
9) Add 100 ul of TE (10:0.1) and 50 ul of phenol. Vortex for 10 seconds and stand for 5 minutes. Vortex once again for 10 seconds. Centrifuge for 3 minutes.
10) Carefully remove the aqueous phase with an Eppendorf pipette and transfer to another Eppendorf tube. (If the aqueous phase has bubbles or is frothy the PEG has not been completely removed by the phenol. (If this is the case, repeat the phenol extraction in step 9).
11) Ethanol precipitate the DNA. Leave overnight.
12) Centrifuge for 15 minutes.
13) Pour off the supernatant, wash pellet with 95% ethanol and dry under vacuum.
14) Resuspend pellet in 30 ul TE (10:0.1) and check a 2 ul aliquot on a mini-gel. Store templates at $^{-}$20°C.

2.22 ^{35}S-LABELLING AND SEQUENCING OF SINGLE-STRANDED (SS) DNA.

The United States Biochemical Corporation's "Sequenase" kit was utilised for DNA sequencing, the protocol being based on the chain termination method (Sanger *et al.*, 1977). This involves the synthesis of a DNA strand by a DNA polymerase *in vitro*, using a single-stranded

template. The polymerase also has the ability to use other nucleotide analogs in sequencing. Synthesis is initiated at the site where an oligonucleotide primer anneals to the template and is terminated by the incorporation of a nucleotide analog that will not support continuation of DNA elongation. The chain termination nucleotide analogs are the 2',3'-dideoxynucleotide 5'-triphosphates (ddNTPs). These lack the 3'-OH group necessary for chain extension. dNTPs and one of the four ddNTPs are mixed in such a way that enzyme-catalysed polymerisation will be terminated in a small fraction of the population of chains at each site where the ddNTP can be incorporated. Four separate reactions, each with one of the ddNTPs, will provide the sequence information. A radioactively labelled nucleotide (in this case [α-^{35}S]-dATP) is included in the sequencing reaction to enable visualisation by autoradiography after separation by high-resolution electrophoresis on a polyacrylamide gel.

The polymerase provided in the "Sequenase" package is a genetic variant of the bacteriophage T7 polymerase created by *in vitro* genetic manipulation (Tabor and Richardson, 1987). This modification completely removes the 3'-5' exonuclease activity of the wild-type T7 DNA polymerase.

The initial part of this procedure involves annealing a synthetic oligonucleotide primer to the template. The oligonucleotide provided in the kit is suitable for use with M13-based vectors. (A different primer, 5'-CGTTCAGGACGCTACTT-3', [Schofield and Watson, 1986] was used when determining Tn5 insertion sites). DNA synthesis is carried out in two steps. First the primer is extended using limiting concentrations of the deoxynucleotide triphosphates (dNTPs), including the radioactively labelled dATP. This reaction continues until complete incorporation of the labelled nucleotide into DNA chains, which are randomly distributed in size from several to hundreds of nucleotides long. In the second step, the concentration of the dNTPs is increased and one of the ddNTPs is added to each of the four reaction tubes. DNA synthesis progresses until the chains incorporate a ddNTP. During this step the chains are extended by an average of several dozen

nucleotides. The reactions are then terminated by the addition of EDTA and formamide, denatured by heating and then separated by electrophoresis on polyacrylamide gels. More detailed descriptions of the reaction buffers and other reagents used in this procedure can be found in the booklet accompanying the "Sequenase" kit.

2.23 POLYACRYLAMIDE GEL ELECTROPHORESIS.

Polyacrylamide gels are used to analyse fragments of DNA less than 1 kb in length. For DNA sequencing, 4.0% acrylamide gels were cast. This allowed DNA fragments between approximately 30 and 400 nucleotides in length to be separated. Sequencing gels were poured between glass plates of either 36 cm in length by 20 cm in width, or on BRL MODEL S2 sequencing gel apparatus where the measurements equal 38 cm by 36 cm. The plates were separated by 0.4 mm spacer bars and run in a vertical position at 1,500 Volts, 30 mAmps and 45 Watts. Shark-tooth combs were employed to provide loading lanes.

To set up the apparatus, first wash the glass plates with detergent in warm water and rinse well. Handle the plates by the edges so as not to transfer grease from fingers onto the clean surfaces. Rinse the plates with ethanol and leave to dry. Wearing gloves and operating in a fume hood, coat the inner surface of the smaller front plate with a 2% solution of dimethyldichlorosilane in 1,1,1,-trichloroethane. The silicon provides a very smooth surface and thus encourages the gel to release from the front plate while adhering to the back plate.

Place the back plate on the bench, inner surface upper-most. Wipe the spacers with ethanol and, when dry, arrange along the sides of the back plate. Lay the front plate, silicon-side down, directly on top of the back plate and spacers and align the bottom edges. Bind the sides and bottom edges with masking tape or electrical tape to form a watertight seal.

2.23.1 SEQUENCING REAGENTS.

1) Acrylamide Mix.

urea	288.0 g
acrylamide	34.2 g
bis-acrylamide	1.8 g

Make up to about 500 ml and warm to dissolve (not greater than 40°C). Add 20 g Amberlite (Sigma A7393) and stir gently for 30 minutes. Remove the resin by filtration through a sintered glass funnel (#2) with suction. Add 10x sequencing TBE and make up to 600 ml with distilled H_2O. De-gas on a water pump and store at 4°C.

2) Ammonium persulphate. (10% stock). 500 mg/ 500 ul H_2O.

3) TEMED: N,N,N^1,N^1 - tetramethylethylenediamine (store at 4°C).

4) Formamide dye mix. Gently stir 100 ml of formamide with 5 g Amberlite MB1 (mixed bed resin) for 30 minutes. Remove resin by filtration. Add 0.03 g xylene cyanol FF, 0.03 g bromophenol blue and Na_2EDTA to 20 mM (i.e. 0.74 g). Store at room temperature.

To prepare acrylamide:

Pre-warm the 40 ml acrylamide mix to room temperature. Add 240 ul 10% NH_4 persulphate and 24 ul TEMED. (For the larger BRL plates, use 60 ml acrylamide, 360 ul 10% NH_4 persulphate and 36 ul TEMED).

2.23.2 GEL PREPARATION.

1) Pour gel ensuring that there are no air bubbles.
2) Insert the clean comb, flat edge down, making sure that the depth does not exceed 4 mm.
3) Allow gel to set for a minimum of 2 hours or, if storing overnight, seal the top of the plates in "Gladwrap".
4) Wash off the surface acrylamide with warm H_2O.
5) Cut the tape off the bottom of the plates.
6) Add water around the comb, lift and remove carefully .
7) Assemble for sequencing.
8) Fill the top and bottom gel box chambers with sequencing TBE.
9) Insert the comb, shark-toothed edge down. Aim to keep the wells as flat as possible.

2.23.3 **RUNNING SEQUENCING GELS.**

1) Thaw the ^{35}S-labelled extended primers and leave at 70°C for 2 minutes.
2) Load a 3 ul sample into a well, avoiding air bubbles in the process.
3) Gels were run at 1,500 V, 30 mA and 45 W. Short gels were run for 2.5 hours and long ones for 4 hours.
4) Gels were then fixed for 15 minutes in a solution comprising 10% ethanol and 10% glacial acetic acid in 80% H_2O.
5) These were vacuum-dried at 80°C for one hour on a Bio-Rad Slab Dryer, model 483, exposed to Fuji Nif RX 100 X-ray film overnight and then developed in a Kodak RP X-OMAT processor, model M6B.

3.0 RESTRICTION MAPPING.

The 233 Tn5 mutation had been mapped to the 7.1 kb *Eco* RI fragment approximately 1.6 kb from the right-hand-side of the map, (see Figure 3.1). As smaller, more manageable-sized pieces were required for sequencing, it was necessary to construct a restriction map for the fragment. Careful selection of the enzymes used would also enable a more accurate determination of the Tn5 insertion sites in the 7.1 kb fragment.

3.1 MAPPING STRATEGY.

Large scale plasmid preparations of pPN354 and of pPN301 were made as outlined in 2.10.1. Trial digests of the plasmids were performed using the following enzymes; *Taq* I, *Pst* I, *Bam* HI, *Alu* I, *Kpn* I, *Sal* I and *Bgl* II. The six base pair cutters, *Bam* HI and *Sal* I cleaved the plasmids into fragments of a size suitable for subcloning. There is one *Bam* HI and one *Sal* I restriction site in pBR328. In addition, each enzyme cuts Tn5 once asymmetrically, yielding two fragments of 2.7 kb and 3.0 kb in both cases (Jorgensen *et al.*, 1979). These characteristics suggested the selection of *Sal* I and *Bam* HI as mapping enzymes.

Single and double digests were set up for pPN354 and pPN301 (see 2.11 for digest protocol). Samples were checked for digest completion on a mini-gel and then run on a 0.7% agarose mini-gel against a lambda *Hind* III standard (see Figure 3.2).

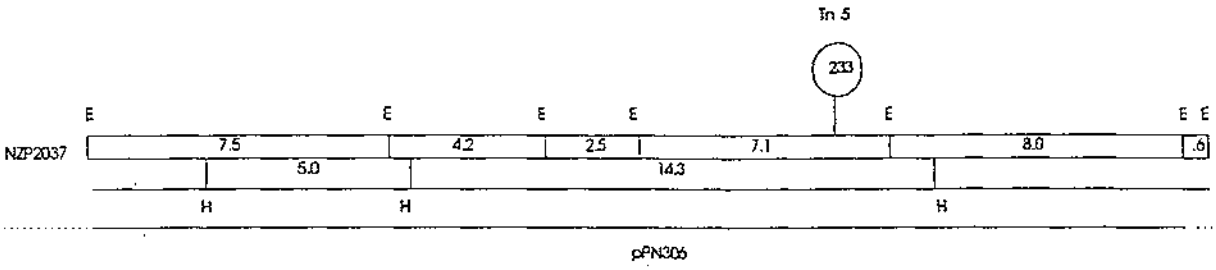
Four other *R. loti* strains, PN4016, PN4019, PN4047 and PN4053, bearing Tn5 insertions in the 7.1 kb *Eco* RI chromosomal fragment were also mapped. However, the strategy used in this case was slightly different to that used to locate the 233 Tn5 and is discussed in section 3.2.

The fragments into which the Tn5 had inserted could be determined by comparing the restriction patterns of the pPN354 and pPN301 single digests (see Table 3.1). Because the Tn5, which is 5.7 kb in length (Berg *et al.*, 1982), possesses one *Bam* HI and one *Sal* I restriction

Figure 3.1: An *Eco* RI (E) and *Hind* III (H) restriction enzyme map of the *nod* region of *R. loti* NZP2037. The approximate location of the Tn5 insertion in the *Nod*$^-$ strain PN233 is indicated. The region sub-cloned into pLAFR1 and designated pPN306 is also illustrated. Dashed lines represent vector DNA.

Figure 3.2: An agarose gel of *Sal* I (S), *Bam* HI (B) and *Eco* RI (E) single and double digests.

1. lambda *Hind* III standard (in kb).
2. pPN301 *Sal* I cut.
3. pPN301 *Bam* HI, *Sal* I cut.
4. pPN301 *Bam* HI cut.
5. pPN354 *Sal* I cut.
6. pPN354 *Bam* HI, *Sal* I cut.
7. pPN354 *Sal* I cut.
8. pPN301 *Eco* RI cut.
9. pPN301 *Eco* RI, *Bam* HI cut.
10. pPN354 *Eco* RI, *Sal* I cut.
11. pPN354 *Eco* RI, *Sal* I cut.


Tn 5
233
E
E
E
E
E
E E
NZP2037
7.5
4.2
2.5
7.1
8.0
.6
5.0
14.3
H
H
H
pPN306

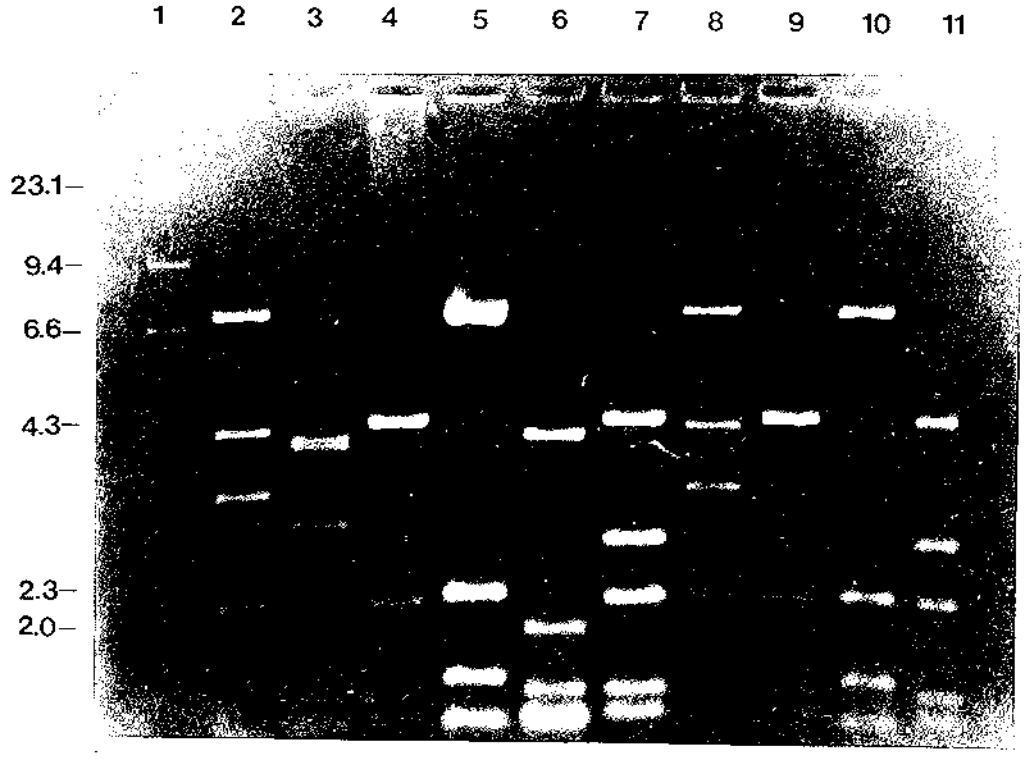

1
2
3
4
5
6
7
8
9
10
11
23.1
9.4
6.6
4.3
2.3
2.0

Table 3.1: Restriction fragments from *Bam* HI and *Sal* I cut pPN354 and pPN301 (kb).

pPN354 *Sal* I	pPN301 *Sal* I	pPN354 *Sal* I & *Bam* HI	pPN301 *Sal* I & *Bam* HI	pPN301 *Bam* HI	pPN354 *Bam* HI
7.1	7.1				
				4.2 x 3	4.2
	4.0	4.0	4.0		
			3.8		
	3.2				
			2.8		2.8
2.3	2.3				
				2.2	2.2
		1.9	1.9		
1.45					
		1.35	1.35	1.35	1.35
1.2	1.2	1.2 x 3	1.2 x 2	1.2	1.2
		0.35	0.35		
		0.3	0.3 x 2		
		0.25	0.25		
		0.2	0.2	0.2	0.2

site (Jorgensen *et al.*, 1979), the pPN301 *Sal* I or *Bam* HI digests will exhibit the loss of one restriction fragment and the appearance of two new fragments. That is, the 1.45 kb *Sal* I fragment in pPN354 was absent from the pPN301 digested with *Sal* I, but two new fragments, which were 4.0 and 3.2 kb in length, had appeared. The sum of the sizes of the two new fragments should be approximately equal to the total achieved by adding 1.45 kb to the 5.7 kb of insert DNA.

4.0 + 3.2 = 7.2 kb; 1.45 + 5.7 = 7.15 kb

From this, it was concluded that the Tn5 had inserted into the 1.45 kb *Sal* I fragment and by using the same logic, into the 2.8 kb *Bam* HI fragment.

Double digests were utilised to establish a restriction map for the 7.1 kb fragment. A single *Bam* HI fragment containing a *Sal* I site, is replaced by two smaller fragments in a double digest. The same applies for a *Sal* I fragment possessing a *Bam* HI site. Therefore, novel fragments in a double digest are the products of internal *Bam* HI sites in *Sal* I fragments and *vice versa*. Some of the *Bam* HI fragments did not contain internal *Sal* I sites and thus could not be ordered.

Other information which was used to develop the map was:

1) pBR328, when cut with *Sal* I, gives fragments of 1.7 kb and 3.2 kb. Therefore, these are the smallest sized fragments in the pPN354 digest that can contain vector DNA. Consequently, the 7.1 kb and 2.3 kb *Sal* I fragments in the pPN354 *Sal* I digest carry vector DNA.
2) pBR328, when cut with *Bam* HI, produces fragments of 1.45 kb and 3.5 kb in length, which are the smallest sized fragments in the *Bam* HI pPN354 digest that can contain vector DNA. The 4.2 kb and the 2.8 kb or 2.2 kb fragments must therefore contain vector DNA.
3) It had been determined that the 1.45 kb *Sal* I and 2.8 kb *Bam* HI fragments contained the insert DNA in pPN301. Therefore, these two fragments must overlap.
4) The Tn5 had been located approximately 1.6 kb from the right-hand-side *Eco* RI site. The transposon must also be inserted in such a position in the 1.45 kb *Sal* I fragment as to yield two fragments of 4.0 kb and 3.2 kb when pPN301 is digested with *Sal* I. Likewise, it

is necessary for the Tn*5* to be in a given position in the 2.8 kb *Bam* HI fragment to result in two 4.2 kb fragments arising from a pPN301 *Bam* HI digest.

The restriction map for the 7.1 kb *Eco* RI fragment (see Figure 3.3) was thus constructed from this body of information and utilised for the determination of Tn*5* insertion site mapping.

3.2 Tn*5* INSERTION SITE MAPPING.

The *R. loti* strains PN4016, PN4019, PN4047 and PN4053 each had a chromosomal Tn*5* insertion in the 7.1 kb *Eco* RI fragment. Four *E. coli* strains carrying these Tn*5* inserts on pPN306 (see Figure 3.1) were constructed using Tn*5* mutagenesis. These *R. loti* and *E. coli* strains were kindly made available by C. Pankhurst, D.S.I.R., Palmerston North.

The 1.5 kb inverted repeats on either side of Tn*5* contain a *Hin*d III restriction site. Consequently, when Tn*5*, which possesses no other *Hin*d III sites, is digested with this enzyme one internal 3.3 kb fragment and two 1.2 kb or greater end fragments are generated. This feature can be utilised to map the position of Tn*5* in the 7.1 kb *Eco* RI fragment.

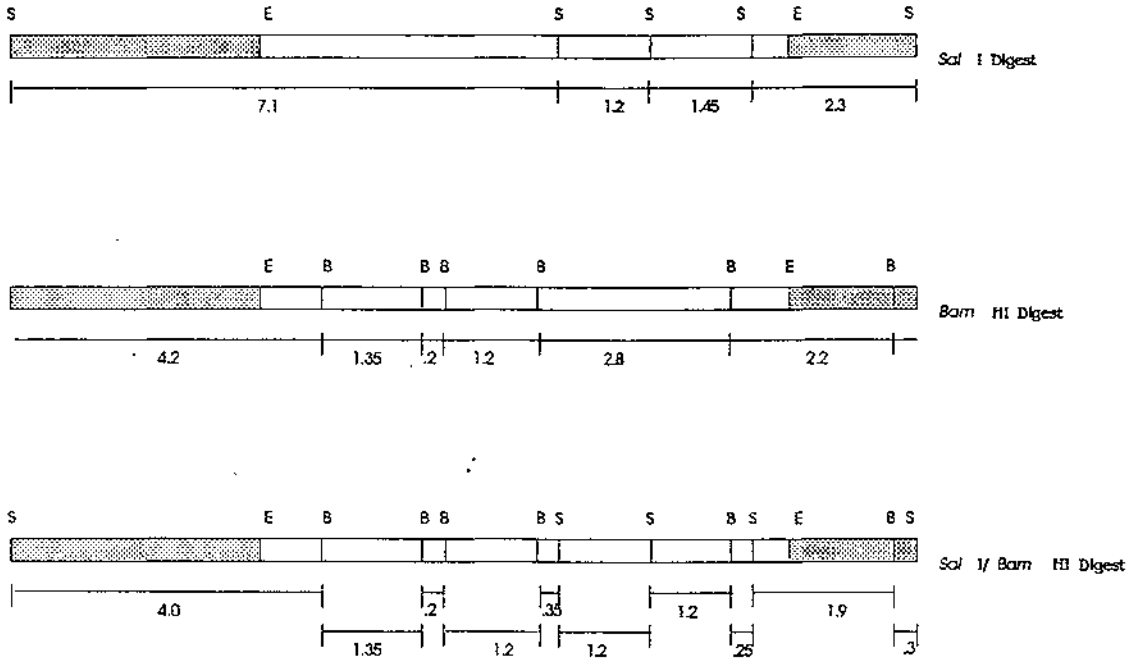
pPN306 carrying Tn*5* insertions, when digested with *Hin*d III, will exhibit three fragments whose sizes are constant. These are: the vector plus part of the 9.6 kb *Hin*d III fragment and approximately 3.25 kb of DNA from the far left-hand-side of the restriction map, the adjacent 5.0 kb *Hin*d III fragment and the internal 3.3 kb Tn*5* fragment. The two other fragments will vary in size depending on the position of the Tn*5* in the 14.3 kb *Hin*d III fragment (see Table 3.2). One can determine the approximate insertion site by subtracting 1.2 kb (the Tn*5* *Hin*d III -cut terminal ends) from the sizes of the two variable-lengthed fragments.

For example, the variable fragments in the *Hin*d III digest of pPN306 that carried the 4053 Tn*5* inserts, were 14.5 kb and 2.2 kb in length. Subtraction of 1.2 kb from 14.5 gives 13.3, which is the distance the 4053 insertion site is from the left-hand-side of the *Hin*d III site on the 14.3 kb fragment, as opposed to the right-hand

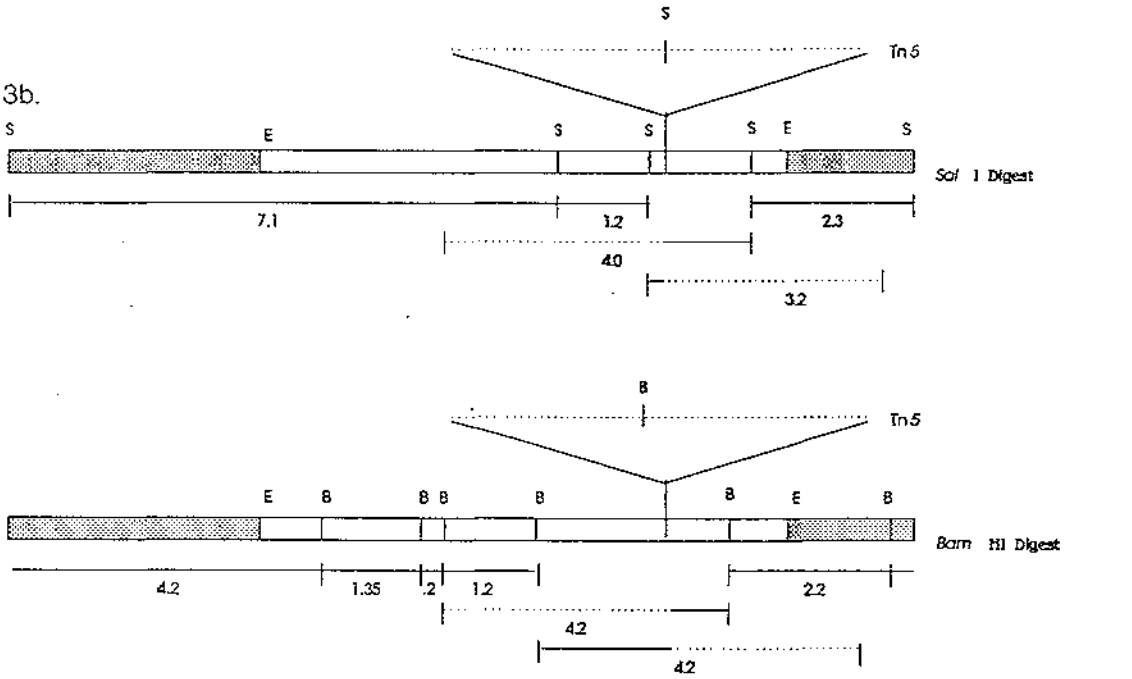
Figure 3.3: *Sal* I and *Bam* HI restriction enzyme maps for the 7.1 kb *Eco* RI fragment.
Shaded boxes represent the vector pBR328 while open boxes represent the inserted *R. loti* chromosomal DNA.
a. pPN354 *Sal* I, *Bam* HI and *Sal* I/*Bam* HI double digests.

b. pPN301 *Sal* I, *Bam* HI and *Sal* I/*Bam* HI double digests. The point of Tn5 insertion is represented as the bar connecting the triangle to the insert DNA. Tn5 DNA is represented by dashed lines.

3a.


S E S S S E S
Sal I Digest
7.1 1.2 1.45 2.3
E B B B B E B
Bam HI Digest
4.2 1.35 .2 1.2 2.8 2.2
S E B B B S S B S E B S
Sal I/ Bam HI Digest
4.0 .2 .35 1.2 1.9
1.35 1.2 1.2 .25 .3

3b.


S
Tn5
S E S S S E S
Sal I Digest
7.1 1.2 2.3
4.0
3.2
B
Tn5
E B B B B E B
Bam HI Digest
4.2 1.35 .2 1.2 2.2
4.2
4.2

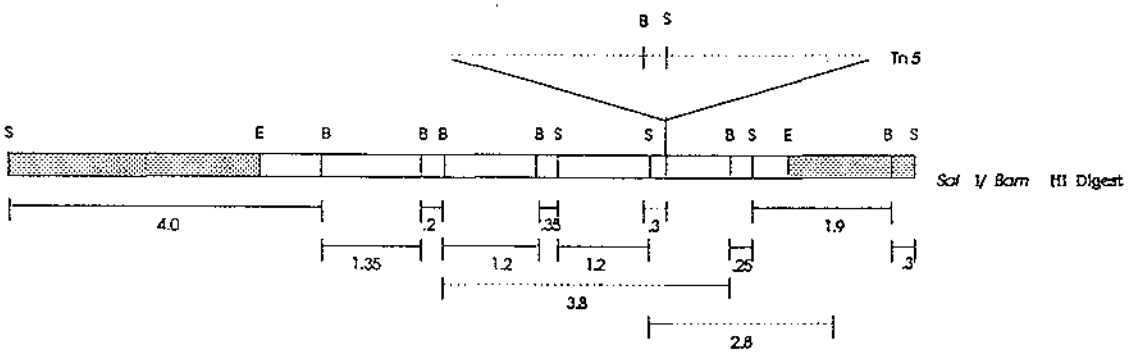

B S
Tn5
S E B B B S S B S E B S
Sal I/ Bam HI Digest
4.0 .2 .35 .3 1.9
1.35 1.2 1.2 .25 .3
3.8
2.8

Table 3.2: *Hind* III digests of the Tn*5* insertions in pPN306 (kb).

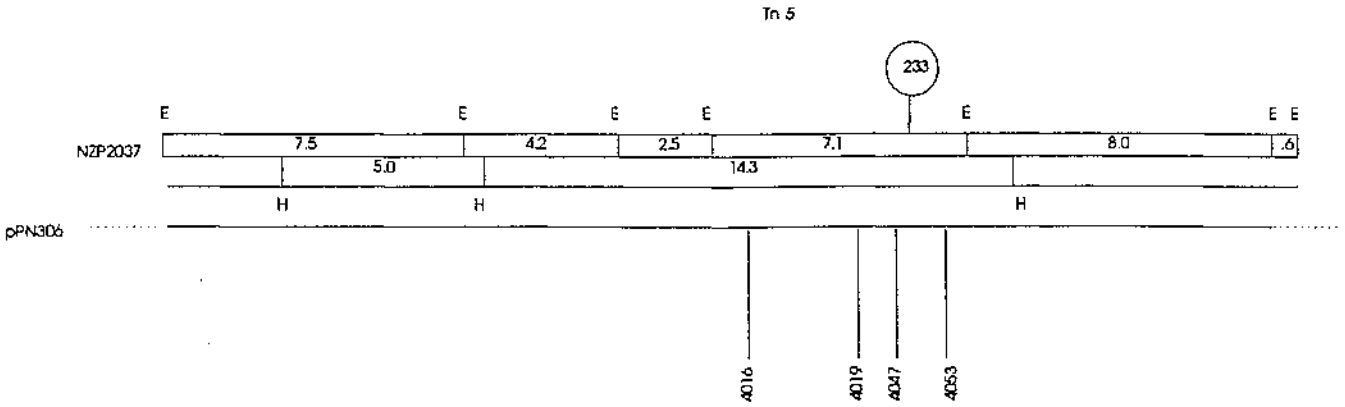
4016	4019	4047	4053	
32.0	32.0	32.0	32.0	
5.0	5.0	5.0	5.0	
3.3	3.3	3.3	3.3	
8.4	12.2	13.0	14.5	} variable-lengthed
8.4	4.5	3.6	2.2	} fragments
57.1	57.0	56.9	57.0	totals

-side, as this would put the Tn5 outside the 7.1 kb *Eco* RI fragment. The same process was used to determine the approximate locations of the other three Tn5 inserts (see Figure 3.4).

3.3 DISCUSSION.

Bam HI and *Sal* I restricted the 7.1 kb DNA into segments of a size suitable for sequencing and, of these, the 233 Tn5 insertion site was located on the 1.45 kb *Sal* I and 2.8 kb *Bam* HI fragments. Restriction mapping placed the 4047 Tn5 insertion on the adjacent 1.2 kb *Sal* I fragment. As this insertion, like that of 233, has a Nod^- phenotype (see Table 4.2), it suggested that the gene affected may extend into the 1.2 kb *Sal* I fragment. This information together with that gleaned from complementation experiments (Chapter 4.0), suggested that the 1.45 kb and 1.2 kb *Sal* I and the overlapping 2.8 kb *Bam* HI fragments were suitable candidates for sequencing.

Figure 3.4: Tn5 insertion sites in pPN306 as determined from *Hind* III restriction digests.


Tn 5
233
E
E
E
E
E
E E
NZP2037
7.5
4.2
2.5
7.1
8.0
.6
5.0
14.3
H
H
H
pPN306
4016
4019
4047
4053

4.0 PLANT COMPLEMENTATION EXPERIMENTS.

The first step in the characterisation of the PN233 Nod$^-$ mutant was to determine whether the Tn5 had insertionally inactivated one of the common *nod* genes.

4.1 COMMON *NOD* GENE COMPLEMENTATION TESTS (EXPERIMENT 1).

Four *R.l.* bv. *trifolii* strains that were each mutant for either *nod* A,B,C or D were available. It was decided to transfer, via triparental crosses, the plasmid pPN305 (see Figure 4.1) into each of the *R.l.* bv. *trifolii nod* mutants to discover whether this 22.7 kb region of the wild-type NZP2037 genome could complement any of the above mutations.

A *R.l.* bv. *trifolii nod* cosmid, pPN26, was used as a control in the complementation tests. A loopful of each of the triparental crosses (see 2.4) was streaked out for single colony isolates on S10 Neo Tet (Neomycin and Tetracycline) plates. Two colonies from each cross were cultured and then used separately to inoculate plates containing two *T. repens* seedlings. The clover seeds were sterilised, germinated and grown on FM plates as outlined in section 2.5. The inoculated seedlings were placed in a growth cabinet, set at 22°C, for two weeks, after which they were examined for nodule formation. Wild-type *R.l.* bv. *trifolii* and *R. loti* were also inoculated onto *T. repens* seedlings as controls. The results of the experiment are shown in Table 4.1.

The experiment demonstrated the ability of pPN305 to complement the *nod*C mutation in *R.l.* bv. *trifolii* (see Figure 4.2). This indicates the functional presence of the *R. loti nod*C gene somewhere on the 22.7 kb fragment carried by pPN305. Given the usual clustering of *nod*DABCIJ genes in many rhizobia, it was surprising that none of the other *nod* gene mutations was complemented. In order to home in on the area of interest and to investigate the complementation results further, another set of complementation tests was conducted. This is discussed in the following section.

Figure 4.1: Plasmids pPN305 and pPN25.
Dashed lines represent the vector pLAFR1.

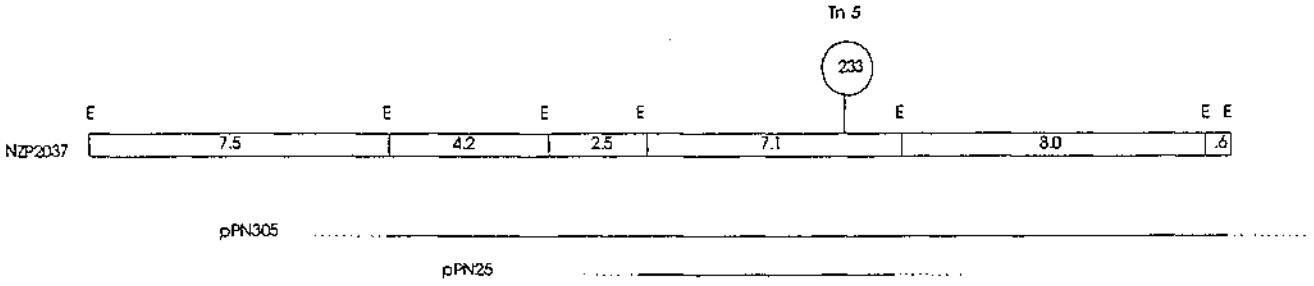

Tn 5
233
E
E
E
E
E
E E
NZP2037
7.5
4.2
2.5
7.1
8.0
.6
pPN305
pPN25

Figure 4.2: Complementation of a *R.l.* bv. *trifolii* *nod*C mutant by pPN305.

1. ANU277 (*nod*C)/pPN305
2. ANU851 (*nod*D)/pPN305
3. ANU252 (*nod*A)/pPN305
4. ANU249 (*nod*B)/pPN305

Nodules and healthy growth can be seen on the plant inoculated with the ANU277/pPN305 exconjugant.

1 2 3 4

Table 4.1: Complementation of *R.l.* bv. *trifolii* common *nod* genes.

Crosses/Strains	Nodule phenotype
ANU249 (*nod*B$^-$)/pPN305	Nod$^-$
ANU252 (*nod*A$^-$)/pPN305	Nod$^-$
ANU277 (*nod*C$^-$)/pPN305	Nod$^+$
ANU851 (*nod*D$^-$)/pPN305	Nod$^-$
ANU249 (*nod*B$^-$)/pPN26	Nod$^+$
ANU252 (*nod*A$^-$)/pPN26	Nod$^+$
ANU277 (*nod*C$^-$)/pPN26	Nod$^+$
ANU851 (*nod*D$^-$)/pPN26	Nod$^+$
NZP2037	Nod$^-$
PN100	Nod$^+$

KEY: Nod$^+$ = nodulation; Nod$^-$ = no nodulation.

Table 4.2: The phenotypes of four other *R. loti* Tn*5* insertions (4016, 4019, 4047 and 4053) in the 7.1 kb *Eco* RI region.

	Phenotype on:	
Strain:	*L. pedunculatus*	*L. corniculatus; L. tenuis*
PN4016	Nod$^+$ Fix$^+$	Nod$^+$ Fix$^+$
PN4019	Nod$^+$ Fix^{+*}	Nod$^+$ Fix$^+$
PN4047	Nod$^-$	Nod$^-$
PN4053	Nod^d Fix$^-$	Nod$^+$ Fix$^+$

KEY: Nod$^+$ Fix$^+$ = normal nodulation; Nod$^+$ Fix^{+*} = 30-50% Fix$^+$ of wild-type; Nod$^-$ = no nodulation; Nod^d Fix$^-$ = slightly delayed nodulation but no fixation.

4.2 COMPLEMENTATION OF THE Tn*5* MUTATIONS (EXPERIMENT 2).

The 233 Tn*5* insertion site had been mapped to the 7.1 kb *Eco* RI fragment. The phenotypes of the four other *R. loti* strains, PN4016, PN4019, PN4047 and PN4053, whose approximate sites of Tn*5* insertions were determined in section 3.2, are outlined in Table 4.2.

Further experiments were conducted to determine whether complementation occurred when pPN305, pPN25 (see Figure 4.1) and pPN26 were crossed into the original Tn*5* mutant (PN233), PN4047, PN4053 and the *R.l.* bv. *trifolii* mutant, ANU277.

Complementation experiments conducted by Evans and Downie (1986) demonstrated that the effects of Tn*5* mutations depended on whether the insertion was on the indigenous symbiotic plasmid, or on DNA cloned into the pLAFR1 vector. In response to this observation and, to determine whether the duplication of the 22.7 kb fragment in complementation tests had an effect on nodulation, pPN305 was crossed into NZP2037 as a control.

For this experiment the seedlings were placed on FM slopes in glass tubes (see 2.5.2). Four seedlings were used in a complementation test. Two single colony isolates for each triparental cross were cultured separately and used to inoculate a pair of seedlings each, in a complementation test. Two uninoculated *T. repens* and two uninoculated *Lotus* plants, as well as two clover and two *L. pedunculatus* seedlings inoculated with their respective microsymbionts, were grown alongside the complementation test plants as controls. Results of the experiments were recorded six weeks after inoculation (see Table 4.3 and Figure 4.3).

4.3 DISCUSSION.

The test results highlighted an unusual aspect of the 233 mutation. While the smaller 7.1 kb *Eco* RI fragment complemented the Tn*5* mutant, the larger 22.7 kb fragment, of which the 7.1 kb was a sub-clone, could not induce nodulation. Slight swellings, with no bacterial invasion or fixation were observed. Both fragments could, however, complement the surrounding mutations, i.e. 4047 and 4053. The presence of the 22.7 kb fragment in NZP2037 and the fact that pPN305

Table 4.3: Complementation results following the introduction of plasmids into the Tn5 mutants.

	Tested on *L. pedunculatus*				Tested on *T. repens*
Strains	NZP2037	PN233	PN4047	PN4053	ANU277
Plasmids pPN305 (22.7 kb)	Nod^{d}	Nod^{-T}	Nod^{d}	Nod^{d}	Nod^{+}
pPN25 (7.1 kb)	ND	Nod^{d}	Nod^{d}	Nod^{d}	Nod^{-}
pPN26	ND	Nod^{-T}	leaky, Nod±	leaky, Nod±	Nod^{+} (see Table 4.1)

Key: Nod^{d} = slight delay in nodulation; Nod^{-T} = ineffective tumorous growths; ND = not done; Nod± = some plants formed effective nodules, others did not.

Figure 4.3: Photographs and micrographs of effective and ineffective complementations.

a. From left to right; *L. pedunculatus* inoculated with NZP2037/pPN305, PN4047/pPN305 and PN4053/pPN305 exconjugants. The 22.7 kb fragment complemented the PN4047 and PN4053 Tn*5* mutants and did not adversely effect the wild-type NZP2037.

b. From left to right; *L. pedunculatus* (x2) uninoculated, inoculated with PN233/pPN305 (x2), NZP2037 (x2) and PN233/pPN25 (x2) exconjugants. The 7.1 kb, but not the 22.7 kb fragment, complemented the PN233 mutant. Ineffective tumorous swellings like those formed by the PN233/pPN305 exconjugant can be seen more clearly in Figure 4.3 f.

c. A micrograph of an effective nodule formed on *L. pedunculatus* by the PN233/pPN25 exconjugant. C = cortex, B = bacteroid-filled cells. magnification = x150

d. A micrograph of an ineffective tumorous swelling formed on *L. pedunculatus* by the PN233/pPN305 exconjugant. T = tannins; C = cortex; S = starch granules. magnification = x100.

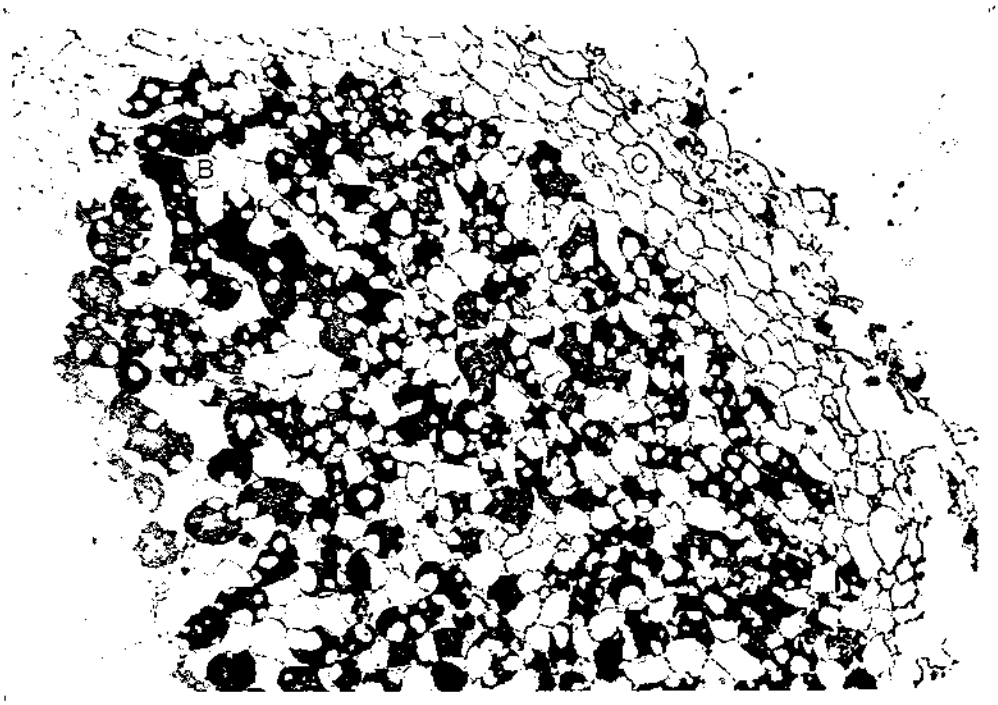
e. From left to right; an uninoculated *T. repens*, *T. repens* inoculated with the exconjugant ANU277/pPN25 and with the *R.l.* bv. *trifolii* wild-type, PN100.

f. From left to right; *L. pedunculatus* inoculated with the PN233/pPN26 exconjugant, with the wild-type NZP2037 and an uninoculated *L. pedunculatus*. The tumorous swellings produced by the PN233/pPN26 exconjugant are visible as pink blotches on the root system.

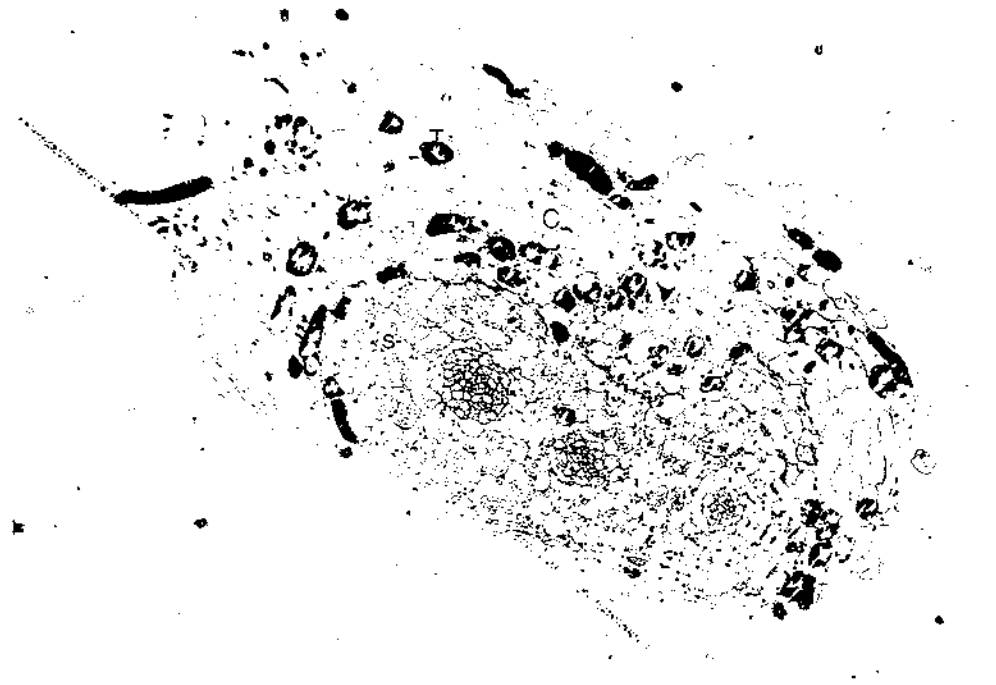
a

b

c



d



e

f

complemented the other Tn5 mutations, indicate that the phenomenon is unlikely to be a gene dosage effect. One possibility is that another gene or genes, located on the 22.7 kb fragment outside the region encompassed by the 7.1 kb *Eco* RI sub-clone, is interacting negatively with the protein product of this particular mutation.

The 22.7 kb fragment from *R. loti* was able to complement the *R.l.* bv. *trifolii* ANU277 *nod*C mutation, indicating the presence of *nod*C somewhere on the fragment. The 7.1 kb sub-clone was unable to complement. It was possible that the Tn5 in PN233 had inserted into *nod*C and that the 7.1 kb region did not contain the whole gene sequence. However, the *R.l.* bv. *trifolii nod* cosmid pPN26 did not complement the *R. loti* PN233 strain either. One would expect this cosmid to complement the Tn5 mutation as it carries all the nodulation genes.

The two other Tn5 mutant strains, PN4047 and PN4053, into which pPN26 had been transferred, displayed a leaky phenotype on *L. pedunculatus*, with some plants developing nodules and others tumorous swellings. This suggests that the genetic background of the host plant may be the deciding factor in determining whether complementation, and hence nodulation, is achieved. As to why the presence of the pPN26 plasmid resulted in a poorer nodulating ability than the Tn5 insertion alone in PN4053 is unclear. It is possible that this represents a gene dosage effect, but then one would expect similar results in the the crosses where pPN26 was transferred into the *R.l.* bv. *trifolii* common *nod* gene mutants, unless complementation failure is due to the *L. pedunculatus* background.

PN233 behaved differently from PN4047 and PN4053 in the presence of pPN26 as complementation was not observed in any plants, once again emphasising the unusual nature of PN233 in comparison with the other Tn5 mutant strains. Some of the questions raised in this section can only be answered in conjunction with information gained from sequence data and, hence, will be dealt with in the final discussion chapter.

The slight delay observed in nodule initiation in the tests may well have been caused by the presence of pLAFR1, the duplication of the 22.7 kb in the rhizobial cell, or a combination of the two. This is supported by the delayed nodule initiation exhibited in the NZP2037/pPN305 exconjugant.

Although the results were recorded six weeks after inoculation, the plants were not disposed of until a couple of weeks later. During this time, a nodule developed on one *Lotus* plant that had been inoculated with PN233 carrying pPN305. The appearance of a nodule in such a cross, after an extended period of unsuccessful nodule initiation, had been observed before at a low frequency (D.B. Scott, personal communication). In these instances, it is possible that a double recombination event had occurred in the bacterium, resulting in the replacement of the DNA region carrying the Tn*5* insert with the wild-type DNA introduced by the pPN305 cosmid (Scott *et al.*, 1985).

Positive identification of the gene into which the Tn*5* carried by PN233 had inserted necessitated sequencing that region of DNA. The area delineated by the Nod^+ Fix^+ phenotype of 4019 and the "leaky" 4053 mutation, in conjunction with the restriction mapping and complementation experiments, determined the selection for sequencing of the region encompassed by the 1.2 and 1.45 kb *Sal* I and the 2.8 kb *Bam* HI fragments. This is discussed in the following chapter.

5.0 DNA SEQUENCING OF THE *R. LOTI* NODULATION GENES.

The original Tn5 insert in strain PN233 was located in the 1.45 kb *Sal* I fragment at the far left-hand-side (see Figure 3.3). The phenotypes of the other neighbouring Tn5 mutations suggested that the gene may have extended into the adjacent 1.2 kb *Sal* I fragment. Therefore, it was decided to sub-clone these two fragments for sequencing purposes. As the 2.8 kb *Bam* HI fragment overlapped the *Sal* I site, this was also sub-cloned to confirm the sequence through the *Sal* I restriction site.

5.1 FRAGMENT SUB-CLONING.

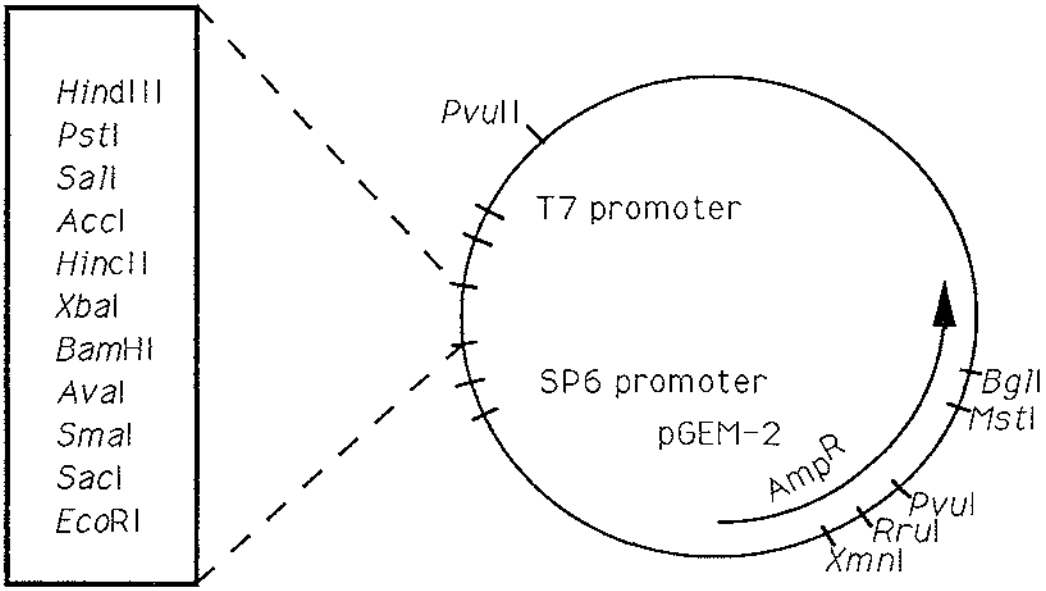
A stock of pPN354 (pBR328 + 7.1 kb *Eco* RI fragment) was made using the alkaline lysis method (see 2.10.1). Two digests were carried out, one with *Sal* I, the other with *Bam* HI, and small samples checked on a mini-gel to confirm digest completion. The remainder was loaded onto a 1%, 100 ml SeaPlaque gel and the digests run alongside a lambda standard at 40 V for 16 hours. The 2.8 kb *Bam* HI and the 1.2 kb and 1.45 kb *Sal* I fragments were excised from the gel and electroluted (see 2.14.1).

The *Sal* I and *Bam* HI fragments were sub-cloned into pGEM-2 (see Figure 5.1). pGEM-2, which is a small, high copy number plasmid with a convenient multiple cloning site cassette bounded by *Eco* RI and *Hind* III sites, was developed by Promega Biotec (Melton *et al.*, 1984). HB101 was transformed with pGEM-2 and a large scale plasmid preparation was made (2.10.1). Samples were then cut with *Sal* I or *Bam* HI, the vector CAP-treated and the 1.2 kb and 1.45 kb *Sal* I and the 2.8 kb *Bam* HI fragments ligated into the prepared vectors.

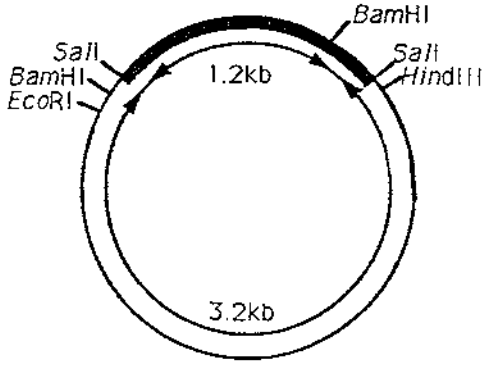
HB101 was transformed with the resulting constructs and then plated out on selective medium to obtain colonies. Ampicillin-resistant, single colony isolates were cultured and small scale rapid boil preparations made (see 2.10.3). The plasmids were then cut with *Sal* I or *Bam* HI and run on a mini-gel to check for the presence of the correct inserts.

Figure 5.1: The vector pGEM-2.
Amp^R = ampicillin resistance gene.
The multiple cloning sites are boxed.

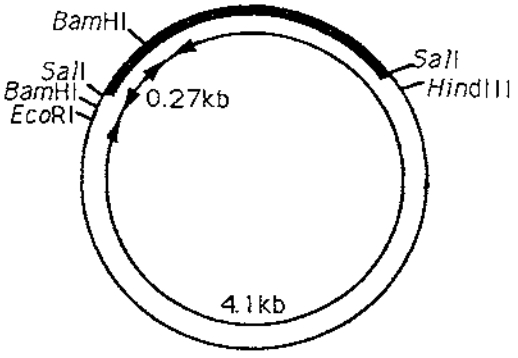
Figure 5.2: Plasmids pJ14011 and pJ14012 that bear the 1.45 kb *Sal* I fragment in opposite orientations.



pJ14011



pJ14012



5.2 ORIENTATION OF THE *SAL* I AND *BAM* HI INSERTS.

It had already been deduced from the restriction map of the 7.1 kb *Eco* RI fragment that the 1.45 kb *Sal* I fragment contained a *Bam* HI site located near the far right-hand-end with respect to the map (see Figure 3.3). This permitted the orientation of the 1.45 kb fragments, ligated into pGEM-2, to be determined. Depending on fragment orientation, *Bam* HI restriction would produce fragments of approximately 1.2 kb and 3.2 kb or, 0.27 kb and 4.1 kb in size. The two orientations were designated #11 and #12 and the plasmids labelled pJ14011 and pJ14012 respectively (see Figure 5.2).

As *Bam* HI inserts could not be orientated using *Sal* I sites, *Pst* I, *Pvu* I, *Sma* I, *Ava* I and *Sph* I digests were conducted to ascertain whether these restriction enzymes would be useful for orientating the 2.8 kb fragment. Of these endonucleases, *Sph* I proved suitable. Two *Sph* I sites were found in the *Bam* HI fragment. In one orientation, restriction with *Sph* I resulted in 0.7 kb, 1.8 kb and 3.2 kb fragments being liberated; in the other, 1.2 kb, 1.8 kb and 2.7 kb were formed. The two orientations were designated #3 and #6, as found in plasmids pJ28003 and pJ28006 respectively (see Figure 5.3).

Sph I digests were also employed to determine the orientation of the *Bam* HI 2.8 kb inserts in relation to those of the 1.45 kb *Sal* I fragments in pGEM-2. If the 2.8 kb fragment was in pGEM-2 in the same orientation as in 1.45 kb #12, then pJ14012, when cut with *Sph* I, would yield fragments of 1.0 kb and 3.35 kb in length, and if in the opposite orientation, 1.4 kb and 2.95 kb (see Figure 5.4). *Sph* I digests demonstrated that 1.45 #12 and 2.8 #3 (i.e. pJ14012 and pJ28003) were similarly aligned, as were 1.45 #11 and 2.8 #6 (i.e. pJ14011 and pJ28006).

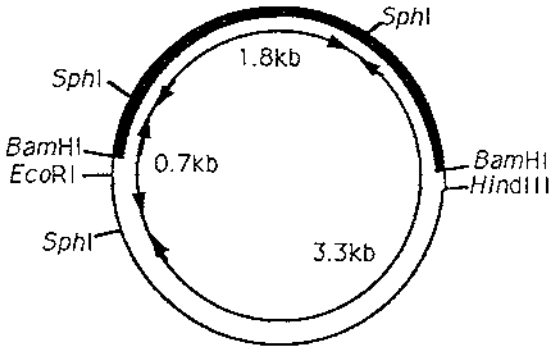
Pvu II digestion allowed the two different orientations of the 1.2 kb *Sal* I inserts in pGEM-2 to be identified, (1.2 #9 in pJ12009, 1.2 #10 in pJ12010; see Figure 5.5). Their relationship to the orientations of the 1.45 kb *Sal* I and 2.8 kb *Bam* HI inserts however, was not clarified until sequencing was initiated. 1.2 #10 was found to match the orientations of 1.45 #12 and 2.8 #3, while the orientation of 1.2 #9 was similar to that of 1.45 #11 and 2.8 #6.

Figure 5.3: Plasmids pJ28003 and pJ28006 bearing the 2.8 kb *Bam* HI fragment in opposite orientations.

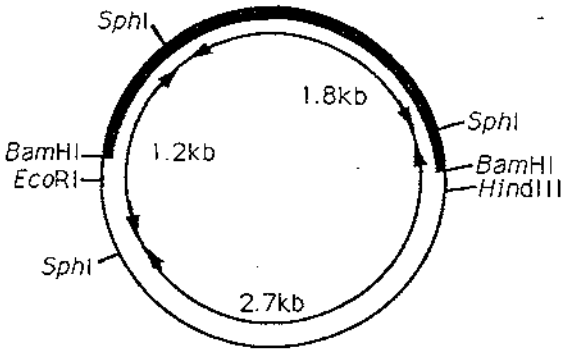
Figure 5.4: Determination of the orientation of the 1.45 kb *Sal* I fragments in relation to the 2.8 kb *Bam* HI fragments using *Sph* I sites.

Figure 5.5: Plasmids pJ12009 and pJ12010 bearing the 1.2 kb *Sal* I fragments in opposite orientations.

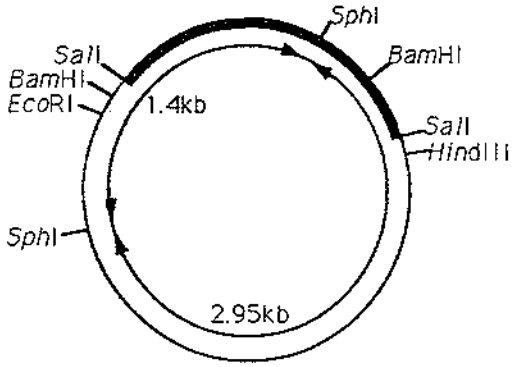
pJ28003



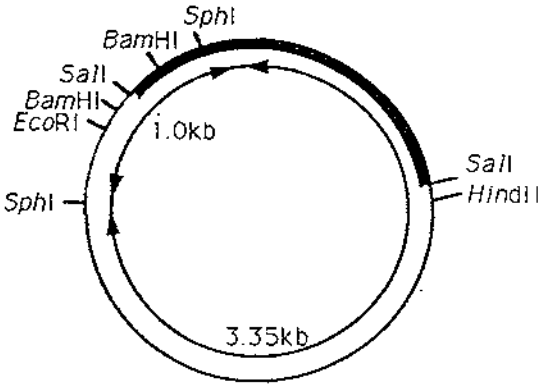
pJ28006



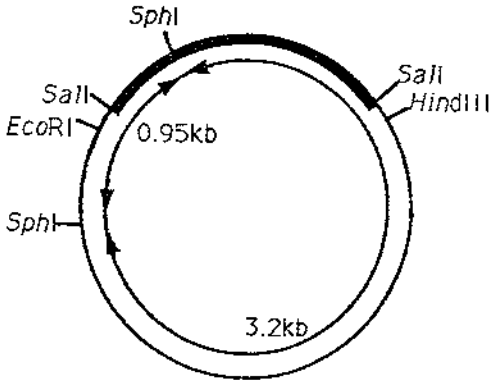
pJ14011



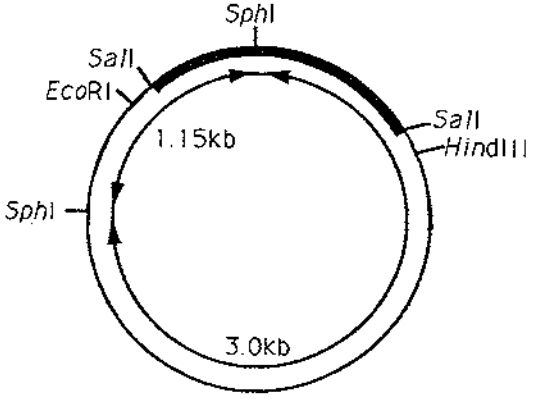
pJ14012



pJ12009



pJ12010



5.3 *BAL* 31 DIGESTION OF THE *SAL* I AND *BAM* HI FRAGMENTS.

Cloning the *Sal* I and *Bam* HI fragments in both orientations enabled reading the DNA sequence for both strands using the one vector. pJ14011, pJ14012, pJ28003, pJ28006, pJ12009 and pJ12010 preparations were linearised by *Hind* III digestion. Trial digests were performed with *Bal* 31 to determine the average number of nucleotides removed per minute for each DNA stock (see 2.16). *Bal* 31 digests were then scaled-up and samples removed at time intervals corresponding to the removal of approximately 250 nucleotides from both ends of the DNA molecule.

The samples were next subjected to *Eco* RI digestion and the resulting vector and insert fragments run on a SeaPlaque gel (see 2.16). The smaller 1.2 kb and 1.45 kb *Sal* I fragments could be easily separated from those of the vector. These fragments were excised from the gel and ligated into *Eco* RI/*Sma* I cut M13mp18 using both ligation methods (2.15.2.1; 2.15.2.2). On average, it is expected that 50% of the DNA molecules exposed to *Bal* 31 treatment will be blunt-ended at any particular time. Hence, *Sma* I, a blunt-end cutter, was selected as the other enzyme for vector preparation. DNA sequence up to the *Sal* I sites was obtained by cloning the 1.2 kb and 1.45 kb fragments into *Sal* I cut and CAP-treated M13mp18.

Because the 2.8 kb *Bam* HI fragment was almost identical in size to the pGEM-2 vector, the digested insert fragments could not be reliably separated from those of the vector on a SeaPlaque gel. Consequently, after *Bal* 31 digestion, pJ28003 and pJ28006 were *Eco* RI digested and the fragments shot-gun cloned into *Eco* RI/*Sma* I cut vector. M13mp18 derivatives carrying the 2.8 kb *Bam* HI fragments were identified by plaque hybridisation using [^{32}P]-labelled probes made from the 2.8 kb *Bam* HI DNA (2.19).

5.4 DETERMINATION OF THE Tn*5* INSERTION SITES.

To identify the 233 Tn*5* insertion site, pPN301 (pBR328 + 7.1 kb *Eco* RI + Tn*5*) was first cut with *Eco* RI. The 12.8 kb band carrying the 7.1 kb *Eco* RI fragment and the Tn*5* insert was separated on a 0.6%

SeaPlaque gel and recovered by excision and purified by phenol/ freezing. This fragment was then subjected to *Sal* I restriction and the resulting fragments separated on a 1% SeaPlaque gel. From the restriction digests in Chapter 3 (see Table 3.1), it was known that the Tn*5* insert was included in the 3.2 kb and 4.0 kb *Sal* I fragments. The 3.2 kb band was re-isolated from the SeaPlaque gel and the purified DNA then digested with *Hind* III. This resulted in the liberation of two fragments: one, the internal Tn*5* *Sal* I/*Hind* III fragment of approximately 1.8 kb, the other, approximately 1.25 kb in length, bearing *R. loti* DNA and Tn5 terminal repeat sequence. These fragments were separated and the 1.25 kb fragment re-isolated and purified in the usual manner. The fragment was ligated into *Sal* I/*Hind* III cut M13mp18. Competent cells were then transformed with the ligation mixture and single-stranded DNA preparations made from the clear plaques.

Medium scale plasmid preparations (2.10.2) were made of the pPN306 plasmids carrying the Tn5 insertions corresponding to the 4019, 4047 and 4053 mutations. The plasmids were digested with *Eco* RI and the fragments separated on a 0.6% SeaPlaque gel. The 12.8 kb band was excised and purified. The approximate locations of the Tn5 sites had been estimated from *Hind* III restriction mapping (see 3.2). From this information, it was predicted that neither *Sal* I or *Hind* III cutting would produce sufficient band separation of the resulting fragments, in all cases, to enable them to be excised cleanly from the gel. Therefore, the purified 12.8 kb *Eco* RI fragment was digested with *Sal* I and *Hind* III and shot-gun cloned into the prepared vector. The Tn*5* primer, which will hybridise to the end of the Tn5 inverted repeats, identified the vectors carrying the correct inserts, as only these would chain extend DNA in the sequencing reactions.

The Tn*5* primer, 5'-CGTTCAGGACGCTACTT-3', hybridises to a region of DNA 15 bp upstream from the ends of the Tn*5* inverted repeats (Auerswald *et al.*, 1981). The sequencing reaction will proceed out of the Tn*5* and into the sequence in which the transposon has inserted. The transition point between Tn*5* and known wild-type DNA is the insertion site and this was checked by comparison with the wild-type sequence.

5.5 SEQUENCE ANALYSIS.

The 1.45 kb fragment and its accompanying *Bal* 31 derivatives were sequenced first. The DNA sequence was assembled with the aid of the Genetics Computer Group Sequence Analysis Software Package, Version 6.1 (Devereux *et al.*, 1984). GAP and BESTFIT were employed to compare the similarity of one sequence to another, while the OVERLAP, FRAMES and GELASSEMBLE programs were utilised to compile the DNA sequence. (Note: GAP makes an optimal alignment between two sequences by inserting gaps to maximise the number of matches, using the algorithm of Needleman and Wunsch, 1970; BESTFIT is similar to GAP but uses the "local homology" algorithm of Smith and Waterman, 1981; OVERLAP compares two sequences in both orientations using the algorithm of Wilbur and Smith, 1983; FRAMES shows open reading frames for the six possible translation frames of the DNA sequence; and GELASSEMBLE is a multiple sequence editing program for assembling fragments into a contiguous sequence, based on the method of Staden, 1980).

Comparison of data from the 1.45 kb fragment with that of other *nod* gene sequences held in the GenBank Genetics Sequence Data Bank files (Bilofsky *et al.*, 1986), revealed that the fragment contained most of the *nodC* gene and the start of *nodI*. Extrapolation from known *nodC* sequences indicated that approximately 50 bp of the 5′ region of the gene must be present on the adjacent 1.2 kb *Sal* I fragment.

Bearing this in mind, the 1.2 kb fragment was *Bal* 31 digested with the aim of focusing in on the right-hand-side of the fragment, with respect to the restriction map, as this region contained the start of the *nodC* gene. *Bal* 31 digestions of the 2.8 kb *Bam* HI fragment were calculated to cover the *Sal* I restriction site and the surrounding region.

Apart from the 5′region of the *nodC* gene, computer analysis revealed that a sequence homologous to the 3′-end of the *nodA* gene was situated about 50 bp upstream from the *nodC* start codon. Rather than sequence the entire *nodA* gene, it was decided to check whether a promoter like sequence (*nod* box) was positioned upstream of the start of *nodA*. The *nodA* gene sequences in the GenBank files average about

610 bp in length. Taking this into consideration, further *Bal* 31 digests were designed to cover the area predicted to encompass the start of the *nod*A gene and the upstream promoter region or, in the absence of the latter, to determine whether another *nod* gene was present in this putative operon. Sequence data revealed that a *nod* box was positioned approximately 100 bp upstream of the putative *nod*A start codon.

The fragments used to sequence the 1.2 kb and 1.45 kb regions are represented in Figures 5.6 and 5.7 respectively. Sequence for the region of DNA covering the *nod* box and the start of *nod*A is shown in Figure 5.8 while sequence extending from the end of *nod*A to the *Sal* I site in *nod*I is presented in Figure 5.9. The salient features of each sequence and the amino acid translations for the proposed coding regions are also shown.

5.5.1 A REGION UPSTREAM OF THE *NOD* BOX DEMONSTRATES A SIMILARITY TO THE START OF *NOD*D.

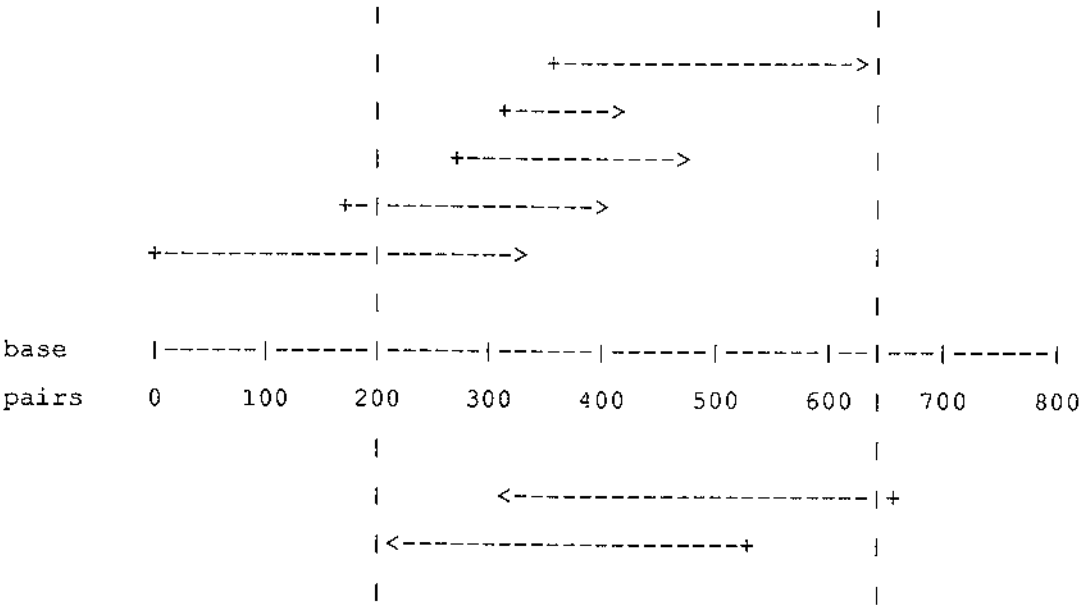
As the common *nod* genes are generally arranged in the *nod*D-*nod* box-*nod*ABCIJ grouping, the sequence upstream of the *nod* box (including that sequenced only on one strand) was checked for homology to *nod*D sequences. A short region of 32 bp, including a small deletion, just upstream of the *nod* box was found to share similarity to *R. meliloti* $nodD_1$, after which, the similarity abruptly stopped (see Figure 5.10). The ATG for this sequence was also in a similar position to those representing the start codon for other *nod*D genes.

5.5.2 *NOD* BOX ANALYSIS.

The *nod* box sequences, which are located upstream of *nod*A and available through GenBank, were compared in Figure 5.11. The downstream, loosely conserved sequence AT(T)AG, found by Spaink *et al.* (1987a) to affect *nod* box efficiency, was included. The *nod* box for each species was compared to the others and the percentage identity for each pair is presented in Figure 5.12. As expected, the *nod* boxes show a high degree of conservation. The AT(T)AG "motif" was present in each of the species however, in *R. loti*, the terminal "G" was absent.

Figure 5.6: Sequencing strategy for the *nod* box and *nod*A 5'-region of *R. loti*. The diagram represents the fragments used to sequence part of the 1.2 kb *Sal* I fragment containing the *nod* box and the start of the *nod*A gene. The region bounded by vertical dashed lines denotes the area where both strands have been sequenced. A map of the region is illustrated below.

Figure 5.6:


base
pairs
0
100
200
300
400
500
600
700
800

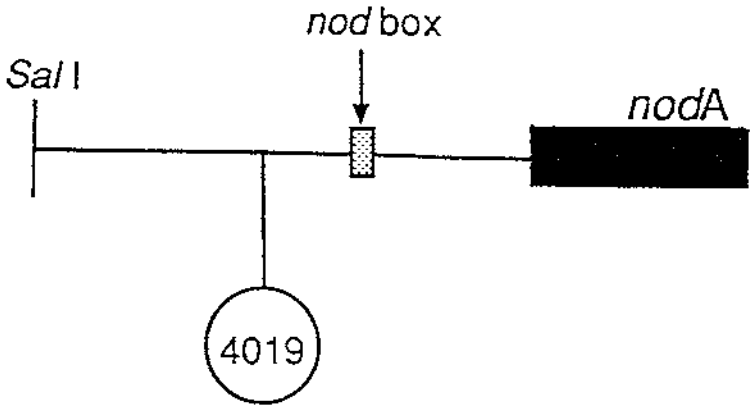
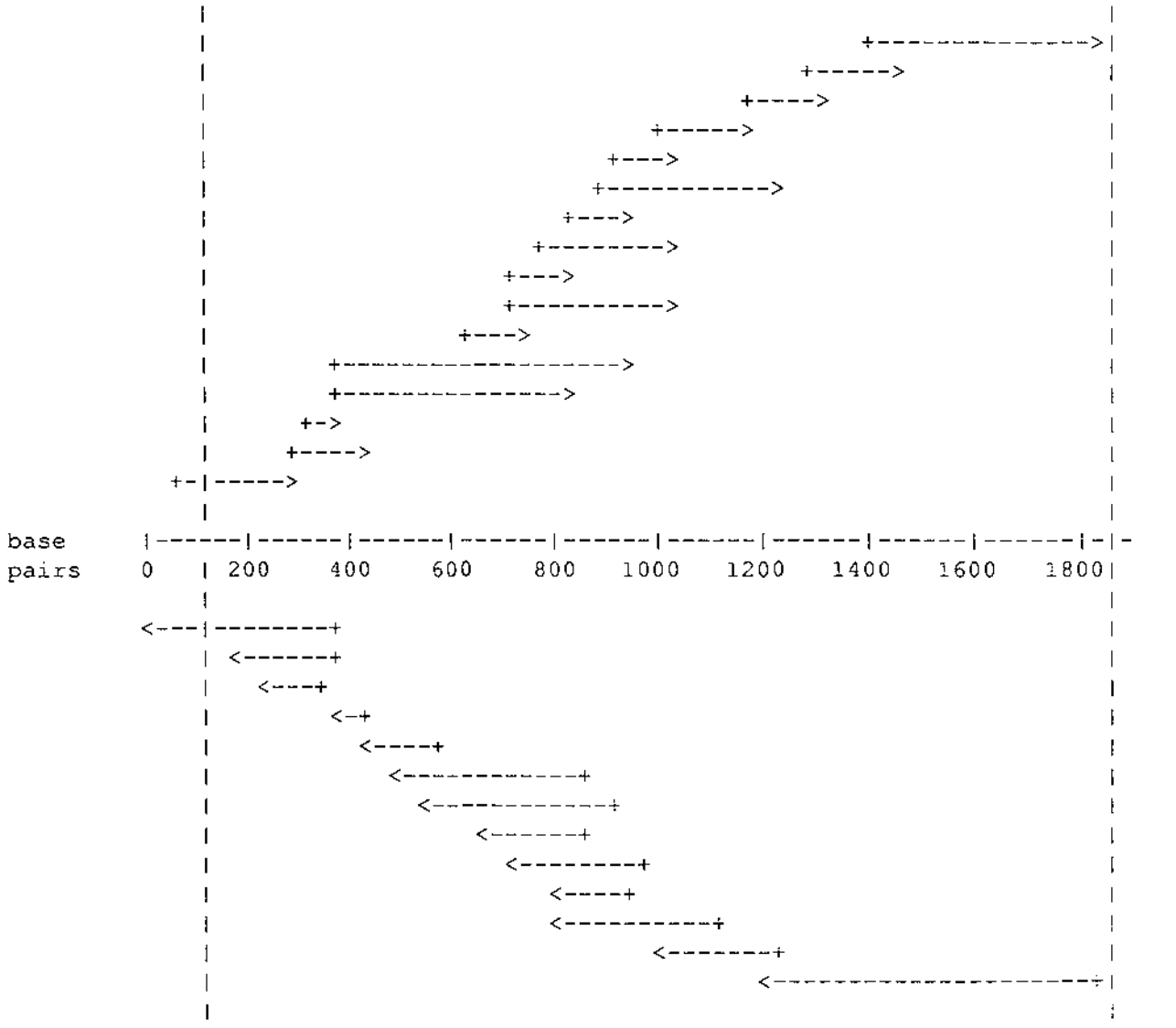

nod box
Sal I
nodA
4019

Figure 5.7: Sequencing strategy for the region of DNA carrying the 3'-end of the *nodA* gene, the *nodC* gene and the 5'-end of the *nodI* gene. The region bounded by dashed vertical lines denotes the area where both strands have been sequenced and the readings are clear. A map of the region is illustrated below.

Figure 5.7:


base pairs
0
200
400
600
800
1000
1200
1400
1600
1800

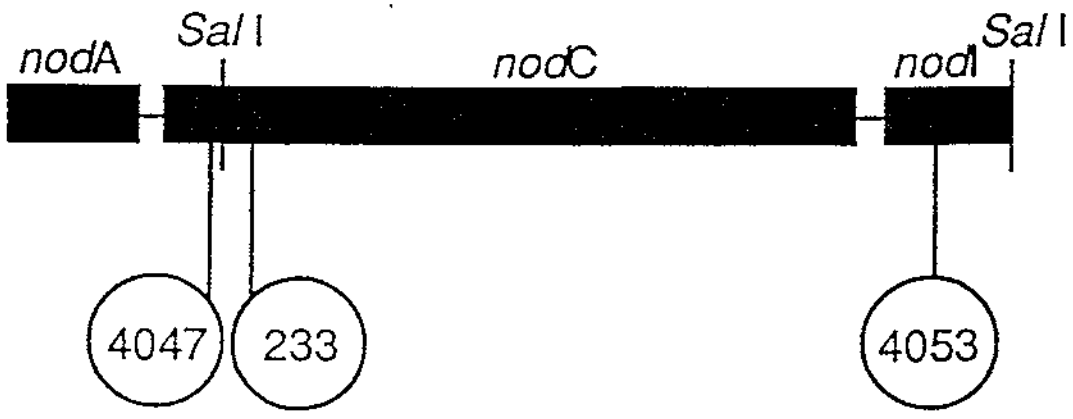

nodA
SalI
nodC
nodI
SalI
4047
233
4053

Figure 5.8: DNA sequence of the *R. loti nod* box and the 5'-end of *nod*A. This figure depicts the 4019 Tn*5* insertion site, the *nod* box sequence, the *Sph* I restriction site and the putative amino acid sequence for the start of the *R. loti* NZP2037 *nod*A gene. S-D represents a possible Shine-Dalgarno sequence while ▼ represents the Tn*5* insertion site.

Amino Acid Codes.

A = Alanine	N = Asparagine
C = Cysteine	P = Proline
D = Aspartic	Q = Glutamine
E = Glutamic	R = Arginine
F = Phenylalanine	S = Serine
G = Glycine	T = Threonine
H = Histidine	V = Valine
I = Isoleucine	W = Tryptophan
K = Lysine	Y = Tyrosine
L = Leucine	* = terminator
M = Methionine	

Figure 5.9: DNA sequence of the *R. loti nod*A 3'-terminal, *nod*C and *nod*I 5'-terminal. The insertion sites for the 4047, 233 and 4053 Tn*5* mutations, the *Sal* I, *Sph* I and *Bam* HI restriction sites and the putative amino acid sequences for the 3'-end of *nod*A, *nod*C and the start of *nod*I are shown. Numbered arrows represent possible stem-loop structures, (STEMLOOP, a GCG package, version 6.1) while S-D represents possible Shine-Dalgarno sequences and ▼ represents Tn*5* insertion sites.

Figure 5.8:

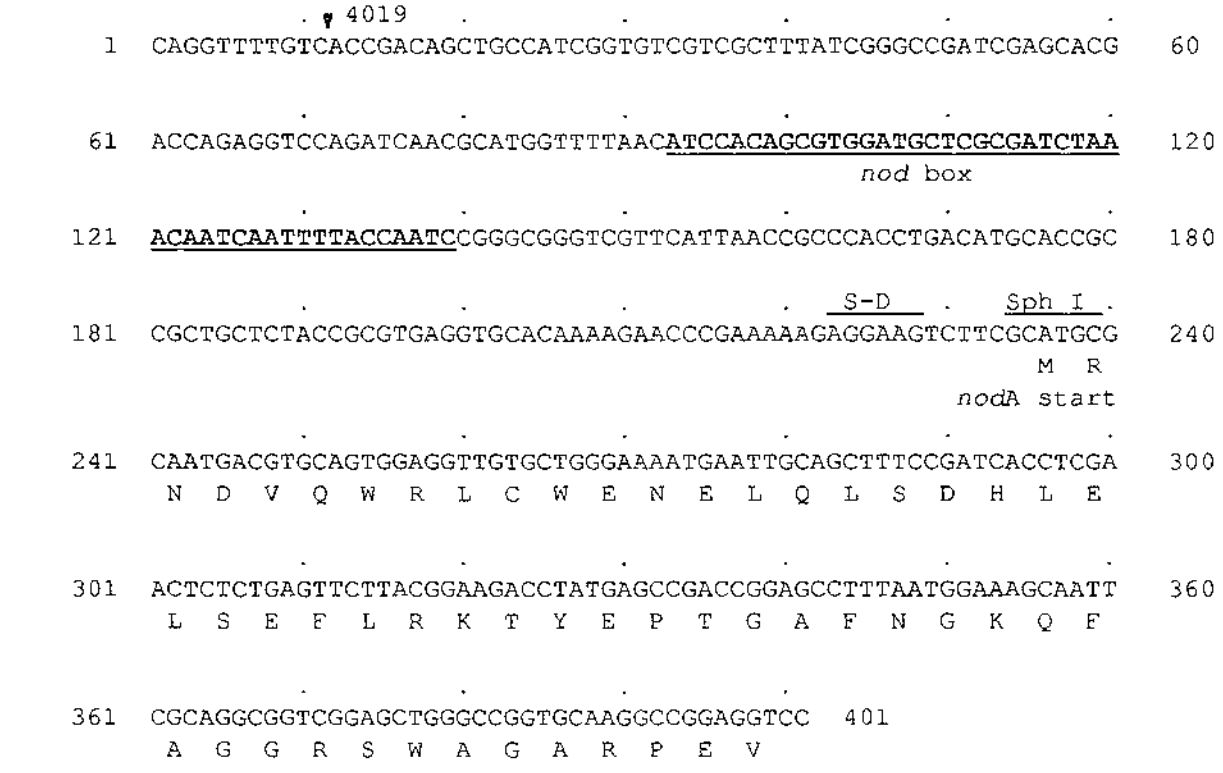
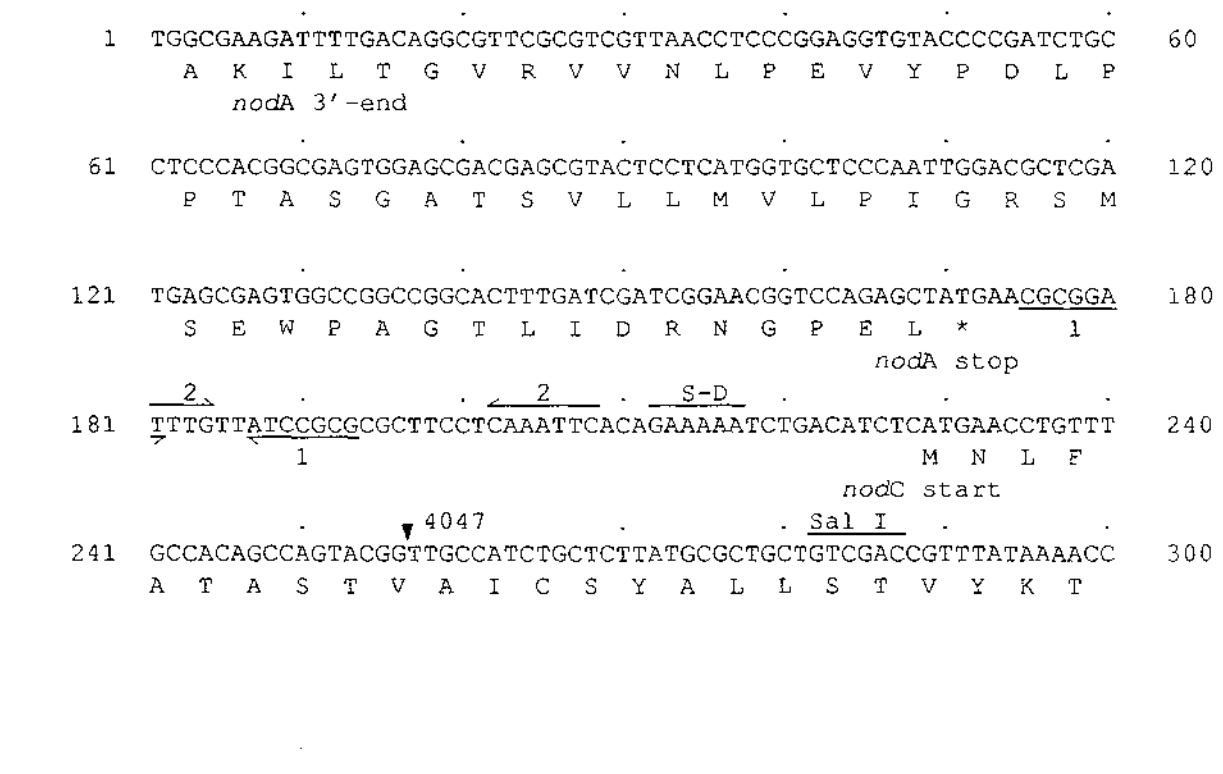


Figure 5.9:



```
                                    ▼ 233
301  GCGCAGGTGTTTTATACCCTGCCTACAAACGTAGCGCCGACGTCGGGCGACCCGGCCGGC  360
     A  Q  V  F  Y  T  L  P  T  N  V  A  P  T  S  G  D  P  A  G

361  GGTGAGCCTTGGCCGAGCGTCGATGTCATTATCCCGTCCTACAACGAGGCGCCTCGCACC  420
     G  E  P  W  P  S  V  D  V  I  I  P  S  Y  N  E  A  P  R  T

421  CTGTCCGACTGCCTGGCATCCATTGCAAGTCAGGAATACGCCGGAAAACTGCAGGTCTAT  480
     L  S  D  C  L  A  S  I  A  S  Q  E  Y  A  G  K  L  Q  V  Y

481  GTTGTTGATGACGGTTCTGCAAACCGCGATGCCCTCGTAGGTGTACAAGAGGAATACGGG  540
     V  V  D  D  G  S  A  N  R  D  A  L  V  G  V  Q  E  E  Y  G

541  CACGACCCGAGGTTCAACTTCATTGCGCTCCCAAAGAATGTCGGAAAGCGAAAGGCGCAG  600
     H  D  P  R  F  N  F  I  A  L  P  K  N  V  G  K  R  K  A  Q

601  ATTGCCGCGGTTCGCCGCTCGTGCGGCGATTTGGTGCTCAATGTAGATTCGGACACGATA  660
     I  A  A  V  R  R  S  C  G  D  L  V  L  N  V  D  S  D  T  I

661  CTCGCGCCGGACGTCGTCACAAGGCTTGCGCTAAAGATGCAAGATCAAGCGATCGGCGCG  720
     L  A  P  D  V  V  T  R  L  A  L  K  M  Q  D  Q  A  I  G  A

721  GCCATGGGCCAGTTGGCGGCTAGCAACCGCAGTGAAACTTGGCTGACGCGGTTGATCGAC  780
     A  M  G  Q  L  A  A  S  N  R  S  E  T  W  L  T  R  L  I  D

781  ATGGAGTACTGGCTCGCCTGCAACGAAGAGCGGGCTGCACAGGCTCGATTCGGTGCCGTC  840
     M  E  Y  W  L  A  C  N  E  E  R  A  A  Q  A  R  F  G  A  V

841  ATGTGTTGCTGCGGACCGTGTGCCATGTACCGGCGATCCGCGCTTCTTTCGGTGCTGGAT  900
     M  C  C  C  G  P  C  A  M  Y  R  R  S  A  L  L  S  V  L  D

901  CAATACGAGACGCAACGCTTTCGAGGGAAGCCGAGCGACTTCGGTGAGGATCGCCACCTC  960
     Q  Y  E  T  Q  R  F  R  G  K  P  S  D  F  G  E  D  R  H  L

961  ACGATCCTTATGCTGAAAGCAGGCTTCCGAACCGAGTATGTCCCGGAGGCCGTCGCGGCA 1020
     T  I  L  M  L  K  A  G  F  R  T  E  Y  V  P  E  A  V  A  A
```

```
1021  ACAGTCGTTCCCGACAGGATAGGTCCCTATCTGCGTCAACAACTTCGCTGGGCCCGGAGC  1080
      T  V  V  P  D  R  I  G  P  Y  L  R  Q  Q  L  R  W  A  R  S

1081  ACTTTCCGTGACACGCTGCTATCACTCCACCTGTCGCGCGGCCTTAATATTTATCTCACA  1140
      T  F  R  D  T  L  L  S  L  H  L  S  R  G  L  N  I  Y  L  T

1141  TTGGACGTGATTGGCCAGAACCTTGGCCCATTATTGCTCTCTTTGTCGGTCCTGGCGGGG  1200
      L  D  V  I  G  Q  N  L  G  P  L  L  L  S  L  S  V  L  A  G

                                                Sph I
1201  CTCGCGCAATTCGTAACGACAGGTACTGTGCCTTGGACGGCATGCCTGACGCTTGCAGCC  1260
      L  A  Q  F  V  T  T  G  T  V  P  W  T  A  C  L  T  L  A  A

1261  ATGACCATAGTTCGCTGCAGCGTGGCAGCGTTTCGTGCGCGCCAACTTCGATTTCTCGGA  1320
      M  T  I  V  R  C  S  V  A  A  F  R  A  R  Q  L  R  F  L  G

1321  TTCTCGCTCCACACACTCATCAACATCTCTCTCTTGCTCCCATTGAAAGCATACGCGCTG  1380
      F  S  L  H  T  L  I  N  I  S  L  L  L  P  L  K  A  Y  A  L

1381  TGTACATTGAGCAATAGCGATTGGCTGTCGCGAAGCTCTGCGACCAAGGTGGCCAGACAT  1440
      C  T  L  S  N  S  D  W  L  S  R  S  S  A  T  K  V  A  R  H

                                Bam HI
1441  CGCGCGCGCTTTCAAAAGCCGACCTTGGTAGGATCCGAAGCAACTTACAGCGAACAGCAA  1500
      R  A  R  F  Q  K  P  T  L  V  G  S  E  A  T  Y  S  E  Q  Q

                                    3  >< 4           < 3
               2 >          4 >  < 2          S-D
1501  TAGTCACGTGATCTCAGGCGGAAGAGTTTTGAGCCAGGACCAGAAAGATCAAGTGTTGAT  1560
      *            1               1                          M  M
nodC stop                                              nodI start
                                4053
1561  GCGCGAGTTGGACCCTAAGGATTTGCGGCGGCCCGAGACTGGTCAGATCGAACGGGAGTC  1620
       R  E  L  D  P  K  D  L  R  R  P  E  T  G  Q  I  E  R  E  S

1621  TCACGAGCAATCAAGCGCGAAAAGCTCCGTGTCTGACTCTGCCTCGACCGTGGCGGTCGA  1680
       H  E  Q  S  S  A  K  S  S  V  S  D  S  A  S  T  V  A  V  D

                                                Sal I
1681  TTTTACAGGCGTAACCAAGTCGTATGGGAACAAGGTCGTCGTCGAC  1726
       F  T  G  V  T  K  S  Y  G  N  K  V  V  V  D
```

5.5.3 COMPARISON OF THE *NOD*A SEQUENCES.

While comparing the start of the *nod*A nucleotide sequences available via GenBank, and the translated amino acid sequences, it was noticed that neither the *R.l.* bv. *viciae* nucleotide or amino acid sequence matched the comparable sequences in other species very well. However, when comparing the reported *R.l.* bv. *viciae nod*A coding sequence with the *R. meliloti* sequence (Figure 5.13), there was a high degree of similarity between the two. One difference though, was the insertion of a space in the *R. meliloti* sequence by the GAP program to optimise matching. When comparing the calculated sequence lengths for the various *nod*A genes, *R.l.* bv. *viciae* was also found to be approximately 120 bp shorter than the rest and that the reported start codon for this species was around 100 bp downstream from that of the other *nod*A genes. This explained the lack of homology when comparing the 5'regions of the *R.l.* bv. *viciae* nucleotide and amino acid sequences with the other species.

If the *R.l.* bv. *viciae nod*A is about 120 bp shorter than the other *nod*A sequences, it would indicate an approximate 25% reduction in gene size. As this is a "common" *nod* gene, i.e. highly conserved, it is unlikely that this gene could complement other *nod*A mutations in this form. Consequently, the nucleotide sequence around and including the *R.l.* bv. *viciae nod*A was compared to *R. meliloti nod*A. (see Figure 5.14).

The surrounding *nod*A sequence for *R.l.* bv. *viciae* shows a high degree of similarity to *R. meliloti nod*A. Deletion of a "T" at position 174 and addition of a space between positions 618 and 619 in the *R.l.* bv. *viciae nod*A sequence creates an new open reading frame that is extended by 91 bp in the 5'-region and by 26 bp in the 3'-region. The nucleotide and amino acid sequences for *R.l.* bv. *viciae* then match the *R. meliloti nod*A extremely well, including lining up with the start and stop codons. From this data it was concluded that, although the published *R.l.* bv. *viciae* sequence may be correct, it is more likely that there exists two errors in the sequence, i.e. the extra "T" and the missing nucleotide base. The presence of the extra base causes a new stop codon to come into frame about 70 bp

Figure 5.10: A comparison between *R . meliloti nod*D_1 nucleotide sequence and the region upstream of the *nod* box in NZP2037.
R. meliloti nodD_1 (Gottfert *et al*, 1989)
NZP2037 (this study).
(In this and all the following sequence alignments "|" = matching bases or amino acids while "." represents a gap inserted to optimalise matches).

Figure 5.11: Alignment of *Rhizobium* and *Bradyrhizobium nod* box sequences and an associated conserved sequence (highlighted).
The sequences chosen are those upstream of the *nod*A genes (except for *R. loti* NZP2213 where the *nod* box is the one upstream of *nod*B).
(1) *R.l.* bv. *viciae* (Shearman *et al.*, 1986)
(2) *R.l.* bv. *trifolii* (Schofield and Watson, 1986)
(3) *R. meliloti* (Egelhoff *et al.*, 1985)
(4) *R. loti* NZP2037 (this study)
(5) *R. loti* NZP2213 (C. Young and D.B. Scott, personal communication)
(6) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 (Scott, 1986)
(7) *nod* box consensus sequence (Spaink *et al.*, 1987a)
N = any base; Y = pyrimidine; U = purine; . = a base

Figure 5.12: A comparison of the *nod* box sequences using the GAP program. (% identity).

Figure 5.10:

```
R. meliloti
strain 41        .         .         .         .         .         .
nodD1            ATGCGTTTTAGGGGCCTAGATCTAAACCTCCTCGTCGCGCTCGACGCACTGATGACC
                 |||||||           |||||  |||| || |||| ||||||    |     |
R. loti          ATGCGTT...........GATCTGGACCT.CTGGTCGTGCTCGATCGGCCCGATAAA
NZP2037
```

Figure 5.11:

```
                 .         .         .         .         .         .
(1) AATATCCATTCCATAGATGATTGCCATCCAAACAATCAATTTTACCAATCTTTCGGATCACTTATAGAAAA
       |||||  |  ||||||||||| |||||||||||||||||||||||
(2) TCGATCCACGCTGTAGATGATTGCGATCCAAACAATCAATTTTACCAATCTTTCGGAGTGCTTATTAGAAA
       |||||    | |||||||| |  |||||||||||||||||||||||
(3) GGCATCCATATCGCAGATGATCGTTATCCAAACAATCAATTTTACCAATCTTGCAGAGTCCTATTAGAGAA
       |||||  |  ||  ||||  |||  |||  ||||||||||||||||||||
(4) AACATCCACAGCGTGGATGCTCGCGATCTAAACAATCAATTTTACCAATCCGGGCGGGTCGTTCATTAACC
       ||||  ||  ||||||||||  |  |||  |||||||  |||||  ||  ||
(5) TCTATCCGCAACGTGGATGCTTGTCATCGAAACAATCGATTTTGCCGATTCGACAGAACGTCCCATAGGAA
       ||||     |||||||  ||  ||||||||||||||||||  ||    |
(6) TCTATCCATCGTGTGGATGTATTCTATCGAAACAATCGATTTTACCAGATTGCGGAATGTTGGATAGCAAA

(7)    ATCCAYNNYUYUGATGNNNNYNATCNAAACAATCUATTTTACCATCY  1-13 bp  AT(T)AG
```

Figure 5.12:

	Brady-rhizobium	*R. meliloti*	*R.l.* bv. *trifolii*	*R.l.* bv. *viciae*	*R. loti* NZP2213
R. loti NZP2037	72.3%	83.0%	85.1%	80.9%	78.7%
R. loti NZP2213	70.2%	72.3%	72.3%	72.3%	
R.l. bv. *viciae*	70.2%	68.1%	89.4%		
R.l. bv. *trifolii*	72.3%	83.0%			
R. meliloti	68.1%				

Figure 5.13: An alignment of *R. meliloti* and *R.l.* bv. *viciae* published *nod*A sequences using the GAP program. While the sequences are very similar, an additional gap (.) was introduced into the *R. meliloti* sequence to match the DNA sequences optimally. The *R.l.* bv. *viciae* sequence is also approximately 120 bp shorter than the average length of the other *nod*A gene sequences present in GenBank. The numbers are those published with the sequences in GenBank.

(1) *R. meliloti* (Egelhoff *et al.*, 1985)

(2) *R.l.* bv. *viciae* (Rossen *et al.*, 1984)

Figure 5.13:

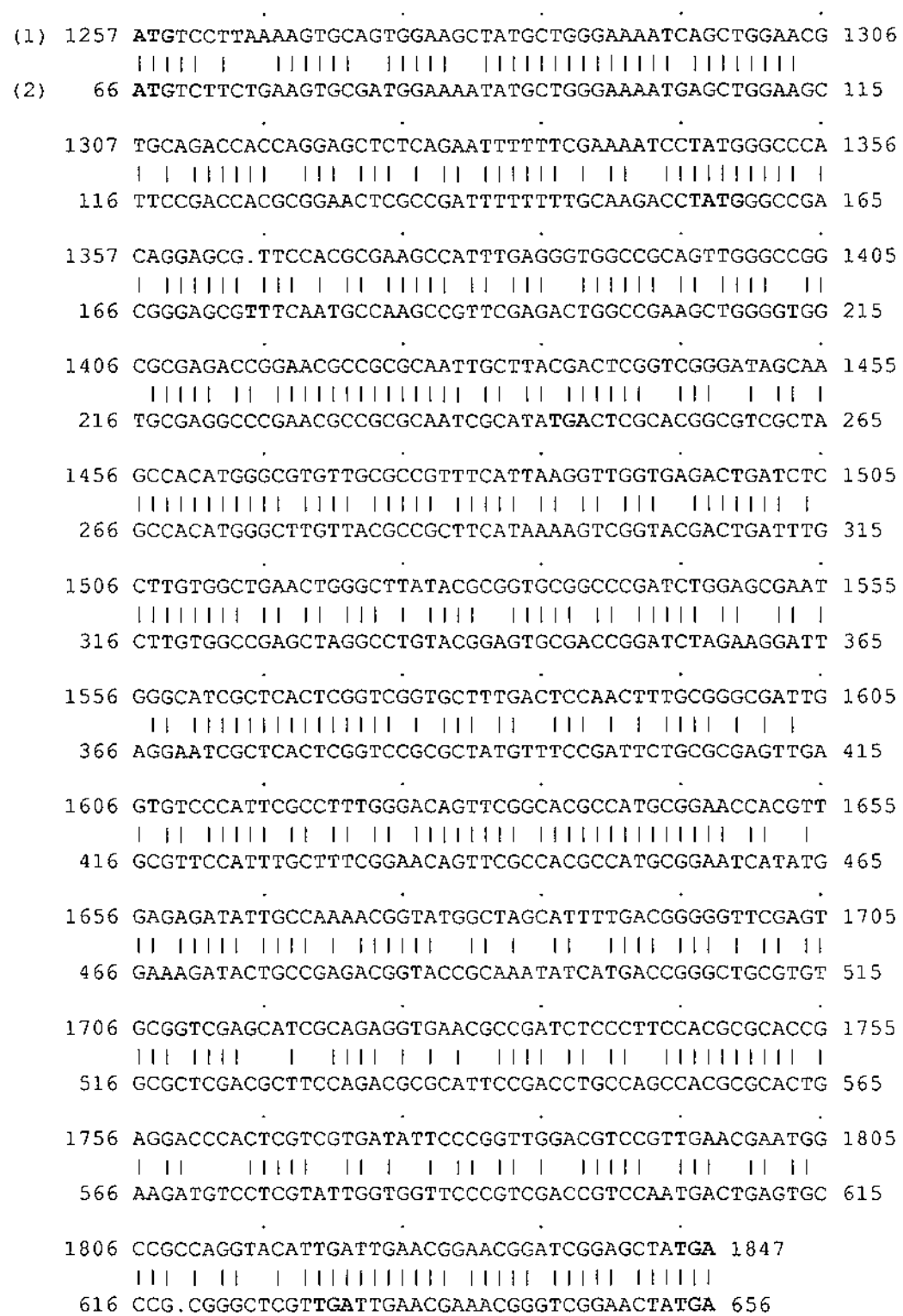

(1) 1307 TGCAGACCACCAGGAGCTCTCAGAATTTTTTCGAAAATCCTATGGGCCCA 1356
(2) 157 ATGGGCCGA 165
1357 CAGGAGCG.TTCCACGCGAAGCCATTTGAGGGTGGCCGCAGTTGGGCCGG 1405
166 CGGGAGCGTTTCAATGCCAAGCCGTTCGAGACTGGCCGAAGCTGGGGTGG 215
1406 CGCGAGACCGGAACGCCGCGCAATTGCTTACGACTCGGTCGGGATAGCAA 1455
216 TGCGAGGCCCGAACGCCGCGCAATCGCATATGACTCGCACGGCGTCGCTA 265
1456 GCCACATGGGCGTGTTGCGCCGTTTCATTAAGGTTGGTGAGACTGATCTC 1505
266 GCCACATGGGCTTGTTACGCCGCTTCATAAAAGTCGGTACGACTGATTTG 315
1506 CTTGTGGCTGAACTGGGCTTATACGCGGTGCGGCCCGATCTGGAGCGAAT 1555
316 CTTGTGGCCGAGCTAGGCCTGTACGGAGTGCGACCGGATCTAGAAGGATT 365
1556 GGGCATCGCTCACTCGGTCGGTGCTTTGACTCCAACTTTGCGGGCGATTG 1605
366 AGGAATCGCTCACTCGGTCCGCGCTATGTTTCCGATTCTGCGCGAGTTGA 415
1606 GTGTCCCATTCGCCTTTGGGACAGTTCGGCACGCCATGCGGAACCACGTT 1655
416 GCGTTCCATTTGCTTTCGGAACAGTTCGCCACGCCATGCGGAATCATATG 465
1656 GAGAGATATTGCCAAAACGGTATGGCTAGCATTTTGACGGGGGTTCGAGT 1705
466 GAAAGATACTGCCGAGACGGTACCGCAAATATCATGACCGGGCTGCGTGT 515
1706 GCGGTCGAGCATCGCAGAGGTGAACGCCGATCTCCCTTCCACGCGCACCG 1755
516 GCGCTCGACGCTTCCAGACGCGCATTCCGACCTGCCAGCCACGCGCACTG 565
1756 AGGACCCACTCGTCGTGATATTCCCGGTTGGACGTCCGTTGAACGAATGG 1805
566 AAGATGTCCTCGTATTGGTGGTTCCCGTCGACCGTCCAATGACTGAGTGC 615
1806 CCGCCAGGTACATTGATTGAACGGAACGGATCGGAGCTATGA 1847
616 CCGCGGGCTCGTTGA 630

Figure 5.14: A comparison of the published *R. meliloti* and *R.l.* bv. *viciae* *nod*A genes and surrounding DNA sequences using the GAP program. The DNA displays a high degree of similarity. With the deletion of the "T" at position 174 (Figure 5.13) and the addition of a gap between positions 618 and 619 (Figure 5.13), the *nod*A sequence in *R.l.* bv. *viciae* would begin and end in the same position as *R. meliloti*. The numbers are those published with the sequences in GenBank. Highlighted features are discussed in the text (section 5.5.3)

(1) *R. meliloti* (Egelhoff *et al.*, 1985)

(2) *R.l.* bv. *viciae* (Rossen *et al.*, 1984)

Figure 5.14:


(1) 1257 ATGTCCTTAAAAGTGCAGTGGAAGCTATGCTGGGAAAATCAGCTGGAACG 1306
(2) 66 ATGTCTTCTGAAGTGCGATGGAAAATATGCTGGGAAAATGAGCTGGAAGC 115
1307 TGCAGACCACCAGGAGCTCTCAGAATTTTTTCGAAAATCCTATGGGCCCA 1356
116 TTCCGACCACGCGGAACTCGCCGATTTTTTTTGCAAGACCTATGGGCCGA 165
1357 CAGGAGCG.TTCCACGCGAAGCCATTTGAGGGTGGCCGCAGTTGGGCCGG 1405
166 CGGGAGCGTTTCAATGCCAAGCCGTTCGAGACTGGCCGAAGCTGGGGTGG 215
1406 CGCGAGACCGGAACGCCGCGCAATTGCTTACGACTCGGTCGGGATAGCAA 1455
216 TGCGAGGCCCGAACGCCGCGCAATCGCATATGACTCGCACGGCGTCGCTA 265
1456 GCCACATGGGCGTGTTGCGCCGTTTCATTAAGGTTGGTGAGACTGATCTC 1505
266 GCCACATGGGCTTGTTACGCCGCTTCATAAAAGTCGGTACGACTGATTTG 315
1506 CTTGTGGCTGAACTGGGCTTATACGCGGTGCGGCCCGATCTGGAGCGAAT 1555
316 CTTGTGGCCGAGCTAGGCCTGTACGGAGTGCGACCGGATCTAGAAGGATT 365
1556 GGGCATCGCTCACTCGGTCGGTGCTTTGACTCCAACTTTGCGGGCGATTG 1605
366 AGGAATCGCTCACTCGGTCCGCGCTATGTTTCCGATTCTGCGCGAGTTGA 415
1606 GTGTCCCATTCGCCTTTGGGACAGTTCGGCACGCCATGCGGAACCACGTT 1655
416 GCGTTCCATTTGCTTTCGGAACAGTTCGCCACGCCATGCGGAATCATATG 465
1656 GAGAGATATTGCCAAAACGGTATGGCTAGCATTTTGACGGGGGTTCGAGT 1705
466 GAAAGATACTGCCGAGACGGTACCGCAAATATCATGACCGGGCTGCGTGT 515
1706 GCGGTCGAGCATCGCAGAGGTGAACGCCGATCTCCCTTCCACGCGCACCG 1755
516 GCGCTCGACGCTTCCAGACGCGCATTCCGACCTGCCAGCCACGCGCACTG 565
1756 AGGACCCACTCGTCGTGATATTCCCGGTTGGACGTCCGTTGAACGAATGG 1805
566 AAGATGTCCTCGTATTGGTGGTTCCCGTCGACCGTCCAATGACTGAGTGC 615
1806 CCGCCAGGTACATTGATTGAACGGAACGGATCGGAGCTATGA 1847
616 CCG.CGGGCTCGTTGATTGAACGAAACGGGTCGGAACTATGA 656

downstream from it, which is probably why that reading frame was discarded as being correct. The extra base also creates a new ATG start codon, 17 bp upstream from itself, that is in-frame with the greater portion of the *nod*A sequence. The gap at the tail-end of the sequence creates a new, in-frame stop codon, (TGA), that shortens the open reading frame.

Taking this information into consideration, the altered *R.l.* bv. *viciae* sequence was employed for the *nod*A nucleotide and amino acid sequence comparisons for the initial and tail-end portions of the gene. Figures 5.15 - 5.18 demonstrate the high degree of conservation in both the nucleotide and deduced amino acid sequences for the start of the various *nod*A genes. The major difference is the extra 39 bp present at the start of the *Bradyrhizobium* sequence. The 3'-ends of the gene sequences (Figures 5.19 - 5.22) are once again very similar; only marginally less so, than the 5'-terminals of the genes.

5.5.4 *NOD*C COMPARISONS.

The *R. loti* NZP2037 *nod*C gene is predicted to be 1275 bp long (424 amino acids). Once again, being a "common" *nod* gene, it is shown to be very similar to the other *nod*C genes (Figures 5.23 - 5.26) and is so at both the nucleotide and translated amino acid sequence levels. The 5' and 3' sections of the sequences are the most variable. The most highly conserved region of *nod*C is that coding for the unusual cysteine-rich cluster, which suggests that this area of the gene is very important for its correct functioning. *Nod*C will be discussed further in the next chapter.

5.5.5 COMPARISONS OF THE 5'-REGIONS OF *NODI*.

The far right-hand-side of the 1.45 kb *Sal*I fragment showed considerable homology to the *R.l.* bv. *viciae* *nod*I gene sequence in GenBank. Two possible start codons were present 99 bp apart. The authors (Evans and Downie, 1986) chose to present the second ATG as being the initiation codon for the *nod*I. When aligning the *nod*I nucleotide sequence for *R. loti* NZP2037 to the *R.l.* bv. *viciae* sequence, no start codon was found in the same region as that of the

Figure 5.15: An alignment of the beginning of a number of *nod*A nucleotide sequences.

(1) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 (Scott, 1986)

(2) *R. loti* NZP2037 (this study)

(3) *R.l.* bv. *trifolii* (Schofield and Watson, 1986)

(4) *R. meliloti* (Egelhoff *et al.*, 1985)

(5) *R.l.* bv. *viciae* (Rossen *et al.*, 1984) [Note: (T) is the base removed from the published *nod*A sequenced to enhance gene sequence similarities.

Figure 5.15:

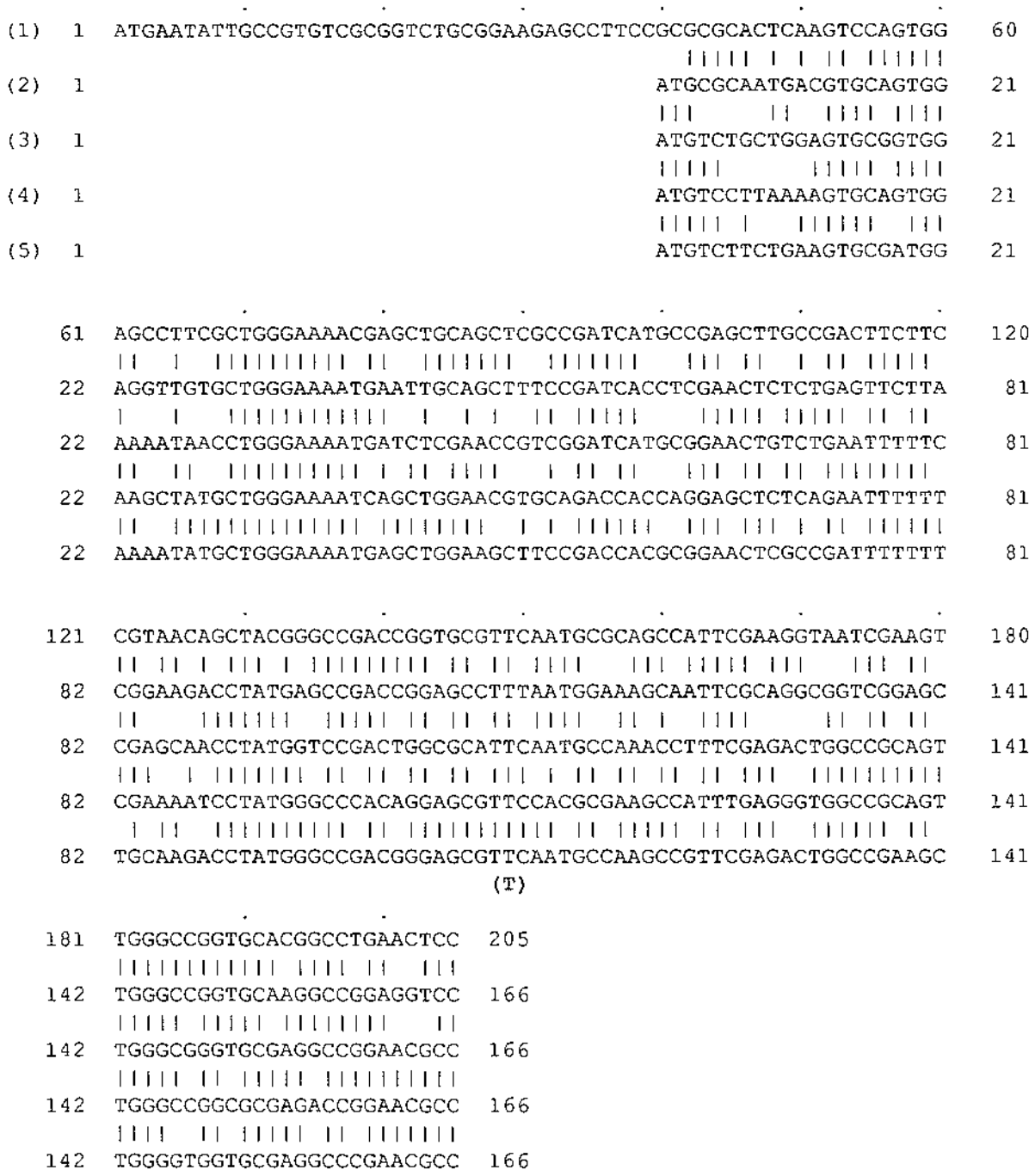

(1) 1 ATGAATATTGCCGTGTCGCGGTCTGCGGAAGAGCCTTCCGCGCGCACTCAAGTCCAGTGG 60
(2) 1 ATGCGCAATGACGTGCAGTGG 21
(3) 1 ATGTCTGCTGGAGTGCGGTGG 21
(4) 1 ATGTCCTTAAAAGTGCAGTGG 21
(5) 1 ATGTCTTCTGAAGTGCGATGG 21
61 AGCCTTCGCTGGGAAAACGAGCTGCAGCTCGCCGATCATGCCGAGCTTGCCGACTTCTTC 120
22 AGGTTGTGCTGGGAAAATGAATTGCAGCTTTCCGATCACCTCGAACTCTCTGAGTTCTTA 81
22 AAAATAACCTGGGAAAATGATCTCGAACCGTCGGATCATGCGGAACTGTCTGAATTTTTC 81
22 AAGCTATGCTGGGAAAATCAGCTGGAACGTGCAGACCACCAGGAGCTCTCAGAATTTTTT 81
22 AAAATATGCTGGGAAAATGAGCTGGAAGCTTCCGACCACGCGGAACTCGCCGATTTTTTT 81
121 CGTAACAGCTACGGGCCGACCGGTGCGTTCAATGCGCAGCCATTCGAAGGTAATCGAAGT 180
82 CGGAAGACCTATGAGCCGACCGGAGCCTTTAATGGAAAGCAATTCGCAGGCGGTCGGAGC 141
82 CGAGCAACCTATGGTCCGACTGGCGCATTCAATGCCAAACCTTTCGAGACTGGCCGCAGT 141
82 CGAAAATCCTATGGGCCCACAGGAGCGTTCCACGCGAAGCCATTTGAGGGTGGCCGCAGT 141
82 TGCAAGACCTATGGGCCGACGGGAGCGTTCAATGCCAAGCCGTTCGAGACTGGCCGAAGC 141
(T)
181 TGGGCCGGTGCACGGCCTGAACTCC 205
142 TGGGCCGGTGCAAGGCCGGAGGTCC 166
142 TGGGCGGGTGCGAGGCCGGAACGCC 166
142 TGGGCCGGCGCGAGACCGGAACGCC 166
142 TGGGGTGGTGCGAGGCCCGAACGCC 166

Figure 5.16: A comparison of the start of the *nod*A sequences using the GAP program. (% identity).

Figure 5.17: An alignment of the predicted amino acid sequences for the start of the *nod*A genes using the GAP program. [Note: see Figure 5.15 for sequence references for this and following comparisons].

(1) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289

(2) *R. loti* NZP2037

(3) *R. meliloti*

(4) *R.l.* bv. *viciae*

(5) *R.l.* bv. *trifolii*

Figure 5.18: A comparison of the N-terminal of the predicted *nod*A amino acid sequences using the GAP program. (% similarity; % identity).

Figure 5.16:

	Brady-rhizobium	*R. meliloti*	*R.l. bv. trifolii*	*R.l. bv. viciae*
R. loti	72.3%	68.7%	65.1%	69.2%
R.l. bv. viciae	68.1%	76.5%	80.1%	
R.l. bv. trifolii	65.1%	72.9%		
R. meliloti	67.5%			

Figure 5.17:

```
(1)  1  MNIAVSRSAEEPSARTQVQWSLRWENELQLADHAELADFFRNSYGPTGAFNAQPFEGNRS  60
(2)  1  MRNDVQWRLCWENELQLSDHLELSEFLRKTYEPTGAFNGKQFAGGRS              47
(3)  1  MSLKVQWKLCWENQLERADHQELSEFFRKSYGPTGAFHAKPFEGGRS              47
(4)  1  MSSEVRWKICWENELEASDHAELADFFCKTYGPTGAFNAKPFETGRS              47
(5)  1  MSAGVRWKITWENDLEPSDHAELSEFFRATYGPTGAFNAKPFETGRS              47

    61  WAGARPEL  68
    48  WAGARPEV  55
    48  WAGARPER  55
    48  WGGARPER  55
    48  WAGARPER  55
```

Figure 5.18:

	Brady-rhizobium	*R. meliloti*	*R.l. bv. trifolii*	*R.l. bv. viciae*
R. loti	76.4% ; 65.5%	81.8% ; 69.1%	80.0% ; 65.5%	80.0% ; 63.6%
R.l. bv. viciae	72.7% ; 65.5%	81.8% ; 72.7%	90.9% ; 81.8%	
R.l. bv. trifolii	70.9% ; 63.6%	81.8% ; 76.4%		
R. meliloti	76.4% ; 69.1%			

Figure 5.19: An alignment of the 3'-end of *nod*A nucleotide sequences.
[Note: (*) = the inserted gap in the *R.l.* bv. *viciae* sequence].
(1) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 (Scott, 1986)
(2) *R. loti* NZP2037 (this study)
(3) *R. meliloti* (Egelhoff *et al*, 1985)
(4) *R.l.* bv. *trifolii* (Schofield and Watson, 1986)
(5) *R.l.* bv. *viciae* (Rossen *et al*, 1984)

Figure 5.20: Comparative similarities for the 3'-end *nod*A nucleotide sequences as determined by the GAP program. (% identity).

Figure 5.19:

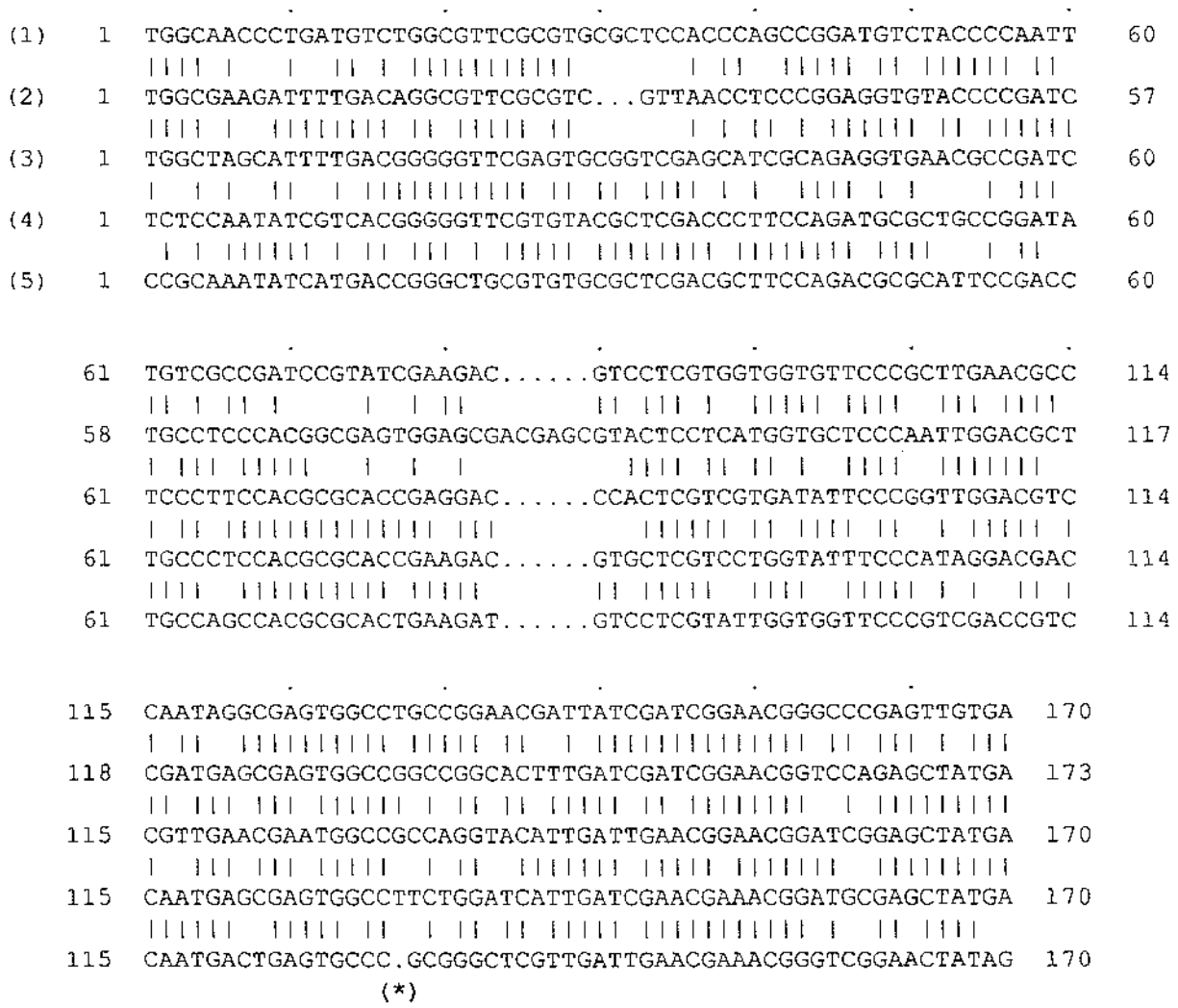


Figure 5.20:

	Brady-rhizobium	*R. meliloti*	*R.l. bv. trifolii*	*R.l. bv. viciae*
R. loti	68.3%	70.1%	65.9%	62.0%
R.l. bv. viciae	61.5%	65.7%	73.4%	
R.l. bv. trifolii	65.3%	71.2%		
R. meliloti	63.5%			

Figure 5.21: An alignment of the predicted amino acid sequences for the C-terminals of the *nod*A proteins. (*): X = the unknown amino acid created by the addition of a gap in the *R.l.* bv. *viciae* sequence.

(1) *Bradyrhizobium* sp. (*Parasponia*) strain ANU289

(2) *R. loti* NZP2037

(3) *R. meliloti*

(4) *R.l.* bv. *trifolii*

(5) *R.l.* bv. *viciae*

Figure 5.22: A comparison of the predicted amino acid sequences for the C-terminals of the *nod*A proteins. (% similarity; % identity).

Figure 5.21:

```
                .         .         .         .         .
(1)    1  ATLMSGVRVRSTQPDVYPNLSPIRIED..VLVVVFPLERPIGEWPAGTIIDRNGPEL*  56
(2)    1  AKILTGVRV.VNLPEVYPDLPPTASGATSVLLMVLPIGRSMSEWPAGTLIDRNGPEL*  57
(3)    1  ASILTGVRVRSSIAEVNADLPSTRTED..PLVVIFPVGRPLNEWPPGTLIERNGSEL*  56
(4)    1  SNIVTGVRVRSTLPDALPDMPSTRTED..VLVLVFPIGRPMSEWPSGSLIERNGCEL*  56
(5)    1  ANIMTGLRVRSTLPDAHSDLPATRTED..VLVLVVPVDRPMTECXAGSLIERNGSEL*  56
                                                    (*)
```

Figure 5.22:

	Brady-rhizobium	*R. meliloti*	*R.l.* bv. *trifolii*	*R.l.* bv. *viciae*
R. loti	78.2% ; 56.4%	76.4% ; 56.4%	72.7% ; 60.0%	70.9% ; 52.7%
R.l. bv. *viciae*	71.4% ; 55.4%	78.6% ; 60.7%	82.1% ; 73.2%	
R.l. bv. *trifolii*	75.0% ; 57.1%	82.1% ; 64.3%		
R. meliloti	76.8% ; 55.4%			

Figure 5.23: An alignment of published *nod*C nucleotide sequences using the GAP program.

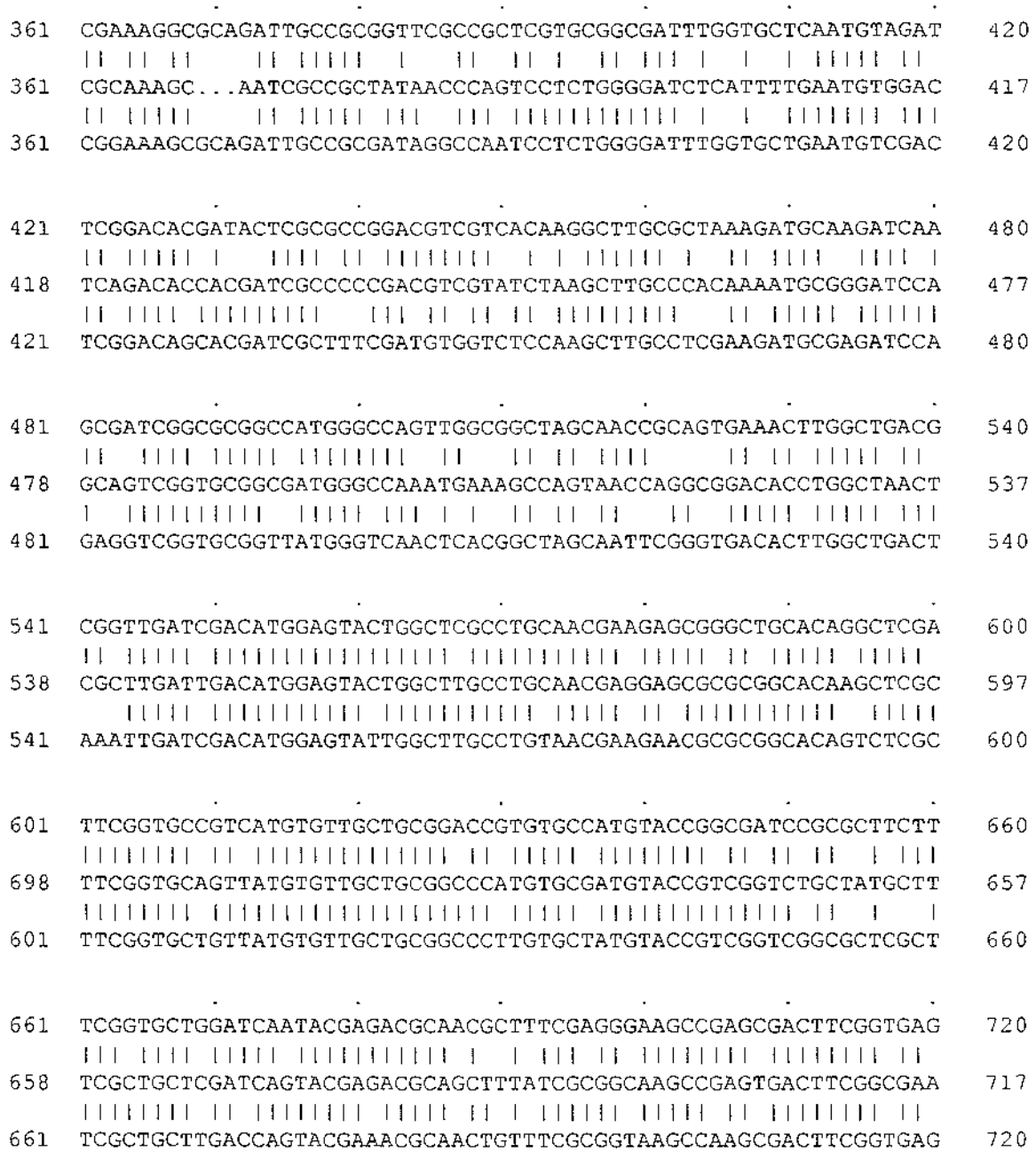
(1) *R. loti* NZP2037 (Collins-Emerson *et al*, 1990)

(2) *R.l.* bv. *viciae* (Rossen *et al*, 1984)

(3) *R. meliloti* (Egelhoff *et al*, 1985)

Figure 5.23:


(1) 1 ATGAACCTGTTTGCCACAGCCAGTACGGTTGCCATCTGCTCTTATGCGCTGCTGTCGACC 60
(2) 1 ATGACCCTGCTCGCAACAACCAGCATCGCCGCCATCTCGCTTTATGCAATGCTCTCCACC 60
(3) 1 ATGTACCTGCTTGACACAACCAGCACCGCCGCTATCTCAATCTACGCGCTGCTCTTGACC 60
61 GTTTATAAAACCGCGCAGGTGTTTTATACCCTGCCTACAAACGTAGCGCCGACGTCGGGC 120
61 GTTTACAAGAGCGCGCAGGTCTTTCATGCTAGGCGGACAACGATTTCAACAACACCTGCG 120
61 GCCTACAGGAGCATGCAAGTCCTATATGCTCGGCCGATAGACGGTCCAGCAGTGGCGGCA 120
121 GACCCGGCCGGCGGTGAGCCTTGGCCGAGCGTCGATGTCATTATCCCGTCCTACAACGAG 180
121 AAAGACATTGAAACCAACCCCGTGCCAAGCGTTGATGTCATCGTGCCGTGCTTCAACGAG 180
121 GAACCGGTCGAGACCCGCCCTCTGCCAGCCGTGGATGTTATCGTCCCCAGCTTCAATGAG 180
181 GCGCCTCGCACCCTGTCCGACTGCCTGGCATCCATTGCAAGTCAGGAATACGCCGGAAAA 240
181 GACCCAATCGTTCTTTCGGAATGCCTCGCGTCGCTTGCGGAGCAAGATTATGCCGGAAAA 240
181 GACCCAGGCATCCTCTCGGCGTGCCTCGCGTCCATTGCAGACCAGGATTATCCTGGAGAA 240
241 CTGCAGGTCTATGTTGTTGATGACGGTTCTGCAAACCGCGATGCCCTCGTAGGTGTACAA 300
241 TTGCGCATCTATGTAGTCGACGACGGTTCCAAAAATCGCGACGCGGTTGTGGCTCAGCGC 300
241 TTGCGAGTCTATGTCGTTGATGATGGTTCTCGGAACCGCGAGGCCATTGTGCGTGTACGC 300
301 GAGGAATACGGGCACGACCCGAGGTTCAACTTCATTGCGCTCCCAAAGAATGTCGGAAAG 360
301 GCTGCCTATGCAGACGATGAGAGATTCAACTTCACAATTCTCCCTAAAAATGTTGGAAAG 360
301 GCCTTCTATTCGCGCGATCCGAGGTTCAGCTTCATTCTGCTCCCAGAGAACGTCGGAAAG 360



```
721  GATCGCCACCTCACGATCCTTATGCTGAAAGCAGGCTTCCGAACCGAGTATGTCCCGGAG  780
718  GATCGCCATTTGACGATTCTCATGCTGAGCGCAGGCTTTCGAACTGAGTATGTTCCGAGC  777
721  GACCGCCATCTGACGATTCTCATGTTGAAGGCAGGCTTTCGAACTGAGTACGTTCCAGAC  780

781  GCCGTCGCGGCAACAGTCGTTCCCGACAGGATAGGTCCCTATCTGCGTCAACAACTTCGC  840
778  GCCATCGCGGCGACAGTCGTTCCAGACACGATGGGTGTTTATCTACGCCAACAACTACGG  837
781  GCCATAGTGGCAACCGTCGTCCCGGATACGCTGAAACCATATCTGCGCCAACAACTGCGT  840

841  TGGGCCCGGAGCACTTTCCGTGACACGCTGCTATCACTCCACCTGTCGCGCGGCCTTAAT  900
838  TGGGCACGCAGCACCTTTCGGGATACTTTGCTTGCGCTTCCTGTACTGCCTGGTCTCGAT  897
841  TGGGCACGCAGCACGTTCCGTGACACGTTTCTAGCGCTCCCTCTGTTGCGCGGCCTCAGC  900

901  ATTTATCTCACATTGGACGTGATTGGCCAGAACCTTGGCCCATTATTGCTCTCTTTGTCG  960
998  CGGTATCTCACGCTGGACGCAATCGGGCAAAATGTCGGCCTGCTACTTCTTGCGCTGTCG  957
901  CCTTTTCTCGCATTTGACGCGGTCGGACAGAATATCGGGCAACTGTTGCTCGCCCTTTCG  960

961  GTCCTGGCGGGGCTCGCGCAATTCGTAACGACAGGTACTGTGCCTTGGACGGCATGCCTG  1020
958  GTATTGACAGGAATTGGCCAGTTTGCGCTGACCGCCACACTGCCCTGGTGGACGATCCTG  1017
961  GTGGTCACGGGTCTTGCGCATCTCATAATGACCGCCACAGTGCCATGGTGGACAATTTTG  1020

1021 ACGCTTGCAGCCATGACCATAGTTCGCTGCAGCGTGGCAGCGTTTCGTGCGCGCCAACTT  1080
1018 GTCATTGGATCCATGACTCTTGTACGCTGTAGCGTGGCTGCCTATCGCGCCCGCGAACTT  1077
1021 ATTATTGCGTGCATGACCATTATACGCTGCAGCGTCGTAGCATTGCATGCTCGCCAACTT  1080
```

```
1081  CGATTTCTCGGATTCTCGCTCCACACACTCATCAACATCTCTCTCTTGCTCCCATTGAAA  1140
1078  AGGTTTTTGGGTTTTGCTCTCCACACGCTCGTGAACATCTTTCTCTTAATTCCCTTGAAG  1137
1081  AGATTTCTTGGCTTCGTTCTGCACACACCCATCAACCTCTTTCTCATACTTCCGCTGAAA  1140

1141  GCATACGCGCTGTGTACATTGAGCAATAGCGATTGGCTGTCGCGAAGCTCTGCGACCAAG  1200
1138  GCCTATGCCCTTTGTACCCTATCCAATAGCGATTGGCTGTCGCGCGGATCAGTCGCGATT  1197
1141  GCTTATGCGTTGTGTACATTGTCCAATAGCGACTGGCTGTCACGCTACTCCGCGCCAGAA  1200

1201  GTGGC...CAG......ACATCGCGCGCGCTTTCAAAAGCCGACCTTGGTAGGATCCGAA  1251
1198  GCGCC...CACTGTTGGGCAGCAGGGCGCTACCAAAATGCCAGGGCGGGCTACATCTGAA  1254
1201  GTACCAGTCAGCGGGGGAAAGCAGACCCCAATTCAAACCTCCGGTCGAGTGACACCTGAC  1260

1252  GCAACTTACAGCGAACAGCAATAG  1275
1255  ATTGCCTATAGTGGCGAGTGA     1275
1261  TGCACTTGCAGCGGCGAGTGA     1284
```

Figure 5.24: A comparison of the *nodC* nucleotide sequences using the GAP program. (% identity).

Figure 5.24:

	R. loti	*R.l.* bv. *viciae*
R. meliloti	70.0%	70.0%
R.l. bv. *viciae*	67.5%	

Figure 5.25: An alignment of the predicted amino acid sequences for the *nodC* proteins.

(1) *R. loti*

(2) *R.l.* bv. *viciae*

(3) *R. meliloti*

Figure 5.25:

```
(1)   1  MNLFATASTVAICSYALLSTVYKTAQVFYTLPTNVAPTSGDPAGGEPWPSVDVIIPSYNE   60
(2)   1  MTLLATTSIAAISLYAMLSTVYKSAQVFHARRTTISTTPAKDIETNPVPSVDVIVPCFNE   60
(3)   1  MYLLDTTSTAAISIYALLLTAYRSMQVLYARPIDGPAVAAEPVETRPLPAVDVIVPSFNE   60

      61  APRTLSDCLASIASQEYAGKLQVYVVDDGSANRDALVGVQEEYGHDPRFNFIALPKNVGK   120
      61  DPIVLSECLASLAEQDYAGKLRIYVVDDGSKNRDAVVAQRAAYADDERFNFTILPKNVGK   120
      61  DPGILSACLASIADQDYPGELRVYVVDDGSRNREAIVRVRAFYSRDPRFSFILLPENVGK   120

     121  RKAQIAAVRRSCGDLVLNVDSDTILAPDVVTRLALKMQDQAIGAAMGQLAASNRSETWLT   180
     121  RKA.IAAITQSSGDLILNVDSDTTIAPDVVSKLAHKMRDPAVGAAMGQMKASNQADTWLT   179
     121  RKAQIAAIGQSSGDLVLNVDSDSTIAFDVVSKLASKMRDPEVGAVMGQLTASNSGDTWLT   180

     181  RLIDMEYWLACNEERAAQARFGAVMCCCGPCAMYRRSALLSVLDQYETQRFRGKPSDFGE   240
     180  RLIDMEYWLACNEERAAQARFGAVMCCCGPCAMYRRSAMLSLLDQYETQLYRGKPSDFGE   239
     181  KLIDMEYWLACNEERAAQSRFGAVMCCCGPCAMYRRSALASLLDQYETQLFRGKPSDFGE   240

     241  DRHLTILMLKAGFRTEYVPEAVAATVVPDRIGPYLRQQLRWARSTFRDTLLSLHLSRGLN   300
     240  DRHLTILMLSAGFRTEYVPSAIAATVVPDTMGVYLRQQLRWARSTFRDTLLALPVLPGLD   299
     241  DRHLTILMLKAGFRTEYVPDAIVATVVPDTLKPYLRQQLRWARSTFRDTFLALPLLRGLS   300

     301  IYLTLDVIGQNLGPLLLSLSVLAGLAQFVTTGTVPWTACLTLAAMTIVRCSVAAFRARQL   360
     300  RYLTLDAIGQNVGLLLLALSVLTGIGQFALTATLPWWTILVIGSMTLVRCSVAAYRAREL   359
     301  PFLAFDAVGQNIGQLLLALSVVTGLAHLIMTATVPWWTILIIACMTIIRCSVVALHARQL   360

     361  RFLGFSLHTLINISLLLPLKAYALCTLSNSDWLSRSS...ATKVARHRARFQKPTLVGSE   417
     360  RFLGFALHTLVNIFLLIPLKAYALCTLSNSDWLSRGSVA.IAPTVGQQGATKMPGRATSE   418
     361  RFLGFVLHTPINLFLILPLKAYALCTLSNSDWLSRYSAPEVPVSGGKQTPIQTSGRVTPD   420

     418  ATYSEQQ*   424
     419  IAYSGE*    424
     421  CTCSGE*    426
```

Figure 5.26: A comparison of the predicted *nod*C amino acid sequences using the GAP program. (% similarity; % identity).

Figure 5.27: An alignment of the *nod*I nucleotide sequences using the GAP program.

(1) *R. meliloti* (Egelhoff *et al.*, 1985)

(2) *R. loti* NZP2037 (this study)

(3) *R. loti* NZP2213 (D.B. Scott and C. Young, personal communication)

(4) *R.l.* bv. *viciae* (Evans and Downie, 1986)

ATG = codon selected by Evans and Downie for the *R.l.* bv. *viciae* *nod*I start. **TTG** = start codons selected for *nod*I based on sequence comparisons and on appropriately placed Shine-Dalgarno-like (underlined) sequences.

Figure 5.26:

	R. loti	*R.l.* bv. *viciae*
R. meliloti	79.0% ; 66.5%	83.3% ; 70.8%
R.l. bv. *viciae*	81.0% ; 69.4%	

Figure 5.27:

```
                  +1
(1)  -10  GGAAAATCAATTGTCAGATCGTGAGATGGCCCAAGAGGCTCCGCGGTGGCTTGAGCCGAG    47
(2)  -14  AGAAAGATCAAGTGTTGATGCGCGAGTTGGACCCTAAGGATTTGCGGCGGCCCGAGACTGG    44
(3)  -14  AGAAACATCAAGTGTTGAAGCGTAAGTTGGGCCCAGAGGAATTGCGGCGGCTCGAGACTCC    44
(4)  -15            GGGCAATTAACGAATTTGAAGACGACGATCGCTGATCGGCACCAGGATCA    35

      48  TTCGTTC......GAGTGGAAGGACCAAACAGGTCTAGCCGCGAAGACCGCAATACCCGG   104
      45  TCAGATC......GAACGGGAGTCTCACGAGCAATCAAGCGCGAAAAGCTCCGTGTCTGA   101
      45  TGCGATC......GAACGGGAGTCTCACGGGCAAACAAGCCCGAAAAGCTCCGTGCCTGA   101
      36  TCTCATCCTATCGGAGCGGCAGCACCAATGGAAATTTAAGGCAATGGATTCGCCTTCTGG    95

     105  CGCCAAACCAACCGTGGCAATCGATGTTGCCAGCGTAACAAAGTCATACGGTGACAAACC   164
     102  CTCTGCCTCGACCGTGGCGGTCGATTTTACAGGCGTAACCAAGTCGTATGGGAACAAGGT   161
     102  CTCTGCGTCAACCGTGGCAGTCGATTTTGCCGGGGTAACCAAGTCGTATGGGAACAAGAT   161
      96  CTCATTGTCCCCAGTCGCAATCGACCTGGCCGGCGTTTCGAAATCATATGGTGGTAAAAT   155

     165  TGTAATCAAC   174
     162  CGTCGTCGAC   171
     162  CGTTGTCGAC   171
     156  CGTCGTCAAT   165
```

second ATG codon for *R.l.* bv. *viciae*. The 5'-terminal of *R. meliloti* *nod*I was also found at the 3'-end of the *nod*ABC sequence submitted for this organism (Egelhoff *et al.*, 1985). A similar alignment revealed an absence of a start codon matching the *R.l.* bv. *viciae* at the proposed initiation site. The same has been observed for the *R. loti* NZP2213 *nod*I gene (Young *et al.*, 1990). Consequently, upstream sequences for *nod*I gene regions were included for investigation. A comparison of the four nucleotide sequences is displayed in Figures 5.27 - 5.28.

R. meliloti and the two *R. loti* sequences in Figure 5.27 share a greater identity at the 5'-end of the sequences than the *R.l.* bv. *viciae* does to any of the other sequences. There appears to be no easily identifiable start codon for some of the *nod*I sequences however, the rarely used "TTG" initiation codon could be considered a candidate. Although the second "TTG" in the *R.l.* bv. *viciae*, the other aligned TTGs in the *R. loti* strains and the ATG in *R. meliloti* could all be considered as possible initiation codons, the absence of strong Shine-Dalgarno sequences upstream, (based on the polypurine hexamer AGGAGG consensus sequence in *E. coli*; (Lewin, 1983)), suggests that the start of the sequences as highlighted in Figure 5.27 presents the best option. The in-frame GTG and ATG that flank the TTG in *R. loti* NZP2037 at least make this area a good candidate for the initiation of the *nod*I gene. A comparison of the predicted amino acid sequences is presented in Figures 5.29 - 5.30. The four amino acid sequences begin to share considerable homology approximately 40 amino acids in from their proposed start codons.

5.5.6 LOCATION OF THE Tn5 INSERTION SITES.

Tn*5* generates a 9 bp direct repeat of flanking DNA at the insertion site. The locations of the Tn5 insertions in the DNA sequence were thus standardised by placing the sites at the right-hand-end of the upstream repeated sequences (see Figures 5.8 - 5.9) The 4019 Tn5 insertion is found 82 bp to the left of the *nod* box, the 4047 insertion 29 bp downstream from the *nod*C initiation site, 233 is 102 bp downstream from the start of *nod*C and 4053 is 35 bp downstream from the suggested initiation site of *nod*I.

Figure 5.28: Comparative similarities of the 5'-terminal of the *nodI* nucleotide sequences using the GAP program. (% identity).

Figure 5.29: An alignment of the predicted *nodI* N-terminal amino acid sequences.

(1) *R. meliloti*

(2) *R. loti* NZP2037

(3) *R. loti* NZP2213

(4) *R.l.* bv. *viciae*

Figure 5.30: A comparison of the predicted *nodI* N-terminal amino acid sequences using the GAP program. (% similarity; % identity).

Figure 5.28:

	R.l. bv. *viciae*	*R. loti* NZP2213	*R. loti* NZP2037
R. meliloti	51.9%	60.2%	56.7%
R. loti NZP2037	53.5%	86.0%	
R. loti NZP2213	52.8%		

Figure 5.29:

```
(1)  1  LSDREMAQEAPRWLEPSSF..EWKDQTGLAAKTAIPGAKPTVAIDVASVTKSYGDKPVIN  60
(2)  1   LMRELDPKDLRRPETGQI..ERESHEQSSAKSSVSDSASTVAVDFTGVTKSYGNKVVVD  60
(3)  1   LKRKLGPEELRRLETPAI..ERESHGQTSPKSSVPDSASTVAVDFAGVTKSYGNKIVVD  60
(4)  1       LKTTIADRHQDHLILSERQHQWKFKAMDSPSGSLSPVAIDLAGVSKSYGGKIVVN  60
```

Figure 5.30:

	R.l. bv. *viciae*	*R. loti* NZP2213	*R. loti* NZP2037
R. meliloti	52.8% ; 32.1%	61.4% ; 36.8%	56.1% ; 33.4%
R. loti NZP2037	56.6% ; 41.5%	84.2% ; 75.4%	
R. loti NZP2213	56.6% ; 41.1%		

5.6 DISCUSSION.

Sequencing of the DNA in which the 233 Tn*5* had inserted, confirmed the plant complementation experimental evidence that the mutation was in the region of the *R. loti nodC* gene. *Nod*C was found to be located between the *nod*A and *nod*I genes. Because of their close proximity and the presence of a single *nod* box upstream of these genes, it is likely that they form an operon. The fragments of the *nod*A sequence and the *nod*C sequence for *R. loti* demonstrate a relatively high degree of homology, at both the nucleotide and putative amino acid sequence levels, with other rhizobial species. The striking difference between the *Bradyrhizobium nod*A and the other sequences is the extra 39 bp at the start of the gene. This region does not show similarity to the DNA upstream of the other *nod*A sequences, so the different start site for *Bradyrhizobium* is unlikely to be the result of a simple point mutation or deletion shifting the start site to another ATG codon upstream of the original. One possibility is that in *Bradyrhizobium*, which is quite distantly related to the fast-growing rhizobia (B.W.D. Jarvis, personal communication), *nod*A represents an ancestral form of the gene that underwent a deletion resulting in the shorter *nod*A genes seen in the more closely related rhizobia. Alternatively, the *Bradyrhizobium nod*A has acquired an insertion resulting in the additional sequence at the start of the gene. The apparent 6 bp insertion in the 3'-terminal of the *R. loti nod*A is also of interest; the mechanism by which it may have been incorporated into this gene is not immediately obvious.

In contrast to *nod*A and *nod*C, the available *nod*I sequences display variability at the putative 5'-terminal. Such divergence could indicate that conservation of the amino acid sequence in this region of the gene is not of great importance or, that this terminal of the protein is host-specific. ATG codons in the *R. meliloti* and *R. loti* sequences are not located in the same region as the proposed start of the *R.l.* bv. *viciae nod*I gene. In fact, the general absence of ATG codons in these published sequences suggests the use of a more rarely used start codon, such as TTG. This is not without precedence, as the *nod*B for *Bradyrhizobium* sp. (*Parasponia*) strain ANU289 is reported to begin with a GTG (Scott, 1986). *Nod*I protein analysis may be necessary to establish the exact initiation site. Until then, it is unlikely that the start codon can be identified with any certainty.

One striking feature of the *R. loti* common *nod* gene arrangement is that it differs from the typical *nod*D-*nod* box-*nod*ABCIJ cluster shared by the organisms investigated to date - the ORFs designated *nod*K (Scott, 1986) and *nod*Y (Nieuwkoop *et al.*, 1987), which are upstream of the *nod*A in some bradyrhizobia, being taken into consideration. In contrast to the quite varied intervals between the *nod* box and the start codon for the downstream *nod* gene, the ATG codon for the *nod*D genes investigated to date is only 5-10 bp to the left of the *nod* box, suggesting that the distance separating the *nod* box may be important. Closer inspection of this region upstream of the *R. loti* NZP2037 *nod* box highlighted an ATG codon in an appropriate position for a *nod*D start site and a short sequence demonstrating similarity to the start of the *R. meliloti nod*D_1 gene after which, the homology ceased. The 4019 Tn5 insertion (82 bp to the left of the *nod* box) yielding a Nod^+ Fix^+ phenotype also indicated that it was unlikely that a functional *nod*D was being produced in this region. This sequence fragment may be the remains of an ancestral *nod*D-*nod* box-*nod*ABCIJ cluster for *R. loti*.

Other evidence that lends support to this possible ancestral gene arrangement came to light after analysis of the DNA in another region of *R. loti* NZP2213 (Scott *et al.*, in preparation) that hybridisided to the *nod*D, *nod* box and *nod*B probes. A comparison of the DNA sequence with the *nod*D genes of other species revealed little similarity to the start of the other genes and no appropriately placed ATG codons. However, a region starting 19 bp to the left of the *nod* box associated with *nod*B in NZP2213 and extending for about another 250 bp, did match well with an area beginning 53 bp into the *R. meliloti nod*D. The region from NZP2037 that displayed similarity to the start of the *R. meliloti nod*D_1 could be inserted into most of the "gap" separating the *nod* box and the beginning of *nod*D in NZP2213 (see Figure 5.31). The region between the areas of homology pinpoints the probable location of DNA breakage during the proposed ancestral chromosome re-arrangement. The tapering-off of similarity between the putative NZP2213 *nod*D sequence and that of *R. meliloti* suggests that the former is also a truncated and non-functional gene.

The region of NZP2037 that displayed similarity to the *R. meliloti nodD*$_1$ was compared to what would normally be the remaining downstream sequence of a *nodD* in this position, i.e. the region further to the left of the *nod* box as viewed conventionally. Within this sequence there exists a region, approximately 140 bp downstream, that shares similarity with the putative *nodD* fragment in NZP2037 and the 5'-terminal of the *R. meliloti nodD*$_1$. What is significant about this observation is that the sequence similarity of the downstream region abruptly terminates at the position where the NZP2213 *nodD* sequence picks up similarity with the *nodD*$_1$ (see Figure 5.32). It is possible that the similarity between the two regions of DNA resulted at sometime in a mismatching of the two areas during crossing-over and resulted in the formation of a truncated gene. Events such as this may have played a role in the proposed re-arrangement of *nod* gene clusters in *R. loti*.

Further sequence information would be required before proposing a mechanism by which the *nodD* and *nodB* genes in *R. loti* were isolated from the common *nodD-nod* box-*nodABCIJ* cluster. However, the repetition of the "CTCGxCGC" motif (highlighted) in the *R. meliloti nodD* coding region (see Figure 5.31) should be noted. Repeated sequences such as these have often been found in regions that have undergone genetic re-organisation. It is possible that mismatching of DNA in this area may have played a role in the proposed re-arrangement of the *R. loti* common *nod* gene cluster. The results of the sequencing work will be discussed in depth in conjunction with the plant complementation results in the following chapter.

Figure 5.31: An alignment between *R. meliloti nodD*$_1$ and sequences upstream of the *R. loti* NZP2037 *nodA nod* box and the *R. loti* NZP2213 *nodB nod* box.

(1) *R. loti* NZP2213 (C. Young and D.B. Scott, personal communication)

(2) *R. meliloti nodD*$_1$ (Gottfert *et al.*, 1989)

(3) *R. loti* NZP2037 (this study)

The *nod* box terminal depicted in the diagram is the complement of what is usually considered the 5'-end in relation to the *nodABCIJ* cluster. Lower case a,t,g or c represents the possibly of A,T,G or C respectively in the sequence. m = A or C; v = A,C or G; s = C or G."CTCGXCGC" = a repeated motif (discussed in text).

Figure 5.31:

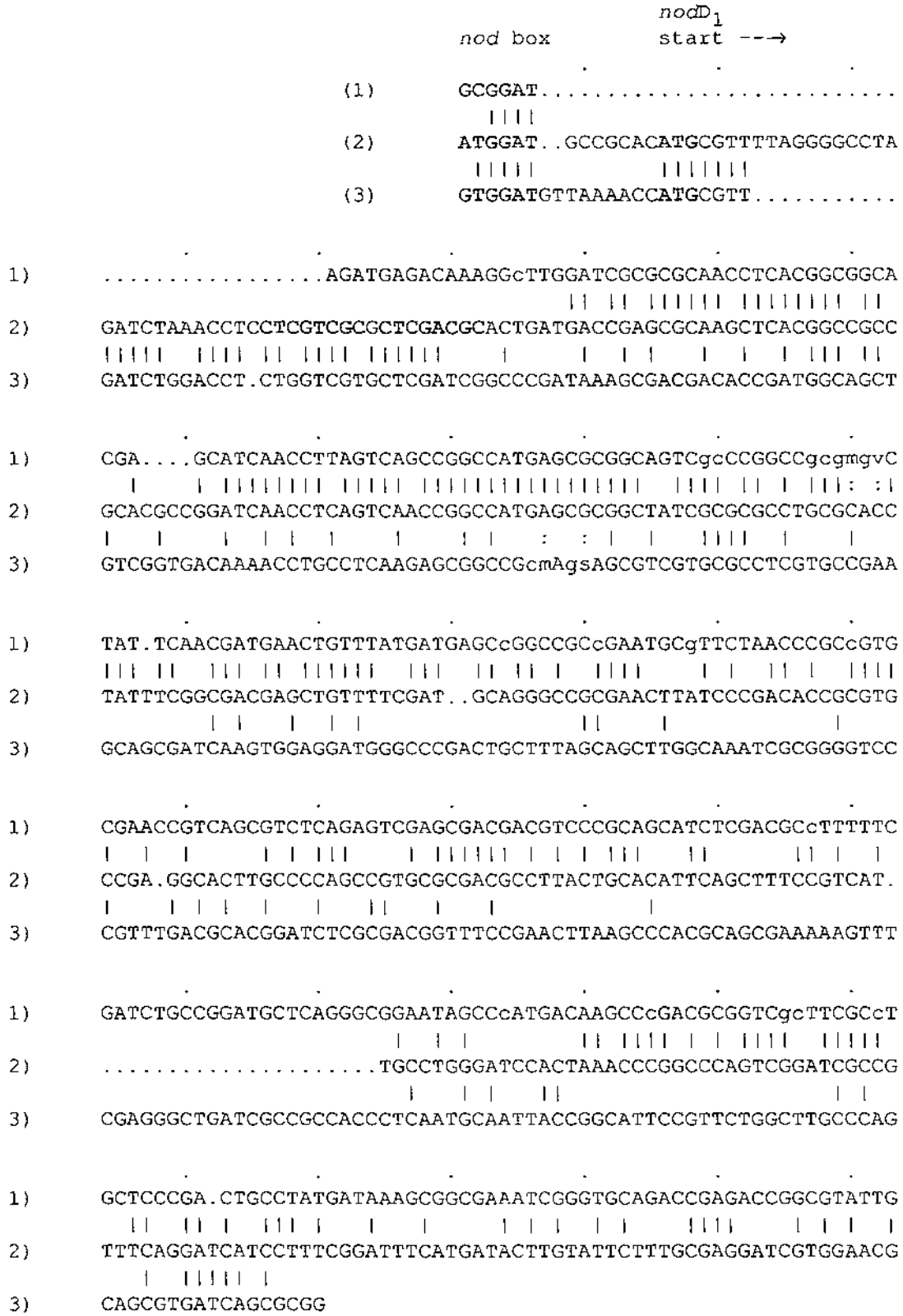

nodD1
nod box
start --→
(1) GCGGAT.....
(2) ATGGAT..GCCGCACATGCGTTTTAGGGGCCTA
(3) GTGGATGTTAAAACCATGCGTT.....
1) .....AGATGAGACAAAGGcTTGGATCGCGCGCAACCTCACGGCGGCA
2) GATCTAAACCTCCTCGTCGCGCTCGACGCACTGATGACCGAGCGCAAGCTCACGGCCGCC
3) GATCTGGACCT.CTGGTCGTGCTCGATCGGCCCGATAAAGCGACGACACCGATGGCAGCT
1) CGA....GCATCAACCTTAGTCAGCCGGCCATGAGCGCGGCAGTCgcCCGGCCgcgmgvC
2) GCACGCCGGATCAACCTCAGTCAACCGGCCATGAGCGCGGCTATCGCGCGCCTGCGCACC
3) GTCGGTGACAAAACCTGCCTCAAGAGCGGCCGcmAgsAGCGTCGTGCGCCTCGTGCCGAA
1) TAT.TCAACGATGAACTGTTTATGATGAGCcGGCCGCcGAATGCgTTCTAACCCGCcGTG
2) TATTTCGGCGACGAGCTGTTTTCGAT..GCAGGGCCGCGAACTTATCCCGACACCGCGTG
3) GCAGCGATCAAGTGGAGGATGGGCCCGACTGCTTTAGCAGCTTGGCAAATCGCGGGGTCC
1) CGAACCGTCAGCGTCTCAGAGTCGAGCGACGACGTCCCGCAGCATCTCGACGCcTTTTTC
2) CCGA.GGCACTTGCCCCAGCCGTGCGCGACGCCTTACTGCACATTCAGCTTTCCGTCAT.
3) CGTTTGACGCACGGATCTCGCGACGGTTTCCGAACTTAAGCCCACGCAGCGAAAAAGTTT
1) GATCTGCCGGATGCTCAGGGCGGAATAGCCcATGACAAGCCcGACGCGGTCgcTTCGCcT
2) .....TGCCTGGGATCCACTAAACCCGGCCCAGTCGGATCGCCG
3) CGAGGGCTGATCGCCGCCACCCTCAATGCAATTACCGGCATTCCGTTCTGGCTTGCCCAG
1) GCTCCCGA.CTGCCTATGATAAAGCGGCGAAATCGGGTGCAGACCGAGACCGGCGTATTG
2) TTTCAGGATCATCCTTTCGGATTTCATGATACTTGTATTCTTTGCGAGGATCGTGGAACG
3) CAGCGTGATCAGCGCGG

Figure 5.32: A comparison of *R. meliloti nodD*$_1$ and *nodD*-like sequences in *R. loti* NZP2037 and NZP2213.

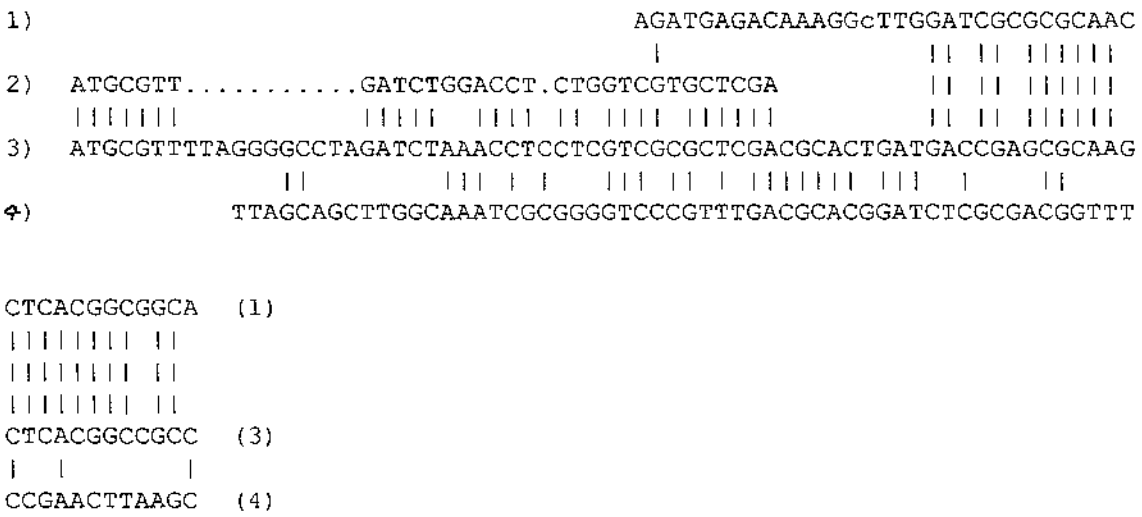
(1) *R. loti* NZP2213 possible *nodD* remnant (C. Young, personal communication)

(2) *R. loti* NZP2037 possible *nodD* fragment.

(3) *R. meliloti nodD*$_1$ (Gottfert *et al.*, 1989)

(4) *R. loti* NZP2037, further upstream (to the left) of (2).

Figure 5.32:


1) AGATGAGACAAAGGcTTGGATCGCGCGCAAC
2) ATGCGTT.....GATCTGGACCT.CTGGTCGTGCTCGA
3) ATGCGTTTTAGGGGCCTAGATCTAAACCTCCTCGTCGCGCTCGACGCACTGATGACCGAGCGCAAG
4) TTAGCAGCTTGGCAAATCGCGGGGTCCCGTTTGACGCACGGATCTCGCGACGGTTT
CTCACGGCGGCA (1)
CTCACGGCCGCC (3)
CCGAACTTAAGC (4)

6.0 DISCUSSION.

6.1 ORGANISATION OF THE *NOD* GENES.

The *R. loti nod*ACIJ arrangement differs from the conventional *nod*ABCIJ grouping in rhizobia investigated to date (reviewed by Long, 1989). Other experimental work (Scott *et al.*, in preparation) suggests that the *R. loti nod*B and a *nod*D like sequence lie approximately 10 kb downstream from *nod*ACIJ. This *nod*D however, may be a non-functional fragment. Both genes are read, or would have read in the case of *nod*D, divergently from a *nod* box that separates them. This gene arrangement raises questions as to the mechanism by which *nod*B was separated from the conventional *nod* gene cluster, which is thought to comprise an operon. Further information would be required before meaningful speculation, concerning the mode of gene cluster re-organisation, could be made. However, it has been demonstrated that DNA reiteration is a characteristic of the Rhizobiaceae genome and, thus conceivably may have played a role in the alterations within the *nod* gene clusters as this characteristic facilitates genetic rearrangements (Flores *et al.*, 1987). Some of these reiterated sequences have been shown to be clustered around the *nif* genes in *B. japonicum* USA 110. The NodD protein is known to bind to the *nod* box. This has been demonstrated using gel retardation (Fisher *et al.*, 1988) and by DNAse I footprinting techniques (Kondorosi *et al.*, 1988). NodD, in conjunction with plant inducers, activates *nod*ABCIJ and host-specificity gene transcription. As *nod*B is preceded by a *nod* box its production is probably co-ordinated with the *nod*ACIJ genes via the interaction of the NodD protein, the plant inducer and the *nod* box. A more detailed discussion of the *nod* box and its possible relationship to the *nod* genes will be presented later in this chapter.

6.2 ANALYSIS OF THE Tn*5* INSERTION SITES AND THEIR PHENOTYPES.

The 4019 Tn5 was found to be located 82 bp upstream from the *nod* box. Its Nod^+ Fix^+ phenotype indicates that the insertion site is sufficiently removed from the *nod* box so as not to interfere with its

functioning. The conventional "common" *nod* gene arrangement includes a *nod*D gene located immediately upstream from the *nod* box that is read divergently from the *nod*ABCIJ grouping. The Nod^+ Fix^+ phenotype of PN4019 and the absence of homology between the sequence upstream of the *R. loti nod* box and other *nod*D genes, indicates that a functional *nod*D gene is probably not present in this region.

The Tn*5* inserts carried by both PN233 and PN4047 fall within the *nod*C coding region. 4047 lies 29 bp in from the start of *nod*C while 233 is located 102 bp downstream from the start site, i.e. only 73 bp separate the two insertions. Their Nod^- phenotype is therefore due to interference with NodC protein production.

The 4053 Tn*5* was found to have inserted 36 bp downstream from the proposed TTG start codon for *nod*I (this study; Young *et al.*, 1990). The phenotype of this mutation depends on the host plant. On *L. pedunculatus* the Tn*5* caused a slight delay in nodulation whereas, in *L. corniculatus* and *L. tenuis*, the phenotype of the Tn*5* insertion is Nod^+ Fix^+. Therefore, Tn*5* mutations in *nod*I and *nod*J do not totally block nodulation. In some species they appear to be unnecessary. For example, in *B. japonicum* a strain with a deletion in *nod*IJ was Nod^+ (Gottfert *et al.*, 1989). Amino acid sequence homologies suggest that NodI and NodJ proteins are membrane-associated and involved in active transport (Evans and Downie, 1986). Thus, the possibility exists that another transport system may be able to substitute for the NodIJ proteins and take up the substrate usually transported by these gene products or, that the substrate may enter the cell more slowly through passive diffusion. Either possibility could explain the slight delaying effect of mutations in *nod*I and *nod*J.

The differing responses of the host plants to the PN4053 mutant may be explained if there is another gene downstream of *nod*J that is required for *L. pedunculatus* nodulation. As Tn*5* insertions usually have a polar effect on downstream genes in the same operon, the 4053 insertion would probably inactivate that gene. This situation would be akin to the presence of the *nod*X gene in *R.l.* bv. *viciae* strain TOM, which is required to nodulate Afghanistan peas (Davis *et al.*, 1988).

6.3 PLANT COMPLEMENTATION TEST RESULTS.

With the sequencing data in hand, some of the questions raised in Chapter 4, concerning the plant complementation tests, can be answered. It was determined that pPN305 bears the *nod* box and *nod*ACI (and through indirect evidence, *nodJ;* C. Young, personal communication) genes while pPN25 carries complete copies of the *nod* box and the full sequences of only *nod*AC. In experiment 1, (see 4.1), the *R.l.* bv. *trifolii nod*C mutant (ANU277) was fully complemented by pPN305. This was due to complete sequences of *nod*C, as well as *nod*IJ, being present on the plasmid and compensating for the *nod*C Tn*5* insert and, presumably, the inactivation of the downstream *nod*IJ in the *R.l.* bv. *trifolii* host. The *nod*B mutant (ANU249), on the other hand, could not be complemented as pPN305 does not carry *nod*B because of the different arrangement of the *nod* genes in *R. loti.* Although *nod*A is encompassed by pPN305, complementation was not achieved in ANU252 because the polar effect of the Tn*5* insertion prevented a NodB protein being produced. As *nod*B is not encoded by pPN305, complementation failed. Likewise, the absence of a *nod*D gene on pPN305 led to nodulation failure in the ANU851/pPN305 exconjugant.

To facilitate analysis of the results from the second complementation experiment (see 4.2) Table 6.1 lists the *nod* genes that are predicted to be present and expressed and the number of copies expected in each exconjugant. It also incorporates the complementation results.

From the results of the NZP2037/pPN305 exconjugant, it can be seen that multiple copies of the *nod* genes do not appear to have any detrimental effect. The slight delay in nodulation may be due to the presence of the pLAFR1-derived plasmid as discussed in Chapter 4. In the remainder of the crosses involving pPN305, all the "common" *nod* genes were present in the rhizobial cell and, with the exception of PN233, full complementation occurred. Since each of the "common" *nod* genes was present and, as sequencing of the Tn*5* insertion sites had demonstrated that the Tn5 inserts carried by PN233 and PN4047 were both within the *nod*C coding region, the differences in their ability to form an effective nodule is difficult to explain.

Table 6.1: Complementation of *R. loti* and *R.l.* bv. *trifolii nod* mutations (Experiment 2) including the number and type of *nod* genes expected to be expressed.

Plasmids	Host *L. pedunculatus* Strains NZP2037	PN233	PN4047	PN4053	Host *R.l.* bv. *trifolii* Strain ANU277
pPN305 (*nod*ACIJ)	D B A A C C I I J J Nod^{d}	D B A A C I J Nod^{-T}	D B A A C I J Nod^{d}	D B A A C C I J Nod^{d}	D A A B C I J Nod^{+}
pPN25 (*nod*AC)	N-D.	D B A A C Nod^{d}	D B A A C Nod^{d}	D B A A C C Nod^{d}	D A A B C Nod^{-}
pPN26 (*nod*DABCIJ)	N-D.	D D B B A A C I J Nod^{-T}	D D B B A A C I J Leaky, Nod $^{+/-}$	D D B B A A C C I J Leaky, Nod $^{+/-}$	D D A A B B C I J Nod^{+} see Table 4.1

Key: Nod^{d} = a slight delay in nodule initiation; Nod^{-T} = tumorous growths on the root but no bacteria observed within the tissue; Nod^{+} = nodulation; leaky, Nod $^{+/-}$ = some plants nodulated and others did not; N-D = not done.

Crosses involving the transfer of pPN25 into the recipient strains appear to indicate that *nodI* and *nodJ* and any co-transcribed downstream genes that may be present are not essential for nodulation in *L. pedunculatus*. In *R.l.* bv. *trifolii* however, the *nodI* and *nodJ* and/or a downstream gene are apparently required for *T. repens* nodulation.

The crosses involving the transfer of pPN26 into *R. loti* recipients also generated interesting results. pPN26 would be expected to complement the PN233 mutant by supplying *nodCIJ*. However, this resulted only in tumorous growths. The comparison between PN233 and PN4047, the two *nodC* mutants, once again demonstrates a significant difference in their responses. It may be that a gene carried elsewhere on the plasmid, possibly a host-specificity gene, is interfering with the PN233/pPN26 exconjugant, but then one would expect the same in PN4047 and PN4053. The "leaky" phenotype displayed by the PN4047/pPN26 and PN4053/pPN26 exconjugants, with some plants nodulating and others not, suggests that the host plant genotype may be influencing this response. One possibility is that host-specificity genes on pPN26 may be slightly altering *R. loti*-generated signal molecules. The response of the host plant may thus depend on the host's genetic background. Another possibility is that the *R.l.* bv. *trifolii nodD* does not respond well to the *L. pedunculatus* flavonoids.

Certain observations can be made from comparing the data in Table 6.1. Firstly, a straight-forward gene dosage effect does not appear to be responsible for the unexpected complementation results as NZP2037/pPN305, with its doubling of the *nod* genes carried on pPN305 and the unbalanced dosages present in the PN4047/pPN305 and PN4053/pPN305 exconjugants, are able to complement. One cannot explain why neither pPN305 nor pPN26 fully complement the PN233 mutant, but that the smaller pPN25 plasmid does, nor why PN233 and PN4047, the two *nodC* mutants, behave differently. Also of note is the response of PN233, PN4047 and PN4053 to both pPN305 and pPN26, i.e. they failed to fully complement in PN233 but could successfully elicit nodulation when carried in the PN4047 and PN4053 strains.

The difference in the response of the PN233 and PN4047 mutants remains the most puzzling problem. The Tn*5* insertions in these strains would be expected to block NodC production and that of any genes downstream of *nod*C, that are co-transcribed. For the 233 and 4047 mutations to behave differently, it is proposed that the PN233 strain is actually producing a mutant NodC product whose interaction with another gene or genes is responsible for the unusual complementation results.

Jacobs *et al.* (1985) carried out complementation tests wherein plants were inoculated with *Rhizobium* strains carrying genomic Tn*5* insertions at one location in *nod*C and another plasmid-borne Tn*5* at a different site on homologous DNA. The results revealed that *nod*C, at least in *R. meliloti*, behaved as two distinct complementation groups in that *trans* complementation was exhibited, i.e. intergenic complementation. Analysis of the sequence data from the region between the two complementation groups did not reveal chi-sites (recombinational hot spots). The genetic behaviour of *nod*C is therefore likely to be due to the NodC product having two functional domains (Jacobs *et al.*, 1985). Analysis of the putative amino acid sequence for the NodC protein predicts two major domains (John *et al.*, 1988). It is possible that these domains may be able to reconstitute an active molecule if produced separately. This may explain the intergenic complementation exhibited by *nod*C. It has been proposed that the multiple, in-frame ATG sites in the middle of the *nod*C coding sequence may allow translation re-initiation (Jacobs *et al.*, 1985). (There are six, in-frame methionine codons in identical positions in the three amino acid sequences compared in Figure 5.25. They lie between positions 156 and 248). However, there are no strong, appropriately placed Shine-Dalgarno sequences upstream of these methionine codons. If PN233 was re-initiating translation at one of these sites and thus producing a mutant, truncated NodC product, one would expect the same for PN4047. Thus, this is unlikely to account for the differences between the PN233 and PN4047 mutants.

Not all Tn5 mutations have a polar effect (Ruvkun *et al.*, 1982; Corbin *et al.*, 1983; de Bruijn and Lupski, 1984; Rostas *et al.*, 1986). It was reported by Corbin *et al.* (1983) that Tn5 insertions near a particular site in two *R. meliloti* strains could promote transcription of a gene. This indicates that the region of Tn5 insertion may influence promoter activity. Berg *et al.* (1980) found weak non-polarity of some Tn5 insertions in the *lacZ* gene of *E. coli*. Mulligan and Long (1985) reported that a promoter reading out of Tn5 appeared to function in *R. meliloti*. There is a TTG codon positioned near the end of the Tn5 sequence that is in-frame with the rest of *nod*C in PN233. Upstream, in an appropriate location, is a Shine-Dalgarno sequence (see Figure 6.1). It may be possible that a mutant NodC protein is being produced from this start site. The proposal that a TTG site is being recognised as the start codon for *nod*I, may support this. Also, unusual start sites are not without precedent. *Nod*B in *Bradyrhizobium* sp. (strain ANU289) is predicted to begin with a GTG (Scott, 1986).

Another possibility is that the truncated N-terminal region of the NodC protein is the cause of complementation failure in PN233. However, such a protein would mainly comprise what is predicted to be the signal sequence responsible for the correct targeting of the gene product. Therefore, the N-terminal region is thought less likely to be the cause of complementation interference than a mutant NodC protein produced from re-initiation within or around the Tn5 insertion site. Table 6.2 was constructed to compare the *nod* genes present and their number in each cross as well as the complementation results, on the assumption that a mutant NodC product is produced in PN233 through re-initiation of transcription and, which in turn, permits expression of downstream genes.

Assuming that PN233 produces a mutant NodC product, then it would appear that complementation failed in cases where two copies of *nod*C, one mutant and one wild-type, were present with duplicate copies of *nod*IJ and any other gene(s) carried on pPN305. Analysis of the predicted amino acid sequences for *nod*I and *nod*J suggests that they

Figure 6.1: A TTG codon in the 233 Tn5 insertion is in-frame with the *nodC* coding region.
Bold represents Tn5 terminal repeat sequence. A Shine-Dalgarno-like sequence is underlined. *** is the TTG codon which is in-frame with the downstream *nodC* sequence. In-frame codons are separated by commas. The *nodC* amino acid translations are shown.

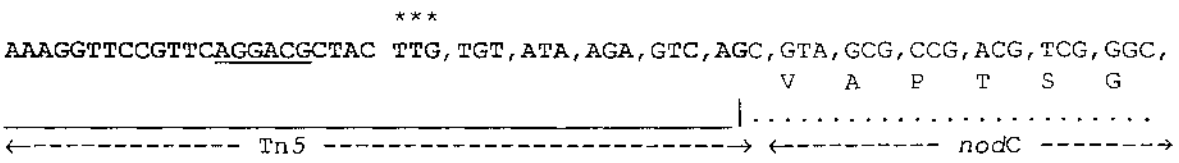

***
AAAGGTTCCGTTCAGGACGCTAC TTG,TGT,ATA,AGA,GTC,AGC,GTA,GCG,CCG,ACG,TCG,GGC,
V A P T S G
Tn5
nodC

Table 6.2: Complementation results of *R. loti* and *R.l.* bv. *trifolii nod* mutations (Experiment 2) assuming PN233 is producing a mutant NodC (C^*).

Plasmids	Hosts *L. pedunculatus* Strains NZP2037	PN233	PN4047	PN4053	Host *R.l.* bv. *trifolii* Strain ANU277
pPN305 (*nod*ACIJ)	D B A A C C I I J J Nod^d	D B A A C^*C I I J J Nod^{-T}	D B A A C I J Nod^d	D B A A C C I J Nod^d	D A A B C I J Nod^+
pPN25 (*nod*AC)	N-D.	D B A A C^*C I J Nod^d	D B A A C Nod^d	D B A A C C Nod^d	D A A B C Nod^-
pPN26 (*nod*DABCIJ)	N-D.	D D B B A A C^*C I I J J Nod^{-T}	D D B B A A C I J Leaky, Nod $^{+/-}$	D D B B A A C C I J Leaky, Nod $^{+/-}$	D D A A B B C I J Nod^+ see Table 4.1

Key: Nod^d = a slight delay in nodule initiation; Nod^{-T} = tumorous growths on the root but no bacteria observed within the tissue; leaky, Nod $^{+/-}$ = some plants nodulated and others did not; N-D = not done.

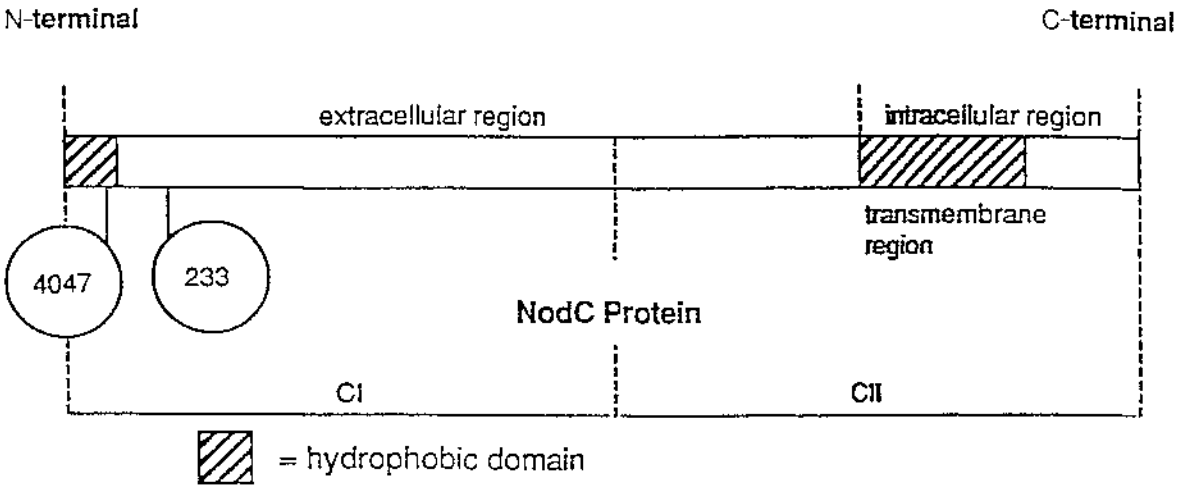
are involved in a membrane transport system and are therefore unlikely to be interacting with NodC whose structure indicates that it may function as a cell surface receptor (John *et al.*, 1988). For this reason it is thought likely that it is another gene(s) carried on pPN305, rather than *nod*IJ, that is interacting with the putative mutant NodC product and preventing nodulation in the PN233/pPN305 and PN233/pPN26 exconjugants.

Alternatively, complementation failure may not be due to the mutant NodC as such, but the fact that re-initiation of translation allows the production of proteins from co-transcribed downstream genes creating an inbalance in the ratio of functional NodC product to NodIJ and any co-transcribed downstream genes. That is, while two copies of *nod*C in combination with single copies of *nod*IJ and any co-transcribed genes permits nodulation, the reverse situation, i.e. 1x *nod*C + 2x *nod*IJ and downstream genes, interferes with the normal nodulation process.

6.4 *NOD* GENE PRODUCTS.

From the predicted amino acid sequence, the NodC protein shows similarities to cell surface receptors. It is thought to comprise two hydrophobic domains that are separated by a hydrophilic, extracellular domain, and followed by a C-terminal hydrophilic intracellular segment. The N-terminal hydrophobic region is thought to correspond to a signal sequence required for the proper targeting of the protein, while the large hydrophobic segment near the C-terminal is thought to act as a membrane anchor (John *et al.*, 1988; see Figure 6.2). The border between the two complementation groups lies in the large, extracellular hydrophilic domain at a position approximately 120 bp upstream from the border of the C-terminal hydrophobic region (Jacobs *et al.*, 1985). This suggests that both the extracellular and membrane-located domains perform distinct functions.

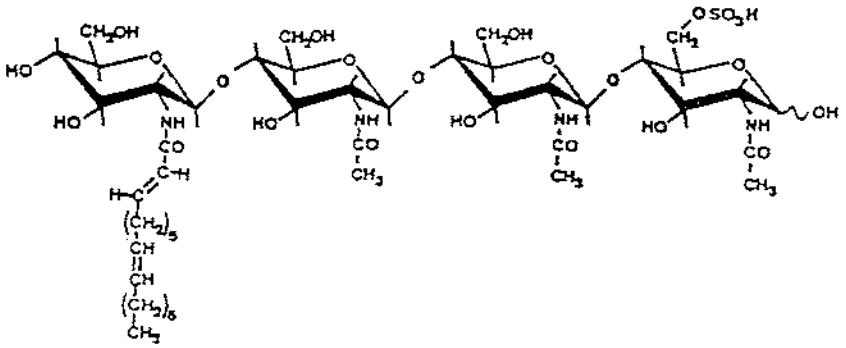
Figure 6.2: The proposed structure for the NodC protein (John *et al.*, 1988). The relative positions of the 4047 and 233 Tn*5* insertions sites are illustrated.


N-terminal
C-terminal
extracellular region
intracellular region
transmembrane region
4047
233
NodC Protein
CI
CII
= hydrophobic domain

In *R. meliloti*, the NodC protein is required for the production of an alfalfa-specific nodulation signal, NodRm-1, which is an N-acyl-tri-N-acetyl-β-1,4-D-glucosamine tetrasaccharide bearing a sulphated group on carbon 6 of the reducing sugar moiety (Lerouge *et al.*, 1990; see Figure 6.3). It will elicit root hair curling in alfalfa. N-acetyl-D-glucosamine is involved in the biosynthesis of two major components of the cell wall in Gram-negative bacteria. These are peptidoglycan in the periplasmic space and lipid A in the outer membrane (Lerouge *et al.*, 1990). It would be useful to determine whether Nod signals are products of a synthetic pathway originating with N-acetyl-D-glucosamine, or if they are involved with the degradation of cell wall macromolecules.

It is known that oligosaccharides, which are derived from cell wall degradation, can act as regulatory molecules (Albersheim and Darvill, 1985) and are active in the nanomolar range (Tran Thanh Van *et al.*, 1985). In tobacco explants, oligosaccharides derived from sycamore cell walls have been shown to regulate morphogenesis (Tran Thanh Van *et al.*, 1985). Abe *et al.* (1984) reported that clover root hair infection was stimulated by lectin-binding oligosaccharides that were derived from polysaccharide depolymerase-treated capsular and extracellular polysaccharides derived from *R.l.* bv. *trifolii*. Also, small, cell wall components (e.g. oligosaccharides) released from either damaged plant cell walls or those of a pathogen have been shown to be involved in eliciting the hypersensitive or wounding response and phytoalexin production in plants (Albersheim and Darvill, 1985). The plant/*Rhizobium* symbiosis raises questions concerning bacterial evasion of plant defence mechanisms, or conversely, plant recognition of compatible rhizobia. In order to generate a symbiotic nodule, the invading bacteria must somehow evade or subvert the normal host defence mechanisms and, on the other hand, the plant must be able to recognise compatible rhizobia. It has been proposed that infection thread formation represents a modified host defence response. The potential conceptual and mechanistic links with plant disease resistance/susceptibility and the *Rhizobium*/host symbiosis is what provides much of the impetus for nitrogen fixation investigations.

Figure 6.3: The alfalfa-specific nodulation signal, NodRm-1 (Lerouge *et al.*, 1990).



It has been suggested that the host-specificity genes *nod*FEL in *R.l.* bv. *viciae* are involved in the biosynthesis of signal molecules. NodF shows sequence similarity to acyl-carrier proteins (Shearman *et al.*, 1986) while NodE shares sequence similarity with a β-ketoacyl synthase (Downie, 1989; Sherman *et al.*, 1989). Based on sequence comparisons, *nod*L is thought to encode an acetyl-transferase; it possibly acetylates a sugar (Downie, 1989). That is, NodFEL proteins share homologies with enzymes that are involved in metabolism of acetate via acetyl-CoA (Downie, 1989). Of interest in this matter is the specific alterations in the acetylation pattern of exopolysaccharides from hybrid recombinant *R.l.* bv. *viciae* carring *R.l.* bv. *trifolii* host specificity genes. These alterations changed the host range of the recipient *R.l.* bv. *viciae* (Philip-Hollingsworth *et al.*, 1989). Also, the acetylation and acylation of the alfalfa-specific nodulation signal, NodRm-1, should be noted (see Figure 6.3). It is possible that acetylation patterns of polysaccharides and/or nodulation signals may be effected by host-specificity genes such as *nod*FEL and could be involved with host-plant recognition of compatible rhizobia.

Given a plant's ability to detect the slight differences between active and inactive oligosaccharides (Albersheim and Darvill, 1985), it is possible that small alterations in acetylation patterns may be sufficient to enable distinction between compatible and heterologous rhizobial species by the host. It is conceivable that *nod*ABC are responsible for the production of a common product in all rhizobia that is then "individualised", for example, by decoration with acetyl groups in a unique pattern, by host-specificity genes, enabling the plant to distinguish between signals from homologous and heterologous rhizobia.

6.5 REFLECTIONS ON THE PRESENCE OF A PUTATIVE PN233 *NODC* PRODUCT.

Assuming that the PN233 strain was producing a mutant NodC by re-initiation of transcription and, based on the proposed domain structure of the protein, the PN233 NodC product would most likely be

missing the N-terminal hydrophobic region that is thought to act as a signal sequence for ensuring correct targeting of the protein to the membrane (see Figure 6.2). As a consequence, the truncated protein may remain within the cytosol. This, together with multiple copies of another gene or genes on the pPN305 plasmid, excluding the region encompassed by pPN25, may be having a detrimental effect.

While NodA and NodB proteins are present in virtually constant amounts during nodule development, NodC concentration increases (John *et al.*, 1988). Protein isolated from nodules of different plant species, indicates that NodC has been processed from a 46.8 kDa protein to one of about 34 kDa; somewhat larger in *Glycine* (soya bean) (John *et al.*, 1988). Whether the truncated protein plays a role is not known but the fact that it appears to be a general phenomenon indicates that it may have an important function in nodule development.

If NodC does possess two functional domains (see Figure 6.2) it is possible that only one of these would be inactivated in a PN233 truncated protein. NodC is thought to exist in the membrane as at least a dimer (John *et al.*, 1988). Assuming that a truncated NodC can complex with the wild-type and that both are being produced in the same cell, as expected in some of the exconjugants, the mutant protein would titrate out the number of wild-type dimers or complexes formed. For example, given the random coupling of two proteins, only one quarter would comprise a pair of wild-type NodC proteins. It is also possible that mutant complexes may be able to perform one of the functions of the NodC protein. For example, if the 34 kDa truncated protein isolated from nodules is performing a function, the PN233 protein may contain sufficient information to fulfil that role. Alternatively, the 233 mutation may prevent proper processing of the protein. It is possible that the gene(s) that is interacting with the putative mutant NodC is involved with the protein's processing.

The production of a mutant NodC however, may not be the direct cause for complementation failure. Re-initiation of translation of the co-ordinately transcribed downstream genes would create an imbalance of these proteins in relation to the one wild-type copy of NodC. This may be responsible for the Nod$^-$ phenotype of the PN233/pPN305 and PN233/pPN26 transconjugants. These factors should be taken into consideration in any further investigation of the mutant. On this note it is suggested that one of the first steps would be to elucidate whether a mutant *nod*C RNA or protein is being produced in PN233 by the Northern or Western blotting technique. If a protein is produced then it's cellular location should be ascertained and it's state of processing determined.

6.6 THE *NOD* BOX.

Computer analysis of the variable-lengthed sequences separating the *nod* box and the downstream start codons for the inducible *nod* genes available in GenBank did not highlight any significant similarities. The object of this exercise was to attempt to identify where the Pribnow box, (-10 sequence [TATAAT consensus sequence in *E. coli*]) and the -35 sequence (TTGACA consensus sequence in *E. coli*), that are generally necessary for recognition by the RNA polymerase, are situated. Other researchers (Rostas *et al.*, 1986) addressed this issue and were unable to identify these sites. Spaink *et al.* (1987a) investigated the promoter region of the *R. leguminosarum* Sym plasmid pRL1JI. They found that complete promoter activity for the inducible *nod*A is carried on a fragment 72 bp in size that encompasses the *nod* box and an additional 21 bp downstream of it. In this region they identifed a loosely-conserved sequence, AT(T)AG, that was generally about 13 bp downstream of the 3'-terminal of the *nod* box. Removal of this sequence strongly reduced the promoter activity and it was therefore suggested that this forms part of the inducible promoter.

The author would like to propose that the AT(T)AG sequence may actually be the Pribnow box and that the -35 sequence is contained within the 3'-end of the *nod* box consensus sequence, possibly the TTACCA hexamer. Transcription for a number of *nod* genes has been reported to start 6-11 bp downstream from the AT(T)AG "motif" (Fisher *et al.*, 1987b; 1988; de Maagd *et al.*, 1989) indicating that it is situated in an appropriate position to serve as the -10 sequence (see Figure 6.4).

NodD is known to bind to the *nod* box both in the presence and absence of root exudate (Hong *et al.*, 1987). It has been demonstrated with gel retardation experiments that NodD-*nod* box binding requires both halves of the *R. meliloti nod* box and footprinting experiments identified the regions protected from DNAse I activity by the bound NodD protein (Kondorosi *et al.*, 1989; see Figure 6.5). The symmetry displayed by the protected areas suggests that a dimer may be binding to the *nod* box. This may indicate a situation analogous to that of the *lac* promoter, in which gene expression is governed by the co-operative binding of proteins to the promoter. Such an arrangement allows for the rapid switching off and on of gene expression in response to relatively small changes in concentration of the regulatory molecule (see Figure 6.6).

In contrast to the variable distance between the start sites for the *nod* genes downstream from the *nod* box, the ATG codons for the *nod*Ds (to the left) are about 5-10 bp from the 5'-end of the *nod* box consensus sequence. It would appear that the distance separating *nod*D and the *nod* box is significant and may play a role in expression of this gene. In *R.l.* bv. *viciae* NodD production is constitutive and manifests autoregulation (Rossen *et al.*, 1985). It is possible that the binding of the NodD protein to the *nod* box could prevent further *nod*D transcription hence autoregulating its own production. In *R. meliloti* the transcription start site for *nod*D is 8 bp downstream from the 3'-end of the *nod* box (Fisher *et al.*, 1987a). It is quite possible in this case that the complement of the proposed -10 sequence for the genes downstream from the *nod* box is also the Pribnow box for *nod*D on the opposite strand; (i.e. in this case, TCTAAT).

Figure 6.4: Some transcription initiation sites for genes downstream from the *nod* box (de Maagd *et al.*, 1989).
The *nod* box and the associated conserved sequence are highlighted while the transcription start sites are indicated by *.

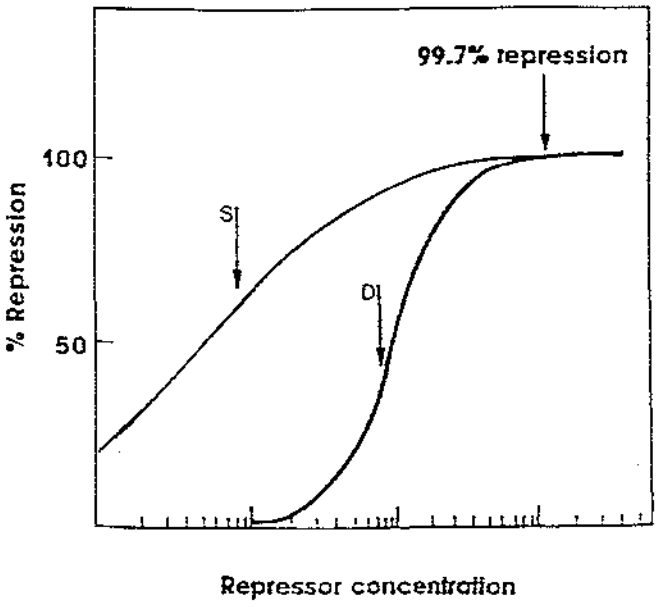
Figure 6.5: Areas of the *R. meliloti nod* box protected from DNAse I digestion by NodD binding. Bold represents the protected bases (Kondorosi *et al.*, 1989).

Figure 6.6: Co-operative binding of the lambda repressor to the operator allows gene expression to be rapidly switched "off" and "on" (based on Ptashne, 1986).
As can be seen from the diagram, switching from an "on" state to an "off" state is very gradual with a promoter that is controlled by a single (S) repressor molecule. However, switching from "on" to "off" states is much more rapid when the system is controlled by binding of dimers (D) to the promoter as only a relatively small change in repressor concentration is needed to change from an "off" to an "on" state and *vice versa*.

3′-terminal of
the *nod* box

```
                                        **
CAATCAATTTTACCAATCTTTCGGATCACTTATAGAAAACCCGG      R.l. bv. viciae  nodA
                                        **
CAATCAATTTTACCAATGATGCCATATGATCCATAGCAGGGCAG      R.l. bv. viciae  nodF
                                         *
CAATGGATTTCACTAATTCGCTCTTGGAAAAGATAAGGGGCACA      R.l. bv. viciae  nodO
```

```
          .....................nod box.....................
          |                                                 |
5′-GTGCGGCATCCATATCGCAGATGATCGTTATCCAAACAATCAATTTTACCAATCTTGCAGAG  3′
3′-CACGCCGTAGGTATAGCGTCTACTAGCAATAGGTTTGTTAGTTAAAATGGTTAGAACGTCTC  5′
```



It is known that NodD, together with plant inducers, is required for the expression of other *nod* genes downstream of the *nod* box. Binding of the NodD to the *nod* box may cover an RNA polymerase binding site in the *nod* box that is required for *nod* gene transcription and hence regulation of their expression. The inducer may cause a conformational change in the NodD protein that results in the exposure of the -35 and -10 sites allowing formation of an open complex with RNA polymerase and transcription of the *nod* genes, possibly even enhancing their expression. This may be supported by the finding that mutations in the carboxyl-terminal region of the NodD protein affect *nod* gene activation and this is also the region that interacts with the flavonoid inducers (Burn *et al.*, 1989). It is possible that the flavonoids that act as inhibitors, bind with NodD but do not trigger the appropriate conformational changes to allow expression of the other *nod* genes. A model for this system is outlined in Figure 6.7. It has also been observed in a number of *R. meliloti* strains that regulation of *nod* gene expression is under both positive and negative control (Kondorosi *et al.*, 1989). A repressor protein that binds to a 33 bp region of DNA just downstream from the *nod* box affects gene expression, as it overlaps the transcriptional start sites for *nodA* and $nodD_1$.

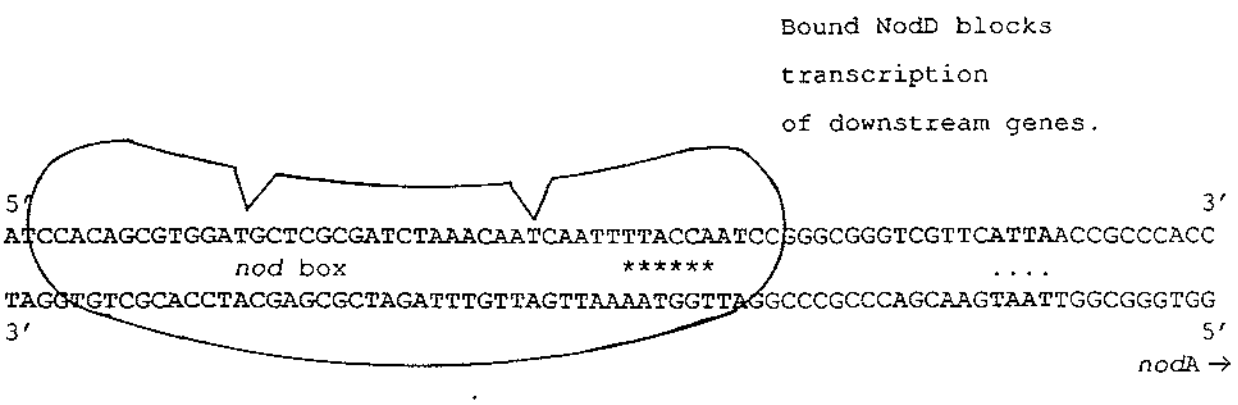
The *nod* box associated with *nodO* in *R.l.* bv. *viciae*, is poorly conserved and the induced gene shows only 23% the activity of the induced *nodA* promoter. However, when multiple copies of *nodD* are present *nodO* expression is increased by 650% (de Maagd *et al.*, 1989). Given the comparatively poor conservation of the *nod* box for that gene it is possible that the NodD protein displays a lower binding constant for this *nod* box, leading to a low level of expression of *nodO*. Multiple copies of *nodD* would likely increase the quantity of NodD in the cell thereby probably elevating the frequency of occupation of the *nod* box by NodD and hence possibly permitting amplified *nodO* expression in the presence of inducers.

Figure 6.7: A model for *nod* gene expression in *R. loti*.

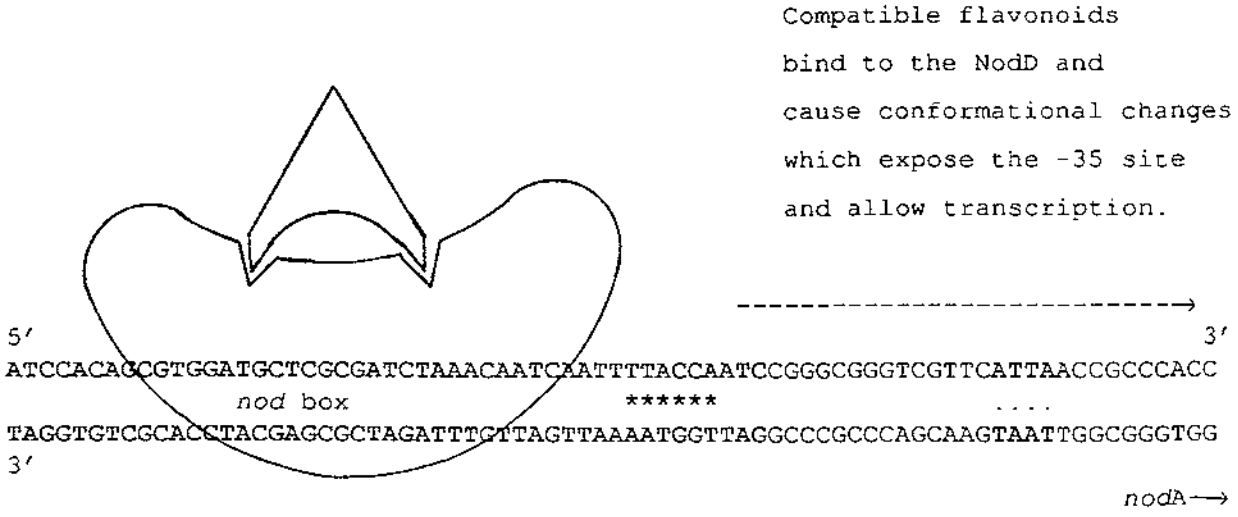
a. In the absence of flavonoids, NodD is bound (possibly as a dimer) to the *nod* box and obscures the -35 site (***) for the downstream *nod* genes and therefore prevents their transcription. (....) represents the proposed - 10 site.

b. Plant flavonoids bind to NodD and cause it to undergo a conformation change which exposes the -35 site (***) allowing transcription of the downstream *nod* genes.

a.


Bound NodD blocks
transcription
of downstream genes.
5′
3′
ATCCACAGCGTGGATGCTCGCGATCTAAACAATCAATTTTACCAATCCGGGCGGGTCGTTCATTAACCGCCCACC
nod box
******
....
TAGGTGTCGCACCTACGAGCGCTAGATTTGTTAGTTAAAATGGTTAGGCCCGCCCAGCAAGTAATTGGCGGGTGG
3′
5′
nodA →

b.


Compatible flavonoids
bind to the NodD and
cause conformational changes
which expose the -35 site
and allow transcription.
5′
3′
ATCCACAGCGTGGATGCTCGCGATCTAAACAATCAATTTTACCAATCCGGGCGGGTCGTTCATTAACCGCCCACC
nod box
******
....
TAGGTGTCGCACCTACGAGCGCTAGATTTGTTAGTTAAAATGGTTAGGCCCGCCCAGCAAGTAATTGGCGGGTGG
3′
nodA⟶

As can be seen from the *R. meliloti* strains carrying a repressor protein, there is variability in the control of *nod* gene expression among species and strains and more information will need to be acquired before a clear picture of the *nod* gene regulation can be developed. Monitoring the effect of site-specific mutations and small deletions in the *nod* box may be useful in further investigations for determining the precise function of the different regions within the *nod* box.

6.7 FURTHER INVESTIGATION POSSIBILITIES.

There are a number of areas where further investigation may help to resolve some of the questions raised by the project concerning *nod*C. Firstly, the phenotypes of wild-type *R. loti* bearing the 233 Tn*5* insertion on plasmids carrying the 7.1 kb and 22.7 kb *Eco* RI fragments respectively, should be tested to determine whether the mutation has the same effect as in the original configurations. Next, the region on the 22.7 kb fragment that is responsible for the failure to complement in PN233 transconjugants should be identified. Initially, to verify whether a co-transcribed gene downstream of *nod*C is responsible for the complementation failure, pPN306 carrying the 4053 Tn5 insertion which is in *nod*I, could be crossed into PN233. Assuming that the Tn*5* insertion displays the usual polar effect, it would prevent the production of genes downstream of *nod*C that are co-transcribed. If the gene(s) that are responsible for the unusual complementation results seen in PN233 is co-transcribed with *nod*C, then the 4053 Tn5 insertion should prevent translation of one of the two copies of the gene(s) in the transconjugant resulting in a 1:1 ratio of wild-type NodC to the theorised protein(s). If the inability of the 22.7 kb fragment to complement in the PN233 is a result of there being one functional copy of *nod*C to two copies of *nod*IJ and any co-transcribed genes, then it is predicted that the PN233/ pPN306::4053 Tn*5* transconjugant should be able to form normal nodules. Such a result would focus the region of investigation on the 22.7 kb insert very rapidly. (It should be kept in mind however, that the pPN306 carries a slightly larger chromosomal

insert than pPN305 as it possess an extra 7.5 kb *Eco* RI fragment, see Figure 3.1., and this could possibly be a complicating factor in the interpretation of such an experiment).

If this experiment was not fruitful, the next step would be to progressively delete sections of the insert DNA, possibly by conducting partial digests of the chromosomal insert DNA and re-ligating appropriate fragments, followed by complementation experiments. This would isolate the region of interest. This process could be used in conjunction with Tn*5* mutagenesis of the 22.7 kb fragment. Complementation tests after each step would be conducted to identify the region of interest. The phenotype of the Tn*5* mutation(s) in the newly identified gene(s) could next be characterised alone, followed by their interaction with the other strains i.e. NZP2037 and PN4047. The next step would be to sub-clone the responsible fragment and perform in-depth analysis of this region by sequencing the DNA and identifying the gene(s). Such experiments may provide information on the interaction between *nod* genes.

Secondly, it should be determined whether PN233 is producing a mutant NodC protein. This may be approached from a number of angles. Firstly, one could make maxicells (possiby *E. coli*, however *Rhizobium* maxicells may have to be developed as the *Rhizobium* promoters may not be recognised by in *E. coli*) that carry the 7.1 kb fragment; the NZP2037, PN4047 and PN233 strains would be compared. These fragments would then be analysed for protein production. The production of a mutant PN233 protein could thus be established and its size determined by running polyacrylamide gels.

Northern blotting techniques by which DNA probes are hybridised to mRNA could be utilised to identify the production of an altered PN233 NodC bearing mRNA. The translation re-initiation site in the mutant could be established by studying the 5′-terminal of the transcript with a primer extension procedure. This information may be of value in understanding gene regulation in *Rhizobium*, for example, by identifying what is acting as the promoter sequence in this gene and whether or not a rare start site is being employed. Determination

of the 5'-end of the mRNA would also allow deduction of exactly how much of the native *nod*C protein the PN233 mutant is missing and, in conjunction with other data, it may provide clues on the roles of the different domains of the wild-type protein.

Another avenue of investigation would be to determine possible antigenic sites on the PN233 protein, make peptides to match and use these to developed antibodies to the NodC protein. Isolation of the processed NodC proteins produced by the wild-type and PN233 strains, from root nodules and tumours respectively, would be attempted; their sizes would be compared on protein gels to determine if the mutation prevents or changes the processing of the PN233 NodC protein. Also, the Immunogold technique of visualising protein localisation in cells, could be employed to determine whether the mutant protein is being targeted to the correct location within the cell. The experiments just outlined could help to elucidate the role of NodC in the plant-microbe dialogue.

One area where further experimentation may provide a wealth of information concerning *nod* gene regulation would involve *nod* box manipulation. Possible approaches include studying the effects of point mutations, small deletions and progressive *Bal* 31 digestion of the *nod* box on gene expression. It should be noted that, in one orientation the *nod* box appears to regulate constitutive expression (i.e. *nod*D) while, in the other, genes like *nod*ABCIJ, are induced. Such experimentation may elucidate the function of the highly conserved regions within the *nod* box. Regulatory elements similar to the *nod* box have also been found in other soil micro-organisms, for example *Agrobacterium* (D. White, D.S.I.R., personal communication) and *Pseudomonas* (Schell and Sukordhaman, 1989). Therefore a common theme may exist and it may be worthwhile keeping this, and any research in those areas, in mind while studying *nod* box structure and function. It could be expected that the results of the investigations outlined above, together with that gleaned from the other research in the area, would directly extend the exiting knowledge and understanding of biological nitrogen fixation and plant-microbe communication.

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.

ALBERSHEIM, P. and DARVILL, A.G. (1985). Oligosaccharins. Scientific American. September p. 44-50.

APPLEBAUM, E.R., THOMPSON, D.V., IDLER, K. and CHARTRAIN, N. (1988). *Rhizobium japonicum* USA 191 has two *nod*D genes that differ in primary structure and function. J. Bacteriol. 170: 12-20.

APPLEBY, C.A. (1974). Leghemoglobin. In: The biology of nitrogen fixation. (editor A. Quispel). North Holland/American Elsevier Publishing Company: Amsterdam. p. 521-554.

AUERSWALD, E-A., LUDWIG, G. and SCHALLER, H. (1980). Structural analysis of Tn*5*. Cold Spring Harbor Symp. Quant. Biol. 45: 107-113.

AUSUBEL, F.M., BUIKEMA, W.J., EARL, C.D., KLINGENSMITH, J.A., NIXON, B.T. and SZETO, W.W. (1985). Organization and regulation of *Rhizobium meliloti* and *Parasponia Bradyrhizobium* nitrogen fixation genes. In: Nitrogen fixation research progress. (editors H.J. Evans, P.J. Bottomley and W.E. Newton). Martinus Nijhoff Publishers: Dordrecht, Netherlands. p. 165-171.

BANFALVI, Z. and KONDOROSI, A. (1989). Production of root hair deformation factors by *Rhizobium meliloti* nodulation genes in *Escherichia coli*: *Hsn*D (*Nod*H) is involved in the plant host-specific modification of the NodABC factor. Plant Mol. Biol. 13: 1-12.

BARONDES, S.H. (1981). Lectins: Their multiple endogenous cellular functions. Ann. Rev. Biochem. 50: 207-231.

BAUER, W.D. (1981). Infection of legumes by rhizobia. Ann. Rev. Plant Physiol. 32: 407-449.

BECKING, J.H. (1975). Root nodules in non-legumes. In: The development and function of roots. (editors J.G. Torrey and D.T. Clarkson). Academic Press: London. p. 508-566.

BENTON, W.D. and DAVIS, R.W. (1977). Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. Science 196: 180-182.

BERG, D.E., JOHNSRUD, L., MCDIVITT, L., RAMABHADRAN, R. and HIRSCHEL, B.J. (1982). Inverted repeats of Tn*5* are transposable elements. Proc. Natl. Acad. Sci. USA. 79: 2632-2635.

BERG, D.E., WEISS, A. and CROSSLAND, L. (1980). Polarity of Tn5 insertion mutations in *Escherichia coli*. J. Bacteriol. 142: 439-446.

BERGERSEN, F.J. (1974). Formation and function of bacteroids. In: The biology of nitrogen fixation. (editor A. Quispel). North Holland/ American Elsevier Publishing Company: Amsterdam. p. 473-498.

BERGERSEN, F.J. (1980). Leghaemoglobin, oxygen supply and nitrogen fixation. In: Nitrogen fixation. (editors W.D.P. Stewart and J.R. Gallon). Academic Press: London. p. 139-160.

BERGERSEN, F.J. and GOODCHILD, D.J. (1973). Aeration pathways in soybean root nodules. Aust. J. Biol. Sci. 26: 729-740.

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

BHUVANESWARI, T.V., PUEPPKE, S.G. and BAUER, W.D. (1977). Role of lectins in plant - microorganism interactions: I. Binding of soybean lectin to rhizobia. Plant Physiol. 60: 486-491.

BHUVANESWARI, T.V., TURGEON, B.G. and BAUER, W.D. (1980). Early events in the infection of soybean (*Glycine max* L. Merr) by *Rhizobium japonicum*: I. Localization of infectible root cells. Plant Physiol. 66: 1027-1031.

BILOFSKY, H.S., BURKS, C., FICKETT, J.W., GOAD, W.B., LEWITTER, F.I., RINDONE, W.P., SWINDELL, C.D. and TUNG, C-S. (1986). The GenBank genetic sequence databank. Nuc. Acids Res. 14: 1-4.

BISSELING, T., VAN DEN BOS, R.C. and VAN KAMMEN, A. (1986). Host-specific gene expression in legume root nodules. In: Nitrogen fixation Vol. 4: Molecular biology. (editors W.J. Broughton and A. Puhler). Clarendon Press: Oxford. p. 280-312.

BOHLOOL, B.B. and SCHMIDT, E.L. (1974). Lectins: A possible basis for specificity in the *Rhizobium* - legume root nodule symbiosis. Science. 185: 269-271.

BOISTARD, P., BATUT, J., DAVERAN, M-L., DAVID, M., GARNERONE, A-M., LI, R.Y. and KAHN, D. (1988). The regulatory pathway of *nif* and *fix* genes in *Rhizobium meliloti*. In: Molecular genetics of plant - microbe interactions 1988. (editors R. Palacios and D.P.S. Verma). APS Press: Minnesota, USA. p. 96-100.

BOLIVAR, F., RODRIGUES, R., GREENE, P.J., BETLACH, M., HENNECKE, H.L., BOYER, H.W., CROSA, J. and FALKOW, S. (1977). Construction and characterisation of new cloning vehicles: II. A multipurpose cloning system. Gene. 2: 95-113.

BOWES, B., CALLAHAM, D. and TORREY, J.G. (1977). Time-lapse photographic observations of morphogenesis in root nodules of *Comptonia peregrina* (Myricaceae). Amer. J. Bot. 64: 516-525.

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.

BRILL, W.J. (1979). Regulation of nitrogen fixation. In: A treatise on dinitrogen fixation. Sections I and II: Inorganic and physical chemistry and biochemistry. (editor R.W.F. Hardy, F. Bottomley and R.C. Burns). John Wiley and Sons: New York. p. 765-798.

BRINK, B.A., TUROWSKI, D. and NOEL, K.D. (1988). Genetics and symbiotic role of lipopolysaccharide in two strains of *Rhizobium leguminosarum* having different host ranges. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 470.

BUCHANAN-WOLLASTON, A.V., BERINGER, J.E., BREWIN, N.J., HIRSCH, P.R. and JOHNSTON, A.W.B. (1980). Isolation of symbiotically defective mutants in *Rhizobium leguminosarum* by insertion of the transposon Tn5 into a transmissible plasmid. Mol. Gen. Genet. 178: 185-190.

BURN, J.E., HAMILTON, W.D., WOOTTON, J.C. and JOHNSTON, A.W.B. (1989). Single and multiple mutations affecting properties of the regulatory gene *nodD* of *Rhizobium*. Mol. Microbiol. 3: 1567-1577.

BURNS, R.C. (1979). Mechanism of dinitrogen reaction. In: A treatise on dinitrogen fixation. Sections I and II: Inorganic and physical chemistry and biochemistry. (editor R.W.F. Hardy, F. Bottomley and R.C. Burns). John Wiley and Sons: New York. p. 491-514.

BURRIS, R.H., ARP, D.J., HAGEMAN, R.V., HOUCHINS, J.P., SWEET, W.J. and TSO, M-Y. (1981). Mechanism of nitrogenase action. In: Current perspectives in nitrogen fixation. (editors A.H. Gibson and W.E. Newton). Aust. Acad. Sci: Canberra. p. 56-66.

CALLAHAM, D., NEWCOMB, W., TORREY, J.G. and PETERSON, R.L. (1979). Root hair infection in Actinomycete-induced root nodule initiation in *Casuarina*, *Myrica*, and *Comptonia*. Bot. Gaz. 140 (suppl.): S1-S9.

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.

CANTER CREMERS, H.C.J., SPAINK, H.P., WIJFJES, A.H.M., PEES, E., WIJFFELMAN, C.A., OKKER, R.J.H. and LUGTENBERG, B.J.J. (1989). Additional nodulation genes on the Sym plasmid of *Rhizobium leguminosarum* biovar *viciae*. Plant Mol. Biol. 13: 163-174.

CARLSON, R.W., KALEMBASA, S. and GARCIA, F. (1988). The expression of *R. phaseoli* CE3 lipopolysaccharide (LPS) O-antigen in *R. leguminosarum* 128C53. In: Nitrogen fixation: hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 471.

CARROLL, B.J. and GRESSHOFF, P.M. (1983). Nitrate inhibition of nodulation and nitrogen fixation in white clover. Z. Pflanzenphysiol. 110: 77-88.

CHANDLER, M.R. (1978). Some observations on infection of *Arachis hypogaea* L. J. Exp. Bot. 29: 749-755.

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.

COHEN, S.N., CHANG, A.C.Y. and HSU, L. (1972). Non-chromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. USA. 69: 2110-2114.

COLLINS, J.M. (1983). Nodule initiation in white clover. Honours thesis, Australian National University: Canberra, Australia.

CORBIN, D., BARRAN, L. and DITTA, G. (1983). Organization and expression of *Rhizobium meliloti* nitrogen fixation genes. Proc. Natl. Acad. Sci. USA. 80: 3005-3009.

CRAIG, A.S. and WILLIAMSON, K.I. (1972). Three inclusions of rhizobial bacteroids and their cytochemical character. Arch. Mikrobiol. 87: 165-171.

CROW, V.L., JARVIS, B.D.W. and GREENWOOD, R.M. (1981). Deoxyribonucleic acid homologies among acid-producing strains of *Rhizobium*. Int. J. Syst. Bact. 31: 152-172.

DART, P.J. (1974). The infection process. In: The biology of nitrogen fixation. (editor A. Quispel). North Holland/American Elsevier Publishing Company: Amsterdam. p. 406-419.

DART, P.J. (1975). Legume root nodule initiation and development. In: The development and fuction of roots. (editors J.G. Torrey and D.T. Clarkson). Academic Press: London. p. 468-499.

DART, P.J. (1977). Infection and development of leguminous nodules. In: A treatise on dinitrogen fixation. Section III: Biology (editors R.W.F. Hardy and W.S. Silver). John Wiley and Sons: New York. p. 367-472.

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

DAZZO, F.B. (1980). Determinants of host specificity in the *Rhizobium*-clover symbiosis. In: Nitrogen fixation. Vol 2.: Symbiotic associations and cyanobacteria. (editors W.E. Newton and W.H. Orme-Johnson). University Park Press: Baltimore. p. 165-187.

DAZZO, F.B. and BRILL, W.J. (1978). Regulation by fixed nitrogen of host-symbiont recognition in the *Rhizobium*-clover symbiosis. Plant Physiol. 62: 18-21.

DAZZO, F.B. and GARDIOL, A.E. (1984). Host-specificity in *Rhizobium*-legume interactions. In: Genes involved in microbe - plant interactions. (editors D.P.S. Verma and T.H. Hohn). Springer-Verlag: New York. p. 3-25.

DAZZO, F.B. and HUBBELL, D.H. (1975a). Antigenic differences between infective and noninfective strains of *Rhizobium trifolii*. Appl. Microbiol. 30: 172-177.

DAZZO, F.B. and HUBBELL, D.H. (1975b). Cross-reactive antigens and lectin as determinants of symbiotic specificity in the *Rhizobium*-clover association. Appl. Microbiol. 30: 1017-1033.

DAZZO, F.B., NAPOLI, C.A. and HUBBELL, D.H. (1976). Adsorption of bacteria to roots as related to host specificity in the *Rhizobium* - clover symbiosis. Appl. Environ. Microbiol. 32: 166-171.

DAZZO, F.B. and TRUCHET, G.L. (1984). Attachment of nitrogen-fixing bacteria to roots of host plants. In: Current developments in biological nitrogen fixation. (editor N.S. Subba Rao). Edward Arnold Ltd.: London. p. 65-99.

DAZZO, F.B., TRUCHET, G.L. and KIJNE, J.W. (1982). Lectin involvement in root-hair tip adhesion as related to the *Rhizobium* - clover symbiosis. Physiol. Plant. 56: 143-147.

DAZZO, F.B., URBANO, M.R. and BRILL, W.J. (1979). Transient appearance of lectin receptors on *Rhizobium trifolii*. Curr. Microbiol. 2: 15-20.

DEBELLE, F., ROSENBERG, C., VASSE, J., MAILLET, F., MARTINEZ, E., DENARIE, J. and TRUCHET, G. (1986). Assignment of symbiotic developmental phenotypes to common and specific nodulation (*nod*) genetic loci of *Rhizobium meliloti*. J. Bacteriol. 168: 1075-1086.

DEBELLE, F. and SHARMA, S.B. (1986). Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host-specificity of nodulation. Nuc. 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., WIJFJES, A.H.M., SPAINK, H.P., RUIZ-SAINZ, J.E., WIJFFELMAN, C.A., OKKER, R.J.H. and LUGTENBERG, B.J.J. (1989). *NodO*, a new *nod* gene of the *Rhizobium leguminosarum* biovar *viciae* Sym plasmid pRL1JI, encodes a secreted protein. J. Bacteriol. 171: 6764-6770.

DENARIE, J., ROSENBERG, C., BOISTARD, P., TRUCHET, G. and CASSE-DELBART, F. (1981). Plasmid control of symbiotic properties in *Rhizobium meliloti*. In: Current perspectives in nitrogen fixation. (editors A.H. Gibson and W.E. Newton). Aust. Acad. Sci: Canberra. p. 137-141.

DEVERALL, B.J. (1982). Introduction. In: Phytoalexins. (editors J.A. Bailey and J.W. Mansfield). Blackie and Son Ltd.: Glasgow. p. 9.

DEVEREUX, J., HAEBERLI, P. and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nuc. Acids Res. 12: 387-395.

DILWORTH, M.J. and APPLEBY, C.A. (1979). Leghemoglobin and *Rhizobium* hemoproteins. In: A treatise on dinitrogen fixation. Sections I and II: Inorganic and physical chemistry and biochemistry. (editor R.W.F. Hardy, F. Bottomley and R.C. Burns). John Wiley and Sons: New York. p. 691-764.

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. USA. 77: 7347-7351.

DOWNIE, J.A. (1989). The *nodL* gene from *Rhizobium leguminosarum* is homologous to the acetyl transferases encoded by *lacA* and *cysE*. Mol. Microbiol. 3: 1649-1651.

DUDLEY, M.E., JACOBS, T.W. and LONG, S.R. (1987). Microscopic studies of cell divisions induced in alfalfa roots by *Rhizobium meliloti*. Planta. 171: 289-301.

EADY, R.R. and SMITH, B.E. (1979). Physico-chemical properties of nitrogenase and its components. In: A treatise on dinitrogen fixation. Sections I and II: Inorganic and physical chemistry and biochemistry. (editor R.W.F. Hardy, F. Bottomley and R.C. Burns). John Wiley and Sons: New York. p. 399-490.

ECONOMOU, A., HAMILTON, W.D.O., JOHNSTON, A.W.D. and DOWNIE, J.A. (1990). Exported proteins encoded by *nod* genes from *Rhizobium leguminosarum* biovar *viciae*. In: 8th International congress on nitrogen fixation. May 20-26, 1990/Knoxville, Tennessee, USA. E-18.

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.

ELKAN, G.H. (1981). The taxonomy of the Rhizobiaceae. Int. Rev. Cytol. Suppl. 13: p. 1-14.

EVANS, H.J., PUROHIT, K., CANTRELL, M.A., EISBRENNER, G., RUSSELL, S.A., HANUS, F.J., and LEPO, J.E. (1981). Hydrogen losses and hydrogenases in nitrogen-fixing organisms. In: Current perspectives in nitrogen fixation. (editors A.H. Gibson and W.E. Newton). Aust. Acad. Sci: Canberra. p. 84-96.

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.

FAUCHER, C., MAILLET, F., VASSE, J., ROSENBERG, C., VAN BRUSSEL, A.A.N., TRUCHET, G. and DENARIE, J. (1988). *Rhizobium meliloti* host range *nod*H gene determines production of an alfalfa-specific extracellular signal. J. Bacteriol. 170: 5489-5499.

FIRMIN, J.L., WILSON, K.E., ROSSEN, L. and JOHNSTON, A.W.B. (1986). Flavonoid activation of nodulation genes in *Rhizobium* reversed by other compounds present in plants. Nature. 324: 90-92.

FISHER, R.F., BRIERLEY, H.L., MULLIGAN, J.T. and LONG, S.R. (1987a). Transcription of *Rhizobium meliloti* nodulation genes: Identification of a *nod*D transcription initiation site *in vitro* and *in vivo*. J. Biol. Chem. 262: 6849-6855.

FISHER, R.F., EGELHOFF, T.T., MULLIGAN, J.T. and LONG, S.R. (1988). Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. Genes Dev. 2: 282-293.

FISHER, R.F., SWANSON, J.A., MULLIGAN, J.T. and LONG, S.R. (1987b). 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*. Appl. Env. Micro. 49: 1432-1435.

FLORES, M., GONZALEZ, V., BROM, S., MARTINEZ, E., PINERO, D., ROMERO, D., DAVILA, G. and PALACIOS, R. (1987). Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. J. Bacteriol. 169: 5782-5788.

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.

GLOUDEMANS, T., MOERMAN, M., VAN BECKUM, J., GUNDERSEN, J., VAN KAMMEN, A. and BISSELING, T. (1988). Identification of plant genes involved in the *Rhizobium leguminosarum* - pea root hair interaction. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 611-616.

GOTTFERT, M., LAMB, J.W., GASSER, R., SEMENZA, J. and HENNECKE, H. (1989). Mutational analysis of the *Bradyrhizobium japonicum* common *nod* genes and further *nod* box-linked genomic DNA regions. Mol. Gen. Genet. 215: 407-415.

GOTZ, R., EVANS, I.J., DOWNIE, J.A. and JOHNSTON, A.W.B. (1985). Identification of the host-range DNA which allows *Rhizobium leguminosarum* strain TOM to nodulate cv. Afghanistan peas. Mol. Gen. Genet. 201: 296-300.

GRAY, J.X. and Rolfe, B.G. (1990). Exopolysaccharide production in *Rhizobium* and its role in invasion. Mol. Microbiol. 4: 1425-1431.

GRESSHOFF, P.M. and DOY, C.H. (1974). Derivation of a haploid cell line from *Vitis vinifera* and the importance of the stage of meiotic development of anthers for haploid culture of this and other genera. Z. Pflanzenphysiol. 73: 132-141.

GRESSHOFF, P.M. and MOHAPATRA, S.S. (1981). Legume cell and tissue culture. In: Tissue culture of economically important crop plants. (editor A.N. Rao). Singapore University, Singapore Press: Singapore. p. 11-24

GULASH, M., AMES, P., LAROSILIERE, R.C. and BERGMAN, K. (1984). Rhizobia are attracted to localised sites on legume roots. Appl. Env. Micro. 48: 149-152.

GYORGYPAL, Z., IYER, N. and KONDOROSI, A. (1988). Three regulatory *nod*D alleles of diverged flavonoid-specificity are involved in host-dependent nodulation by *Rhizobium meliloti*. Mol. Gen. Genet. 212: 85-92.

HALVERSON, L.J. and STACEY, G. (1986). Signal exchange in plant-microbe interactions. Microbiol. Reviews. 50: 193-225.

HENIKOFF, S., HAUGHN, G.W., CALVO, J.M. and WALLACE, J.C. (1988). A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA. 85: 6602-6606.

HENNECKE, H., FISCHER, H-M., GUBLER, M., THONY, B., ANTHAMATTEN, D., KULLIK, I., EBELING, S., FRITSCHE, S. and ZURCHER, T. (1988a). Regulation of *nif* and *fix* genes in *Bradyrhizobium japonicum* occurs by a cascade of two consecutive gene activation steps of which the second one is oxygen sensitive. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 339-344.

HENNECKE, H., MEYER, L., GOTTFERT, M. and FISCHER, H-M. (1988b). Genetics of the *Bradyrhizobium japonicum* - soybean symbiosis: Recent developments on genes for nodulation, bacteroid respiration, and regulation of nitrogen fixation. In: Molecular genetics of plant - microbe interactions 1988. (editors R. Palacios and D.P.S. Verma). APS Press: Minnesota, USA. p. 118-123.

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

HONG, G-F., BURN, J.E. and JOHNSTON, A.W.B. (1987). Evidence that DNA involved in the expression of nodulation (*nod*) genes in *Rhizobium* binds to the product of the regulatory gene *nod*D. Nuc. Acids Res. 15: 9677-9690.

HONMA, M.A. and AUSUBEL, F.M. (1987). *Rhizobium meliloti* has three functional copies of the *nod*D symbiotic regulatory genes. Proc. Natl. Acad. Sci. USA. 84: 8558-8562.

HOOYKAAS, P.J.J., VAN BRUSSEL, A.A.N., DEN DULK-RAS, H., VAN SLOGTEREN, G.M.S. and SCHILPEROORT, R.A. (1981). Sym plasmid of *Rhizobium trifolii* expressed in different rhizobial species and *Agrobacterium tumefaciens*. Nature. 291: 351-353.

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 product. EMBO J. 6: 841-848.

INNES, R.W., KUEMPEL, P.L., PLAZINSKI, J., CANTER-CREMERS, H., ROLFE, B.G. and DJORDJEVIC, M.A. (1985). Plant factors induce expression of nodulation and host-range genes in *Rhizobium trifolli*. Mol. Gen. Genet. 201: 426-432.

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. Bact. 32: 378-380.

JOHN, M., SCHMIDT, J., WIENEKE, U., KRUSSMANN, H-D. and SCHELL, J. (1988). Transmembrane orientation and receptor-like structure of the *Rhizobium meliloti* common nodulation protein NodC. EMBO J. 7: 583-588.

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

KAHN, D., DAVID, M., BATUT, J., DAVERAN, M-L., GARNERONE, A-M., HERTIG, C., PAQUES, F., RUO YA, L. and BOISTARD, P. (1988). Cascade activation of *nif* genes in *Rhizobium meliloti*. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 357-361.

KHAN, M.M.A., BAUER, W.D. and GRAHAM, T.L. (1990). Isoflavones and their glucosides from host plants that induce nodulation gene expression in rhizobia are also potent and specific chemoattractants. In: 8th International congress on nitrogen fixation. May 20-26, 1990/Knoxville, Tennessee, USA. B-12.

KNIGHT, C.D., ROSSEN, L., ROBERTSON, J.G., WELLS, B. and DOWNIE, J.A. (1986). Nodulation inhibition by *Rhizobium leguminosarum* multicopy *nod*ABC genes and analysis of early stages of plant infection. J. Bacteriol. 166: 552-558.

KONDOROSI, A. (1986). Molecular biology of symbiotic nitrogen fixation by *Rhizobium meliloti*. In: Nitrogen fixation Vol. 4: Molecular biology. (editors W.J. Broughton and A. Puhler). Clarendon Press: Oxford. p. 245-279.

KONDOROSI, A., KONDOROSI, E., GYORGYPAL, Z., BANFALVI, Z., GYURIS, J., PUTNOKY, P., GROSSKOPF, E., JOHN, M., SCHMIDT, J., CAM HA, D.T., LADOS, M., SLASKA-KISS, K. and SCHELL, J. (1988). *Rhizobium meliloti nod* and *fix* genes controlling the initiation and development of root nodules. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 399-403.

KONDOROSI, E., GYURIS, J., SCHMIDT, J., JOHN, M., DUDA, E., HOFFMANN, B., SCHELL, J. and KONDOROSI, A. (1989). Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. EMBO J. 8: 1331-1340.

LALONDE, M. (1980). Techniques and observations of the nitrogen fixing alnus root nodule symbiosis. In: Recent advances in biological nitrogen fixation. (editor N.S. Subba Rao). Edward Arnold: London. p. 421-434.

LARA, M., ORTEGA, J.L., OLGUIN, B.E. and SANCHEZ, F. (1988). Nodulin expression and nitrogen metabolism in *Phaseolus vulgaris* root nodules. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 617-622.

LAW, I.J. and STRIJDOM, B.W. (1977). Some observations on plant lectins and *Rhizobium* specificity. Soil Biol. Biochem. 9: 79-84.

LEGOCKI, R.P. and SZALAY, A.A. (1984). Molecular biology of stem nodules. In: Genes involved in microbe - plant interactions. (editors D.P.S. Verma and T.H. Hohn). Springer-Verlag: New York. p. 255-268.

LEONG, S.A., DITTA, G.S. and HELINSKI, D.R. (1982). Heme biosynthesis in *Rhizobium*. Identification of a cloned gene coding for δ-amino-levulinic acid synthase from *Rhizobium meliloti*. J. Biol. Chem. 257: 8724-8730.

LEROUGE, P., ROCHE, P., FAUCHER, C., MAILLET, F., TRUCHET, G., PROME, J.C. and DENARIE, J. (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature. 344: 781-784.

LEWIN, B. (1983). Genes. John Wiley and Sons: New York. p. 180, 538-540.

LI, D. and HUBBELL, D.H. (1969). Infection thread formation as a basis of nodulation specificity in *Rhizobium* - strawberry clover associations. Can. J. Microbiol. 15: 1133-1136.

LIBBENGA, K.R. and BOGERS, R.J. (1974). Root-nodule morphogenesis. In: The biology of nitrogen fixation. (editor A. Quispel). North Holland/ American Elsevier Publishing Company: Amsterdam p. 431-467.

LIBBENGA, K.R. and HARKES, P.A.A. (1973). Initial proliferation of cortical cells in the formation of root nodules in *Pisum sativum* L. Planta. 114: 17-28.

LIE, T.A. (1964). Nodulation of leguminous plants as affected by root secretions and red light. H. Veenman En Zonen N.V: Wageningen. p. 2.

LITTLE, S., HYDE, S., CAMPBELL, C.J., LILLEY, R.J. and ROBINSON, M.K. (1989). Translational coupling in the threonine operon of *Escherichia coli* K-12. J. Bacteriol. 171: 3518-3522.

LONG, S.R. (1989). *Rhizobium* - legume nodulation: Life together in the underground. Cell. 56: 203-214.

LONG, S.R. and EHRHARDT, D.W. (1989). New route to a sticky subject. Nature. 338: 545-546.

LUDWIG, R.A. and DE VRIES, G.E. (1986). Biochemical physiology of *Rhizobium* dinitrogen fixation. In: Nitrogen fixation Vol. 4: Molecular biology. (editors W.J. Broughton and A. Puhler). Clarendon Press: Oxford. p. 50-69.

LUGTENBERG, B.J.J, VAN BRUSSEL, A.A.N., OKKER, R.J.H., RECOURT, K., SCHLAMAN, H.R.M., SPAINK, H.P., WIJFFELMAN, C.A. and ZAAT, S.A.J. (1988). Mechanism of activation of *Rhizobium nodD* gene. In: Molecular genetics of plant - microbe interactions 1988. (editors R. Palacios and D.P.S. Verma). APS Press: Minnesota, USA. p. 79-83.

MACGREGOR, A.N. and ALEXANDER, M. (1972). Comparison of nodulating and non-nodulating strains of *Rhizobium trifolii*. Plant And Soil. 36: 129-139.

MAIER, R.J. and BRILL, W.J. (1976). Ineffective and non-nodulating mutant strains of *Rhizobium japonicum*. J. Bacteriol. 127: 763-769.

MANIATIS, T, FRITSCH, E.F. and SAMBROOK, J. (1982). Molecular cloning: A laboratory manual. Cold Spring Harbor. Cold Spring Harbor Laboratory Press: New York.

MARTINEZ, E., ROMERO, D. and PALACIOS, R. (1990). The *Rhizobium* genome. Critical Reviews In Plant Sciences. 9: 59-93.

MARX, J.L. (1985). How rhizobia and legumes get it together: Formation of nitrogen-fixing nodules on legume roots requires an intricate interplay between the plant and the nodulating rhizobial bacterium. Science 230: 157-158.

MAURO, V.P., NGUYEN, T., KATINAKIS, P. and VERMA, D.P.S. (1985). Primary structure of the soybean nodulin-23 gene and potential regulatory elements in the 5'-flanking regions of nodulin and leghemoglobin genes. Nuc. Acids Res. 13: 239-249.

MELTON, D.A., KRIEG, P.A., REBAGLIATI, M.R., MANIATIS, T., ZINN, K. and GREEN, M.R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nuc. Acids Res. 12: 7035-7056.

MIKIKO, A., 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.

MILLER, J.H. (1972). Experiments in molecular genetics. Cold Spring Harbor. Cold Spring Harbor Laboratory Press: New York. p. 4-9.

MILLER, R.C. and BOWLES, D.J. (1982). A comparative study of the localization of wheat-germ agglutinin and its potential receptors in wheat grains. Biochem. J. 206: 571-576.

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

NAGL, W. and RUCKER, W. (1974). Shift of DNA replication from diploid to polyploid cells in cytokinin-controlled differentiation. Cytobios. 10: 137-144.

NAPOLI, C.A. and HUBBELL, D.H. (1975). Ultrastructure of *Rhizobium*-induced infection threads in clover root hairs. Appl. Microbiol. 30: 1003-1009.

NEEDLEMAN, 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. (1976). A correlated light and electron microscopic study of symbiotic growth and differentiation in *Pisum sativum* root nodules. Can. J. Bot. 54: 2163-2186.

NEWCOMB, W. (1980). Control of morphogenesis and differentiation of pea root nodules. In: Nitrogen fixation Vol 2. Symbiotic associations and cyanobacteria. (editors W.E. Newton and W.H. Orme-Johnson). University Park Press: Baltimore. p.87-101.

NEWCOMB, W. and PETERSON, R.L. (1979). The occurrence and ontogeny of transfer cells associated with lateral roots and root nodules in Leguminosae. Can. J. Bot. 57: 2583-2602.

NEWCOMB, W., PETERSON, R.L., CALLAHAM, D. and TORREY, J.G. (1978). Structure and host-actinomycete interactions in developing root nodules of *Comptonia peregrina*. Can. J. Bot. 56: 502-531.

NEWCOMB, W., SIPPELL, D. and PETERSON, R.L. (1979). The early morphogenesis of *Glycine max* and *Pisum sativum* root nodules. Can. J. Bot. 57: 2603-2616.

NIEUWKOOP, A.J., BANFALVI, Z., DESHMANE, 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.

NORMARK, S., BERGSTROM, S., EDLUND, T., GRUNDSTROM, T., JAURIN, B., LINDBERG, F.P. and OLSSON, O. (1983). Overlapping genes. Ann. Rev. Genet. 17: 499-525.

NUTMAN, P.S. (1948). Physiological studies on nodule formation: 1. The relation between nodulation and lateral root formation in red clover. Ann. Bot. N.S. 12: 81-96.

NUTMAN, P.S. (1959). Some observations on root-hair infection by nodule bacteria. J. Exp. Bot. 10: 250-263.

NUTMAN, P.S. (1981). Hereditary host factors affecting nodulation and nitrogen fixation. In: Current perspectives in nitrogen fixation. (editors A.H. Gibson and W.E. Newton). Aust. Acad. Sci.: Canberra. p. 194-204.

O'HARA, M.J. (1989). Analysis of genomic rearrangement and plasmid conjugation of an inoculant strain of *Rhizobium leguminosarum* bv. *trifolii*. Ph. D. thesis, Massey University, New Zealand.

OLD, R.W. and PRIMROSE, S.B. (1981). Principles of gene manipulation: An introduction to genetic engineering. Blackwell Scientific Publications: London. p. 82.

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.

PARNISKE, M., AHLBORN, B. and WERNER, D. (1990). Glyceollin resistance of *Bradyrhizobium japonicum* is specifically induced by isoflavonoids from *Glycine max*. In: 8th International congress on nitrogen fixation. May 20-26, 1990/Knoxville, Tennessee, USA. B-16.

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

PHILIP-HOLLINGSWORTH, S., HOLLINGSWORTH, R.I. and DAZZO, F.B. (1989). Host-range related structural features of the acidic extracellular polysaccharides of *Rhizobium trifolii* and *Rhizobium leguminosarum*. J. Biol. Chem. 264: 1461-1466.

PHILLIPS, D.A. (1971). A cotyledonary inhibitor of root nodulation in *Pisum sativum*. Physiol. Plant. 25: 482-487.

PHILLIPS, D.A and TORREY, J.G. (1970). Cytokinin production by *Rhizobium japonicum*. Physiologica Plantarum. 23: 1057-1063.

PHILLIPS, D.A and TORREY, J.G. (1972). Studies on cytokinin production by *Rhizobium*. Plant Physiol. 49: 11-15.

POSTGATE, J.R. (1982). The fundamentals of nitrogen fixation. Cambridge University Press: Cambridge, UK. p. 103-130.

PTASHNE, M. (1986). A genetic switch: Gene control and phage lambda. Blackwell Scientific Publications and Cell Press: Australia. p. 29.

PUEPPKE, S.G. (1983). *Rhizobium* infection threads in root hairs of *Glycine max* (L.) Merr., *Glycine soja* Sieb. and Zucc., and *Vigna unguiculata* (L.) Walp. Can. J. Microbiol. 29: 69-76.

PURCHASE, H.F. (1958). Restriction of infection threads in nodulation of clover and lucerne. Aust. J. Biol. Sci. 11: 155-161.

RANGA RAO, V. and KEISTER, D.L. (1978). Infection threads in the root hairs of soybean (*Glycine max*) plants inoculated with *Rhizobium japonicum*. Protoplasma. 97: 311-316.

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.

RIDGE, R.W. and ROLFE, B.G. (1985). *Rhizobium* sp. degradation of legume root hair cell wall at the site of infection thread origin. Appl. Env. Micro. 50: 717-720.

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.

ROLFE, B.G. and GRESSHOFF, P.M. (1988). Genetic analysis of legume nodule initiation. Ann. Rev. Plant Physiol. Plant Mol. Biol. 39: 297-319.

ROLFE, B.G. and SHINE, J. (1984). *Rhizobium*-Leguminosae symbiosis: The bacterial point of view. In: Genes involved in microbe - plant interactions. (editors D.P.S. Verma and T. Hohn). Springer-Verlag: New York. p. 95-128.

ROSSEN, L., JOHNSTON, A.W.B. and DOWNIE, J.A. (1984). DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nod*AB and C required for root hair curling. Nuc. 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 *nod*ABC 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.

SADOWSKY, M.J. and BOHLOOL, B.B. (1983). Possible involvement of a megaplasmid in nodulation of soybeans by fast-growing rhizobia from China. Appl. Env. Micro. 46: 906-911.

SADOWSKY, M.J., BOHLOOL, B.B. and KEYSER, H.H. (1987). Serological relatedness of *Rhizobium fredii* to other rhizobia and to bradyrhizobia. Appl. Env. Micro. 53: 1785-1789.

SANGER, F., COULSON, A.R., HONG, G.F., HILL, D.F. and PETERSEN, G.B. (1982). Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. 162: 729-773.

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

SCHELL, J., JOHN, M., SCHMIDT, J., WINGENDER-DRISSEN, R., SIMONS, A., METZ, B., OSTERGAARD JENSEN, E., HOFFMAN, H-J., WELTERS, P. and DE BRUIJN, F.J. (1988). Regulation of gene expression in plants with special emphasis on the nodulation process. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 591-598.

SCHELL, M.A. and SUKORDHAMAN, M. (1989). Evidence that the transcription activator encoded by the *Pseudomonas putida nah*R gene is evolutionarily related to the transcription activators encoded by the *Rhizobium nod*D genes. J. Bacteriol. 171: 1952-1959.

SCHERES, B., VAN DE WIEL, C., ZALENSKY, A., HORVATH, B., SPAINK, H., VAN ECK, H., ZWARTKRUIS, F., WOLTERS, A-M., GLOUDEMANS, T., VAN KAMMEN, A. and BISSELING, T. (1990). The ENOD12 gene product is involved in the infection process during the pea-*Rhizobium* interaction. Cell. 60: 281-294.

SCHMIDT, J., JOHN, M., WIENEKE, U., KRUSSMANN, H-D. and SCHELL, J. (1986). Expression of the nodulation gene *nod*A in *Rhizobium meliloti* and localization of the gene product in the cytosol. Proc. Natl. Acad. Sci. USA. 83: 9581-9585.

SCHOFIELD, P.R., DJORDJEVIC, M.A., ROLFE, B.G., SHINE, J. and WATSON, J.M. (1983). A molecular linkage map of nitrogenase and nodulation genes in *Rhizobium trifolii*. Mol. Gen. Genet. 192: 459-465.

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. Nuc. Acids Res. 14: 2891-2903.

SCHOLLA, M.H. and Elkan, G.H. (1984). *Rhizobium fredii* sp. nov., a fast-growing species that effectively nodulates soybeans. Int. J. Syst. Bact. 34: 484-486.

SCHWEDOCK, J. and LONG, S.R. (1990). ATP sulphurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. Nature. 348: 644-647.

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.

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

SCOTT, D.B., YOUNG, C.A., COLLINS-EMERSON, J.M., TERZAGHI, E.A., LEWIS, P.E. and PANKHURST, C.E. Mutational and structural analysis of *Rhizobium loti* nodulation genes. In preparation.

SCOTT, K.F. (1986). Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium* sp. (*Parasponia*). Nuc. Acids Res. 14: 2905-2919.

SHEARMAN, C.A., ROSSEN, L., JOHNSTON, A.W.B. and DOWNIE, J.A. (1986). The *Rhizobium leguminosarum* nodulation gene *nodF* encodes a polypeptide similar to acyl-carrier protein and is regulated by *nodD* plus a factor in pea root exudate. EMBO J. 5: 647-652.

SHERMAN, D.H., MALPARTIDA, F., BIBB, M.J., KIESER, H.M., BIBB, M.J. and HOPWOOD, D.A. (1989). Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tu22. EMBO J. 8: 2717-2725.

SMITH, T.F. and WATERMAN, M.S. (1981). Comparison of biosequences. Advances In Applied Mathematics 2: 482-489.

SOUTHERN, E. (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. (1987a). Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Plant Mol. Biol. 9: 27-39.

SPAINK, H.P., WIJFFELMAN, C.A., PEES, E., OKKER, R.J.H. and LUGTENBERG, B.J.J. (1987b). *Rhizobium* nodulation gene *nod*D as a determinant of host specificity. Nature. 328: 337-340.

SPRENT, J.I. (1979). The biology of nitrogen-fixing organisms. McGraw-Hill Book Company Ltd.: UK. p. 19-22.

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

STANLEY, J., LONGTIN, D., MADRZAK, C. and VERMA, D.P.S. (1986). Genetic locus in *Rhizobium japonicum* (*fredii*) affecting soybean root nodule differentiation. J. Bacteriol. 166: 628-634.

STOWERS, M.D. and EAGLESHAM, A.R.J. (1984). Physiological and symbiotic characteristics of fast-growing *Rhizobium japonicum*. Plant And Soil. 77: 3-14.

STRUHL, K. (1985). A rapid method for creating recombinant DNA molecules. Biotechniques 3: 452-453.

SURIN, B.P. and DOWNIE, J.A. (1988). Characterization of the *Rhizobium leguminosarum* genes *nod*LMN involved in efficient host-specific nodulation. Mol. Microbiol. 2: 173-183.

SUTHERLAND, I.W. (1985). Biosynthesis and composition of Gram-negative bacterial extracellular and wall polysaccharides. Ann. Rev. Micro. 39: 243-270.

SUTTON, W.D. (1983). Nodule development and senescence. In: Nitrogen fixation Vol. 3: Legumes. (editor W.J. Broughton). Clarendon Press: London. p. 144-212.

SWANSON, J.A., TU, J.K., OGAWA, J., SANGA, R., FISHER, R.F. and LONG, S.R. (1987). Extended region of nodulation genes in *Rhizobium meliloti* 1021 I. Phenotypes of Tn5 insertion mutants. Genetics. 117: 181-189.

TABOR, S. and RICHARDSON, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA. 84: 4767-4771.

TAYLOR, J.M., ILLMENSEE, R. and SUMMERS, J. (1976). Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. Biochem. Biophys. Acta. 442: 324-330.

THURING, R.W.J., SANDERS, J.P.M. and BORST, P. (1975). A freeze-squeeze method for recovering long DNA from agarose gels. Anal. Biochem. 66: 213-220.

TJEPKEMA, J.D. (1981). Aspects of oxygen regulation in legume, *Parasponia* and actinorhizal nodules. In: Current perspectives in nitrogen fixation. (editors A.H. Gibson and W.E. Newton). Aust. Acad. Sci.: Canberra. p. 268.

TOROK, I., KONDOROSI, E., STEPKOWSKI, T., POSFAI, J. and KONDOROSI, A. (1984). Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nuc. Acids Res. 12: 9509-9524.

TORREY, J.G. and CALLAHAM, D. (1978). Determinate development of nodule roots in actinomycete-induced root nodules of *Myrica gale*. Can. J. Bot. 56: 1357-1364.

TRAN THANH VAN, K., TOUBART, P., COUSSON, A., DARVILL, A.G., GOLLIN, D.J., CHELF, P. and ALBERSHEIM, P. (1985). Manipulation of the morphogenetic pathways of tobacco explants by oligosaccharins. Nature. 314: 615-617.

TRINICK, M.J. (1980). Relationships amongst the fast-growing rhizobia of *Lablab purpureus, Leucaena leucocephala, Mimosa* spp., *Acacia farnesiana* and *Sesbania grandiflora* and their affinities with other rhizobial groups. J. Appl. Bact. 49: 39-53.

TRINICK, M.J. (1982). Competition between rhizobial strains for nodulation. In: Nitrogen fixation in legumes. (editor J.M. Vincent). Academic Press: Sydney. p. 234.

TRUCHET, G. (1978). Sur l'etat diploide des cellules du meristeme des nodules radiculaires des legumineuses. Annales Des Sciences Naturelles, Botanique. 12 Serie. 19: 3-38.

TRUCHET, G., BARKER, D.G., CAMUT, S., DE BILLY, F., VASSE, J. and HUGUET, T. (1989). Alfalfa nodulation in the absence of *Rhizobium*. Mol. Gen. Genet. 219: 65-68.

TSIEN, H.C., DREYFUS, B.L. and SCHMIDT, E.L. (1983). Morphogenesis of stem nodules of *Sesbania rostrata*. In: Abstracts of the North American *Rhizobium* Conference. Ithaca: New York. p. 6.

TURGEON, B.G. and BAUER, W.D. (1982). Early events in the infection of soybean by *Rhizobium japonicum*. Time course and cytology of the initial infection process. Can. J. Bot. 60: 152-161.

VERMA, D.P.S. and NADLER, K. (1984). Legume-*Rhizobium* symbiosis: Host's point of view. In: Genes involved in microbe - plant interactions. (editors D.P.S. Verma and T. Hohn). Springer-Verlag: New York. p. 57-93.

VINCENT, J.M. (1970). The cultivation, isolation and maintenance of rhizobia. In: A mannual for the practical study of root-nodule bacteria. (IP3 Handbook no.15). Blackwell Scientific Publications: Oxford. p. 3.

VINCENT, J.M. (1977). *Rhizobium*: General microbiology. In: A treatise of dinitrogen fixation Section III: Biology. (editors R.W.F. Hardy and W.S. Silver). John Wiley and Sons: New York. p. 277-366.

VINCENT, J.M. (1980). Factors controlling the legume - *Rhizobium* symbiosis. In: Nitrogen fixation Vol. 2: Symbiotic associations and cyanobacteria. (editors W.E. Newton and W.H. Orme-Johnson). University Park Press: Baltimore. p. 103-123.

VINCENT, J.M. (1982). Nitrogen fixation in legumes. (editor J.M. Vincent). Academic Press: Sydney. p. 112.

WANG, S-P. and STACEY, G. (1990). A divergent *nod* box sequence is essential for *nodD*$_1$ induction in *B. japonicum*. In: 8th International congress on nitrogen fixation. May 20-26, 1990/Knoxville, Tennessee, USA. E-36.

WARREN WILSON, J. (1978). The position of regenerating cambia: Auxin/sucrose ratio and the gradient induction hypothesis. Proc. R. Soc. B. 203: 153-176.

WEINMAN, J.J., DJORDJEVIC, M.A., SARGENT, C.L., DAZZO, F.B. and ROLFE, B.G. (1988). A molecular analysis of the host range genes of *Rhizobium trifolii*. In: Molecular genetics of plant - microbe interactions. (editors R. Palacios and D.P.S. Verma). APS Press: Minnesota. p. 33-34.

WERNER, D., BASSARAB, S., HUMBECK, C., KAPE, R., KINNBACK, A., MELLOR, R.B., MORSCHEL, E., PARNISKE, M., PAUSCH, G., ROHM, M., SCHENK, S., THIERFELDER, H., THYNN, M., WETZEL, A. and WOLFF, A. (1988). Nodule proteins and compartments. In: Nitrogen fixation: Hundred years after. (editors H. Bothe, F.J. de Bruijn and W.E. Newton). Gustav Fischer: New York. p. 507-515.

WHITFELD, P.L., SEEBURG, P.H. and SHINE, J. (1982). The human pro-opiomalanocortin gene: Organisation, sequence, and interspersion with repetitive DNA. DNA 1: 113-143.

WIJFFELMAN, C.A., PEES, E., VAN BRUSSEL, A.A.N., OKKER, R.J.H. and LUGTENBERG, B.J.J. (1985). Genetic and functional analysis of the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Arch. Microbiol. 143: 225-232.

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

WINSOR, B.A.T. (1989). A nod at differentiation: The *nod*D gene product and initiation of *Rhizobium* nodulation. T.I.G. 5: 199-201.

WIPF, L. and COOPER, D.C. (1940). Somatic doubling of chromosomes and nodular infection in certain Leguminosae. Amer. J. Bot. 27: 821-824.

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.

YAO, P.Y. and VINCENT, J.M. (1976). Factors responsible for the curling and branching of clover root hairs by *Rhizobium*. Plant And Soil. 45: 1-16.

YOUNG, C., COLLINS-EMERSON, J.M., TERZAGHI, E.A. and SCOTT, D.B. (1990). Nucleotide sequence of *Rhizobium loti nod*I. Nuc. Acids Res. 18: 6691.

ZAAT, S.A.J., SCHRIPSEMA, J., WIJFFELMAN, C.A., VAN BRUSSEL, A.A.N. and LUGTENBERG, B.J.J. (1989). Analysis of the major inducers of the *Rhizobium nod*A promoter from *Vicia sativa* root exudate and their activity with different *nod*D genes. Plant Mol. Biol. 13: 175-188.

ZURKOWSKI, W. and LORKIEWICZ, Z. (1979). Plasmid-mediated control of nodulation in *Rhizobium trifolii*. Arch. Microbiol. 123: 195-201.