Construction and Utilisation of a Bidirectional Reporter Vector in the Analysis of Two *nod*-Boxes in *Rhizobium loti*

A thesis presented in partial fulfilment of the requirements for the degree of
Masters of Science
in Molecular Genetics at Massey University

Simon Keith Parry
1993

ABSTRACT

The *nod*-box is a 47bp *cis*-acting regulatory region which has been conserved amongst every species of *Rhizobium* studied to date. In species such as *R. meliloti* and *R. leguminosarum*, the *nod*-box has been shown to promote constitutive activity towards the regulatory *nod*D gene, and flavonoid-inducible expression towards the divergently-transcribed *nod*ABCIJ operon. This bidirectional regulation of the so-called common *nod* genes was not observed in *R. loti*. A previous analysis of this species had shown that its *nod*-box promoted inducible activity towards the truncated '*nod*D' gene, as well as the *nod*ACIJ operon. It was the unusual arrangement of these *R. loti nod* genes that had initially aroused interest in this bacteria.

To further investigate the role of the *nod*-box in the regulation of the *R. loti* common *nod* genes, a bidirectional reporter vector (pSPV4) was constructed. This novel vector allowed the promoter activity of a cloned *nod*-box-containing fragment to be concurrently measured in either direction using the same culture of cells. To achieve this construct, the *gus*A gene from pRAJ260 was blunt-end ligated into pUC21. An in-frame ribosome binding site (rbs) was cloned upstream of the *gus*A coding sequence to facilitate transcriptional fusions. The rbs and *gus*A gene were later excised as a functional unit and blunt-end ligated into pMP220 alongside the ß-galactosidase reporter gene but in the opposite orientation. Hence, both reporter genes could be divergently transcribed from a common regulatory region cloned into the multiple cloning site that separated the genes.

The fragments of DNA that were eventually cloned into the bidirectional vector were generated through the polymerase chain reaction. Each DNA insert contained the *nod*-box bracketed by differing lengths of flanking region. Once these PCR-generated fragments had been sequenced in pUC118 and subcloned into pSPV4, the resulting constructs were transformed into *R. loti* cells by electroporation. As the electroporation of these cells had not previously been reported, the conditions for this procedure were established and optimised.

The results obtained from the bidirectional reporter assays disagreed with those observed in the earlier assays by Teo (1990). Neither the *nod*ACIJ *nod*-box of NZP2037 nor the *nod*B *nod*-box of NZP2213, showed bidirectional inducible expression. In fact, both *nod*-boxes showed constitutive expression in the '*nod*D' direction and inducible expression in the opposite direction. This indicates that the control of the *nod* genes in *R. loti* is fundamentally the same as that seen in other fast-growing *Rhizobium* species. Three regulatory elements affecting the levels of *nod* gene expression have tentatively been identified outside the *nod*-box sequence, though the results indicating their presence may simply be due to spacing differences between the *nod*-box and the reporter gene.

ACKNOWLEDGEMENTS

Primarily, I should like to thank my chief supervisor Associate Professor Eric Terzaghi for his advice, encouragement and patience at all stages of the research. Hopefully, the inexplicable experimental results that I regularly presented him with did not contribute too much to his early 'retirement'. Similarly, sincere thanks must go also to my co-supervisor Shashi Sharma (CRI AgResearch Grasslands) who conceived the notion of a bidirectional reporter vector and provided much assistance in its development.

Thanks must also go to the Department of Microbiology and Genetics for providing the facilities and funding necessary for the project to be completed. I am also indebted to Professor Barry Scott for the bacterial strains he kindly supplied, and for his willingness to provide advice throughout the project. Acknowledgement is also given to Drs Brian Mansfield, Rosie Bradshaw and Hale Nicholson for their discussions and strategic advice at various stages of the work. Thanks also goes to Trish McLenachan and Carolyn Young for technical guidance.

Paul Hirst and David Harte helped to keep me going during the more difficult stages of my work and their support was much appreciated.

My parents have also been a great source of support, both emotionally and financially. Their help has been has been gratefully acknowledged.

Finally, thanks must go to my girlfriend Rhonda Stanway for her love, friendship, proof-reading ability and cookery skills she demonstrated throughout the course of this project. Rhonda's patience and tolerance were also greatly appreciated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 - INTRODUCTION

1.1 SYMBIOTIC RELATIONSHIP BETWEEN *Rhizobium* AND LEGUMES

In many communities the most stable and persistent associations between species are those based on obligative mutualism. Since the interaction is mandatory for their survival, the existence of both species in a given habitat is essential. Because of this requirement, a variety of intricate structural or behavioural adaptations have evolved whereby the partnership is maintained throughout the generations. In other cases the relationship is perpetuated simply through the high probability of mutual encounters resulting from a substantial population density or vast offspring numbers contributed by one or both members.

There is a well-characterised mutualistic relationship between *Rhizobium* and leguminous dicotyledons. In this case the *Rhizobium* stimulates the legumes to develop root nodules, which the bacteria infect and inhabit. Ultimately the two symbionts cooperate via metabolic exchanges, enhancing their mutual survival. More specifically the bacteria reduce atmospheric nitrogen into ammonia which is then exported to the plant for assimilation into protein and other nitrogenous compounds. The plant on the other hand supplies photosynthates which provide the energy necessary for bacterial multiplication and nitrogen fixation. Studies of mutants have shown that the major carbon-based energy sources exported by the plant to the bacteroids are dicarboxylic acids, such as succinate, malate and fumarate (Ronson *et al.*, 1981; Engelke *et al.*, 1987; Yarosh *et al.*, 1989). Salminen and Streeter (1992), using C^{14} labelling, were able to demonstrate that a significant proportion of the carbon imported into the *Rhizobium leguminosarum* bacteroids was diverted into the production of glutamate via the tricarboxylic acid (TCA) cycle, and the fixed nitrogen exported to the plant in this form.

Although the aforementioned mutual benefits are considerable, they are not strictly obligatory. In fact, both parties to the symbiosis may exist independently of the other. The *Rhizobium* can be found as a free-living soil organism though its existence is more sensitive to external environmental factors than its nodule-dwelling counterparts. Similarly, nodulation is not always imperative for the survival of legumes in nitrogen-rich soil, but it does appear to confer selective advantages. This is borne out by the observation (reviewed by Beringer *et al.*, 1979), that leguminous seeds contain a significantly higher protein content than the seeds of non-legumes (25% compared to 10% in wheat). Conceivably the extra requirement of seeds for nitrogen may have been an important factor in the selection for nitrogen-fixing nodules in leguminous plants.

1.2 *Rhizobium*-INDUCED NODULATION OF LEGUMINOUS PLANTS

An important part of the symbiotic relationship between *Rhizobium* and the legumes is the initiation of the formation of the nodule. The process that causes the plant to undergo the transition from normal root growth to the development of nodules is fascinating in that it represents a situation where the *Rhizobium* exerts a regulatory influence on the plant genome. This results in the differentiated meristematic tissue becoming developmentally active again, and redifferentiating into nodule tissue. If this important regulatory process could be understood, it would be of great use in gene manipulation studies, both in a pure and applied context.

1.2.1 Brief Summary of Nodule Formation

The development of nitrogen-fixing nodules may be attributed to a series of complex interactions between the legume and compatible rhizobia (Newcomb, 1981). In general, the sequence of events can be likened to a two-way conversation between the plant and the bacteria.

1.2.1.1 Chemotaxis of Rhizobia to Legume Roots

The movement of the rhizobia to the legume roots is an important first step in the formation of a nitrogen-fixing nodule. Interestingly, a number of other organisms, such as *Azospirillum*, exhibit motility towards roots due to specific attractants diffused therefrom (Bashan and Lavanony, 1987). It is very likely that the rhizobia follow a similar form of dispersal as they have been shown to be motile and chemotactic towards nutrients (Bowra and Dilworth, 1981; Bergman *et al.*, 1988) and root exudates (Gitte *et al.*, 1978; Caetano-Anolles *et al.*, 1988).

The attraction exhibited by the *Rhizobium* in response to nutrients enables the bacteria to reach new sources of energy substrate. Indeed it has been reported that wild-type *R. leguminosarum* bv. *trifolii* grew considerably better than non-motile mutants of the same variety when inoculated into sterile soil (Barnet, 1991). Clearly the wild-type organisms gained a selective advantage over their non-motile and non-chemotactic competitors due to their ability to migrate to improved nutritional environments.

Leguminous roots release secondary plant metabolites, such as rhizopines, into the immediate surroundings of the rhizosphere, where the wild-type rhizobia are able to utilise these products as a carbon and energy source (Murphy *et al.*, 1987). One recently investigated metabolite, trigonelline, has also been implicated as an energy source for the rhizobia in both the rhizosphere and the nodule itself (Boivin *et al.*, 1990). In this case

the *trc* genes found on the *Rhizobium* pSym plasmid are induced by the trigonelline to produce proteins that can catabolise the plant product. The evolution of inducible genes able to catabolise these specific plant products has conferred enormous competitive advantages over other bacteria.

As mentioned earlier, nutrients are not the only compounds to which rhizobia are attracted. Root exudates also elicit a positive chemotactic response. Caetano-Anolles *et al.* (1988) and Kape *et al.* (1991) have identified specific flavonoid compounds contained within the root exudate as being the chemical attractants. Presumably the combined chemotactic pressure exerted by the plant-derived nutrients and flavonoid compounds draws the rhizobia towards the leguminous roots.

1.2.1.2 Induced Rhizobial Response to Leguminous Flavonoid Compounds

The flavonoid compounds also have a regulatory role to play in the formation of the nodule. They are indirectly responsible for regulating the expression of the *Rhizobium* nodulation (*nod*) genes. The actual process by which the *nod* genes are controlled appears to be a departure from the phage λ paradigm of gene regulation and constitutes the primary subject of investigation for this thesis. The *nod* genes code for a variety of enzymes which are involved in the synthesis of Nod factors (Redmond *et al.*, 1986; Peters and Long, 1988). These morphogens play an important role in inducing the formation of nodule primordia in the root cortex.

The chemical structure of Nod factors synthesised by several *Rhizobium* strains has recently been examined and analysed. The basic conformation of these factors appears to be a ß-1,4-linked oligomer of N-acetylglucosamine bearing an N-acyl substitution on the non-reducing end (Lerouge *et al.*, 1990; Truchet *et al.*, 1991; Spaink *et al.*, 1991). Slight differences in the length of the oligomeric backbone and the fatty acyl substitutions result in a large family of possible Nod factors. Further variation in the Nod factors can be observed at the reducing end of the compound. This terminus may be modified by the addition of either a sulphate, fucose or methylfucose group depending on the species. The minor differences between the families of Nod factors synthesised by each rhizobial strain are thought to be at least partially responsible for host range specificity. Indeed Roche *et al.* (1991), and Spaink *et al.* (1991), have shown that these modifications are critical for the host-specific induction of nodule primordia. The idea that simple chemical structures can induce profound developmental changes is not new. Albersheim and Darvill (1985), working extensively with oligosaccharides demonstrated that these simple compounds can act as regulatory molecules at very low concentrations.

The synthesised Nod factors are released from the *Rhizobium* and constitute a molecular response to the plants initial flavonoid signal. The diffusible properties of the Nod factors were examined by Ames and Bergman (1981), and Mellor *et al.* (1987). They found that non-motile mutants retained the ability to induce nodulation. It therefore follows that the movement of the rhizobia towards the legume is not essential for nodulation to occur. The message from the bacteria to the plant is instead delivered by the diffusible Nod factor.

The means by which the Nod factor signals the plant to form a nodule is not well understood. Recently, however, Ehrhardt *et al.* (1992), measured transmembrane potential change in single infectible root hair cells in response to *Rhizobium meliloti* extracellular Nod factor NodRm-IV(S). It was found that a rapid depolarisation of transmembrane potential occurred when the Nod factor was present. A second messenger, such as auxin or cytokinin, may be responsible for transducing the signal for nodule formation to the root interior (Hirsch *et al.*, 1989). Perhaps depolarisation of the membrane and changes in plant hormone levels are somehow involved in a Nod-factor-induced signal cascade. A considerable amount of further experimentation will be required before this facet of nodulation can be elucidated.

1.2.1.3 Formation of the Nodule

The rapidly expanding root hair cells are susceptible to *Rhizobium* Nod factor-induced deformation (Bhuvaneswari *et al.*, 1981). Although the curling can be induced by bacteria-free lysates, it is most severe when the rhizobia is bound. Van Batenburg *et al.* (1986), suggests this process can be attributed to the *Rhizobium* provoking a local stimulation in the rate of plant wall expansion.

The *Rhizobium* and the root hair are able to maintain a tight attachment to one another through the formation of cellulose fibrils. These compounds assist in embedding the *Rhizobium* into the mucilaginous material protecting the surface of the hair cells. Furthermore, the binding may be strengthened by some calcium-dependent proteins (adhesins) produced by rhizobia (Smit *et al.*, 1989). Lectins produced by the host bind to both the plant and compatible rhizobia. Dazzo and Gardiol (1984), and Diaz *et al.* (1989), proposed that this lectin-mediated attachment may be an important step early in the recognition of rhizobial strains that can form a nodule. This hypothesis has been reinforced by recent experiments by Philip-Hollingsworth *et al.* (1989), working with *R. leguminosarum*. When the microsymbiont *R. leguminosarum* 300 received the *nod* genes specifying the host range of the clover microsymbiont *R. leguminosarum* ANU843, its *in situ* binding of clover lectin (trifoliin A) increased greatly to a level comparable to that of the clover symbiont ANU483. Furthermore, when Diaz *et al.* (1989), transformed

white clover with pea lectin genes, they observed that the clover's roots were susceptible to nodulation by the pea microsymbiont 248.

The attachment of the bacteria to the plant root elicits a sophisticated developmental response in the legume beyond that of root hair curling. This deformation coincides with the initiation of cell division in the sub-epidermal layer of the root cortex (Newcomb, 1981; Dudley *et al.*, 1987; Guinel and LaRue, 1991). The activation of the root cortical cells leads in turn to the formation of the nodule meristem, whose growth pattern is vastly dissimilar from the neighbouring lateral root cells at the same temporal stage (Dudley *et al.*, 1987). The infected plant cells are then stimulated to produce cell wall sheathes termed 'infection threads' (Callaham and Torrey, 1981). The infection thread, which can be descibed as an inwardly growing tunnel, invades a cell in the outermost layer of the root cortex before infecting subsequent cell layers in a pre-determined manner. These plant-derived threads provide access for the rhizobia into the nodule primordia. When the bacteria have reached the end of the sheath they get released into the plant cells by endocytosis (Dart, 1977; Goodchild, 1978).

At this stage of nodule development the rhizobia become enveloped in host-derived peribacteroid membranes (Robertson *et al.*, 1978; Bauer, 1981), which are antigenically similar to the plant cell plasma membrane (Perotto *et al.*, 1991). The intracellular rhizobia then differentiate into pleiomorphic forms called bacteroids which are able to fix nitrogen and exchange metabolites with the differentiated nodule tissue (Robertson *et al.*, 1978; Verma and Long, 1983).

The formation of nodules is accompanied by the nodule-specific expression of the plant nodulin genes. Early nodulin genes are expressed when the plant becomes infected and the nodule begins to form, whereas the more numerous late nodulin genes are first expressed around the onset of nitrogen fixation (reviewed by Franssen *et al.*, 1989).

1.3 THE NODULATION (*nod*) GENES OF *Rhizobium*

The rhizobial genes responsible for inducing the plant to undergo the transition from normal root growth to the initiation of nodules are the nodulation (*nod*) genes. Two different classes of *nod* genes have been identified in *Rhizobium*, both of which are required for the nodulation of the specific host legume.

The first group of nodulation genes, called the common *nod* genes, are so named because they appear to be functionally interchangeable between rhizobia. Mutations in one *Rhizobium* species can be complemented by cloned wild-type genes from another

species. Furthermore, these cloned genes are insufficient to extend the host range of any recipient species. The best characterised common nodulation genes found in *Rhizobium* are the *nod*ABC and *nod*D genes. Mutations in the *nod*DABC genes of *R. leguminosarum* bv. *viciae* (Downie *et al.*, 1985) and bv. *trifolii* (Djordjevic *et al.*, 1987) resulted in legume roots exhibiting neither hair curling (Hac$^-$) nor nodule formation (Nod$^-$). These genes are also necessary for the induction of cortical cell division in the plant root (Dudley *et al.*, 1987).

Based on the observations that the *nod*ABC genes are required in these pivotal steps of nodule formation, Long (1989) suggested a common antecedent for these events. Support for this proposal came from Marvel *et al.* (1987). They showed that the *nod*ABC gene products were even required in the nodulation of *Parasponia*, the only non-legume genus whose members form *Rhizobium*-induced nodules.

Biochemical and immunological studies have shown that NodA and NodB are localised in the cytosol (reviewed by Denarie *et al.*, 1992). It is thought that these cytosolic proteins may be involved in the synthesis of the amino sugar backbone of the Nod factor. Alternatively, the NodAB proteins may have a role in the N-acylation of this signal molecule.

The other common *nod* gene involved in the production of the Nod factor is *nod*C. Sequence comparisons have revealed that the *nod*C gene product shows significant homology with chitin synthases from *Saccharomyces cerevisiae* (Bulawa and Wasco, 1991). This suggests that NodC is the major protein responsible for synthesising the chitin oligomeric backbone of the Nod factor. Immunological studies have demonstrated that at least part of the NodC protein is located in both of the two bacterial membranes (John *et al.*, 1988; Hubac *et al.*, 1992). This indicates that it could also be involved in the transport of the Nod factor out of the cell.

The common *nod*ABC genes are all under the control of the NodD protein. This positive transcriptional activator has been localised exclusively to the cytoplasmic membrane in *R. leguminosarum* bv. *viciae* (Schlaman *et al.*, 1989). Schlaman *et al.* (1989), proposed that the amphipathic NodD may be anchored in the cytoplasmic membrane with a substantial hydrophilic domain extending into the cytoplasm. The structure and regulatory properties of NodD will be discussed in far greater detail in a following section. It will also become evident that while *nod*D has been classified as a common *nod* gene on the basis of cross-complementation studies, it is also involved in host recognition. At this stage it should merely be noted that the *nod*D gene product mediates the first step in the plant-microbe dialogue by means of interactions with a plant-specific flavonoid (as discussed earlier).

The other common nodulation genes are *nod*I and *nod*J, about which little is known. It has been suggested that their coded proteins form a transport complex in the membrane of the *Rhizobium* (Evans and Downie, 1986).

The second group of nodulation genes consists of the host-specific nodulation (*hsn*) genes. This class is involved in the decoration of the Nod factor and is therefore responsible for the recognition of potentially nodule-bearing host plants. Evidence for this is shown by mutations in the *hsn* genes, which result in expanded host ranges (Kondorosi and Kondorosi, 1986). The entire *hsn* group of genes are unable to complement *nod* mutants where the mutation is located within one of the *hsn* genes from heterologous *Rhizobium* species, suggesting high species specificity. There are also some genes, within this class (including *nod*EF, *nod*H and *nod*LMN), which affect the position and tightness of root haircurling, and the regulation of cell division (reviewed by Long, 1989).

1.4 PHAGE LAMBDA (λ) REGULATORY PARADIGM

Before examining the regulation of the *nod* genes in detail, it would be of great benefit to outline the general features of the λ regulation system. This system illustrates many aspects of gene control that are thought to be essential in the promoter/operator function of almost all the genes studied to date. It is, therefore, a valuable model with which the regulation of the *nod* genes may be compared and contrasted.

1.4.1 Overview of Phage Induction

The bacteriophage λ is a very well studied virus whose biochemistry and genetics have been exhaustively examined and understood. Included in this analysis has been the dissection of the regulatory control features exhibited by the phage. Now, through comparisons with regulatory systems found in other organisms, these features have emerged as a useful model of promoter/operator systems.

Bacteriophage λ possesses two modes of replicating its DNA. In physiologically favourable conditions the λ DNA is incorporated into the DNA of its host where it is passively replicated along with the other elements of the bacterial chromosome. However, when an infected bacterium is exposed to unfavourable conditions, such as ultraviolet (UV) light, a different train of events is set in motion. Once the phage has sensed that the external disturbance is likely to compromise the bacterium, it abandons its previously useful strategy of lysogenic existence and enters the lytic pathway. This mechanism of switching from the lysogenic cycle to the lytic cycle involves the promoter/operator control system, long considered the paradigm of control function. It is the objective of this discussion to describe this paradigm with a view to comparing the superficially similar mode of bidirectional control exhibited by the *nod*-box in *Rhizobium*, at a later stage.

1.4.2 Components of the Genetic Switch

The regulatory proteins responsible for controlling the switch are the λ repressor and the Cro protein which are encoded by the phage genes *c*I and *cro* respectively. These two genes, separated by 80 bp, are transcribed divergently from promoters located in this intergenic region. The *c*I promoter, called P_{RM}, directs the bacterially-encoded RNA polymerase to the left, whereas the *cro* promoter, P_R, orients the polymerase in the opposite direction. Significantly, the *c*I and *cro* promoters are juxtaposed yet do not overlap (Figure 1.1).

Figure 1.1. Genetic Organisation of the Phage λ Control Circuitry

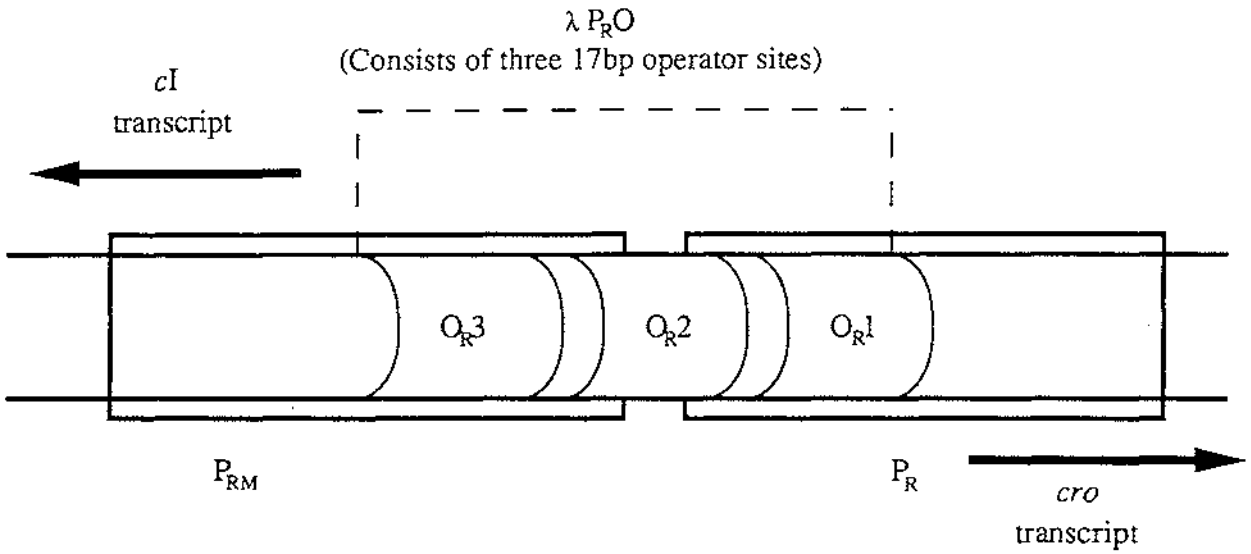


Figure 1.1

The *cis*-acting genetic components of the phage λ regulatory switch. O_R1, O_R2 and O_R3 represent the tripartite operator to which the regulatory proteins bind. Overlapping these sites to varying degrees are the two divergent promoters, P_R and P_{RM}, which provide the RNA polymerase binding sites allowing transcription of these proteins. (After Ptashne).

The 80bp intergenic segment contains another regulatory site in addition to the divergent promoters. This is the right operator of λ, which is comprised of three adjacent protein binding sites (O_R1, O_R2 and O_R3). Each of these 17bp operator sites functionally overlaps either one or the other promoter, or in the case of O_R2, both of the promoters. Sequence analysis has shown that the three operator sites are all imperfect inverted repeats, which are similar to one another. The inverted repeat symmetry is thought to be structurally important in the binding of dimeric proteins.

Although the three protein-binding sites display partial sequence symmetry about an axis through the central base pair, they are not identical. The regulatory proteins can distinguish between them. For example, considering any two operator sites, one might have a higher affinity than the other for Cro. Hence at a specific concentration, it is more likely that a Cro molecule would be found at the site with a higher binding constant, than at the site with a lower binding constant. A hierarchy of affinity can be elucidated for each of the regulatory proteins (Cro and the λ repressor) and for each of their three binding sites (O_R1, O_R2 and O_R3).

The λ repressor protein can exist in a monomeric or dimeric form. Formation of the functional dimeric repressor structure is facilitated by interactions between carboxyl domains of individual monomers. The energetic stability afforded by the dimeric configuration results in the predominance of this favourable species in the lysogen. The other functional unit of the repressor protein is the N-terminal domain which is linked to the C-terminal domain by a 40 amino acid residue connector. While the C-terminal domain of the repressor is responsible for the formation of dimers, the N-terminal domain has a DNA-binding role. Unsurprisingly, this domain contains recognition sites that bind specifically to the operator sites on the integrated phage λ chromosome.

Previous *in vitro* investigations have shown that both types of repressor monomeric domain can exercise their function independently of the other. Firstly, the dissociated N-terminal domains are still capable of binding to DNA operator sites, yet do so less readily than the intact repressor molecule. This difference in affinity is due to the cooperative effect conferred by the dimeric form for inverted repeats. The repressor dimer is thought to induce a conformational change which results in this form having a higher binding constant than the monomeric form (see Figure 1.2a). The C-terminal domains behave in a similar fashion albeit without cooperativity. In the absence of N-termini, these domains are able to form oligomeric chains *in vitro*. This demonstrates that the inter-monomeric binding capacity lies solely in this C-terminal region but does not comment on their *in vivo* activity.

The second regulatory protein, Cro, is considerably smaller than the λ repressor (66 aa compared to 236 aa). In fact the Cro monomer folds into a single globular conformation whose size is similar to the amino domain of the repressor. As with their regulatory counterparts, the Cro monomers have an equilibrium constant that favours the formation of dimers. Predictably, these Cro dimers also exhibit a much greater affinity for DNA operator sites than the monomeric form.

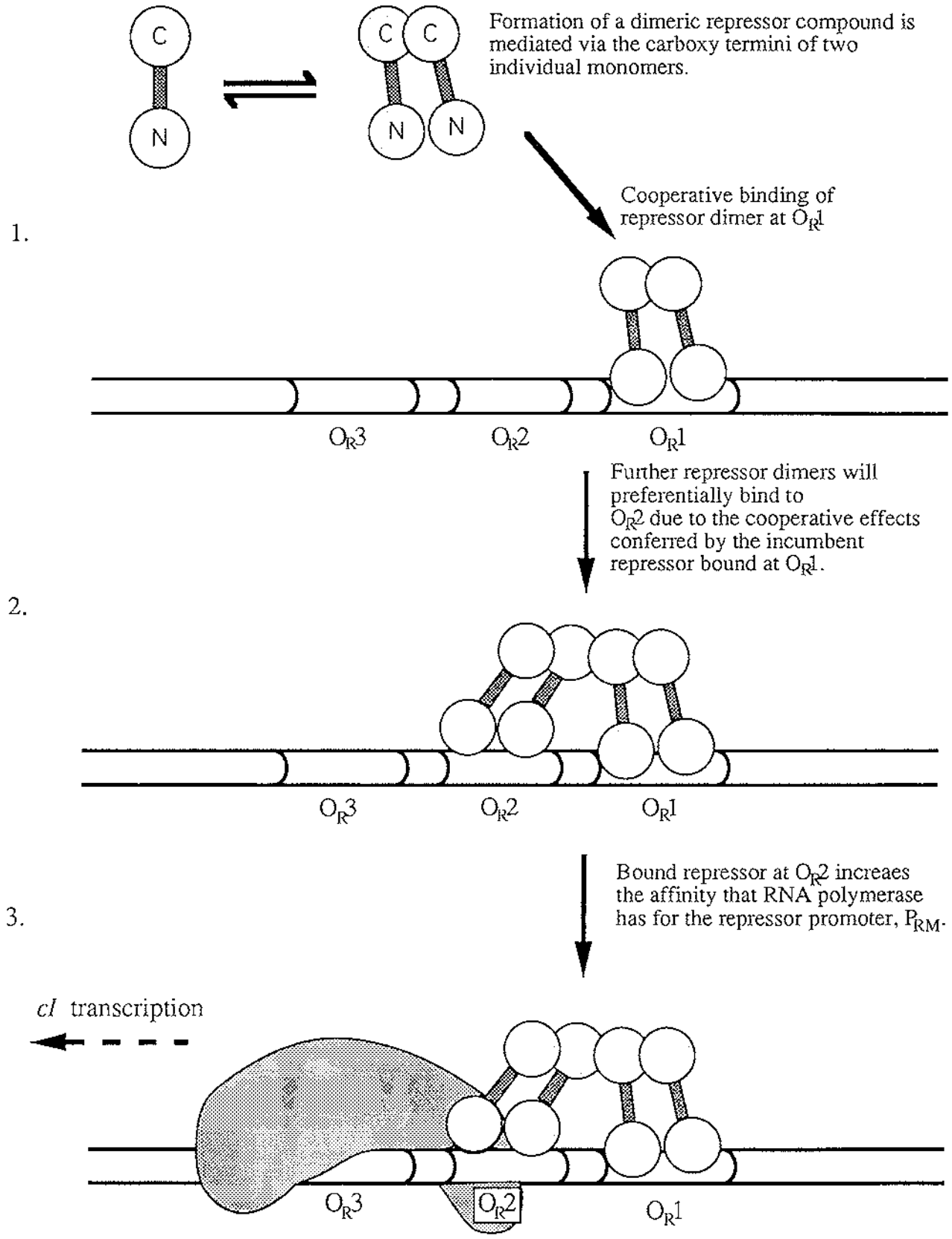
1.4.3 The Lysogenic State

In favourable conditions, it is important for the phage to maintain the lysogenic state. Being the only phage-encoded protein synthesised during this passive phase, the repressor molecule is the primary determinant of its continuation. To effectively maintain lysogeny the concentration of the repressor dimers must be regulated to ensure levels remain within a window of physiological functionality. If the concentration of the repressor dimers drop to an extremely low level, such that P_R is unblocked, premature induction of the lytic phase would certainly follow. On the other hand, if the repressor becomes greatly overexpressed, the phage would be insensitive to external stimuli that would usually induce cell lysis.

Appropriate levels of repressor protein are maintained through an elegant regulatory system. When faced with the triplication of operator binding sites at the λO_R, the repressor dimer binds to the site to which its affinity is the greatest. In this case it is always site 1. The preferential occupation of this site is an important feature in the maintenance of the lysogenic state. A cooperative interaction is thought to be mediated via protein-protein interactions between the bound repressor molecule and a second incoming dimer. This interaction results in the repressor dimer at O_R1 greatly increasing the affinity with which a second dimer can bind to site 2. The connector region of the first repressor is thought to orient the C-terminal regions of the dimer in such a way that they contact the C-terminal domains of the second dimer (see Figure 1.2(b)). This cooperative effect cannot be extended to a third repressor dimer binding to O_R3 because of steric limitations. The lack of a cooperative effect, taken with the low affinity of the repressor dimer for O_R3, means that site 3 is usually unoccupied.

A second effect of dimers binding to O_R1 and O_R2 is one of positive control. Initial contact between the N-terminal domain of the repressor molecule bound to O_R2 and the bacterial RNA polymerase, assists the latter to bind to P_{RM}. This cooperative effect allows the RNA polymerase to preferentially transcribe additional *cI* mRNA molecules. Hence, it would appear that the occupation of these two operator sites by repressor proteins stimulates transcription of their own *cI* gene (see Figure 1.2(c)).

Figure 1.2. Cooperativity Exhibited By the λ Repressor



The binding of repressor dimers to O_R1, and latterly to O_R2, displays negative regulatory control in addition to the positive control just outlined. Indeed, the regulatory suppression of the *cro* transcript by the repressor dimers is an integral feature of the λ switch. The negative control is mediated through the superposition of O_R1 and O_R2 with the Cro promoter, P_R. The binding of these sites by repressor molecules prevents the bacterial RNA polymerase from transcribing the *cro* gene. Therefore, by blocking the P_R promoter, the repressor dimers are effectively abolishing expression of the lytic genes, whilst concomitantly promoting transcription of their own lysogenic *c*I gene. The position of the λ switch is therefore solely dependent on the maintenance of an optimal repressor concentration.

Sometimes the rate of cell division varies, resulting in fluctuating concentrations of repressor molecule. For the integrated phage to remain sensitive to inducing agents, this repressor concentration must be kept below a specific level. In circumstances where the concentration of repressor exceeds this threshold value, the dimers will occupy the O_R3 site which has a higher binding constant than the other sites. This binding physically blocks the RNA polymerase from proceeding to transcribe further repressor molecules. When the cell divides, there is a resulting decrease in the repressor concentration, binding of the O_R3 site is attenuated, and transcription of further repressor molecules will actively proceed. As mentioned earlier, this transcription is enhanced by the cooperative interaction between the two bound dimers at O_R1 and O_R2 with the adjacently bound RNA polymerase.

1.4.4 Switching from the Lysogenic State into the Lytic Cycle

The lysogenic state of the phage could be maintained indefinitely in the absence of external agents. However, irradiating the lysogen with UV light induces a dramatic change in life cycle. In a process yet to be resolved, the RecA protease activity somehow becomes activated and cleaves the connecting region of the repressor molecules. The separation of the amino domain from the carboxyl domain effectively inactivates the repressor because the separated amino domain cannot dimerise. As the monomers have a lower affinity for the DNA than the dimeric form. It is unlikely that any monomeric amino domains will bind to the operator sites. It therefore follows that when the repressor molecules vacate the operator sites there will be too few uncleaved repressor dimers available to replace them.

The resultant absence of repressor molecules bound to the tripartite operator, leads to two important changes. Firstly, there is a decrease in the amount of repressor protein synthesised. The reason for this decline lies in the self-regulatory properties mentioned earlier. As the concentration of the repressor dimers falls, the cooperative effect between the repressor dimer at O_R2 and the RNA polymerase is lost. The expression of the *cI* gene is therefore no longer stimulated, and a negative cascade occurs.

Further disruption of the lysogenic cycle is caused by the initiation of transcription of the previously unexpressed phage genes. Of these, a second phage regulatory protein, Cro, is synthesised, and binds to the vacant operator sites. In a situation opposite to that exhibited by repressor dimers, the Cro dimers have the highest affinity for O_R3. Therefore, once bound, the Cro protein abolishes *cI* synthesis by blocking the polymerase binding site of the repressor. At this point the phage has switched from the lysogenic cycle to the lytic train of events.

1.4.5 Important Features of Phage λ Regulation

The previous sections outlined the series of events which cause phage λ to switch from the lysogenic state to the lytic cycle. The three main points regarding this λ regulatory paradigm are:

1) Operator sites contain inverted repeats which allows dimers to bind, i.e. one monomer subunit per half a binding site.

2) Different affinities for the operator sites by the repressor and the Cro protein results in regulatory proteins binding in a specific order.

3) Cooperativity at several levels enhances binding, and therefore λ regulation.

1.5 REGULATION OF THE *nod* GENES IN *Rhizobium*

The dramatic changes that occur in the transition from normal root growth to nodule primordia are fascinating not only from a developmental viewpoint, but also from a molecular genetic perspective. Considering the complex nature of this system, it is not surprising that the main thrust of recent research has been focused on the molecular level. Numerous investigators have been actively engaged in elucidating the mechanism by which *Rhizobium* induces the formation of nodules on the leguminous plants. While the answer is expected to provide clues about fundamental processes in plant development, it will certainly provide insights into bacterial gene regulation. The evidence accumulated to date on the subject of *nod* gene regulation indicates that a novel control system may be responsible for their transcription. Through comparisons with the phage λ control system outlined in the previous section, it should be possible to highlight differences and similarities between the two bacterial regulatory mechanisms.

1.5.1 Fundamental Features of *nod* Gene Regulation

In *R. meliloti* and *R. leguminosarum* bv. *viciae* the regulatory *nod*D gene is transcribed divergently from the other common *nod* genes. Typically, these are found as an intact *nod*ABCIJ operon. The *nod*D gene is constitutively transcribed in free-living cells, while the *nod*ABCIJ operon is expressed only in the presence of plant-derived inducer molecules and the NodD protein. The upstream region of any inducible *nod* operon contains a highly conserved DNA sequence termed the *nod*-box, which is essential for promoter function. Within this putative regulatory region, the oppositely oriented promoters of *nod*D and *nod*A have been found to overlap. It has been demonstrated that NodD binds to the *nod*-box DNA regardless of the presence of flavonoid inducers. While the NodD-*nod*-box complex is thought to be responsible for inducing transcription of the *nod*ABCIJ operon, the molecular mechanism has not been unambiguously elucidated.

From this brief commentary describing the events that occur at the common *nod* gene locus, it is clear that the physical organisation of the regulatory region shares many similarities with the phage λ control region. In general terms, both systems have inducible genes that are under the control of a divergently transcribed regulatory gene. In each case, the regulatory gene product exerts its control by binding to operator sequences located between the divergent genes. The regulatory switch is thrown when an external agent modifies the regulatory protein in such a way as to allow the inducible genes to be expressed.

A further similarity between the two systems occurs after the induction event. In both cases, the expression of the induced genes is only required for a determinate period of time. With respect to the phage λ situation, the early lytic genes are only required until the vegetative/lytic functions are fully operational. Once that state has been achieved, a downstream promoter takes over. Comparable results have also been observed in the *nod* genes. In this case the *nod*ABCIJ genes are only required transiently to synthesise the backbone of the diffusible Nod factor. Once this signal molecule has initiated a root nodule on a compatible legume, the continued expression of this operon is disadvantageous for the *Rhizobium*. Excessive *nod*ABCIJ gene expression in bacteroids has been shown to result in Fix$^-$ nodules, presumably because of the energetic drain of synthesising unneeded functions. Hence, both the *nod* and the phage systems of regulation must possess the potential to switch off expression of at least some of the inducible genes once their respective proteins have performed their required functions.

Although there are significant similarities between the regulation of phage λ and the regulation of the *nod* genes, one notable difference also exists. This is the apparant lack of a functionally homologous Cro protein in the *nod* system. As mentioned earlier, *cro* is an inducible gene coding for a product that maintains the lytic cycle by blocking the promoter of the *c*I gene (P_{RM}). Accordingly, transcription of the repressor mRNA is totally inhibited. In contrast, the main regulatory gene in the *nod* system, *nod*D, is still expressed after the regulatory switch has been thrown. The bidirectionally versatile nature of the *nod* gene expression is thus in contrast to the two mutually exclusive states of the phage λ switch.

The preceding comparisons have briefly drawn attention to the organisational similarities and differences exhibited by the *nod* control system and the phage λ control system. On a superficial level, at least, there appears to be overwhelming evidence supporting the idea that the regulation of the *nod* genes conforms to the phage λ paradigm of control. However, it should be noted that, at this stage, the three crucial features of λ control have not yet been addressed. It is the objective of the following sections to closely examine the *nod* gene control circuitry with a view to comparing these model control characteristics.

1.5.2 NodD Protein

1.5.2.1 NodD Belongs to the LysR Family of Proteins

In most *Rhizobium* species the transcription of the nodulation genes, except *nod*D, is dependent on plant-derived flavonoids and the *nod*D gene product (Mulligan and Long, 1985). The importance of this protein was confirmed by Tn5 mutagenesis. The *nod*D⁻ genotype conferred significant delays in root hair curling and resulted in the loss of nodulation potential (Downie *et al.*, 1985). Expression of the *nod*D gene was subsequently examined by Rossen *et al.* (1985), and Spaink *et al.* (1987), working with different biovars of *R. leguminosarum*. Using plasmid-borne *nod*D-*lac*Z fusions, it was established that the *nod*D gene was expressed constitutively when placed in a strain lacking an indigenous symbiotic plasmid. The results described here infer strongly that the NodD protein is important in the regulation of the common *nod* genes.

The apparent regulatory role of the *nod*D gene was further substantiated by computer analysis. On the basis of amino acid sequence homology, NodD has been classified as a member of the LysR family of prokaryotic transcriptional regulators (Henikoff *et al.*, 1988). Individual proteins within the LysR family come from a diverse range of bacteria and perform different cellular functions, yet possess many common features. The foremost criterion for belonging to this protein family is the presence of a specific, highly conserved N-terminal motif. A DNA database searching procedure has identified this conserved region to be a putative helix-turn-helix DNA-binding domain (Henikoff *et al.*, 1988; reviewed by Schlaman *et al.*, 1992). This is consistent with the proposed function of the LysR family of proteins, as the helix-turn-helix domain is a common feature of many transcriptional regulators.

While the LysR family all share a specific DNA-binding motif, there are several other features that also characterise them. Physically, the genes that code for these regulatory proteins are transcribed divergently from the genes that they control. This is presumably to allow a finely-tuned one-step induction event to operate. Induction is always mediated via an inducing compound that is thought to interact with the variable C-terminal region of the proteins. *In vitro* studies suggest that the interaction of the inducer with the LysR family of proteins does not alter the binding of these proteins to their respective DNA target sites. Predictably, these target sites all contain a conserved DNA sequence which has been designated the LysR motif (reviewed by Schlaman *et al.*, 1992). In *Rhizobium*, this motif is contained within the *nod*-box sequence. A further common feature of the LysR family of proteins is autoregulation. Although this property has been demonstrated in *R. leguminosarum*, it does not appear to be ubiquitous amongst

all rhizobial *nod*D genes. The diverse range of mechanisms regulating *nod*D in different species will be discussed more fully in a later section. Notwithstanding this slight variation, there are clearly striking similarities between the members of the LysR family and the NodD protein. This provides additional support for the inclusion of NodD in this protein family.

1.5.2.2 Subcellular Location of NodD

Using antibodies raised against a *lacZ-nod*D gene fusion product, Schlaman *et al.* (1989), determined the subcellular location of NodD in *R. leguminosarum* bv. *viciae*. Their results suggested that this amphipathic protein was located exclusively on the inner surface of the cytoplasmic membrane. However, conflicting observations have been made in *R. meliloti*. In this species, considerable amounts of NodD have been isolated in the soluble cytosolic fraction of a biochemical preparation (reviewed by Schlaman *et al.*, 1992). Similar cytosolic localisation of NodD has also been noted in strains of *Rhizobium* that over-express the *nod*D gene. Significantly, in these strains NodD was also found associated with the cytoplasmic membranes indicating that protein concentration may influence the ultimate subcellular localisation of NodD with saturation of membrane sites leading to spill-over into the cytoplasm (Schlaman *et al.*, 1989).

The amino acid sequence of the NodD protein has been compared to sequences stored on a database. In concordance with its putative intra-cellular location, the computer analysis predicted a hydrophobic α-helical structure for the presumed membrane-associated region of the NodD protein. Such tertiary domains commonly interact with the hydrophobic regions within the phospholipid bilayers of cytoplasmic membranes. Furthermore, a number of proline residues have been observed in the integrated helical region of NodD. Found frequently in membrane proteins, these proline residues are thought to act as receptor subunits or transporters (reviewed by Schlaman *et al.*, 1992). Since NodD proteins and flavonoid inducers are both localised in the cytoplasmic membrane, it is possible that their proposed interaction is mediated through this proline-rich receptor site.

While the direct binding between NodD and flavonoid inducers has not yet been conclusively demonstrated, experimental evidence strongly supports their interaction. The outcome of this presumed binding is two-fold. Firstly, mutagenesis experiments suggest that flavonoid-binding induces a conformational change in NodD. Indeed, Spaink *et al.* (1989), demonstrated that a number of different hybrid NodD proteins activate the expression of inducible *nod* genes independently of flavonoids. Assuming the hybrid protein is conformationally different from native NodDs, this result indicates that the tertiary structure of NodD is important for induction. Secondly, the NodD-

flavonoid complex is thought to have a higher binding constant for the *nod*-box than the NodD protein by itself (Kondorosi *et al.*, 1989). If this was true, it might explain why transcription of the *nod*D gene decreases slightly after induction in *R. meliloti*.

1.5.2.3 NodD Binding to the *nod*-Box

The ability of NodD protein to bind to the highly-conserved 47bp *nod*-box sequence (Figure 1.3a) appears to be a fundamental characteristic of *nod* gene regulation. This binding was experimentally demonstrated by Hong *et al.* (1987), using gel electrophoresis binding assays. This group showed that fragments of DNA containing a *nod*-box sequence were electrophoretically retarded in gels following their exposure to cell-free extracts taken from *R. leguminosarum* NodD$^+$ strains. A similar degree of retardation was observed when the *nod*-box fragments were exposed to a cell-free extract taken from *E. coli* cells containing *nod*D. These data provide strong circumstantial evidence that NodD does bind to the regulatory sequences in this DNA. Using a similar technique, Fisher *et al.* (1988), found comparable results in *R. meliloti*. In this case, it was demonstrated that extracts from cells overproducing NodD retarded the electrophoretic migration of the *nod*-box-containing fragments. The same retardation effect was observed when these DNA fragments were exposed to partially purified *R. meliloti* NodD proteins (Fisher *et al.*, 1988; reviewed by Winsor, 1989). Footprinting analyses have subsequently confirmed that the NodD protein does bind to the *nod*-box (Kondorosi *et al.*, 1989).

1.5.2.4 Analysis of the Functional Domains of the NodD Protein

Although binding of the NodD protein to the *nod*-box had been demonstrated, it was not known which region of the protein was responsible for the attachment. *In vitro* mutagenesis of the *nod*D gene was carried out to investigate the functions of the individual NodD domains. Four classes of NodD mutant derivatives were isolated and characterised by Burn *et al.*, (1987). Each exhibited specific regulatory properties that differed from those displayed by wild-type *R. leguminosarum* bv. *viciae*. The amino acid region harbouring the mutation was sequenced and subsequently correlated with other members of the same class. From this analysis, it was possible to assign functions to particular domains of the NodD protein. Interspecies chimeras have also been constructed to assign functions to specific NodD domains. In one case, the NodD protein from a broad host range *Rhizobium* strain (MPIK3030) had its carboxy terminus replaced with the corresponding region from *R. meliloti*. This hybrid protein was only activated by flavonoids produced by the more restricted host range of *R. meliloti*. Furthermore, a *R. meliloti* NodD protein (NodD$_1$) was unable to complement *Rhizobium* MPIK3030 *nod*D$_1$ mutations on a non-host plant. In contrast, the wide host range

NodD$_1$ was able to complement *R. meliloti* C-terminal mutations on the *R. meliloti* host plant alfalfa. These experiments indicate that the C-terminal part of NodD, which is less conserved than the N-terminal end, is involved in determining flavonoid specificity. In addition, this carboxy terminus displays homology to ligand-binding domains of animal steroid receptors that are known to bind flavonoids (reviewed by Denarie *et al.*, 1992).

The mutagenesis results presented above, coupled with the proposed N-terminal helix-turn-helix DNA binding motif, suggests that the NodD protein consists of two separate functional domains. The N-terminal region of the protein appears to bind to the *nod*-box while the carboxy terminus seems to interact with specific inducing compounds. However, results with double-mutants have demonstrated that the C-terminal domain is also involved in DNA binding. Similarly, some mutations that caused modifications in flavonoid specificity were found to map to the N-terminal region (Burn *et al.*, 1989; reviewed by Denarie *et al.*, 1992; reviewed by Schlaman *et al.*, 1992). Therefore, the functions of flavonoid recognition and *nod*-box-binding depend on the overall tertiary structure of NodD.

1.5.2.5 Multiple *nod*Ds

Some species of *Rhizobium* have more than one copy of the *nod*D gene. The respective gene products each respond to specific external stimuli which indicates slight variations in their C-terminal domains or tertiary structures. The most extensively studied multi-gene family is the one found in *R. meliloti*. Originally, this species was thought to contain three *nod*D genes. Each of these genes alone were sufficient to elicit nodules on the *R. meliloti* host-range of plants. A triple *nod*D mutant resulted in a Nod$^-$ phenotype on the same plants (Honma and Ausubel, 1987). Further confirmation of the functionality of each individual *nod*D gene came from *nod*C-*lac*Z fusions. Honma *et al.* (1990), showed that these constructs were induced in the presence of each *nod*D gene and appropriate flavonoids. More recently, a fourth member of the *nod*D gene family was mapped close to *nod*D$_3$. This autoregulatory gene, designated *syr*M, is thought to play a role in the regulation of the *nod*D$_3$ gene (Mulligan and Long, 1989; Honma *et al.*, 1990).

While the *nod*D$_1$, *nod*D$_2$ and *nod*D$_3$/*syr*M genes are functional, they are not exactly equivalent in nodulation (Gyorgypal *et al.*, 1988; Honma and Ausubel, 1987). Studies using purified NodD compounds demonstrated that these proteins have different sensitivities to the root exudates of various hosts. The regulatory protein encoded by *nod*D$_1$ positively controls the common *nod* genes in the presence of the flavone, luteolin (Mulligan and Long, 1985; Peters *et al.*, 1986). The *nod*D$_2$ gene product activates the expression of *nod*ABC in the presence of plant signals other than luteolin. Hence, the

flavonoid spectra recognised by these two *nod*D genes are not identical, suggesting a host-specific role for these genes in nodulation. Moreover, it has been proposed that the *nod*D_3 gene regulates the *nod* gene regulon in response to both the combined nitrogen status and appropriate flavonoid (Dusha *et al.*, 1989; Kondorosi *et al.*, 1991a).

The multiple *nod*D genes found in some *Rhizobium* may confer broad responsiveness to a variety of plant hosts (Gyorgypal *et al.*, 1988; Honma and Ausubel, 1987). Conversely, the relatively narrow host-range of *R. meliloti* implies that the additional *nod*D genes may also be responsible for tightening *nod* gene regulation. Perhaps the presence of a multi-gene family represents an evolutionarily advanced fine-tuning system for host-range determination. If this were the case, then complex interactions between the *nod*D genes would expand the range of compatible hosts, while maintaining stringent control of the *nod* regulon.

Expression of the *R. meliloti nod*D is subject to an intricate control system involving all of the *nod*D gene products and a variety of external stimuli (Mulligan and Long, 1989; Maillet *et al.*, 1990; Davis and Johnston, 1990; Kondorosi *et al.*, 1991a). The complex range of positive, negative and autoregulatory functions exhibited by these *nod*D genes is beyond the scope of this thesis. It is important to note, however, that while induction of *nod* gene expression by plant signals or different environmental or metabolic conditions may be mediated at the *nod*D multigene level, the actual mechanism of *nod* gene induction through *nod*D conformational changes is likely to remain the same.

1.5.3 The *nod*-Boxes of *Rhizobium*

The *nod*-box is a highly conserved 47bp *cis*-acting regulatory element that is thought to be responsible for regulating expression of the *nod* genes. Transcript sizing, fingerprinting and primer extension studies have demonstrated that *nod*-boxes in *R. meliloti* are typically located 26-28bp upstream of the transcriptional start site of the *nod*A gene (Fisher *et al.*, 1987b). Using similar techniques, Fisher *et al.* (1987a), also determined that the trancriptional initiation site of the *nod*D gene lies on the *nod*A side of the *nod*-box. These results imply that the *nod*A regulatory region lies within the leader region of the *nod*D transcript. As *nod*D transcription in *R. leguminosarum* requires the 30bp to the *nod*A side of the *nod*-box, the *nod*D promoter is thought to extend into the leader region of *nod*A (Rossen *et al.*, 1985; Fisher *et al.* 1987a). Hence, while the divergent *nod*D and *nod*A transcription start sites do not overlap, their promoter regions certainly do.

1.5.3.1 Conserved Regions Within the *nod*-Box

The involvement of *nod*-boxes in the coordinated expression of the *nod* genes has long been established (Rostas *et al.*, 1986). Of less certainty is the presence of repetitive consensus sequences contained within the conserved *nod*-box sequence (Figure 1.3a). Footprint and gel retardation studies suggested a modular construction of the *nod*-box and indicated the occurrence of DNA bending between the modules (Fisher and Long, unpublished). Following this train of thought, Wang and Stacey (1991) analysed a presumptive *nod*-box-like sequence from *Bradyrhizobium japonicum*. This divergent, yet functional, *nod*-box sequence was compared with a number of (brady)rhizobial 'classic *nod*-box sequences'. The results of this investigation led to the identification of a putative 9bp directly repeated consensus sequence (Figure 1.3b).

Wang and Stacey (1991), proposed that in most *Rhizobium* species the *nod*-box was composed of four 9bp units, arranged as two paired direct repeats separated by 4bp. The presence of the 4bp 'linker' region led them to believe that the proposed modules were situated on opposite faces of the helix. Synthetic *nod*-box constructs were designed to examine the effect of altering the steric arrangement of the paired repeats on the DNA helix. Insertion of defined sequences into the region between the paired repeats reduced the ability of the NodD protein to bind to the modified *nod*-box. The binding potential was restored, however, when 8bp were inserted. As 8bp loosely corresponds to one complete turn of the DNA helix, it was proposed that the steric relationship between the two paired repeats is of considerable importance in binding NodD (Wang and Stacey, 1991).

Whilst not disputing this steric conclusion, Fisher and Long (unpublished) do not agree with the prediction that the NodD binding sites are found on opposite faces of the DNA helix. Using promoter-oriented phasing and interference experiments, this group found that NodD protein from *R. meliloti* bound to two sites along one face of the DNA helix. Furthermore, footprint analysis revealed that binding of the NodD protein resulted in the small region between the binding sites of the *nod*-box becoming hypersensitive to DNAseI. Fisher and Long (unpublished) propose that the bound NodD protein induces a bend in this hypersensitive region which enables it to make contact with the two binding sites in the promoter DNA for approximately five turns. The important conclusion from this work is that the NodD protein can bind to two sites, located on the same face of the DNA helix, due to the bend that is induced in the *nod*-box by the bound protein. Having not considered the kink, Wang and Stacey (1991), interpreted the region in the middle of the *nod*-box as being a linker sequence that resulted in the NodD protein binding to repeats on opposite faces of the helix.

The consensus repeat sequence of the *nod*-box proposed by Wang and Stacey (1991), has also been challenged. By comparing the *nod*-box sequences of (brady)rhizobia with a divergent *nod*-box from the more distantly-related *Azorhizobium caulinodans*, it was possible to identify the nucleotide sequences that were important for the efficient binding of the NodD protein (Goethals *et al.*, 1992). From this analysis, it was suggested that each *nod*-box is made up of two discrete NodD-binding sites, termed NodD boxes (Figure 1.3c). These sites contain an inverted repeat structure similar to those found in DNA target sites that are symmetrically bound by protein dimers or tetramers. Goethals *et al.*, (1992), propose that NodD dimers bind to each NodD box in a manner that is influenced by the binding status of the other NodD box. This hypothesis is consistent with the NodD footprint results obtained by Fisher and Long (1989), as well as mutagenesis data presented by Goethals *et al.*, (1992). In the latter study, site-specific mutagenesis of the inverted repeats in the NodD box impaired the binding of the NodD protein. The subsequent identification of inverted repeats in the target sites of other members of the LysR family of proteins further substantiates their importance (Goethals *et al.*, 1992). The highly conserved nature of these target sites agrees with the earlier comment regarding the conservation of the helix-turn-helix motifs found in the N-terminal domains of LysR-type proteins.

Significantly, the NodD box model proposed by Goethals *et al.* (1992), has a number of similarities to the λ regulatory model. The presence of inverted repeats in the *nod*-box binding sites is particularly interesting. While these repeats do not appear to be as extensive as those in the λ operator sites, they do tend to indicate that binding of a symmetrical protein configuration occurs. If a NodD dimeric molecule did bind to the NodD boxes, the cooperativity observed between the λ repressor dimers may be mirrored in this system. It is conceivable that binding of the first NodD dimer orientates the *nod*-box in such a way as to facilitate the binding of the second NodD dimer. Could the NodD-induced bend in the *nod*-box region play a role in this orientation? Clearly, there are still a number of questions yet to be resolved. One of the most interesting in terms of the λ comparison, is the observation that one *nod*-box region in *A. caulinodans* contains three consecutive NodD boxes while most other *nod*-boxes contain just two (Goethals *et al.*, 1992). The identification of three consecutive NodD boxes bears an uncanny resemblance to the three operator sites in the λ system. It would certainly be interesting if a hierarchy of binding constants was also found between the NodD boxes in the *nod*-box system.

Figure 1.3. Conserved Sequences Within the *nod*-Box

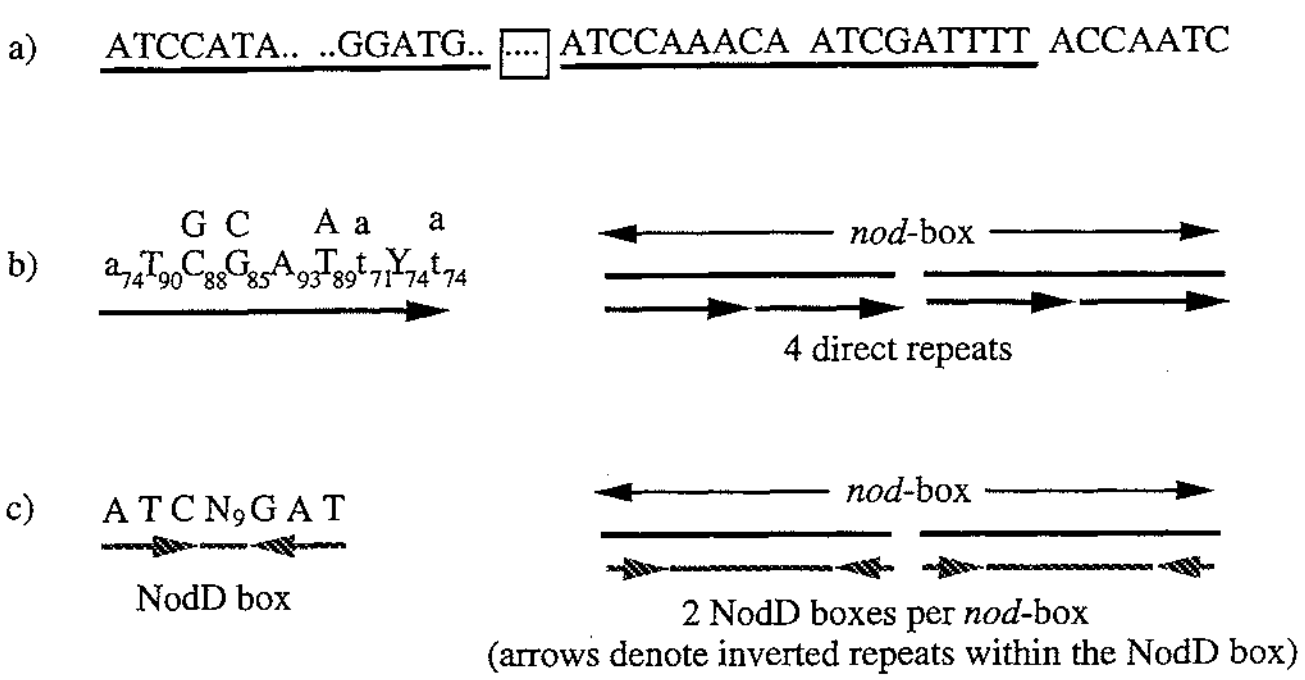


Figure 1.3.

Conserved Sequences Within the *nod*-Box

a) The 'classical consensus sequence' for the 47bp *nod*-box (Wang and Stacey, 1991). An unconserved base is indicated (.). Regions of DNA which are protected in footprinting analyses are underlined. The boxed sequence represents the hypersensitive site (Fisher and Long, 1989, unpublished).

b) The 9bp consensus repeat sequence (Wang and Stacey, 1991). The subscripts denote the percentages of the occurrences of a base at a specific position from the 72 repeats examined (Wang and Stacey, 1991). The capital letters represent highly conserved base pairs. It was proposed that the *nod*-box consists of four direct repeats of this consensus sequence. Each 9bp repeat may be recognised by one subunit of an active NodD tetramer. Another possibility is that each paired repeat may interact with a NodD dimer (Wang and Stacey, 1991).

c) The 15bp consensus NodD box sequence proposed by Goethals *et al.* (1992). In a conventional *nod*-box, this sequence would be repeated twice. The inverted repeats flanking the NodD box are thought to assist binding of a NodD dimer.

1.5.3.2 Regulation of the *nod*D Gene

The *nod*-box is not the only *cis*-acting element controlling the *nod* genes. In many strains of *R. meliloti*, a repressor binding site has been identified through gel retardation studies. In these studies, a 33bp fragment containing the translational initiation site of *nod*A and *nod*D1 was retarded by extracts from *R. meliloti* strain AK631 but not from 1021. Footprinting analyses in AK631 confirmed the binding of the NolR repressor to this site (Kondorosi *et al.*, 1989). Computer analysis of this repressor protein revealed a DNA-binding domain comprising a helix-turn-helix motif which exhibits homology to NodD and other proteins in the LysR family. Like a number of the LysR proteins, NolR binds to its operator site as a dimeric unit (Kondorosi *et al.*, 1991b). Mutations in *nol*R have been found to cause slight delays in the nodulation of host plants (Kondorosi *et al.*, 1991b). This suggests that the NolR protein has a role in fine-tuning the regulation of the *nod* genes. It is also interesting to note that strains containing a functional *nol*R gene express the *nod* genes at a significantly lower level than *nol*R$^-$ strains. Currently, there is no idea as to how NolR brings this about.

While most strains examined from *R. meliloti* appear to carry *nod*D genes that are subject to repression by NolR (Kondorosi *et al.*, 1989), other strains from this species, such as 1021, appear to contain *nod*D genes that are unregulated (Mulligan and Long, 1985; reviewed by Fisher and Long, 1992). In strains such as 1021, expression of a *nod*D-*lac*Z translational fusion was proportional to the copy number of the *nod*D gene (Mulligan and Long, 1985). With this observation in mind, the *nod*D gene of *R. meliloti* 1021 is thought to be unregulated.

A third mechanism of *nod*D regulation has been observed in *R. leguminosarum*. Using *nod*D-*lac*Z fusions, it was shown that the fused *nod*D gene was expressed constitutively in a *nod*D$^-$ strain, yet repressed when a functional *nod*D gene was present. This suggests that the *nod*D gene is autoregulated in *R. leguminosarum* (Rossen *et al.*, 1985). Although it has not been experimentally proven, the observation of NodD binding to the *nod*-box (which contains part of the *nod*D leader transcript) suggests that some degree of autoregulation must occur in the other species of *Rhizobium* too.

The diversity of *nod*D control mechanisms is further accentuated by that found in the closely related genus *Bradyrhizobium*. In this case, Banfalvi *et al.* (1988), have shown that the expression of *nod*D is induced by flavonoids. The variability in *nod*D gene regulation represents a further mechanism by which plant signals can influence the level of *nod* gene expression in a specific manner. However, in an evolutionary context, the wide range of control mechanisms is puzzling. In the apparent absence of an identifiable

selective force, the reason for such diversity is not immediately obvious. This is especially noteworthy in view of the highly conserved nature of the *nod* genes.

1.5.4 Models for *nod* Gene Regulation

The previous sections have outlined the experimental evidence relevant to the control of the *nod* genes. Assembling these data into a comprehensive model is a daunting task, and one that has not been undertaken by many authors. The two models described below are based fundamentally on those presented by Kondorosi *et al.*, (1989). They have been extended slightly to accommodate recent experimental observations.

1.5.4.1 Model One

The first model, based on *R. meliloti* 1021, appears to represent an extremely simple form of *nod* gene regulation. As discussed in the previous section, the *nod*D gene from this strain is apparently unregulated. Therefore the following model can presumably be considered as the most elementary system as it bears no sophisticated regulatory accessories.

In the early stages of the uninduced state (that is, low NodD concentration), the RNA polymerase is thought to bind to the *nod*D gene promoter and transcribe the *nod*D gene constitutively. This results in the synthesis of large amounts of NodD protein. Recent evidence suggests that much of this protein ends up associated with the cytoplasmic membrane, though the excess remains in the cytosol. The cytosolically-localised NodD binds to the *nod*-box and apparently masks the polymerase binding site of the *nod*ABC genes (see Figure 1.4a). Due to the inclusion of the *nod*-box sequence in the *nod*D leader transcript, binding of the *nod*-box should also prevent transcription of *nod*D. This point will be discussed in Section 1.5.4.3.

In induced cells, a complex forms between the NodD protein and the inducer, probably causing a conformational change that exposes the *nod*ABC polymerase binding site to the RNA polymerase. The NodD-flavonoid complex may interact with the RNA polymerase to assist the latter to bind to the promoter region. Transcription of *nod*ABC can therefore proceed uninhibited. The relatively low level of *nod*ABC transcription still permits divergent transcription of the *nod*D gene. This expression is at a slightly reduced level (see Figure 1.4b) due to the high binding constant between the NodD-flavonoid complex and the *nod*-box. The resulting increased occupancy of the *nod*-box is thought to be responsible for the decrease in *nod*D expression, though this will be discussed further in Section 1.5.4.3.

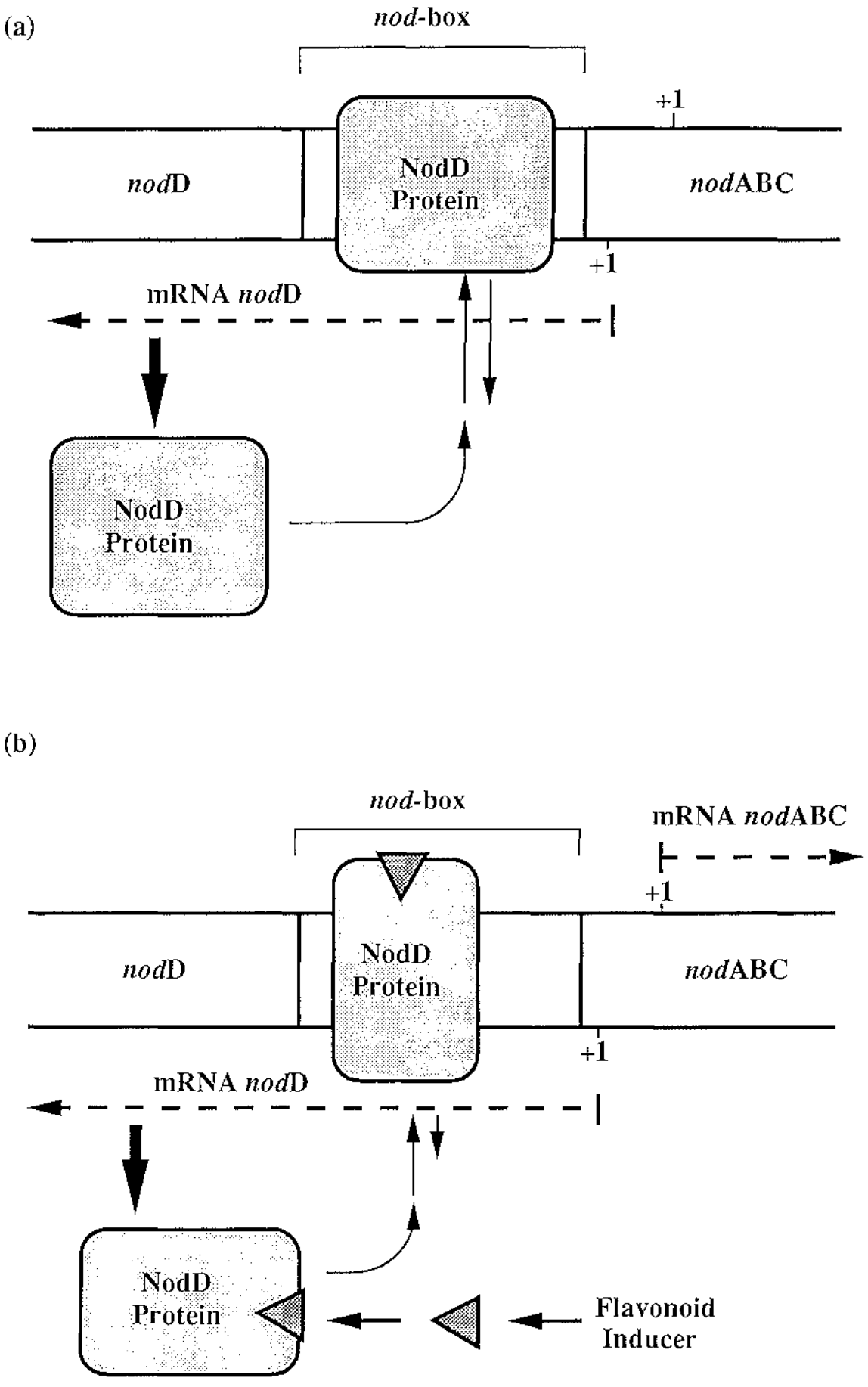
Figure 1.4.

Model One

a) In strain 1021, constitutive expression of the *nod*D regulatory gene results in large quantities of the NodD protein. This protein binds to the *nod*-box and blocks transcription of the divergently-expressed *nod*ABC genes. The mechanism that allows *nod*D transcription to proceed through the NodD-bound *nod*-box has not been resolved (see Section 1.5.3 for a discussion).

b) In the presence of flavonoids, expression of the *nod*ABC genes is induced. This is due to the proposed interaction between the flavonoid inducers and the NodD protein, resulting in the latter becoming conformationally different. This change in tertiary structure becomes evident when the complex binds to the *nod*-box. The flavonoid-NodD complex does not cover the promoter region of the *nod*ABC operon which allows transcription of these genes to occur unimpeded. The NodD-flavonoid complex may also form positive interactions with the RNA polymerase which assist the latter to bind to the *nod*ABC promoter region.

Figure 1.4. Regulation of the *nod* Genes in *R. meliloti* 1021



1.5.4.2 Model Two

The second model, characterised by *R. meliloti* AK631, differs from Model One in that the NolR repressor protein exerts a regulatory effect on the control of the *nod* genes. Unlike Model One, there is no constitutive transcription of the *nod*D gene. This is due to competition for the overlapping promoter region by the NolR repressor dimer and the RNA polymerase. Accordingly, binding of the polymerase to the *nod*D promoter is intermittent leading to a reduced amount of *nod*D transcription. The corresponding low NodD protein concentration and a possible interfering effect of strong repressor binding prevents NodD from consistently occupying the *nod*-box (Figure 1.5a). The competition between the repressor dimer and the RNA polymerase also impedes transcription of the *nod*ABC operon.

Upon induction, a complex presumably forms between the flavonoid inducer and the NodD protein in a similar fashion to Model One. The outcome of this interaction is thought to be an increased binding constant between NodD and the *nod*-box. This promotes binding of the RNA polymerase to the *nod*A promoter as with the 1021 model. The polymerase attachment can subsequently out-compete the repressor for the binding site resulting in transcription of the *nod*ABC operon (see Figure 1.5b). Kondorosi *et al.* (1989), suggest that the inducer is only required for increasing the binding constant of NodD to the *nod*-box. However, it seems highly probable that the inducer must also cause the conformational change that exposes the polymerase binding site, as observed in the 1021 model.

1.5.4.3 Problems with the Models

There are several interesting features of the *nod* gene regulation system that do not get explained in these models. Firstly it has been observed that the location of the *nod*D transcription initiation site lies to the right of the *nod*-box, (Fisher *et al.*, 1987a). This implies that the transcript coding for NodD includes part of this region. The observation that NodD binds to the *nod*-box provides a simple explanation for the autoregulation of *nod*D in *R. leguminosarum.* When the concentration of NodD is low, the RNA polymerase has a higher probability of binding to the *nod*D promoter, resulting in transcription of *nod*D. However, under conditions of high NodD concentration, the NodD protein will out-compete RNA polymerase for the *nod*-box region and repress further transcription of *nod*D. In many respects, the autoregulation of *nod*D is very similar to that seen in phage λ. In this virus, the *c*I gene is autoregulated through binding of its gene product to a specific pattern of operator sites (see Section 1.4). Occupancy of O_R3 by a repressor dimer shuts off expression of the *c*I gene in a

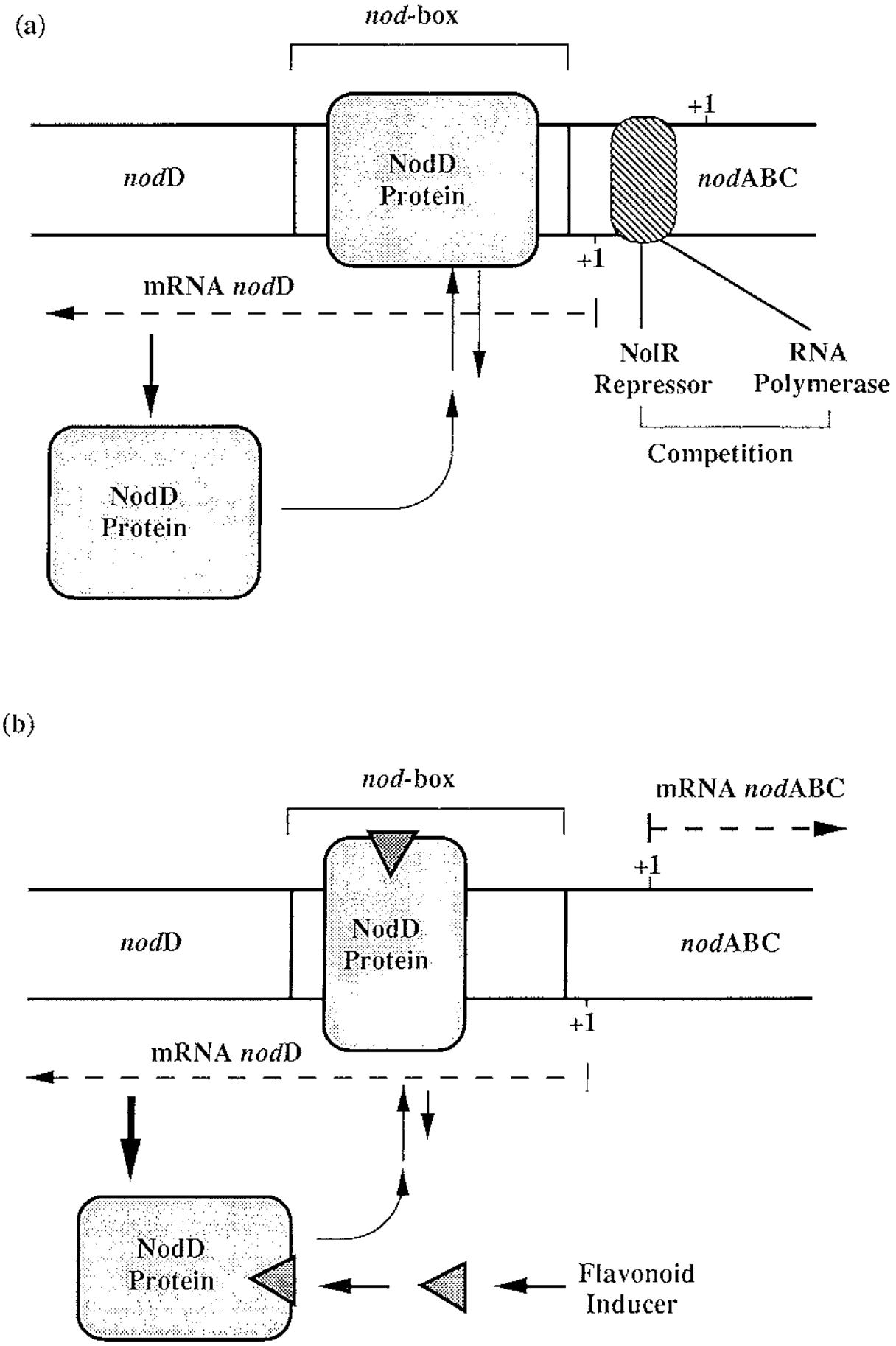
Figure 1.5.

Model Two

a) In *R. meliloti* AK631, expression of the *nod*D gene is regulated by the NolR repressor protein and the RNA polymerase. Both of these proteins bind to the overlapping promoter region of *nod*D. The competition for this site leads to a reduced amount of *nod*D expression and therefore less NodD protein. The low NodD concentration coupled with the strong repressor binding, affects the binding of NodD to the *nod*-box. The competition between NolR and RNA polymerase also prevents expression of the *nod*ABC operon.

b) The NodD-flavonoid complex has a higher binding constant for the *nod*-box than the NodD protein alone. As with Model One, the occupancy of the *nod*-box by the conformationally different NodD-flavonoid complex promotes the binding of RNA polymerase to the *nod*A promoter. The polymerase out-competes the repressor for the binding site resulting in expression of the *nod*ABC genes.

Figure 1.5. Regulation of the *nod* Genes in *R. meliloti* AK631



functionally similar way as NodD binds to the *nod*-box to repress further transcription of *nod*D. While the NodD-bound *nod*-box explains the autoregulatory properties of the *nod*D gene in *R. leguminosarum*, it cannot explain the lack of *nod*D regulation in *R. meliloti* strain 1021 (Model One).

It is a well accepted hypothesis that the RNA polymerase travels along the DNA strand 'reading' the nucleotide sequence and transcribing it into RNA. It therefore follows that there must be no impedance in the form of bound proteins that would hinder the progress of the polymerase. There is no way hitherto known that the polymerase can transcribe the information contained in the DNA unless the two strands can be separated and the coding strand exposed. By observing the binding of the NodD protein to the *nod*-box in 1021, one would immediately assume that this should upset the transcription of *nod*D. However, this disruption has not been observed and the *nod*D transcript is of the length expected if the protein was absent (Kondorosi *et al.*, 1989). The simplest and most elegant explanation for this conundrum is based on the assumption that NodD bound to the *nod*-box is in equilibrium with unbound NodD and RNA polymerase.

In contrast to the divergent negative regulatory effect of NodD binding, the occupancy of the *nod*D promoter by RNA polymerase allows transcription of the *nod*D gene to occur. Assuming NodD cannot interact with the *nod*-box once RNA polymerase has bound, transcription through this region could proceed unimpeded. It could be argued, however, that in the absence of NodD-binding, the *nod*ABC operon would also be transcribed. To counter this legitimate query, one must make the further assumption that the binding constant of RNA polymerase for the *nod*D promoter is higher than its binding constant for the *nod*ABC promoter. This was also suggested in an earlier section, when it was proposed that the binding of RNA polymerase to the promoter region of *nod*ABC may require protein-protein interactions with the induced NodD protein, as is the case of *c*I transcription in λ. Therefore, in the case of *R. meliloti* 1021, the binding constant of the RNA polymerase must be much higher than the binding constant of NodD for the *nod*-box. It seems very wasteful in terms of energy and cellular materials for the *Rhizobium* to incessantly produce NodD protein. Interestingly, hybridisation experiments using the *nol*R gene as a probe, have demonstrated that strain 1021 carries a *nol*R repressor gene (Kondorosi *et al.*, 1991b). It was later shown that this gene did not code for a functional protein in this species. As the repressor protein, NolR, is thought to be required for fine adjustment of *nod* gene expression (Kondorosi *et al.*, 1991b), it is not surprising that the regulatory functions of the *nolR*⁻ strain 1021 are affected. The presence of this mutated gene suggests that this strain arose recently from a *R. meliloti* strain that was under negative control mediated through *nol*R. Perhaps 1021 represents a bacterial strain that is an evolutionary 'trial' presumably destined for eventual extinction.

The presumed requirement of the RNA polymerase to transcribe through the NodD-bound *nod*-box is just one example of the uncertainties still surrounding the regulation of the *nod* genes. Due to the inherent limitations of current *in vitro* procedures, improved *in vivo* techniques will be required to answer this question. Such techniques will not only increase our understanding of this problem, they will also help to unravel several other puzzling regulatory features of the *nod* genes. Of primary interest is the determination of the nature of the interaction between flavonoids and the NodD protein. Demonstration of the hypothesised conformational change in NodD will certainly lend credibility to the current models. Further research may also identify additional proteins that interact with NodD or the *nod*-box to modulate regulation. Preliminary observations in *A. caulinodans* have noted that at least three other proteins, smaller than NodD, bind to the *nod*-box (Goethals *et al.*, 1992). Once again, *in vivo* experimentation will have to be used to ascertain the role of ancillary proteins in the regulation of the *nod* genes. In the past, comparisons between rhizobial species have provided valuable insights into *nod* gene regulation. Continued analysis of these divergent species are particularly useful as they can be used to quickly identify important conserved features of the *nod* control system. With this point in mind, the current project is focused on the *nod*-boxes in *Rhizobium loti*, an organism capable of nodulating the *Lotus* plant.

1.5.5 The Characteristics of *Rhizobium loti*

The mountainous high country of New Zealand's South Island provides an ideal habitat for the *Lotus* plant. This hardy legume is an essential forage plant in these adverse conditions where few other leguminous plants can thrive. The hardiness of the *Lotus* plant can be partly attributed to its ability to form nitrogen-fixing root nodules. Nodulation of the *Lotus* roots is initiated by either the slow-growing bacterium *Bradyrhizobium lotus* or the fast-growing *R. loti*. It is the latter nodulating microorganism that is under investigation in this project.

R. loti are broad host range microbes that can be divided into two taxonomic groups based on their ability to form nitrogen-fixing nodules on *Lotus* roots. The wild-type strain, NZP2037, has the capacity to form effective nodules on an extensive range of *Lotus* species. Conversely, the second group, represented by NZP2213, exhibits a much more restricted host-range. In fact, of the *Lotus* species investigated hitherto, only *L. corniculatus* and *L. tenuis* have been found to be effectively nodulated by this strain of *R. loti*. The remaining *Lotus* species form nodules which are ineffective in their ability to fix nitrogen when infected by the NZP2213 group of *R. loti* strains (Pankhurst *et al.*, 1979, 1987).

In *R. meliloti* and *R. leguminosarum*, the nodulation genes are carried on indigenous megaplasmids (reviewed by Denarie *et al.*, 1992). However, transfer of the megaplasmid of NZP2037 into other strains failed to transmit the information required to nodulate *Lotus* plant roots (Pankhurst *et al.*, 1983). Furthermore, when strains 2037 and 2213 were cured by heat and UV irradiation respectively, the plasmid-free bacteria retained the potential to form effective nodules (Chua *et al.*, 1985). Previous investigations by Pankhurst *et al.* (1987), yielded similar results for the nitrogen fixation (*nif*) genes and the flavolan resistance genes. The absence of all these genes from the indigenous *R. loti* megaplasmid indicates that the symbiotic genes for this species are chromosomally borne. The unusual location of the *nod* genes in *R. loti* makes this species particularly interesting for further research.

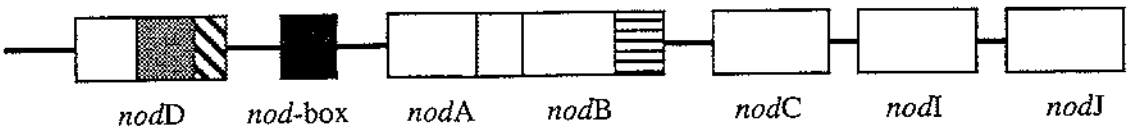
Recent sequence analysis of the common *nod* genes of *R. loti* has exemplified the atypical nature of this species. While the conventional organisation of the *nod* genes takes the form of a single *nod*ABCIJ operon with an adjacent, divergently transcribed *nod*D gene, the arrangement in *R. loti* is considerably different. The most striking alteration is the apparent excision and transposition of the *nod*B gene from its putative ancestral *nod*ABCIJ operon. Scott *et al.* (unpublished), report the location of the *R. loti* *nod*B gene to be adjacent to a second *nod*-box positioned elsewhere on the chromosome, away from the rearranged *nod*ACIJ cluster (Figure 1.6).

Although the *nod*ACIJ operon is located several kilobases from the *nod*B gene, strong evidence exists to suggest that they were once joined. Indeed, sequence analysis has shown that a 46bp C-terminal duplication of *nod*A (ORF3) is immediately upstream of *nod*B, and a 45bp C-terminal duplication of *nod*B is immediately downstream of *nod*A (Scott *et al.*, unpublished). Interestingly, the termination-initiation overlap TATG has been preserved in both cases, even though translational coupling between *nod*A and *nod*B is impossible in *R. loti*, due to the rearrangement described (Normark *et al.*, 1983).

Another disparity has been noted in the arrangement of the *R. loti nod* genes. There is an almost total absence of the *nod*D gene which was expected to be found to the left of the *nod*ACIJ *nod*-box. A 54bp remnant of the *nod*D N-terminus (ORF1) was identified in this position. Consistent with the notion of evolutionary 'splice and paste', a complementary portion of this gene was located upstream of the *nod*B operon. The extensive deletions and nucleotide replacements found in the *nod*D fragments suggest that neither of these remnants are functional. In addition to these deletions, the *nod*D gene fragment associated with the *nod*B gene contains several frame-shift mutations. The accumulation of mutations in this sequence suggests that it is no longer under the selective pressure experienced by functional genes (Scott *et al.*, unpublished).

Figure 1.6. Rearrangement of *R. loti nod* Genes

1. *R. meliloti.*



2. *R. loti.*

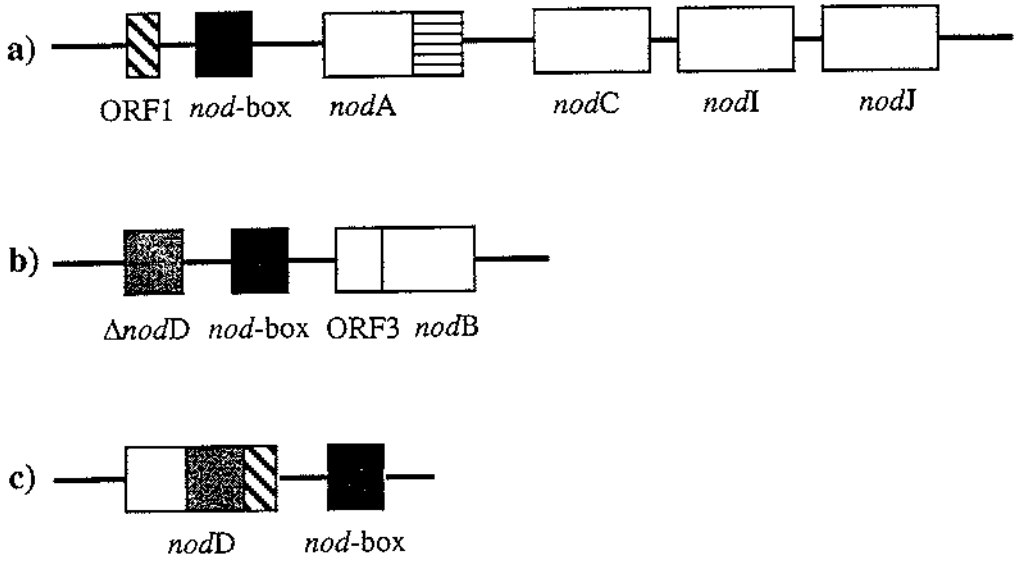


Figure 1.6.

Comparison between the organisation of the common *nod* genes in *R. meliloti* and *R. loti*. The *nod*-boxes are represented by black boxes. Regions that have been evolutionarily rearranged are indicated by different patterns. Assuming the *R. meliloti nod* gene organisation is the ancestral form, there have been at least two mutation events in *R. loti*. The ancestral *nod*D gene has been rearranged, leaving a small C-terminal region of this gene (ORF1) downstream of the *nod*-box (a), and a larger deleted portion (Δ*nod*D) downstream of a second *nod*-box (b). An intact *nod*D gene has been identified next to a third *nod*-box (c). The *nod*B gene in *R. loti* has been excised from the ancestral *nod*ABC operon and is now found to the right of a separate *nod*-box (b). A C-terminal duplication of *nod*A has been identified immediately upstream of *nod*B, and a C-terminal duplication of *nod*B has been found downstream of *nod*A suggesting these two genes were once joined. Diagram taken from Collins-Emerson, (unpublished).

In *R. loti* 2037, a complete *nod*D gene has been identified adjacent to a third, rather divergent *nod*-box (Scott *et al.*, unpublished). When this gene was inactivated by Tn5 mutagenesis, normal nodulation of *L. pedunculatus* was disrupted. The successful cross-complementation of this mutation with the NZP2213 gene confirmed the functionality of this *nod*D gene. While a functional copy of this *nod*D gene is required for nodulation of *L. pedunculatus*, it is not essential for the formation of effective nodules on *L. corniculatus*. It therefore seems likely that this gene has a host-specific role (Scott *et al.*, unpublished). Nodulation of *L. corniculatus* is probably mediated through NodD protein encoded by an uncharacterised *nod*D gene. For a summary of the *nod* gene rearrangements, see Figure 1.6.

Fragmentation of the common *nod* genes has also been observed in type II strains of *R. leguminosarum* bv. *phaseoli*. In this case, the *nod*A gene has been separated from the *nod*BC genes by some 20kb (Vazquez *et al.*, 1991). Although the common *nod* genes from both *R. loti* and this strain of *R. leguminosarum* have been dispersed, the *nod*-box associated with each gene presumably allows their coordinate expression.

1.5.6 Research Aims

Primarily, this investigation set out to examine the regulatory properties conferred by the DNA regions flanking the *nod*-box in *R. loti*. The rearrangement of the common *nod* genes in this species presents an opportunity to study the *nod*-box in the presence of different bracketing regions. Conceivably, it may allow the identification of regulatory elements lying outside the conserved *nod*-box region. If such elements do exist, then it seems highly unlikely that they would be present adjacent to each *nod*-box given the rearranged state of the *R. loti nod* genes. The absence of any regulatory region should manifest itself in the form of irregular *nod* gene control.

Initial results have indicated that the *R. loti* regulatory properties are unusual. Teo (1990), found that the *nod*A *nod*-box was bidirectionally inducible in the presence of appropriate seed exudate. As the *nod*D gene is expressed constitutively in all fast-growing rhizobia studied to date, this result was unexpected. The current investigation will test the validity of this result. Furthermore, a 'deletion' analysis will be performed on the regions immediately flanking the *nod*A and *nod*B nod-boxes. The boundaries of the desired DNA sequences will be specifically delineated by carefully chosen PCR primers. This will result in the production of DNA fragments whose 5' and 3' *nod*-box-flanking regions are markedly different. Using reporter genes to monitor the promoter activity of each *nod*-box-containing fragment, it should be possible to identify any new regulatory elements that may be present.

To facilitate measurement of the activity initiated by the PCR fragments, a bidirectional reporter vector will be designed and constructed. The major advantage conferred by this vector would be its ability to measure the transcription promoted by a cloned fragment simultaneously in both directions. Thus, it is predicted that the improvements in the reporter system will yield significantly more useful results than those obtained using the unidirectional reporter vector pMP220.

A further aim of this project was to design and implement a fast and efficient system for the direct transformation of *R. loti* cells. The current method of triparental mating is extremely time-consuming and laborious. It is therefore very inconvenient to transform these cells routinely. A technique that promises to be vastly superior to this conventional conjugation method is electroporation. Hence, an attempt to electrotransform *R. loti* cells will be undertaken and if successful, the electroporation conditions will be optimised.

Chapter 2 - MATERIALS AND METHODS

2.1 BACTERIAL STRAINS

Bacterial strains and plasmids are listed in Tables 2.1 and 2.2 respectively.

2.2 GROWTH MEDIA

2.2.1 Liquid Media

2.2.1.1 Luria-Bertani Medium (LB) (Miller, 1972)

Composition: 1% (w/v) Bacto tryptone (Difco); 0.5% (w/v) yeast extract (Difco); 0.5% (w/v) NaCl, pH 7.5. Autoclave.

2.2.1.2 Tryptone Yeast Extract (TY) (Beringer, 1974)

Composition: 0.5% (w/v) Bacto tryptone; 0.3% (w/v) Yeast Extract. Autoclave, then add 5ml of sterile 30% (w/v) $CaCl_2.6H_2O$ before use.

2.2.1.3 S40 Medium

1) Salt stock solution (g/L): $MgSO_4.7H_2O$ 25; $CaCl_2.2H_2O$ 2; FeEDTA 1.5; NaCl 20.
2) Trace elements (mg/200ml): $ZnSO_4.7H_2O$ 3; $Na_2MoO_4.2H_2O$ 40; H_3BO_3 50; $MnSO_4.4H_2O$ 40; $CuSO_4.5H_2O$ 4. Add 1ml of 0.2g/L $CoCl_2.6H_2O$ solution.
3) Vitamin stock solution (mg/50ml): Thiamine HCl 50; Ca pantothenate 100; biotin 1.
4) Phosphate stock solution (g/L): K_2HPO_4 100; KH_2PO_4 100.

Add (per 500ml):
- 5ml salts solution
- 0.5ml trace minerals
- 0.5ml vitamins
- 5g MES

and adjust pH to 6.2.

After autoclaving add: 2.5ml phosphate solution, 10ml 20% sucrose, 10ml 10% Na glutamate.

Table 2.1 **Bacterial Strains**

Strain	Relevant Characteristics	Reference or Source
E. Coli		
DH1	F' *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 ($r^- m^+$) *sup*E44 *rel*A1 λ^-	Hanahan (1983); Low (1968)
DH5	*rec*A1 *end*A1 *gyr*A96 *sup*E44 *hsd*R17 (higher transformation frequency than DH1)	Hanahan (1983); Low (1968)
HB101	F' *pro leu thi lac*Y *Str* $r^- m^-$ *Endo*I *Rec*A	Boyer and Roulland-Dussoix (1969)
MC1022	*ara* 13a *(ara leu)* 7697 (*lac*Z) ΔM15 *gal*U *gal*K *Str*A	Casadaban and Cohen (1980)
PN1158	MC1022 / pPN25	B. Scott Massey, P.N., N.Z.
PN1175	DH5 / pMP220	Spaink *et al.* (1987)
PN1233	HB101 / pPN37	Gillman (1990)
SB652	DH5∝ / pTZ19R	S. Sharma AgResearch CRI, P.N., N.Z.
SPA1	MC1022 / pSPA1	This study
SPA2	MC1022 / pSPA2	This study
SPA3	MC1022 / pSPA3	This study
SPB1	MC1022 / pSPB1	This study
SPB2	MC1022 / pSPB2	This study
SPB3	MC1022 / pSPB3	This study
SPB4	MC1022 / pSPB4	This study
SPB5	MC1022 / pSPB5	This study
SPV1	DH1 / pSPV1	This study
SPV2	DH1 / pSPV2	This study
SPV3	DH1 / pSPV3	This study
SPV4	DH1 / pSPV4	This study

Strain	Relevant Characteristics	Reference or Source
E. Coli		
SPA11	DH1 / pSPA11	This study
SPA12	DH1 / pSPA12	This study
SPA13	DH1 / pSPA13	This study
SPB11	DH1 / pSPB11	This study
SPB12	DH1 / pSPB12	This study
SPB13	DH1 / pSPB13	This study
SPB14	DH1 / pSPB14	This study
SPB15	DH1 / pSPB15	This study
SP11	DH1 / pSP11	This study
R. Loti		
NZP2037	Nod^+ Fix^+(*Lotus pedunculatus*)	AgResearch CRI Palm. Nth., NZ
NZP2213	Nod^+ Fix^-(*Lotus pedunculatus*)	AgResearch CRI Palm. Nth., NZ
PN184	NZP2037 / Str^R	Chua *et al* . (1985)
SPA101	PN184 / pSPA11	This study
SPA102	PN184 / pSPA12	This study
SPA103	PN184 / pSPA13	This study
SPB101	PN184 / pSPB11	This study
SPB102	PN184 / pSPB12	This study
SPB103	PN184 / pSPB13	This study
SPB104	PN184 / pSPB14	This study
SPB105	PN184 / pSPB15	This study
SP100	PN184 / pSPV4	This study
SP101	PN184 / pSP11	This study

Table 2.2 **Plasmids**

Plasmid	Relevant Characteristics	Source or Reference
pUC118	3.2kb high copy number vector containing amino-terminal fragment of *lacZ* which displays α-complementation in appropriate hosts. MCS consists of: *Eco*RI *Sst*I *Kpn*I *Sma*I *Bam*HI *Xba*I *Sal*I *Pst*I *Sph*I *Hind*III; Amp^R	Vieira and Messing (1987)
pPN25	pUC118 + 1.4kb *Sal*I fragment from the *nod*A *nod* -box region of NZP2037; Amp^R	Gillman (1990)
pPN37	pMP190 + 4.1 kb *Sal*I fragment from the *nod*B *nod* -box region of NZP2213; Cam^R Str^R	Gillman (1990)
pSPA1	pUC118 + 0.494kb PCR fragment A1; Amp^R	This study
pSPA2	pUC118 + 0.408kb PCR fragment A2; Amp^R	This study
pSPA3	pUC118 + 0.145kb PCR fragment A3; Amp^R	This study
pSPB1	pUC118 + 0.739kb PCR fragment B1; Amp^R	This study
pSPB2	pSPB1 - 0.468kb *Eco*RI fragment; Amp^R	This study
pSPB3	pUC118 + 0.468kb *Eco*RI fragment from pSPB1; Amp^R	This study
pSPB4	pUC118 + 0.552kb PCR fragment from pSPB1; Amp^R	This study
pSPB5	pUC118 + 0.074kb *Eco*RI fragment from pSPB4; Amp^R	This study

Plasmid	Relevant Characteristics	Source or Reference
pUC21	3.2 kb cloning vector; derivative of pUC118; α–complementation; MCS consists of : *Spe*I *Stu*I *Xho*I *Bgl*II *Cla*I *Sph*I *Kpn*I *Sma*I *Eco*RI *Hind*III *Pst*I *Xba*I *Spe*I; Amp^R	Vieira and Messing (1991)
pMP220	10.5 kb IncP, low copy number plasmid containing promoterless *lacZ* reporter gene; MCS consists of: *Hind*III *Bgl*II *Eco*RI *Kpn*I *Xba*I *Pst*I *Sph*I *Hind*III; Tet^R	Spaink *et al*. (1987)
pRAJ260	EMBL9 derivative containing promoterless *gus*A reporter gene bound by *Pst*I sites	Jefferson *et al*. (1986)
pSPV1	pUC21 + 1.8 kb blunt-ended *Pst*I fragment containing *gus*A from pRAJ260; potential bidirectional reporter vector (translational fusions); MCS consists of: *Spe*I *Xho*I *Bgl*II *Cla*I *Sph*I *Kpn*I *Sma*I *Eco*RI *Hind*III *Pst*I *Bam*HI *Xba*I *Spe*I; Amp^R	This study
pSPV2	pSPV1 + 0.199kb *Bam*HI/*Sph*I portable rbs from pMP220; potential bidirectional reporter vector (transcriptionally fused *gus*A gene, potential for *lacZ* translational fusion); MCS consists of: *Spe*I *Xho*I (rbs) *Hind*III *Sph*I *Kpn*I *Sma*I *Eco*RI *Hind*III *Pst*I *Bam*HI *Xba*I *Spe*I; Amp^R	This study
pSPV3	pUC118 + 0.199kb *Bam*HI/*Sph*I portable rbs from pMP220; Amp^R	This study

Plasmid	Relevant Characteristics	Source or Reference
pSPV4	pMP220 + *Spe*I fragment from pSPV2; bidirectional reporter vector (transcriptional fusions); MCS consists of: *Hind*III *Sph*I *Kpn*I *Eco*RI *Hind*III *Pst*I *Xba*I; Tet^R	This study
pTZ19R	pUC118 + 0.48kb *Bam*HI / *Sac*I fragment of *nod*A *nod*-box from *R. meliloti*	S. Sharma AgResearch CRI P.N.,N.Z.
pSPA11	pSPV4 + 0.521kb *Eco*RI / *Xba*I fragment from pSPA1; Tet^R	This study
pSPA12	pSPV4 + 0.435kb *Eco*RI / *Xba*I fragment from pSPA2; Tet^R	This study
pSPA13	pSPV4 + 0.172kb *Eco*RI / *Xba*I fragment from pSPA3; Tet^R	This study
pSPB11	SPV4 + 0.758kb *Kpn*I / *Xba*I fragment from pSPB1; Tet^R	This study
pSPB12	pSPV4 + 0.271kb *Eco*RI / *Xba*I fragment from pSPB2; Tet^R	This study
pSPB13	pSPV4 + 0.099kb *Eco*RI / *Xba*I fragment from pSPB3; Tet^R	This study
pSPB14	pSPV4 + 0.571kb *Kpn*I / *Xba*I fragment from pSPB4; Tet^R	This study
pSPB15	pSPV4 + 0.493kb *Eco*RI / *Xba*I fragment from pSPB5; Tet^R	This study
pSP11	pSPV4 + 0.45kb *Eco*RI / *Xba*I fragment from pTZ19R; Tet^R	This study

2.2.2 Solid Media

Solid media was prepared by adding 15g/L agar (Difco) to liquid media.

2.2.3 Antibiotics

See Table 2.3.

Table 2.3 Antibiotics

Antibiotic	Stock concentration (μg/ml)	Working concentration (μg/ ml)		Storage (°C)
		E. Coli	*R. Loti*	
Ampicillin	10 mg/ml in water	200	-	-20
Streptomycin	50 mg/ml in water	-	200	-20
Tetracycline	10 mg/ml in methanol	15	2	-20

(-) indicates that the antibiotic was not used in this study for that organism.

All antibiotic stock solutions were sterilised by filtration through 0.22μm membrane filters. The required dosage was added aseptically to cooled media to selectively screen for bacteria containing the antibiotic marker of interest.

2.3 GROWTH AND MAINTENANCE OF BACTERIAL STRAINS

Strains of *E. coli* were grown at 37°C in LB media for 16 hours to obtain a saturated solution. A colony of *R. loti* grown at 30°C on TY media took 48 hours to reach the corresponding cell density. All media contained appropriate antibiotics to maintain the identity of the bacterial colonies.

Both the *E. coli* and the *R. loti* colonies were stored at room temperature on sealed agar plates supplemented with appropriate antibiotics. The colonies were restreaked at monthly intervals on fresh antibiotic-containing media to ensure the viability of the desired bacteria.

2.4 ISOLATION OF PLASMID DNA FROM *E. coli*

2.4.1 Large Scale Alkaline Lysis Preparation of Plasmid DNA

(Modified method of Ish-Horowicz and Burke, (1981); Banfalvi *et al.*, (1988)).

1) Solution I: 50mM glucose; 25mM Tris-HCl, pH 8.0; 10mM Na_2-EDTA. Autoclave.
2) Solution II: 0.2M NaOH; 1% Sarcosyl (freshly prepared).
3) Solution III: 60 ml of 5M KOAc, pH 5.8; 11.5ml glacial acetic acid; 28.5ml sterile Milli-Q water.
4) Lysozyme: 50mg/ml of Solution I (freshly prepared).
5) TE (10/1): 10mM Tris-HCl, pH 8.0; 1mM Na_2-EDTA, pH 8.0.
6) TES (10/1/100): 10mM Tris-HCl, pH 8.0; 1mM Na_2-EDTA, pH 8.0; 100mM NaCl
7) NaCl-Isopropanol: 5M NaCl in TE (10/1); then add an equal volume of isopropanol and stir overnight.
8) Cold isopropanol (-20°C)
9) 70% (v/v) ethanol
10) Ethidium bromide: 10mg/ml of sterile water
11) Caesium chloride (CsCl)

A single colony of the bacteria containing the plasmid of interest was inoculated into 5ml of LB containing the appropriate antibiotic and allowed to grow until it had reached late log phase. Upon obtaining an OD_{600} of approximately 0.6, the entire culture was inoculated into 500ml of LB containing the appropriate antibiotic, and incubated for 18 hours at 37°C with vigorous shaking. The bacterial cells were then harvested by centrifugation at 10,000g for 10 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 30ml of Solution I, before 3ml of freshly made lysozyme was added. After 10 minutes at room temperature, 60ml of Solution II was included, mixed by inverting the bottle several times, and left on ice for 10 minutes to lyse the cells. The proteins and chromosomal DNA were then precipitated by mixing in 45ml of Solution III and leaving on ice for 10 minutes. Contaminating debris was separated from the soluble plasmid DNA by centrifuging for 10 minutes at 10,000g. The plasmid DNA was then transferred into new centrifuge tubes, and precipitated using 0.6 volumes of cold isopropanol (room temperature, 20 minutes) before being recovered by centrifugation at 23,000g for 10 minutes. After washing the pelleted DNA with 70% ethanol, it was dried, redissolved in 3.75ml of TE (10/1) and placed at 65°C for 10 minutes. CsCl (1.05g/ml DNA suspension) and ethidium bromide (75µl/ml of 10mg/ml stock) were added to the mixture. The refractive index of the supernatant was checked to ensure that n = 1.3860-1.3920, before it was placed into TV-865 tubes. Once

balanced with TE (10/1) that contained both CsCl and ethidium bromide in the right proportion, the mixture was centrifuged at 55,000rpm for 5 hours in a Sorvall Ultracentrifuge. A hypodermic needle was later plunged into the top of the tubes, while a second needle was inserted just below the lower band of DNA, clearly visible under long wave UV light. This band, containing closed circular plasmid DNA, was withdrawn from the tube into a sterile syringe, and later placed into Eppendorf tubes. The ethidium bromide was subsequently removed by repeated extractions using equal volumes of isopropanol saturated with NaCl followed by vigorous vortexing. Once the top phase contained no pink colour, the DNA in the bottom phase was dialysed against TES (10/1/100). The removal of CsCl was complete after four buffer changes over two days. Finally, a 1/50 dilution of the large scale plasmid preparation was measured using a spectrophotometer to determine the purity of the DNA ($A_{260}/A_{280} > 1.8$) and the concentration of the DNA ($A_{260} = 1$ corresponds to 50mg/ml).

2.4.2 **Rapid Boil Method** (Holmes and Quigley, 1981)

1) HQ-STET Buffer: 8% (w/v) sucrose; 50mM Na_2-EDTA; 50mM Tris-HCl pH 8.0. Autoclave. Add 5% (v/v) Triton X-100.
2) Lysozyme: 10mg/ml of 10mM Tris-HCl, pH 8.0 (freshly prepared).
3) Cold isopropanol (-20°C)
4) 70% (v/v) ethanol

A colony of *E. coli* cells was inoculated into 5ml of LB broth supplemented with appropriate antibiotics, and grown with shaking at 37°C for a period of 16-20 hours. 1.5ml of this culture was pelleted in a microfuge for 5 minutes. The supernatant was discarded and the pellet resuspended in 350µl HQ-STET buffer. 25µl of fresh lysozyme was added to the cell suspension and inverted several times to ensure mixing. The solution was allowed to sit for 30 seconds before being plunged into a boiling water bath for 40 seconds. It was then immediately spun for 10 minutes in a microfuge and the gelatinous pellet removed with a toothpick. An equal volume of cold isopropanol was added to the plasmid DNA in the supernatant. This solution was mixed by inversion and then placed at -70°C for 20 minutes. The plasmid DNA was later spun for 15 minutes in a microfuge, and the supernatant discarded. The resulting pellet was washed with 70% ethanol, spun for 5 minutes, drained and then dried using the Speedvac dryer. The plasmid DNA was resuspended in 50µl distilled water and stored frozen (-20°C) until required.

Note:

1) Antibiotics in LB are essential when the colony is picked from LBAmp plates after electroporation. This is due to the bacteriostatic effect of the ampicillin on the surrounding satellite colonies that get picked with the colony of interest.

2.4.3 Plasmid Isolation Using Magic Minipreps

This is an extremely simple method that has the advantage of being very fast and efficient. It was used to prepare plasmid DNA for use in PCR and sequencing experiments.

1) Magic Minipreps DNA Purification System (Promega)
2) 3ml syringe

The experimental protocol followed was described in the Promega Technical Bulletin that was supplied with the kit.

2.5 PURIFICATION AND DETECTION OF DNA

Unless otherwise referenced, these protocols were based on the descriptions in Maniatis (1989).

2.5.1 Phenol-Chloroform Extraction of DNA

1) Phenol-equilibrated with 10mM Tris-HCl, pH 8.0.
2) Chloroform-isoamyl alcohol (24:1; v/v)

Equal volumes of phenol and chloroform were added to the DNA sample and vortexed until an emulsion formed. This was then centrifuged for 3 minutes in a microfuge. The DNA-containing aqueous phase was pipetted into another Eppendorf tube, while the interface and the organic phase, containing the protein contaminants, were discarded. A final extraction in an equal volume of chloroform was made to remove any residual phenol. Once again the top phase was transferred to another Eppendorf tube. The DNA contained in this solution was recovered by precipitation with ethanol as outlined below (Method 2.5.2).

2.5.2 Precipitation of DNA With Ethanol or Isopropanol

1) 3M Sodium acetate (NaOAc)
2) 95% (v/v) cold ethanol (-20°C)
3) 70% (v/v) ethanol

After estimating the volume of aqueous DNA, the correct ionic environment for precipitating DNA was provided by adding 0.1 volumes of 3M NaOAc. Three volumes of cold 95% ethanol or 0.6 volumes of cold isopropanol were then gently mixed in and the solution placed at -70°C for 20 minutes. The resulting precipitate was spun in a microfuge at 4°C for at least 15 minutes (see Note 1). After the supernatant had been discarded, the pellet was washed with 70% cold ethanol, followed by another centrifugation at 4°C for 5 minutes. The supernatant was then carefully removed and the DNA pellet dried in a Speedvac for 5 minutes. The DNA pellet was later resuspended in sterile water and stored frozen at -20°C.

Notes:

1) Very short DNA fragments (<500bps) were precipitated using 5 volumes of cold 95% ethanol and placed at -70°C for at least 30 minutes. The mixture was then spun in a microfuge for 30 minutes to obtain an efficient yield of DNA.
2) The centrifugation time is the most critical parameter in the recovery of DNA (Zeugin and Hartley, 1985).

2.5.3 Horizontal Agarose Gel Electrophoresis of DNA

1) SeaKem Agarose (FMC)
2) 1x TBE Buffer: 2mM Na_2-EDTA; 90mM boric acid. Adjust to pH 8.2.
3) SDS Dye mix: 20% sucrose; 5mM Na_2-EDTA; 1% (w/v) SDS; 0.2% (w/v) Bromophenol blue.

After preparing 1% (w/v) agarose in 1x TBE, the molten gel solution was poured into a comb-containing mini-gel box (11x14cm) and allowed to solidify for 20 minutes. The gel was then overlaid with 1x TBE buffer, before the comb was extracted and the DNA samples (combined with the dye mix) loaded into the resultant wells. The application of an electric field at 100V for 1 hour caused the DNA fragments to migrate at their size-dependent rates.

Note:

1) For DNA fragments whose size was expected to be less than 1kb, a 1.5% agarose gel was used to assist in their separation.

2.5.4 Gel Staining and Photography

1) Ethidium bromide 25µl/ml.

At the conclusion of electrophoresis, the gel was immersed in a solution of ethidium bromide (25µl/ml) for 15 minutes, and then destained in tap water for 5 minutes before being mounted on a transilluminator. Upon transillumination by a UV light source the gel was observed and then photographed through a red filter at f8 for 1 second with Polaroid 667 instant film.

2.5.5 Extraction of DNA from Seaplaque Agarose (Thuring *et al.*, 1975).

1) 1% (w/v) Seaplaque agarose (FMC)
2) 1x TAE Buffer: 40mM Tris-HCl; 20mM glacial acetic acid; 2mM Na_2-EDTA.
3) Phenol-equilibrated with 10mM Tris-HCl, pH 8.0

A 1% Seaplaque gel was poured, and then cooled at 4°C for 30 minutes to ensure hardening of the agarose. Once the 1x TAE buffer had been poured over the gel and the DNA samples loaded, the gel was electrophoresed for 1-1.5 hours at 80V. The gel was stained in ethidium bromide for 15 minutes and destained in tap water for 5 minutes before being examined under long-wave UV light. The band of interest was excised using a sterile razor blade and blotted to remove surplus buffer. The trimmed slice of agarose was later cut into small pieces, placed in an Eppendorf tube and melted in a heating block at 65°C. After 10 minutes, an equal volume of phenol was added to the melted Seaplaque, the mixture vortexed thoroughly and placed at -70°C for 20 minutes to separate the DNA from the agarose. The frozen sample was then centrifuged for 10 minutes, and the aqueous phase transferred to another Eppendorf tube. This DNA solution was further purified using a phenol/chloroform extraction (Method 2.5.1).

2.6 ENZYMATIC MANIPULATION AND ELECTROPORATION OF PLASMID DNA

2.6.1 Digestion of DNA with Restriction Enzymes

1) 10x Commercial buffer
2) Restriction enzyme: 2-5 U/µg DNA
3) RNAse: 10mg/ml sterile water (Heat at 100°C for 10 minutes to inactivate DNAse).

A reaction volume was comprised of xµl of DNA, 2.5µl 10x commercial buffer and xµl of restriction enzyme, with sterile water added to 25µl. This mixture was then subjected to a short spin, a few flicks and then another brief spin to guarantee complete mixing. The reaction mixture was then incubated for 1 hour at 37°C. In cases where the DNA was obtained from the Rapid Boil Method (2.4.1), it was essential to add 1µl of RNAse to the reaction and incubate for a further 2 minutes. This eliminated the RNA that this method invariably retains, and therefore facilitated analysis of the DNA cleavage pattern. A sample (5µl) of the digest was run on a minigel to confirm the reaction had gone to completion. When the digest was incomplete, more enzyme was added, and the reaction left for a second hour. Once the enzyme had sufficiently cleaved the DNA, it was inactivated either by heating at 65°C for 20 minutes, or by extracting with phenol/chloroform (Method 2.5.1).

2.6.2 End-Filling of Single-Stranded Overhangs

1) T4 Polymerase (BRL)
2) 10x T4 Polymerase Buffer (BRL)
3) 1mM dNTP Mix

The DNA with the single-stranded overhangs was dissolved in 25µl sterile water before 3µl of T4 Polymerase Buffer and 1µl of 1mM dNTPs were added. Finally, 1µl of T4 Polymerase was included and the mixture incubated at 37°C for 15 minutes. The T4 Polymerase could later be inactivated by phenol/chloroform extractions leaving the blunt-ended DNA fragments available for further manipulation.

Note:

1) The Klenow fragment (Boehringer Mannheim) was also found to successfully blunt end DNA fragments when substituted for T4 Polymerase.

2.6.3 Calf-Alkaline Phosphatase (CAP) Treatment of Vector DNA

1) CAP enzyme: 1U/µl (Boehringer Mannheim)
2) 10x CAP Buffer (Boehringer Mannheim)
3) 10% (w/v) SDS
4) Proteinase K (Boehringer Mannheim): 25mg/ml in sterile Milli-Q water
5) 250mM Na_2-EDTA, pH 8.0.

The vector DNA was fully linearised with restriction enzymes, which were subsequently rendered inactive through appropriate means. For every 5µg of vector DNA, 0.5U of CAP was added, and this reaction was then incubated in the presence of 1x CAP buffer at 50°C for 30 minutes. The CAP enzyme was then inactivated by the addition of Na_2-EDTA (5mM), SDS (0.5%) and 50µg/ml Proteinase K. After this mixture had been incubated at 56°C for 30 minutes, the DNA was purified according to Method 2.5.1.

2.6.4 Ligation

1) T4 DNA ligase (400,000U/ml) (NEB): 1/10 dilution with sterile Milli-Q water
2) 10x ligase buffer (NEB)
3) Sterile Milli-Q water

The ligation of homologous cohesive ends was carried out by mixing 100ng of vector DNA, xng insert DNA, 2µl 10x ligase buffer, 1µl diluted T4 DNA ligase and sterile Milli-Q water to make up the volume to 20µl. The reagents were mixed and then incubated at 37°C for 4 hours to ensure ligation. After the reaction was completed, the ligase could be inactivated by heating at 65°C for 10 minutes.

Notes:

1) The amount of insert DNA added was calculated to give 2-3 times as many insert cohesive ends as vector ends.
2) Several modifications from the described protocol were required to facilitate the ligation of blunt end termini. The reaction volume was halved (10µl), while the number of DNA termini were doubled thereby increasing the number of favourable interactions between vector and insert. The reaction was performed at room temperature for 6-16 hours in the presence of 80U of ligase.

2.6.5 Preparation of Competent Cells For Electroporation

1) LB broth
2) Sterile 10% (v/v) Glycerol
3) Sterile Milli-Q water (cold)

A flask containing 1L of LB was aseptically inoculated with 5ml of a fresh overnight culture of the desired *E. coli* strain. These cells were placed at 37°C with vigorous shaking until they had reached mid-log phase (OD_{600} of 0.5-1.0). Once at the appropriate cell density, the flask was transferred into a container of ice for 20 minutes. The salt-rich LB broth was later decanted from the pellet resulting from a centrifugation step (4000g for 15 minutes) carried out in a cold rotor. After the cells were resuspended in 1L of cold sterile water, the centrifugation was repeated, and once again the supernatant was discarded. Any salts still residing in the pellet were removed in the second wash, which was in 500ml of cold water. With the deionisation of the cell solution complete, the cells were resuspended in 20ml of cold 10% glycerol. This volume was reduced to 3ml of 10% glycerol after another centrifugation step (4000g for 15 minutes).

The cells, whose final concentration was approximately 1-3 x 10^{10} cells/ml, were then either used without delay, or snap-frozen by liquid nitrogen in 200µl aliquots. These frozen cells were then stored at -80°C where they remained competent for a period of at least 6 months.

2.6.6 Electroporation of *E. coli*

1) SOC: LB (autoclaved), then added 10mM $MgSO_4$, 10mM $MgCl_2$, 20mM glucose
2) Gene Pulser (Bio-Rad): set at 25µF, 2.5kV
3) Gene Controller (Bio-Rad): set at either 200Ω or 400Ω
4) Cuvettes (0.2cm)
5) LB plates containing antibiotic

The frozen competent cells were gently thawed and placed on ice along with the electroporation cuvettes and the safety chamber slide. After setting the Gene Pulser and Gene Controller (Bio-Rad) to the correct parameters, a mixture containing 2µl of the transforming DNA and 40µl of the competent cells was prepared in an Eppendorf, and placed on ice for 30 seconds. This mixture was then transferred to a chilled electroporation cuvette and the suspension shaken to the bottom. The cuvette, which was placed into the cold safety chamber slide was then slid into position between the contacts in the chamber, and a pulse passed through the suspension. Once

electroporated, the cells were immediately resuspended in 1ml of SOC medium, transferred to an Eppendorf and incubated with shaking at 37°C for 30 minutes. The cells were finally plated on selective medium to identify transformants.

Note:

1) The period between applying the pulse and transferring the cells to recovery medium is extremely important for resuscitating the *E. coli* transformants. Dower *et al.* (1988), have shown that if the recovery medium is added 10 minutes after the shock, less than 25% of the transformed cells survive.

2.6.7 Microdialysis of DNA Samples For Electroporation

Sometimes a DNA sample had an ionic strength that was too high for electroporation to occur without arcing. The most efficient and facile method of removing this excess salt was through microdialysis of the DNA sample.

1) Millipore 0.025µm membrane

A 0.025µm membrane was floated on top of a solution of 10% glycerol and gently agitated using a magnetic stirrer. Ensuring no liquid was on the uppermost surface, the DNA samples were transferred onto the membrane and allowed to dialyse for 30 minutes. The DNA was suitable for electroporation once it had been removed from the membrane using a pipette.

2.6.8 Blue/White Selection

1) 20mg/ml X-gal made up in dimethyl formamide

Using a pipette, xµl (see Note 1) of X-gal was added to each plate before the LB agar had solidified. The X-gal was distributed throughout the media by manual circular rotation of the plate. The cells were then plated and incubated overnight, thereby allowing the white recombinants to be easily distinguished from the blue non-recombinants.

Note:

1) pUC118 required only 30µl of X-gal per plate, whereas pUC21 needed 50µl per plate to enable the blue colour to be visualised.

2.7 POLYMERASE CHAIN REACTION (PCR)

2.7.1 Primers

The 20-mer exact match PCR primers were synthesised by the Separation Science Unit at Massey University, after having been checked for possible primer-dimer and self-hybridisation properties using the computer program Primers.

2.7.2 *In Vitro* Amplification of DNA by PCR

1) 2mM dNTP Mix
2) 10x $Mg2^{+}$-free Buffer (Promega)
3) *Taq* Polymerase (Promega)
4) Primers: 10μM working solution in sterile water
5) 25mM $MgCl_2$ (Promega)
6) Oil (Perkin-Elmer-Cetus)

A 25μl reaction volume was made up of 15μl sterile water, 1.5μl $MgCl_2$, 2.5μl 10x Buffer, 2.5μl 2mM dNTP mixture, 1.25μl of each primer and approximately 20ng of plasmid DNA. 1μl *Taq* polymerase was finally added after the mixture had been heated to 94°C. A single drop of light mineral oil was then overlaid to prevent evaporation of the sample during the repeated cycles of heating and cooling.

The PCR tube was inserted into the PCR machine and the parameters programmed (Table 2.4). Upon conclusion of the thermal cycling, a sample of the amplified DNA was withdrawn and analysed by gel electrophoresis (Method 2.5.3). If the DNA was sufficiently amplified, an extraction, using 30μl of chloroform was performed to remove the oil. Further purification could then be undertaken to remove spurious bands or excess primers.

Notes:

1) In order to optimise the PCR conditions, a number of control runs were set up and analysed for each set of primers used, at various temperatures and DNA concentrations.
2) A positive control of reliably amplifiable DNA was usually run in parallel with the samples of interest to guarantee the authenticity of any negative results.
3) A negative control consisting of everything except the template DNA was occasionally set up to check for freedom from DNA contamination.

Table 2.4. PCR Conditions

Cycle	Denaturation	Annealing	Polymerisation
Programme 1:			
First cycle	95°C for 1 minute	50°C for 2 minutes	72°C for 3 minutes
45 cycles	94°C for 1 minute	50°C for 2 minutes	72°C for 3 minutes
Last cycle	94°C for 1 minute	50°C for 2 minutes	72°C for 5 minutes
Programme 2:			
35 cycles	94°C for 1 minute	50°C for 1 minute	72°C for 1 minute

Soak file: 4°C for both Programme 1 and 2.

2.7.3 T-Tailing pUC118

Attempts to clone PCR products as blunt-ended fragments have been very inefficient, due to the template-independent terminal transferase activity of *Taq* polymerase which results in the addition of a single nucleotide at the 3' end of the fragment (Clark, 1988; Mole *et al.*, 1989). Clark (1988), found that this nucleotide is almost exclusively an adenosine due to the strong preference of the polymerase for dATP. By using *Taq* polymerase to add a single thymidine at the 3' end of each vector fragment, it is possible to create a cloning scheme which has the efficiency of a sticky end ligation (Marchuk *et al.*, 1991).

1) *Taq* polymerase (Promega)
2) 10x *Taq* polymerase buffer (Promega)
3) 2mM dTTP

A 100µl reaction volume consisted of 10µl 10x buffer, 0.6µg of *Sma*I-cut pUC118, 3µl *Taq* polymerase, 10µl dTTP and xµl sterile Milli-Q water. Once the solution had been thoroughly mixed, it was incubated at 70°C for 2 hours. The vector was ready for ligation after it had undergone a phenol/chloroform extraction (Method 2.5.1) and precipitation (Method 2.5.2).

Notes:

1) Vector self-ligation events were blocked by the 3' thymidine overhang.
2) Concatamerisation of the insert cannot occur due to the unphosphorylated 5' end contributed by the oligonucleotide primer, as well as the 3' adenosine overhang added by the *Taq* polymerase during the PCR reaction.

2.8 PROBING

The rapid screening of hundreds of bacterial colonies for plasmids containing a desirable insert was carried out using radioactive probing techniques. Due to the small size of the DNA fragments in this investigation, the probe was end-labelled instead of the more commonly used method of random primed labelling.

2.8.1 End-Labelling of Oligonucleotide Probes

1) T4 Polynucleotide kinase (BRL)
2) T4 Polynucleotide kinase buffer (BRL)
3) γ-^{32}P-ATP
4) TNES: 10mM NaCl; 10mM Tris, pH 8.0; 2mM EDTA; 0.1% SDS
5) Salmon sperm DNA (4mg/ml)

The single-stranded oligonucleotide was labelled by placing 50pmoles in a reaction mixture containing 1µl of T4 polynucleotide kinase, 2µl of 10x buffer, 4µl of γ-^{32}P-ATP and xµl of sterile Milli-Q water to raise the final volume of the mixture to 20µl. Once thoroughly mixed, the Eppendorf containing the reaction volume was incubated at 37°C for one hour, allowing the polynucleotide kinase to radiolabel the oligo. Before the oligo probe was ready for use, 200µl TNES and 100µl salmon sperm DNA were added to the probe mixture.

2.8.2 Preparation of DNA on Nitrocellulose Filters

1) Denaturation Solution: 0.5M NaOH; 1.5M NaCl
2) Neutralisation Solution: 1M Tris-HCl, pH 7.5; 1.5M NaCl
3) 20x SSC: 3M NaCl; 0.3M trisodium citrate; pH 7.0.

2.8.2.1 Colony Streaks

The potential bacterial recombinants were streaked onto gridded nitrocellulose filters and allowed to grow for 5 hours at 37°C by placing the filters on LB plates containing antibiotic. The *E. coli* cells were then lysed by placing the filters (with the colonies uppermost) onto blotting paper soaked in the denaturation solution for 5 minutes, ensuring no liquid was permitted to come into contact with the upper surface of the filters. After briefly blotting the filters they were placed onto fresh blotting paper soaked in neutralisation solution for 5 minutes. The filters were agitated in 2x SSC for 5 minutes before being laid on dry blotting paper. The liberated DNA was finally bound to the nitrocellulose by baking the filters for 2 hours under vacuum at 80°C.

2.8.2.2 Colony Lifts

Approximately 1000 colonies per antibiotic-containing plate were allowed to grow until their diameter was 1-2 mm. The plates were chilled at 4°C for one hour and then overlaid with nitrocellulose filters for 5 minutes. The transferred cells were later laid on blotting paper soaked in denaturation solution for 15 minutes to separate the DNA strands. After denaturation, the filters were placed on blotting paper soaked in neutralisation solution. Finally, the single-stranded DNA was immobilised on the nitrocellulose filters in their relative positions by heating at 80°C under vacuum for one hour.

2.8.3 Hybridisation

1) Prehybridisation Buffer: 6xSSC; 0.5% (w/v) SDS; 0.05% (w/v) NaPyrophosphate and 10x Denhardts solution i.e. 0.1% (w/v) each of Ficoll, polyvinylpyrrolidine (PVP) and bovine serum albumin (BSA).
2) Hybridisation Buffer: 6xSSC; 10x Denhardts solution and 0.05% (w/v) NaPyrophosphate.
3) Wash Solution: 6xSSC and 0.05% NaPyrophosphate

The filters prepared by colony hybridisation were sealed flat in a container with at least 20ml of pre-heated prehybridisation solution per 100cm^2 of membrane. Prehybridisation was performed for at least one hour at 68°C in a shaking waterbath. The filters were subsequently transferred to a hybridisation tube containing hybridisation solution (3ml/100cm^2 membrane) and the freshly labelled probe DNA (see Method 2.8.1). The tube was later incubated overnight in a 68°C hybridisation oven which had vertically rotating holders. The hybridisation solution was then discarded and the membranes

washed 3x 10 minutes with the wash solution at room temperature. One further 2 minute wash with solution pre-heated to 68°C was performed to ensure any nonspecific binding that had occurred was dislodged. The filters were later exposed to Kodak AR film for 16 hours at -20°C in the presence of an intensifying sheet.

2.9 DOUBLE-STRANDED SEQUENCING

The ultimate confirmation of a controlled DNA manipulation is to obtain the sequence of the product. This was accomplished through the double-stranded sequencing of clones ligated into pUC vectors using Sanger's dideoxy chain termination method (Sanger *et al.*, 1977).

2.9.1 **Purification of Plasmid DNA for Double-Stranded Sequencing** (Saunders and Burke, 1990).

1) TE (10/1)
2) CsCl
3) 10% (w/v) Ethidium Bromide
4) Isopropanol equilibriated with 20x SSC
5) 3M NaOAc
6) Cold isopropanol (-20°C)
7) Sterile Milli-Q water
8) 70% (v/v) ethanol

The plasmid DNA to be sequenced was extracted from an overnight culture using either Magic Minipreps (Method 2.4.3) or the rapid boil method (Method 2.4.2). The DNA, resuspended in 100µl TE, was shaken thoroughly with 100mg of CsCl before 10µl of 10% ethidium bromide was mixed in. This solution was spun in a microfuge for 15 minutes to separate the DNA from the precipitated complexes formed between the ethidium bromide and the cellular debris. The DNA was then carefully transferred into a new Eppendorf tube leaving the floating material behind. The solution was later extracted three times using isopropanol equilibrated with 20x SSC. If the supernatant was still pink at this stage, further extractions were undertaken. The clean plasmid DNA was then precipitated by adding 400µl TE, 50µl 3M NaOAc and 720µl isopropanol. This mixture was mixed through inversion and placed at -70°C for 20 minutes. After washing in 70% ethanol the dried pellet was redissolved in 20µl of sterile Milli-Q water. Typically 2-3µl of this solution was used in the sequencing reactions.

2.9.2 Preparation of Single-Stranded Plasmid DNA using Alkaline-Denaturation

1) 1M NaOH
2) 250mM Na_2-EDTA
3) 3M NaOAc
4) 95% (v/v) cold ethanol
5) 70% (v/v) ethanol
6) Milli-Q water

Incubation of the DNA to be sequenced (37°C for 30 minutes) in the presence of 0.2M NaOH and 0.2mM EDTA brought about the separation of the plasmid DNA strands into their single-stranded form. The mixture was precipitated according to Method 2.5.2 and finally resuspended in 40µl of sterile Milli-Q water.

2.9.3 Sequencing Reactions

1) Sequenase Version 2.0 (USB)
2) Sterile Milli-Q water

The reactions were performed in accordance with the brief protocol provided by the manufacturers.

Notes:

1) dGTP Termination Mixtures were preferred to dITP mixtures.
2) Termination reactions were carried out in microtitre dishes instead of Eppendorf tubes, due to the increased ease of handling of the former.

2.9.4 Setting Up Sequencing Gels

1) 10x Sequencing Buffer: 1.35M Tris; 450mM Boric acid; 25mM Na_2-EDTA.
2) 40% Acrylamide/bis-acrylamide (38:2) stock
3) Acrylamide/urea mix: 288g urea was added to 90ml 40% acrylamide/bis-acrylamide stock, and warmed with stirring to dissolve. One teaspoon of Amberlite MB-3 (Sigma A17518) was added and stirred gently for 30 minutes, and then filtered out using a sintered glass funnel. 60ml of 10x Sequencing Buffer was added before distilled water was used to make the final volume up to 600ml.
4) TEMED (N,N,N',N'-tetramethylethylenediamine)
5) 5% (w/v) Ammonium persulphate
6) 2% (v/v) Silane in carbon tetrachloride

The glass plates and spacers were washed in a warm detergent solution, rinsed thoroughly with tap water then finally rinsed with deionised water. After drying, one of the surfaces of one plate was treated with silane solution to prevent the gel from tightly adhering after electrophoresis. The plate was rinsed in deionised water, dried, and rested face down on the spacers which were positioned on the side edges of the second plate. The sides and bottom of the glass plates were bound together using gel-sealing tape to provide a watertight seal. The acrylamide/urea gel was then prepared by placing sequentially in a 250ml side-arm flask: 100ml acrylamide/urea mix, 60µl TEMED and 1.2ml 5% ammonium persulphate. After swirling the solution briefly, the gel mixture was degased for several minutes and immediately poured between the taped glass plates. Any air bubbles were removed (using a thin spacer or comb) as these can interfere with the migration of the DNA samples. At the completion of pouring, the flat sides of two shark's tooth combs were inserted approximately 0.5cm into the gel solution to obtain a level surface upon gel verticalisation. These combs were clamped in position with bulldog clips, and the gel was allowed to polymerise for 30 minutes while placed on a slight incline.

Note:

1) If the gel was to be stored overnight, the top of the gel was wrapped in Gladwrap to prevent dehydration.

2.9.5 Loading and Running Sequencing Gels

The bulldog clips and combs were removed from the plates and the tape peeled from the bottom of the gel mould. Foam rubber inserts were placed onto the exposed region of the spacers, and the gel fixed into the sequencing apparatus. 1x Sequencing Buffer was poured into both the upper and the lower reservoirs of the sequencing apparatus and the combs extracted from the gel. Excess urea and polyacrylamide were washed from the top of the gel before reinserting the shark's tooth comb with its teeth just sticking into the loading surface of the gel. In order to prevent renaturation of the single-stranded DNA during the primary stages of electrophoresis, it was necessary to run the gel at 65W for 30 minutes prior to loading the sequencing reactions. Before loading the DNA samples, the microtitre plate in which they were all sitting was heated at 80°C for 2 minutes. 3.5µl from each of the 4 sequencing reactions of each DNA sample, was loaded into adjacent lanes of the gel in the order GATC. The sequencing gel was then electrophoresed at 65W for the appropriate length of time.

Note:

1) The time for a short run was usually 1.5 hours, while the time for the long run was usually 6 hours.

2.9.6 Autoradiography of Sequencing Gels

1) Kodak XAR-2 or XAR-5 x-ray film

After the electrophoresis run had been stopped, the gel mould was removed from the sequencing stand and laid down on a flat surface. The gel sealing tape was peeled off and the spacers extracted prior to the two plates being prised carefully apart using the end of a metal spatula. The unsilanised plate (to which the gel remained adherent) was laid horizontally such that the gel was uppermost. Any folds or distortions in the gel were gently manipulated out using gloved fingers, before a piece of 3MM paper was superimposed on top. Gentle pressure was applied to this paper before a piece of filter paper was placed on top. The 3MM paper was then pulled upwards from one corner, resulting in the gel being released from the glass plate. After trimming the paper to a size compatible with the gel dryer, the gel was wrapped with Gladwrap and subsequently dried for 1 hour under vacuum in a gel dryer set at 80°C. Upon removal of the Gladwrap, the gel was exposed to the Kodak x-ray film for 16-24 hours at room temperature. The film was devoloped, fixed and washed using standard photographic techniques, before being read and analysed.

2.10 MEASUREMENT OF *nod* GENE ACTIVITY

The phenolic flavonoid compounds that induce the expression of the *nod*ABC genes are contained within the exudate that can be extracted from leguminous seeds. Due to the specificity displayed by the *Rhizobium* for a particular legume(s), it is important to use seeds whose flavonoid inducers are recognised by the bacterial strain.

2.10.1 Preparation of Seed Exudates For *nod* Gene Induction (McSweeney, 1987; Mulligan and Long, 1985).

1) 95% (w/v) Na hypochlorite
2) *Lotus pedunculatus* seeds (ex AgResearch CRI)
3) 95% (v/v) Ethanol

20g of *L. pedunculatus* seeds were soaked in 95% ethanol for 30 minutes. The ethanol was subsequently decanted and replaced with a similar volume of Na hypochlorite for the same period of time. The bleach was then decanted, and the seeds washed 5 times with distilled water. The seeds were soaked overnight in a volume of 150ml in the dark at room temperature. The supernatant was centrifuged at 27,000g for 10 minutes to

remove the insoluble cellular material, and later filter-sterilised to exclude contaminants. The sterile supernatant was stored at -20°C. To obtain greater yields of seed exudate the insoluble material from the centrifugation was soaked in 75ml of water and incubated overnight as before. The following day, the centrifugation step was repeated and the filter-sterilised supernatant was later pooled with the exudate obtained from the previous day's extraction.

2.10.2 *Rhizobium nod* Gene Induction

1) S40 medium
2) TY medium
3) 10x Na phosphate buffer: 0.6M Na_2HPO_4 and 0.4M NaH_2PO_4 adjusted to pH 7.0.
4) Z buffer: 10mM KCl; 1mM $MgSO_4.7H_2O$; 50mM β-mercaptoethanol in 1x Na phosphate buffer, pH 7.0.

The *R. loti* culture containing the *nod-lacZ* and *nod-gus* fusions was allowed to grow at 30°C with shaking for 18 hours. The cells were harvested by centrifugation (speed 5 for 5 minutes in Heraeus Sepatech Megafuge 1.0) and resuspended in an equal volume of S40 medium before resuming the incubation for a further 8 hours.

The addition of 250µl of filter-sterilised seed exudate was preceded by the subculturing of the cells into fresh S40 medium (1/12.5 dilution into a final volume of 2.5ml). After another overnight incubation at 30°C the cells were harvested again (speed 5 for 10 minutes in Heraeus Sepatech Megafuge 1.0) and resuspended in an equal volume of Z buffer.

2.10.3 β-Galactosidase Assay (modified method of Miller, 1972)

1) 1x Na phoshate buffer (see 2.10.2)
2) Z buffer (see 2.10.2)
3) o-nitrophenyl-β-D-galactosidase (ONPG): 4mg/ml in 1x Na phosphate buffer
4) 0.1% (w/v) SDS
5) 1M Na_2CO_3
6) Chloroform

The induced *Rhizobium* culture, suspended in Z buffer, was diluted twofold (1ml final volume) with fresh Z buffer, before 2 drops of chloroform and 1 drop of 0.1% SDS were added. The mixture was vortexed and then incubated for 5 minutes at 28°C. 200µl of ONPG was later added to the mixture, and the time taken for the sample to change from

colourless to pale yellow was recorded. The reaction was stopped by adding 450µl of Na_2CO_3. A 5 minute centrifugation in an Eppendorf, followed by spectrophotometric measurement at OD_{420}, made it possible to ascertain the activity of the β-galactosidase,

$$ie \quad Units = \frac{1000 \; x \; A_{420}}{A_{600} \; x \; volume \; x \; time}$$

2.10.4 β-Glucuronidase Assay (Jefferson *et al.*, 1986; Jefferson, 1987)

1) Z buffer (see 2.10.2)
2) 2mM p-nitrophenyl β-D-glucuronide (PNPG) in Z buffer, plus 0.1% (v/v) Triton X-100.
3) 1M Na_2CO_3
4) Chloroform
5) 0.1% (w/v) SDS

After induction with seed exudate and resuspension in 0.5ml Z buffer (see Method 2.10.2), the *R. loti* cells were permeabilised by the addition of 1 drop of SDS and 1 drop of chloroform followed by thorough vortexing. A 5 minute incubation at 37°C allowed the β-glucuronidase to diffuse out of the cell where it could be conveniently assayed. 0.5ml PNPG was then added and the mixture left at 37°C for one hour, before 400µl of Na_2CO_3 was added to stop the reaction. The cells were later pelleted and the A_{415} measured to determine the amount of enzyme present.

$$Units \; of \; GUS = \frac{1000 \; x \; A_{415}}{A_{600} \; x \; volume \; x \; time}$$

2.11 NON-DENATURING POLYACRYLAMIDE GEL ELECTROPHORESIS

Separation of very small DNA fragments cannot be achieved using agarose gel electrophoresis. Significant improvements in the resolution of these fragments can be provided by using polyacrylamide gels.

1) 10x and 1x TBE buffer, pH 8.0
2) 29:1 (wt/wt) acrylamide/bisacrylamide
3) TEMED
4) 10% (w/v) ammonium persulphate
5) Protean II vertical gel aparatus

The gel casting apparatus was assembled according to the description in the Bio-Rad Protean II Instruction Manual. After these glass plates had been set up, the 20% acrylamide gel mixture was prepared by adding 5ml 10x TBE buffer, 38.3ml 29:1 acrylamide/bisacrylamide and 11.7ml water to a side-arm flask. This solution was thoroughly shaken to ensure complete mixing had occurred. Once the mixture had been degassed under vacuum, 25µl TEMED and 250µl 10% ammonium persulphate were added. The acrylamide gel mix was poured between the glass plates and the appropriate comb inserted. Polymerisation of the gel was complete after 30 minutes, and at this stage the comb was removed. The plates were then attached to the vertical Protean II electrophoresis tank before being immersed in 1x TBE such that the top and bottom of the gel was covered. A syringe was used to remove all the residual bubbles from the wells. This left a flat surface onto which the DNA samples were subsequently loaded. The gel was run at 100V until the desired resolution had been obtained. To visualise the electrophoresed DNA fragments, the gel was stained in ethidium bromide (25µl/ml) for 30 minutes, destained in water for 30 minutes and then placed under short wave length UV light for photography.

2.12 ECKHARDT GEL ELECTROPHORESIS OF *Rhizobium* PLASMIDS (Eckhardt, 1978; Harte, 1991)

Eckhardt gel electrophoresis is a technique whereby large extra chromosomal elements can be extracted from prokaryotic cells. The gentle *in-situ* lysis of log-phase cells reduces the amount of hydrodynamic shear imposed on the DNA and therefore minimises plasmid fragmentation.

1) TBE Buffer
2) 10/1 TE Buffer
3) Proteinase K
4) Lysozyme
5) Solution 1 (in TBE Buffer): 10% (w/v) Ficoll type 400; 0.05% (w/v) Bromophenol Blue; 0.3 U/ml RNAse A; 2mg/ml lysozyme just prior to use.
6) Solution 2 (in TBE Buffer): 10% (w/v) Ficoll type 400; 0.2% w/v) SDS. Add 4.0 mg/ml Proteinase K.
7) Solution 3 (in TBE Buffer): 5% (w/v) Ficoll type 400; 0.2% (w/v) SDS.

The *Rhizobium* culture of interest was inoculated into 5ml of TY broth containing appropriate antibiotics and incubated at 28°C for 48 hours with shaking. A 200µl aliquot was taken from this saturated solution and inoculated into 5ml of fresh TY broth containing the same selective antibiotics. The tube was subsequently placed with

shaking at 28°C for 16 hours to provide a source of log-phase *Rhizobium* cells. 500ml of this overnight culture was pelleted in a microfuge. Using a pipette, the supernatant was carefully removed, and the pellet resuspended completely in 1ml of TE buffer. Once again the cells were pelleted and the supernatant discarded. The cell pellet was gently resuspended in Solution 1 and then immediately transferred to a blot-dried well in a 0.7% agarose gel. Incubation at room temperature for 15 minutes enabled the lysozyme to digest the cell wall of the bacteria. 20µl of Solution 2 was then combined with each cell suspension and incubated at room temperature for a further 15 minutes. Finally, the cell suspension was overlayed with Solution 3 to prevent the DNA from diffusing away. The gel was subsequently electrophoresed at 45V for 40 hours at 4°C.

Notes:

1) The Eckhardt solutions should be stored at 4°C but equilibrated to room temperature before using.
2) The agarose gel should be poured approximately 4 hours before the cell suspension is loaded. This allows for complete-setting to occur and results in improved resolution of plasmid bands.

CHAPTER 3 - ELECTROPORATION OF *Rhizobium loti*

The efficient introduction of DNA into bacteria is a technique of great practical importance in molecular biology. Indeed, research into the genetics of numerous bacterial species is hindered by the lack of a convenient DNA transfer system. In many cases only conjugation or protoplast transformation systems are available. These methods can be both tedious and time-consuming.

Before contemplating the large-scale *in vitro* construction of assayable plasmids in any organism, it is first necessary to devise a simple and efficient strategy for introducing DNA into that organism. Historically, the triparental mating procedure developed by Ditta *et al.* (1980) has been the preferred method for accomplishing this objective in *Rhizobium*. Using this procedure, a plasmid that is not self-transmissible becomes mobilised into a recipient bacterium through the action of the Tra genes carried on a helper plasmid from a third strain. While this conjugation method is reasonably efficient, purification of the desired transformants is extremely laborious and can take several weeks. To overcome these inherent drawbacks, many researchers are utilising new transformation techniques such as electroporation. Indeed, the number of bacteria and fungi that have been successfully electrotransformed is rapidly increasing. However, one bacterial genus that has not been extensively examined in terms of its potential for electroporation is *Rhizobium*. The only *Rhizobium* species that has reportedly been electrotransformed is *R. meliloti* (Bio-Rad Booklet). Despite several attempts, *R. loti* had not yet been successfully transformed in this manner (Terzaghi, pers. comm.).

It was the intention of this preliminary investigation to develop a protocol whereby exogenous DNA could be electroporated into *R. loti* cells stably and efficiently. To achieve this aim, a number of bacterial electroporation methods were examined. As the method outlining electrotransformation of *R. meliloti* was unpublished, the most relevant protocols were considered to be those describing the electroporation of the related genera *Bradyrhizobium* and *Agrobacterium* (Mattanovich *et al.*, 1989; Guerinot *et al.*, 1990; Mersereau *et al.*, 1990; Nagel *et al.*, 1990). Through comparisons of these methods, the outline of a potential protocol was designed for *R. loti*. To complement these fundamental features, a series of experiments were undertaken to optimise the electroporation conditions for this particular species. This chapter outlines these optimisation experiments along with several other considerations that may have altered the electroporation efficiency. The final composite protocol is presented at the end of the chapter. It was this method that was used to routinely transform DNA into *R. loti*.

3.1 INTRODUCTION TO ELECTROPORATION

In order to appreciate the results of the electroporation experiments, it is important to understand the fundamental physics behind the technique.

The electric pulse generated by most electroporation equipment (such as the Bio-Rad Gene Pulser) results from the discharge of a variable capacitor directed through a sample placed between two electrodes. The pulse has the characteristics of an exponential decay given by the expression:

$$\textit{Equation 1:}\quad V_t = V_0 \cdot [e^{-(t/\tau)}]$$

where V_t = the voltage at time, t; V_0 = the initial voltage, and τ = the time constant.

Prior to the discharge, the potential difference across the two electrodes is increased rapidly to a peak value that can be appropriately controlled. τ (the time constant) provides a convenient parameter related to the duration of the pulse, and according to Equation 1, is the time required for the voltage to decline to 1/e (approximately 37%) of its peak value.

Cells are permeabilised when the potential drop across the length of the cell reaches some threshold value. This potential drop is determined by the strength of the electric field (E) and the size of the cells. The field strength is related to the potential difference (V) between the electrodes by:

$$\textit{Equation 2:}\quad E = V/d$$

where 'd' represents the separation of the electrodes, and is usually fixed during the pulse.

By substituting V from Equation 2 for V_t in Equation 1, it can be observed that the electric field follows an exponential decay as well.

3.2 FACTORS AFFECTING ELECTROPORATION EFFICIENCY

3.2.1 Growth Phase of the *R. loti* Cells

The particular stage at which the bacteria are harvested plays an important role in obtaining electrocompetent cells. Miller *et al.* (1988), using *Campylobacter jejuni*, and Powell *et al.* (1988), using *Streptococcus lactis* have shown that cells taken at early- to mid-log phase are the best candidates for electroporation. The slower growing bacteria gathered at late-log phase tend to be relatively unreceptive to this form of transformation. It is therefore extremely important to obtain the bacteria at approximately the optimum growth stage.

In order to ascertain the OD_{600}, at which the bacterial cells are in a suitable growth phase, it is necessary to know the growth kinetics of the bacterial species. As the growth curve for *R. loti* had not hitherto been determined, a preliminary investigation was necessary. Determination of the *R. loti* growth pattern was obtained through spectrophotometric means as described below.

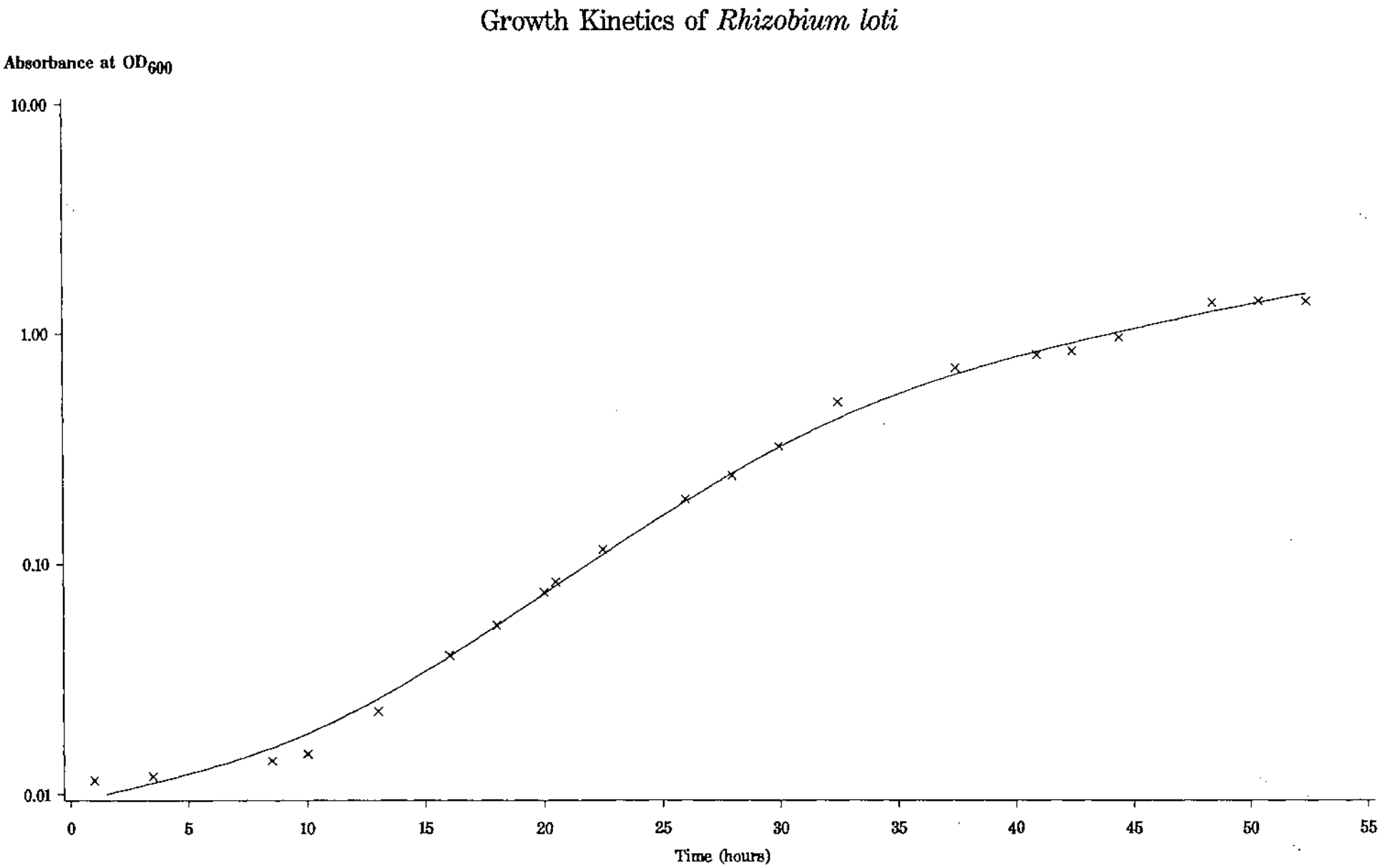
A saturated 5ml culture of *R. loti* strain PN184 was transferred into 500ml of TY containing streptomycin (TYStrep). The flask was subsequently grown at 30°C with shaking to facilitate cellular growth. 1ml aliquots were withdrawn from the flask every few hours and monitored spectrophotometrically at OD_{600} to follow the rhizobial growth. The experiment was repeated five times and a consensus growth curve was attained (see Figure 3.1).

The *R. loti* growth curve illustrated above was used to estimate the OD_{600} at which the cells were passing through early-mid log phase. As mentioned earlier, this growth state represents the phase where the cells are presumed to be at their electrocompetent optimum. In this case, the *R. loti* cells were harvested when the OD_{600} of the growing cells was between 0.45 and 0.65. Cell preparations whose density was less than OD_{600} 0.45 could also have been efficiently transformed but the corresponding low yield of competent cells made harvesting at this stage impractical. Occasionally, cells were harvested at optical densities higher than those considered optimum. These cellular preparations could still be electrotransformed albeit with a diminished efficiency (data not presented). Therefore, while the most favourable cellular optical densities are between 0.45 and 0.65, OD_{600} values of up to 0.80 can still yield reasonably high electroporation efficiencies.

Figure 3.1.

Growth curve of *R. loti*. The curve was obtained by regularly measuring the cell density of a *R. loti* culture inoculated into 500ml of TY containing streptomycin. Growth was monitored spectrophotometrically at OD_{600}.

Figure 3.1


Growth Kinetics of *Rhizobium loti*
Absorbance at OD_{600}
10.00
1.00
0.10
0.01
0
5
10
15
20
25
30
35
40
45
50
55
Time (hours)

3.2.2 Removal of Extracellular Materials From *R. loti*

Many bacteria have extracellular capsules of various compositions that may interfere with DNA uptake. Such capsules can often be removed using chemical or enzymatic means, although this is not always necessary. Indeed, Shigekawa and Dower (1988), have reported that removal of the large muco-polysaccharide capsule of *Klebsiella pneumoniae* did not enhance the efficiency of electroporation at all.

To investigate the effects of the extracellular polysaccharide material on the transformation frequency of *R. loti*, the cells were resuspended in 10/1 TE buffer containing either 0.1% or 0.01% N-laurylsarcosine and then thoroughly vortexed. This treatment removed the excreted exopolysaccharides, theoretically providing the DNA with a less obtructed passage into the cell. The stripping of the *R. loti* glycoprotein also prevents the cells from clumping together, allowing more DNA to come into contact with a greater membrane surface area.

Using identical electroporation conditions, the washed cells were later transformed with selectable plasmid DNA. Analyses of the transformants indicated whether the efficiency of transformation was dependent on the removal of the exopolysaccharides of the *R. loti*. The results are presented in Table 3.1.

The experimental data tabulated above shows conclusively that the removal of extracellular materials using N-laurylsarcosine had no effect on electroporation efficiency. Accordingly, further preparations of competent *R. loti* cells were not stripped of their exopolysaccharide content.

3.2.3 DNA Concentration

Dower *et al.* (1988), examined the effect of DNA concentration on the electroporation efficiency of *E. coli*. Their investigation found that the recovery of transformants increased linearly with DNA concentration over a wide range (10pg/ml to 7.5μg/ml). Hence, it was concluded that the DNA concentration was important in determining the probability of any given cell becoming transformed. This hypothesis was confirmed by the observation that the yield of transformants increases when a higher cell concentration was used in the presence of a fixed concentration of DNA. On a molecular scale this can be explained by the increased number of interactions between the DNA and the cells when the concentration of both is high, resulting in an increased transformation frequency.

The observed linear relationship between DNA concentration and electrotransformation frequency was assumed to be universal amongst bacteria. Therefore, the optimal DNA concentration was not experimentally examined in *R. loti*. In general, the amount of aqueous DNA added to the electroporation mixture was limited only by the ionic strength of this solution. Typically, a final concentration of 5µg DNA/ml competent cells was routinely electroporated in this investigation.

3.2.4 Pre-Pulse Incubation

For efficient electrotransformation to occur, the cells of some organisms must be incubated in the presence of the transforming DNA prior to the pulse. This pre-shock incubation period allows the DNA to bind to the competent cells, thereby facilitating transformation. While the attachment of DNA to cells is a prerequisite for transformation in some bacteria, it is unnecessary in others. Indeed, Dower *et al.* (1988), demonstrated that the pre-shock incubation time had no significant effect on the frequency of transformation in *E. coli*. In fact, lengthy pre-pulse time delays may even be detrimental for transformation if, for example, nucleases are secreted from the bacteria into the cell suspension (Chassey and Flickinger, 1987). The effect of varying the pre-pulse incubation time on transformation had to be determined in *R. loti*. If DNA binding was a critical feature of the electrotransformation of this species, the electroporation efficiencies would be expected to increase as the pre-shock incubation time was lengthened. The results obtained in this investigation are shown in Table 3.2.

These data indicate that the binding of exogenous DNA to *R. loti* cells does not have a major impact on the electroporation efficiency of this species. However, the results did show that a slight increase in the efficiency of transformation was obtained when the DNA was incubated with the competent cells for two minutes prior to pulsing. While this may simply be due to experimental uncertainty, it was regarded as the optimum incubation time nonetheless.

The slightly reduced electroporation efficiency of the five minute incubation probably reflects experimental variation rather than increased nuclease activity. In fact, the overall uniformity of these results suggests that DNA degradation is not an important factor in the electroporation of *R. loti*.

3.2.5 Electrical Variables Affecting Electroporation

By altering specific electrical parameters it is possible to modify the efficiency of electroporation. The two most important electrical variables affecting this efficiency are the initial field strength and the time constant, τ. The field strength may be adjusted by altering the potential to which the capacitor is charged, or, conversely, by varying the distance between the two electrodes. This latter variation can usually be accomplished by using electroporation cuvettes of different dimensions. The effect of modifying the electric field was demonstrated by Miller *et al.*, (1988). Working with *Campylobacter*, this group showed that pulses of a few milliseconds duration produced no detectable transformants at field strengths of less than 5kV/cm, but when the electric field was raised incrementally to 13kV/cm the efficiency of transformation sharply increased.

The time constant, τ, can be expressed as the product of capacitance, C, afforded by the Gene Pulser and the resistance of the circuit, R, (see Equation 3).

Equation 3: $\tau = RC$

According to this equation, τ can be lengthened by increasing the capacitance of the variable capacitor, or by increasing the resistance of the circuit. For bacterial preparations, this circuit consists of the DNA sample in parallel with the resistors of the Pulse Controller. As the ionic strength of the DNA sample is very low, its resistance is considerably higher than that of the parallel resistors. Hence, the parallel resistance of the Pulse Controller determines the total resistance of the circuit. Control of the time constant can therefore be mediated through selection of an appropriate parallel resistance.

There is a direct relationship between the field strength and the time constant on transformation efficiency (Miller *et al.*, 1988). Compensating combinations of field strength and pulse length produced similar efficiencies of transformation and cell death. This compensatory effect does, however, have certain limitations. For example, when the field strength is very low (2.0kV/cm) and very long pulses (900ms) were applied by Dower *et al.* (1988), transformation efficiencies were vastly reduced ($>10^6$/μg DNA).

The effect of altering the electrical variables in a standard electroporation experiment was examined in *R. loti*. Due to the voltage limitations of the electroporation apparatus, only the time constant was experimentally modified. As mentioned earlier, this parameter could be varied by altering the parallel resistance of the Pulse Controller. The results are presented in Table 3.3.

The conclusion drawn from this data is that the optimum time constant (12.8s) is delivered when the parallel resistance is set at 600Ω. Resistances below 600Ω resulted in a reduced transformation efficiency probably because the correspondingly low time constants were not sufficient to 'open' up the transient 'holes' in the *R. loti* membrane. At the other extreme, time constants generated by an 800Ω parallel resistance were so high that many of the cells could not be resuscitated. The increased cell mortality was thought to be responsible for the decreased efficiency of transformation. It can therefore be deduced that the intermediate value of 600Ω produces a time constant that is high enough to efficiently transform *R. loti*, yet low enough to restrict cell mortality.

3.3 PROTOCOL FOR THE ELECTROPORATION OF *Rhizobium loti*

Having analysed all of the major factors that effect the efficiency of electroporation, it was possible to determine a method whose conditions and parameters were fully optimised. This protocol, detailed below, was routinely used for the efficient transformation of *R. loti*.

3.3.1 Preparation of Competent *R. loti* Cells

1) TY broth containing streptomycin (TYStrep)
2) Sterile 10% (v/v) glycerol
3) Sterile Milli-Q water

A single colony of *R. loti* was inoculated into 5ml of TYStrep and grown at 30°C for 2 days with shaking. The resulting saturated solution was aseptically inoculated into a flask containing 1L of TYStrep. These cells were placed at 30°C with vigorous shaking until the OD_{600} was between 0.45 and 0.65 (approximately 30 hours). At this stage the flask was transferred into an ice bucket for 30 minutes.

Table 3.1 The cells were washed in differing strengths of detergent to remove the exopolysaccharides from the *Rhizobium* cells. The results indicate that the washing has affected neither the time constant nor the efficiency of electroporation.

Table 3.2 Variations in the length of pre-pulse incubation time were examined for *R. loti* cells. While the time constant is unaffected by this parameter, minor fluctuations in the electroporation efficiency were noted. The optimal incubation time was 2 minutes.

Table 3.3 The time constant was increased by raising the parallel resistance of the circuit. The result of this increase was an improved efficiency of electroporation. The optimal time constant was achieved when the parallel resistance was set at 600Ω.

Table 3.1 **Effect of Washing Exopolysaccharides from Rhizobial Competent Cells**

Strength of Detergent (% N-laurylsarcosine)	Time Constant (τ) (sec)	Electroporation Efficiency (transformants/ml cells/μg DNA)
0	12.8 ± 0.6	(2.4 ± 0.4) x 10^{-6}
0.01	12.7 ± 0.2	(2.2 ± 0.4) x 10^{-6}
0.1	13.1 ± 0.2	(2.6 ± 0.3) x 10^{-6}

Table 3.2 **Effect of Variable Pre-Pulse Incubation Times on Electroporation Efficiency**

Time (minutes)	Time Constant (τ) (sec)	Electroporation Efficiency (transformants / ml cells / μg DNA)
0	12.9 ± 0.3	(1.7 ± 0.1) x 10^{-6}
1	12.6 ± 0.1	(2.0 ± 0.3) x 10^{-6}
2	12.8 ± 0.6	(2.4 ± 0.4) x 10^{-6}
5	13.0 ± 0.2	(2.1 ± 0.1) x 10^{-6}

Table 3.3 **Effect of Time Constant on Electroporation Efficiency**

Parallel Resistance (Ω)	Time constant (τ) (sec)	Electroporation Efficiency (transformants / ml cells / μg DNA)
200	4.6 ± 0.1	(0.8 ± 0.3) x 10^{-6}
400	8.8 ± 0.2	(1.6 ± 0.3) x 10^{-6}
600	12.8 ± 0.6	(2.4 ± 0.4) x 10^{-6}
800	16.4 ± 0.4	(2.1 ± 0.1) x 10^{-6}

The cells were then pelleted by centrifugation (4000g for 15 minutes) and the supernatant discarded. The *R. loti* cells were later resuspended in 1L of cold water before being pelleted again. Washing of these cells was repeated 4 more times using 500ml of cold water each time. This ensured the complete deionisation of the cell solution and may also have removed excess mucilaginous material from the rhizobial cells. The pellet from this final wash was resuspended in 20ml of cold 10% glycerol. This volume was reduced to 2ml of 10% glycerol after another centrifugation step. The competent cells were then either electroporated immediately or stored at -80°C. These *R. loti* cells could be used at least 4 months after freezing, though their competency was markedly reduced.

3.3.2 Electroporation of *R. loti*

1) TY Broth
2) Gene Pulser (Bio-Rad): set at 25μF, 2.5kV
3) Gene Controller (Bio-Rad): set at 600Ω
4) Cuvettes (0.2cm)
5) TY plates containing antibiotics

A mixture containing 200ng of plasmid DNA and 40μl of competent cells was incubated on ice for 2 minutes. This solution was then transferred into a pre-chilled electroporation cuvette. An electric pulse was passed through the cell suspension after the cuvette had been placed into a cold safety chamber slide and slid between the contacts. 1ml of TY broth was promptly added to the transformation mixture which was later transferred to an Eppendorf tube. The cells were then incubated with shaking at 30°C for 2 hours before being plated on selective media to identify transformants.

Chapter 4 - CONSTRUCTION OF A BIDIRECTIONAL REPORTER VECTOR

4.1 INTRODUCTION

The control and regulation of the putative *R. loti nod*-box (Collins-Emerson, 1991) has been analysed by Teo (1990). This study was primarily undertaken to ascertain whether the constitutive and inducible regulatory functions of the *nod*ACIJ *nod*-box are orientation-dependent (c.f. *R. leguminosarum*). Teo (1990), adopted a strategy which required cloning DNA fragments that included the *nod*-box, into the reporter vector pMP220 (Spaink *et al.*, 1987). To monitor activity in the opposite direction, the same fragments were also cloned in the second orientation. Both sets of clones were later assayed and the mode of regulation deduced.

In order to facilitate the study of the *nod*-box regulatory functions, it was decided to construct a bidirectional reporter vector. This vector was designed in such a way as to contain promoterless reporter genes which could be divergently transcribed from an extensive multiple cloning site (MCS). Such a vector would have the capacity to simultaneously respond to the bidirectional promoter activity conferred by any DNA fragment cloned into this polylinker. The bidirectional vector would therefore have distinct advantages over the single reporter gene method employed by Teo (1990). Firstly, the reporter vector allowed the bidirectional promoter activity of a cloned DNA fragment to be monitored in the same culture of cells. As the precise physiological state of the cells may well be crucial for *nod*-box regulation, the relative promoter activity of the cloned fragment could be directly compared to the corresponding activity in the opposite direction. While this investigation is mainly directed towards the interaction of the NodD protein with the *nod*-box, in the presence and absence of flavonoid, it is important to assess the relative bidirectional regulatory activity resulting from these interactions. The simplicity of the proposed vector would greatly facilitate these studies. Secondly, the convenience of having to clone a set of putative promoter fragments only once is extremely attractive. This point is emphasised when it is considered that the successful ligation, transformation, and assay of the clones need only be performed a single time for each fragment.

4.2 *gusA* AND *lacZ* REPORTER GENES

The past few years have seen a dramatic increase in the development and use of the β-glucuronidase gene (*gusA*) as a reporter gene. It encodes an acid hydrolase responsible for catalysing the cleavage of a wide range of β-glucuronides, including a synthetic chromogenic substrate whose breakdown may be monitored spectrophotometrically. As these substrate molecules are generally soluble in water and manufactured commercially, they provide the scope for inexpensive and convenient enzyme assays. It is the low cost and simplicity of this quantitative enzyme assay that makes it superior to the chloramphenicol acetyltransferase (CAT) fusions previously preferred by reseachers.

There are a number of additional advantages in using GUS as a reporter gene. Most importantly, there is little if any, endogenous β-glucuronidase activity found in a wide range of organisms, including *Rhizobium* (reviewed by Jefferson (1987); pers. comm. Sharma). This eliminates difficulties in detecting low or variable β-glucuronidase reporter activity against high endogenous levels. Secondly, the small size of the β-glucuronidase enzyme and gene results in the *in vitro* construction and analysis of gene fusions which are less cumbersome than other reporter genes currently in use. Another favourable feature of the GUS reporter gene is its histochemical properties. Although not applicable in this investigation, the ability to assess the spatial distribution of gene regulation and expression in histological preparations can be very important.

The reporter gene chosen to be expressed divergently from *gusA* was the β-galactosidase gene (*lacZ*). The effectiveness of *lacZ* as a reporter gene in *Rhizobium* had previously been demonstrated by Spaink *et al.* (1987), who incorporated it into the expression vector pMP220. Indeed, notwithstanding the convenience of the *lacZ* colourometric assay, this reporter gene was primarily selected because it had already been set up for transcriptional fusions in *Rhizobium*. This feature facilitated the construction of the proposed bidirectional reporter vector.

4.3 OVERVIEW OF THE STRATEGY USED TO CONSTRUCT THE BIDIRECTIONAL VECTOR

In order to coherently explain the strategy used to construct this vector, a brief summary of the cloning steps has been provided. It is anticipated that the overview will help the reader to comprehend the latter sections of this chapter where in-depth details of cloning procedures have been included (Figure 4.1).

At this stage, it must only be noted that the strategy involved three basic steps. The first step entailed the cloning of the promoterless *gus*A gene into the extensive MCS of pUC21. This resulted in the plasmid, pSPV1, where the translational start codon of the *gus*A gene was inserted downstream of the numerous enzyme sites in pUC21. The second step of the strategy involved cloning a ribosome binding site (rbs) upstream of the *gus*A gene in pSPV1. The rbs sequence comprised an ATG translational start codon and a Shine-Dalgarno sequence separated by a spacer region. This converted *gus*A into a gene capable of forming transcriptional fusions. Two strategies were employed to do this. Initially, two synthetic oligonucleotides were designed so they would anneal together and form a duplex rbs. However, problems with the oligos meant that an alternate plan had to be devised. This second strategy required cloning the portable rbs from pMP220 into the region immediately upstream of the *gus*A gene. The third cloning step saw the functional *gus*A-rbs-MCS module excised from pSPV2 and cloned into pMP220 which contains a transcriptionally-fused *lac*Z gene. The final product, pSPV4, consisted of *gus*A and *lac*Z reporter genes which were divergently expressed from the extensive MCS of pUC21. An important feature of pSPV4 is the symmetry provided by the two identical ribosome binding sites upstream of the two reporter genes.

4.4 VECTOR CONSTRUCTION

4.4.1 Clone *gus*A Gene Into pUC21

The initial step in the strategy for construction of the bidirectional vector was to insert the *gus*A gene into a vector which contained numerous restriction enzyme sites in its polylinker. After careful consideration, the most appropriate intermediate vector was deemed to be pUC21, a high copy number plasmid originally created to facilitate the cloning and manipulation of the HSV genome (reported by Vieira and Messing, 1991). The extensive MCS required for the GC-rich DNA of that organism was equally well suited for the cloning of the GUS gene. Importantly, the MCS is flanked by *Spe*I enzyme sites. As there are no sites in the *gus*A gene recognised by the *Spe*I enzyme, the entire *gus*A-MCS module could be excised as an *Spe*I-*Spe*I fragment. This was important because pUC21 lacked the genetic elements required to control plasmid replication and transfer in *Rhizobium*. Accordingly, this convenient vector was used only as an intermediary before the *gus*A/MCS cassette was finally transferred into a plasmid compatible with *Rhizobium*.

Figure 4.1.

1) **Clone *gus*A Gene Into pUC21**

The *gus*A gene was blunt-end ligated into the *Stu*I site of pUC21. The resulting clone, pSPV1, had the ATG start codon of *gus*A adjacent to the extensive MCS of pUC21. This allowed a wide selection of enzyme sites into which DNA of interest could be cloned.

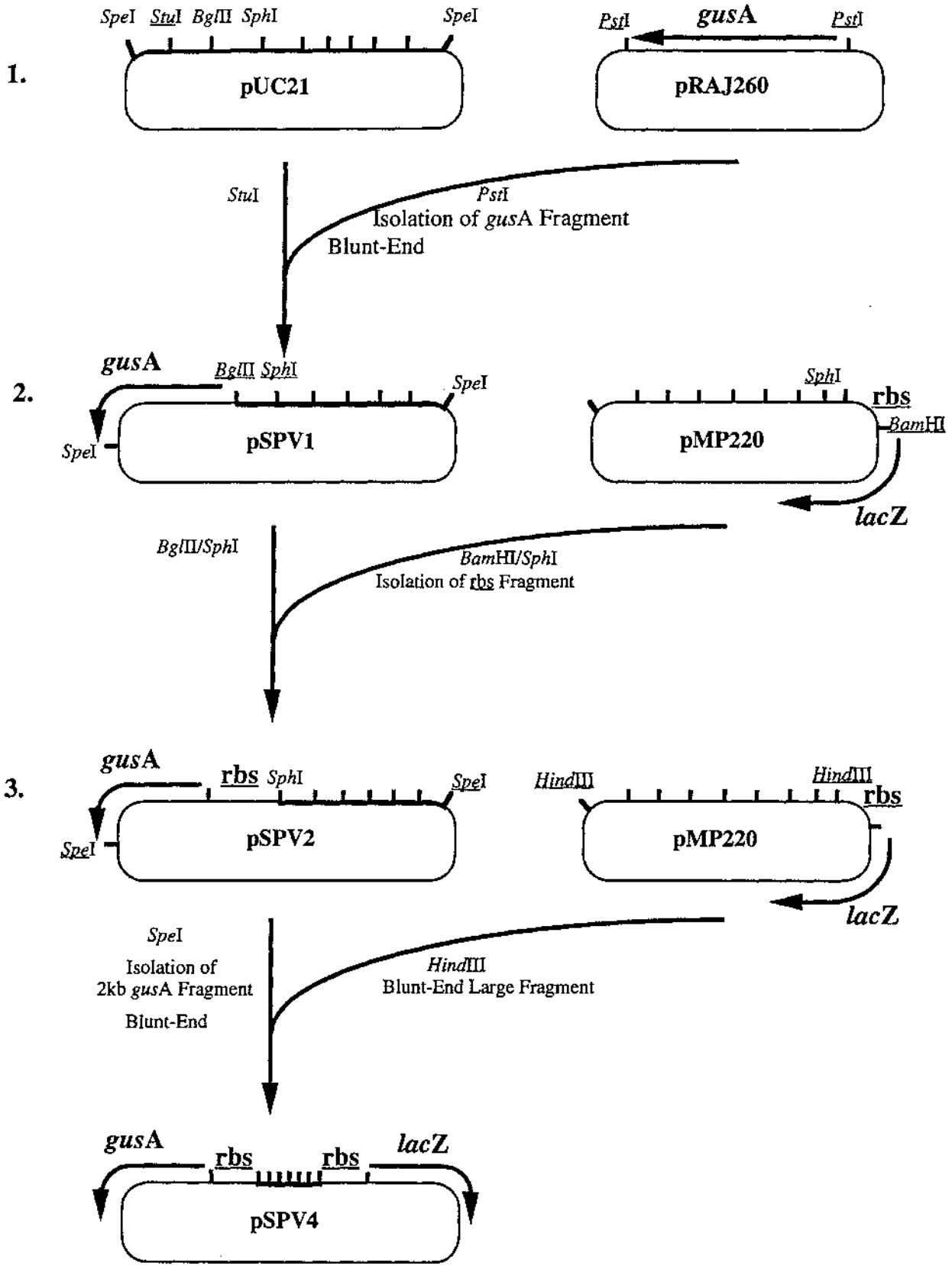
2) **Clone Ribosome Binding Site Into pSPV1**

The portable rbs of the pMP220 *lac*Z gene was excised as a *Bam*HI/*Sph*I fragment and ligated into pSPV1. Cloning this rbs into the *Bgl*II/*Sph*I sites located immediately upstream of the *gus*A gene facilitated formation of transcriptional fusions. The enzymes *Bam*HI and *Bgl*II have compatible ends.

3) **Clone Transcriptionally-Fused *gus*A-rbs-MCS Module Into pMP220**

The functional *gus*A-rbs-MCS unit was isolated from pSPV2 as an *Spe*I fragment and then blunt-ended using the Klenow enzyme. This fragment was ligated into the Klenow-treated *Hind*III sites of pMP220. The orientation of the cloned *Spe*I fragment was such that the *gus*A gene was divergently expressed from the *lac*Z gene of pMP220 with the MCS from pSPV2 positioned between the two. This final bidirectional expression vector was entitled pSPV4.

Figure 4.1. Overview of Bidirectional Vector Construction



4.4.1.1 Cloning Details

To facilitate the transfer of the *gus*A coding region into heterologous environments, Jefferson *et al.* (1986) designed and constructed a set of gene module vectors. The pRAJ series was generated by first separating the *gus*A coding sequence from the regulatory elements and Shine-Dalgarno (SD) region located upstream. The coding sequence was then positioned between various restriction enzyme sites to allow easy manipulations in the construction of new vectors. The vector used in this investigation, pRAJ260, was employed because of the elimination of the *Bam*HI site located in the *gus*A sequence. The presence of this restriction site would have precluded the use of this popular enzyme in any further molecular manipulations.

A large-scale preparation of pRAJ260, followed by its subsequent purification on a CsCl gradient, generated a good yield of clean DNA (donated by Sharma). Trial digests of this pRAJ260 preparation were performed using *Pst*I and *Bam*HI (Method 2.6.1). As expected, the *Pst*I digest liberated a band corresponding to the 1.8kb *gus*A gene, while the *Bam*HI enzyme cleaved only once verifying the absence of a *Bam*HI site internal to the *gus*A gene. Having confirmed that the desired plasmid was present, a second *Pst*I digest was carried out using an increased amount of plasmid DNA. Once the reaction had gone to completion, the digested DNA was electrophoresed on a 1% Seaplaque gel to separate the 1.8kb *gus*A gene from the vector band (Method 2.5.5). The former was then excised from the low melting-point agarose and cleaned using a phenol/chloroform extraction. Unfortunately, utilisation of the *Pst*I sites generated from the earlier digest was impractical due to the central positioning of the *Pst*I site in the pUC21 MCS. Thus, the 3' single-stranded overhangs were removed using the exonuclease activity of the T4 polymerase, leaving a blunt-ended fragment (Method 2.6.2).

These blunt termini were later ligated into the CAP-treated *Stu*I restricted site which is located immediately adjacent to one of the *Spe*I sites in pUC21 (Method 2.6.3 and 2.6.4). The ligation mixture was later electroporated into the cells of *E. coli* strain MC1022 (Method 2.6.6). A number of white, ampicillin-resistant colonies were identified and picked for further analysis. To confirm the presence of the required insert, small scale rapid boil preparations were performed (Method 2.4.2) and the resulting plasmids digested with diagnostic enzyme combinations (Method 2.6.1). The first screen of the plasmid identity was a size check. The potential recombinants were digested with *Eco*RI, an enzyme expected to cut only once within the desired clone. Any plasmids whose size was approximately 5.0kb (3.2kb pUC21 + 1.8kb *gus*A) were digested with additional enzymes. Digestion with *Spe*I established whether or not an insert of the correct size had been cloned into the *Stu*I site of pUC21.

Emphasis has already been placed on the importance of cloning the *gus*A gene such that its start codon is positioned next to the remaining available restriction enzyme sites in the polylinker. While the previous two tests determined the presence of a clone containing an insert corresponding to the *gus*A gene, they did not comment on the orientation of this insert. Hence, the final enzymatic confirmation of the putative recombinant was one ascertaining the orientation of the *gus*A gene within the plasmid. *Eco*RV is an enzyme that cuts the plasmid once in the pUC21 MCS, and twice in the *gus*A gene itself. Fortunately the *Eco*RV sites in the *gus*A gene are asymmetrically located, providing a convenient means by which orientation may be ascertained. Indeed, the DNA band patterns generated by the correctly oriented *gus*A gene (0.24kb + 0.65kb + 4.1kb) differ markedly from the corresponding band distribution of the incorrectly oriented clone (0.24kb + 1.1kb + 3.7kb). One clone, designated pSPV1, was consistent with the results expected from a correctly-oriented *gus*A gene ligated into the *Stu*I site of pUC21 (Figure 4.2 and 4.3).

4.4.1.2 Sequencing the *gus*A Gene From pSPV1

The double-stranded sequence determination employed here was extremely important for two reasons. Firstly, it provided the most direct evidence available for establishing the presence and orientation of the cloned *gus*A gene. Secondly, and most importantly, it gave the exact position of the *gus*A gene ATG start codon relative to the MCS. This information could be utilised to ensure that the positioning of the ribosome binding site (rbs) would give a functional (in-frame) fusion (Figure 4.4). It was therefore essential to obtain absolutely unequivocal sequencing data for this important stretch of DNA.

The putative *gus*A clone, pSPV1, obtained in the previous section was sequenced using the Universal -40 Primer. This synthetic oligonucleotide annealed to the *lac*Z gene of pUC21 allowing the sequence to be read back through the MCS and into the insert. The determined sequence corresponded exactly to that expected from the known sequence of the pRAJ and pUC21 vectors, (see Figure 4.5a).

The opposite end of the inserted *gus*A gene was also sequenced. Using the Universal Reverse Primer, the *Spe*I-3' *gus*A region was sequenced to check for any major genetic rearrangements. The sequence obtained was as predicted and therefore indicated that no unexpected events had occurred.

Figure 4.2. Restriction Enzyme Map of pSPV1

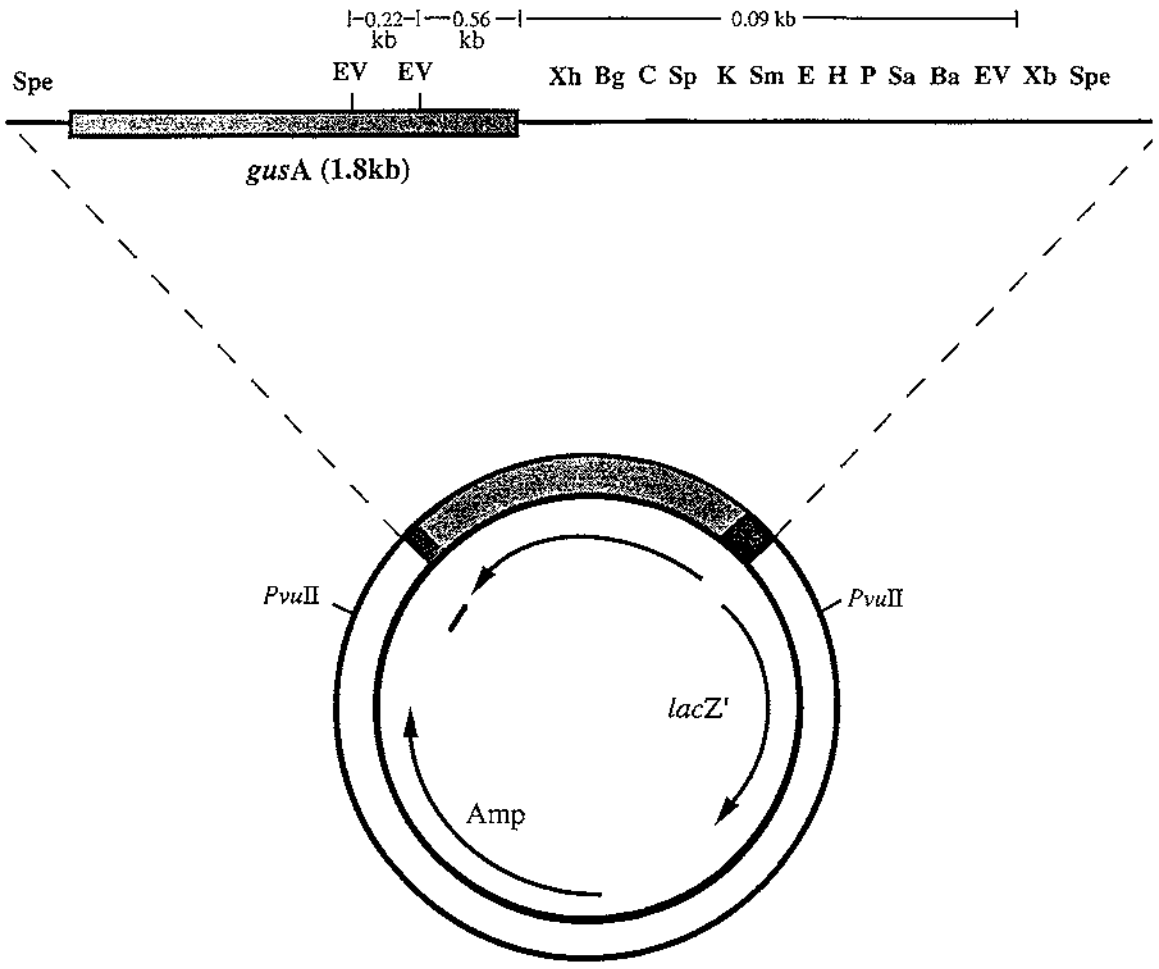
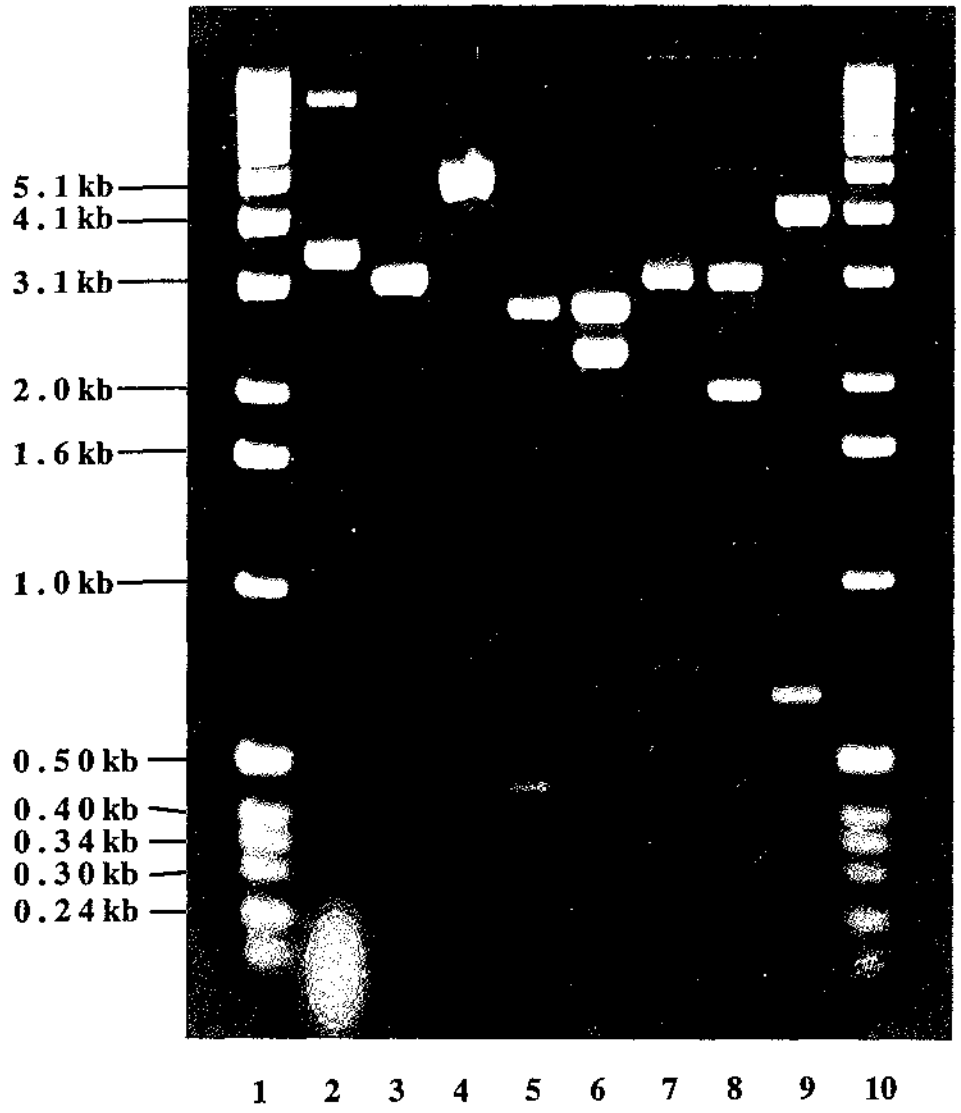


Figure 4.2. Restriction Enzyme Map of pSPV1

The *gus*A coding sequence cloned into the *Stu*I site of pUC21. The shaded region indicates the *gus*A gene which is read in the direction of the arrow. The darkly hatched area denotes the MCS. An enlargement of this sequence has been presented to show the preserved enyzme sites, ie. Spe- *Spe*I; Xh-*Xho*I; Bg-*Bgl*II; C-*Cla*I; Sp-*Sph*I; K-*Kpn*I; Sm-*Sma*I; E-*Eco*RI; H-*Hind*III; P-*Pst*I; Ba-*Bam*HI; EV-*Eco*RV; Xb-*Xba*I.

Figure 4.3. Restriction Analysis of pSPV1



1. 1kb ladder
2. pSPV1 uncut
3. pUC21 *Eco*RI cut
4. pSPV1 *Eco*RI cut
5. pUC21 *Pvu*II cut
6. pSPV1 *Pvu*II cut
7. pUC21 *Spe*I cut
8. pSPV1 *Spe*I cut
9. pSPV1 *Eco*RV cut
10. 1kb ladder

Figure 4.4. Reporter Gene Fusion Between a Portable rbs and the *gus*A Gene

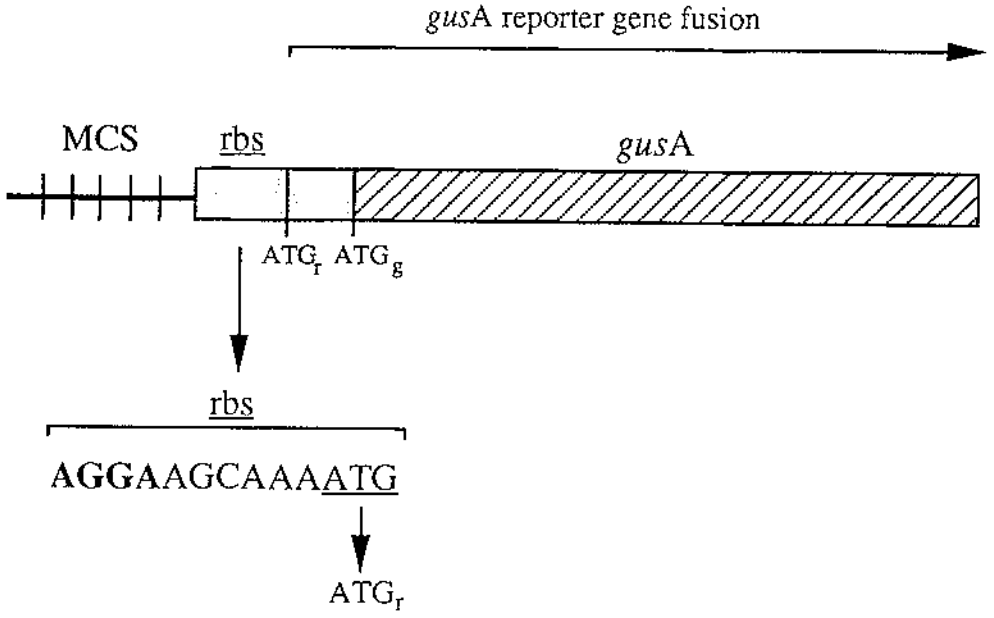


Figure 4.4.

A reporter gene fusion formed between a portable rbs and the *gus*A gene. The shaded region represents the *gus*A coding sequence, while the speckled area denotes a fragment containing an rbs. The DNA sequence of a functional rbs has been presented below the main diagram. It consists of a Shine-Dalgarno sequence (S-D) and an ATG translational start codon (underlined) separated by a spacer region of specific length and identity. The translational start codon of the rbs (ATG_r) is in the same reading frame as the *gus*A translational start codon (ATG_g). This allows functional reporter gene fusions to be formed. Putative promoter elements can be cloned into the MCS which is located upstream of the rbs-reporter gene construct.

Figure 4.5. Sequence of the *gus*A Gene Cloned Into pUC21

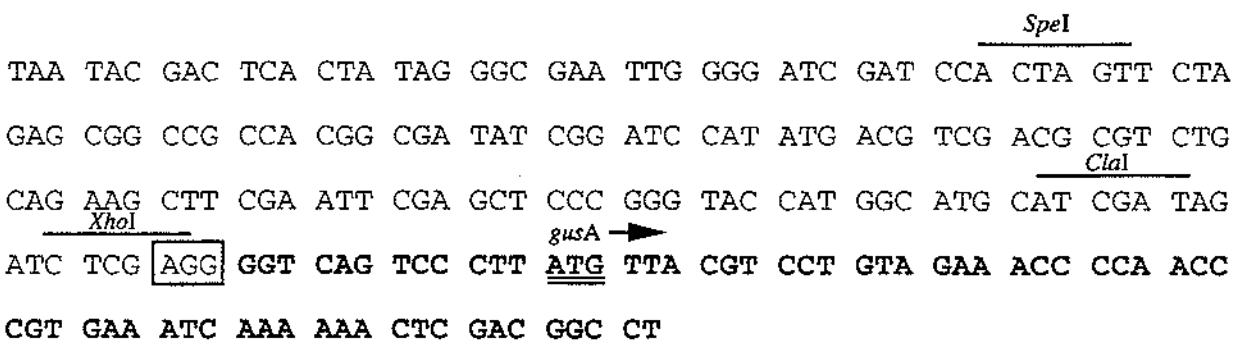


Figure 4.5.

The region spanning the insertion point of the *gus*A gene in pUC21 has been presented. This sequence was required to deduce the reading frame of the *gus*A ATG translational start codon with respect to the pUC21 MCS. The pUC21 sequence is in plain type, while the *gus*A cassette is in bold. The ATG translational start cadon of *gus*A is doubly underlined and the direction of transcription arrowed. Enzyme sites for *Spe*I, *Cla*I and *Xho*I have been indicated. The boxed sequence refers to the cleaved *Stu*I enzyme site into which the blunt-ended *Pst*I fragment from pRAJ260 was inserted.

4.4.2 Cloning of a Ribosome Binding Site into pSPV1

An essential feature of the transcriptional fusion proposed here is the ribosome binding site. These DNA sequences contain a start codon, usually an ATG, which in most cases is preceded by a purine-rich Shine-Dalgarno (SD) region resembling the consensus sequence 5'-GAGGA (Shine and Dalgarno, 1974; Hall *et al.*, 1982). This latter region is thought to assist the 30S ribosomal particle in positioning itself at the correct place with respect to the start codon on the mRNA. The third component of the rbs is a spacer region which separates the SD sequence from the ATG start codon by a specific number of nucleotides.

Interestingly, Hui *et al.* (1984), suggest that the actual SD sequence plays only a minor role in determining translational efficiency. The RNA sequence between the SD region and the start codon plays a major, if not predominant, role in determining the efficiency of the translation initiation process. Indeed, Shepard *et al.* (1982) have observed that variations in the length of this spacer region greatly affect the efficiency of ribosome binding sites, as does the actual nucleotide composition of this region. Stormo *et al.* (1982) concluded that functional ribosome binding sites display a non-random distribution of bases ouside the two conserved regions. Additionally, mutations found between the SD sequence and the start codon affect expression levels while other base changes outside that region have no effect (reviewed by Gold *et al.*, 1981).

A systematic analysis of the sequences in the spacer region of the ribosome binding site has indicated which nucleotides are optimal for high translational efficiencies. Working with a gene expression system in *E. coli*, de Boer *et al.* (1983), investigated the influence exerted by defined nucleotide alterations in this region. By varying the four bases following the SD sequence, it was found that A or T residues are preferential to C or G bases in terms of translational efficiency. A second mutational analysis of the spacer region focused on the three bases preceding the AUG start codon. Hui *et al.* (1984), showed that the most favourable combinations of bases in the -1 triplet are UAA and CUU. Through the comparison of mutants differing only in the position immediately adjacent to the start codon, it was generally found that a U residue maximally enhances translational efficiency. An A residue at the -2 position has a similar effect, though in both cases the degree of enhancement depends on the neighbouring bases. To summarise, it would appear that the predominance of A and T bases spread throughout the spacer region confers higher translational efficiency than the presence of the G and C residues.

4.4.2.1 Design of Two Complementary Oligonucleotides

The production of a vector employing reporter gene fusions from the existing pSPV1 plasmid involved the insertion of a carefully designed sequence. This stretch of DNA had to contain a SD sequence placed a specific number of nucleotides away from an ATG codon which was in-frame with the *gus*A start codon. In order to obtain such a defined sequence of DNA, it was decided to have two complementary oligonucleotides synthesised. These synthetic products (Figure 4.6) were carefully designed to accommodate several important features.

a) Complementary Oligonucleotides Facilitating Sticky-Ended Ligations

The complementary nature of the two oligonucleotides was designed to enable the strands to anneal together yielding double-stranded DNA fragments appropriate for cloning into vector pSPV1. The sites into which this synthetic fragment was to be inserted were those cleaved by restriction enzymes *Xho*I and *Cla*I. The proximal position of *Xho*I to the *gus*A insertion site was the primary reason for the selection of this enzyme. The choice of *Cla*I was based on its distance from *Xho*I. The two sites were separated by enough nucleotides to preclude restriction enzyme interference when preparing the vector for the insert, yet they were close enough to prevent large-scale abolition of restriction enzyme sites.

To facilitate the ligation of the synthetic fragment into these two sites, the oligonucleotides were specifically designed such that single-stranded overhangs would be generated once the two strands were annealed. These 5' overhangs would form cohesive ends complementary to the 5' overhang resulting from the *Xho*I/*Cla*I double-digest of pSPV1. The ensuing ligation event was expected to be considerably more efficient than if the fragment was to be cloned using blunt-ends.

b) Position of ATG in the Ribosome Binding Site

The selection of *Xho*I as the proximal enzyme site into which the synthetic fragment was to be cloned, allowed the correct position of the ATG codon to be determined. As the sequence of the region between the *gus*A ATG start codon and the *Xho*I site had been precisely resolved (Figure 4.5), it was a facile exercise to maintain the integrity of the reading frame. By adding an appropriate number of bases to the oligonucleotides, the synthesised ATG codon was placed in the same reading frame as the *gus*A start codon. Therefore, production of a chimeric GUS protein was possible through the described reporter gene fusion.

c) DNA Sequence of Ribosome Binding Site

As discussed above, the sequence of the SD region, and the spacer region are very important for obtaining high levels of transcription. Since the crucial features of the rbs sequence are still not absolutely understood, it was decided to copy the sequence of an rbs that was known to initiate expression. This sequence, provided by Sharma, incorporated all of the features discussed above and had already been demonstrated to be functional (Sharma and Signer, 1990).

d) Diagnostic Restriction Enzyme Site

In order to identify a successful ligation event between the oligonucleotide duplex and the restricted pSPV1 plasmid, an *Asn*I enzyme site was incorporated into the oligonucleotide sequence. Cleavage of potential clones with this enzyme resulted in distinctive band patterns that allowed recognition of a positive clone. The typical diagnostic method of cutting the inserted fragment back out of the vector could not be used in this case due to the small size of the oligonucleotide duplex.

e) Termination Codons in Every Reading Frame

The possibility of expression becoming initiated at a start codon located upstream of the introduced ribosome binding site may result in failure to synthesise active reporter enzyme. The inclusion of TAA termination codons in all three reading frames was intended to prevent this translational read-through. Hence, translation of the fusions should only be intitiated at the ATG codon situated within the proposed oligonucleotide duplex.

4.4.2.2 Cloning Oligonucleotide Duplex

To preliminarily identify pSPV1 recombinants containing a ligated oligo duplex, colony lifts and colony streaks were employed (Method 2.8). However, after many unsuccessful cloning attempts, the ability of the two oligos to anneal together was questioned. A non-denaturing polyacrylamide gel was used to check the compatibility of the single-stranded oligos when placed together under different annealing conditions. As Figure 4.7 shows, the oligos did not anneal together under any conditions. It was also interesting to note that the two oligos migrated at different rates. This is probably due to conformational differences rather than differences in size. (The ladder in Lane 1 cannot be used to size the oligos as it consists of double-stranded DNA). Lane 6 was intended to check the integrity of the cohesive ends of the expected duplex. Had the oligos annealed successfully, higher molecular size material should have been observed.

Figure 4.6 Synthetic rbs Oligonucleotides

Important features include:

- complementarity between RBS1 and RBS2
- functional rbs formed after annealing
- Shine-Dalgarno sequence (single overline)
- ATG start codon in same reading frame as *gus*A ATG (double overline)
- Termination codons in all three reading frames prior to ATG translational start codon (bold face)
- generation of *Xho*I (C/TCGAG) and *Cla*I (AT/CGAT) ends after annealing
- diagnostic *Asn*I site to check for recombinants carrying this synthetic fragment (boxed)

Figure 4.7. Annealing Potential of Synthetic rbs Oligonucleotides

Synthetic rbs oligonucleotides electrophoresed on a 20% nondenaturing polyacrylamide gel to examine their annealing potential under two different conditions. The double-stranded DNA ladder in Lane 1 was run to size the annealed oligonucleotides. The DNA preparation in Lane 6 was intended to check the *Xho*I/*Cla*I cohesive ends that were designed to form after annealing. As the two oligonucleotides did not anneal (Lanes 2-5), the DNA in Lanes 1 and 6 was redundant.

Figure 4.6. Synthetic rbs Oligonucleotides

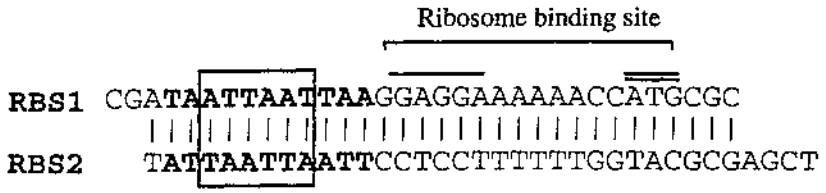
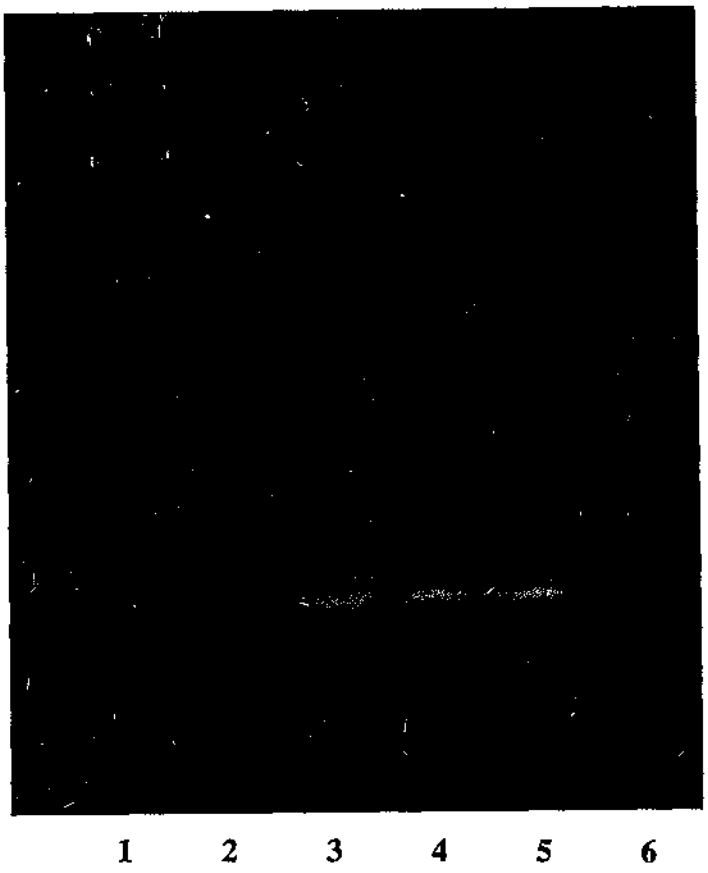


Figure 4.7. Annealing Potential of Synthetic rbs Oligonucleotides



1. M13mp19 *Hae*III + *Bam*HI cut
2. RBS 1 oligonucleotide
3. RBS 2 oligonucleotide
4. RBS 1 + RBS 2 heated to 80°C, then cooled to 25°C
5. RBS 1 + RBS 2 heated to 100°C, then cooled to 4°C
6. RBS 1 + RBS 2 heated to 100°C, then cooled to 4°C; phosphorylated and then ligated.

There was 1µg of each oligonucleotide in the annealing and ligation mixtures. The entire mixture was loaded onto the gel.

4.4.2.3 Cloning the Portable Ribosome Binding Site From pMP220

The inability of the two oligonucleotides to form a sticky-ended duplex led to their eventual rejection. Due to time limitations, synthesis of further oligonucleotides was not considered to be a viable option. Accordingly, this initial strategy was abandoned, and an alternate plan formulated. This second scheme utilised the pre-existing, portable rbs located immediately upstream of the promoterless *lacZ* gene in pMP220 (Spaink *et al.*, 1987). This rbs, cloned from XJ003 (Rossi *et al.*, 1983) was originally derived from the chloramphenicol acetyl transferase gene taken from *E. coli*. In retrospect, the employment of this portable rbs was more appropriate than the one contained in the oligonucleotide duplex because the final construct would be perfectly symmetrical with respect to the MCS. Exactly the same rbs would be located upstream of each of the reporter genes, implying that differences in activity could be attributed solely to promoter strength and not to inefficiencies in ribosome binding or other factors related to assymetry in the 5' untranslated region. Secondly, a previous analysis of the rbs in pMP220 had already confirmed that consistent results were routinely obtained in *Rhizobium* (Spaink *et al.*, 1987). This effectively guaranteed the functionality of the rbs in the organism under investigation.

4.4.2.4 Cloning Details

A large-scale preparation of pMP220 was made as outlined in Method 2.4.1. The purified plasmid DNA was enzymatically digested using the restriction enzymes *Bam*HI and *Sph*I (Method 2.6.1). The *Bam*HI site was an artifact of a previous cloning event (Spaink et al., 1987), and was conveniently located at the portable rbs/*lacZ* boundary. Cleavage at this site separated the ATG associated with the rbs from the ATG start codon of *lacZ*. The rationale behind selecting the *Sph*I site (located in the pMP220 multiple cloning site) as the second cleavage point shall become evident later.

The *Bam*HI/*Sph*I double digest was checked on a 1% agarose gel to confirm liberation of the fragment (Method 2.5.3). Once the presence of the 199bp portable rbs had been identified, the restricted DNA was passed through a phenol/chloroform extraction to eliminate any residual enzymatic activity (Method 2.5.1).

A similar extraction was performed on a *Bgl*II/*Sph*I digested preparation of pSPV1. These enzyme cuts were intended to provide cohesive ends into which the *Bam*HI/*Sph*I rbs fragment could become inserted. Therefore, after the purification of the two plasmid digests, appropriate amounts of the pSPV1 DNA fragments were added to the pMP220 fragments and ligation effected (Method 2.6.4). Once the ligation had gone to

completion, the DNA ligase was heat denatured, and a 'suicide' digest performed on the ligation mix. This involved the addition of an enzyme whose sites were not found in the desired final plasmid construct. In this case, *Bgl*II was the most suitable enzyme, because its site was not reconstituted in successful ligation events involving *Bam*HI and *Bgl*II cohesive ends. Furthermore, as there were no additional *Bgl*II sites either within the pSPV1 plasmid or within the inserted rbs, this enzyme had no effect on the desired product. On the other hand, any ligated product, or uncut plasmid that retained an intact *Bgl*II site was cleaved and thereafter linearised. As stable transformation of DNA in bacteria requires the presence of circularised plasmid, any cleaved DNA molecules were prevented from being maintained in the cells.

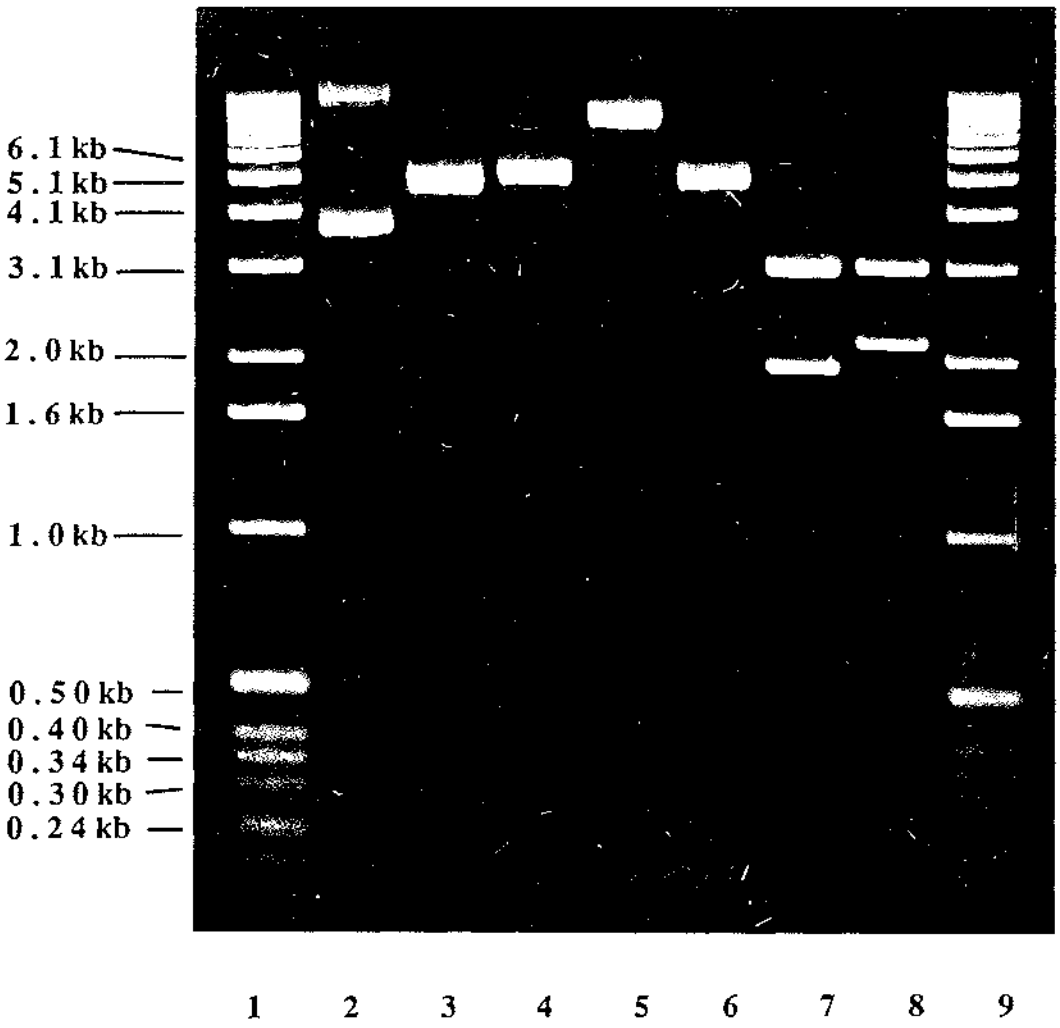
Once the 'suicide' cut had been completed, the reaction mixture was microdialysed to remove excess salt introduced to optimise conditions for the *Bgl*II cleavage. The mixture was then electroporated into DH1 cells. Ampicillin-resistant potential recombinants were later picked, and grown up for further analysis. Mini preparations of the plasmid DNA from the possible clones were generated using the rapid boil technique (Method 2.4.2). These were later double-digested with *Xho*I and *Eco*RI before being electrophoresed on a 1% agarose gel (Method 2.6.1 and 2.5.3). Identification of a 199bp fragment verified the successful cloning of the desired rbs (Figure 4.8). Cleavage at the pUC21-derived *Pvu*II sites also confirmed the presence of an insert cloned into the MCS. The required clone was designated pSPV2.

The extremely high proportion of positive results observed can be attributed to the stringent 'suicide' digest, together with the incompatible pSPV1 vector termini. The incompatible *Bam*HI/*Sph*I ends were not only chosen to eliminate undesirable background colonies, but also to ensure that the orientation of the insert was correct. The actual selection of *Sph*I was due to its convenient location within the MCS of both pSPV1 and pMP220. The use of this enzyme safeguarded the integrity of the MCS of the final construct. Consequently, the number of enzyme sites that were eliminated, or conversely, duplicated, was kept to a minimum.

4.4.2.5 Sequence Determination of Ribosome Binding Site in pSPV2

As emphasised earlier, the reading frame of the ATG contained in the cloned ribosome binding site must be in exactly the same reading frame as the ATG start codon of the *gus*A gene (Figure 4.4). If this is not the case, expression of an assayable GUS fusion will not occur. The vital importance of obtaining the precise sequence of the region flanked by the two ATG sites necessitated the accurate sequencing of the 5' junction of this construct.

Figure 4.8. Restriction Analysis of pSPV2



1. 1kb ladder
2. pSPV2 uncut
3. pSPV1 *Eco*RI cut
4. pSPV2 *Eco*RI cut
5. pMP220 *Sph*I + *Bam*HI cut
6. pSPV2 *Xho*I + *Eco*RI cut
7. pSPV1 *Pvu*II cut
8. pSPV2 *Pvu*II cut
9. 1kb ladder

Figure 4.8.

The agarose gel demonstrates the successful ligation of a 0.199kb portable rbs into pSPV1. For identification of enzyme sites see Figure 4.2.

Obtaining the sequence of this region was also interesting intrinsically. The sequence data presented by Rossi *et al.* (1983), contained significant discrepancies with the nucleotide sequence published by Alton and Vapnek (1979). These disparities could be examined through comparisons with independently generated sequence data such as that obtained in this investigation.

Sequencing of the purified plasmid, pSPV2, was attempted using the Universal -40 Primer whose annealing site was located downstream of the MCS. Regrettably, this binding forced the Sequenase to 'read' through the entire MCS before it was able to reach the region of interest. Therefore, although this strategy generated approximately 250bp of readable data, the essential sequence at the *Bam*HI/*Bgl*II junction was still undecipherable. The cloned *gus*A gene made sequencing from the opposite direction equally futile. To overcome this problem, the newly cloned fragment had to be re-cloned into a sequencing vector without the extra flanking region.

4.4.2.6 Production of pSPV3 Subclone

The plasmid, pSPV2, was isolated from *E. coli* cells using the rapid boil method (Method 2.4.2). This preparation was later digested with *Xho*I and *Eco*RI to excise the DNA region of interest (Method 2.6.1). To prevent the undesirable cloning of the remaining 5kb pSPV2 *XhoI*/*Eco*RI fragment, the reaction mix was also digested with *Sac*II. This resulted in the production of *XhoI*/*Sac*II and *Eco*RI/*Sac*II terminal ends for the large fragment, whilst retaining the *XhoI*/*Eco*RI termini on the region of interest.

To accommodate the *XhoI*/*Eco*RI rbs fragment, a preparation of purified pUC118 was digested with the compatible enzymes *Sal*I and *Eco*RI (Method 2.6.1). After the two digests had been checked on a 1% agarose gel, the enzymes were heat inactivated. A ligation reaction was subsequently set up between the *Eco*RI/*Sal*I fragments of pUC118, and the multiply cleaved pSPV2 preparation (Method 2.6.4). Once this ligation had gone to completion the ligase was inactivated and a 'suicide' digest performed on the ligation mixture. Although many enzymes could have been used, *Sma*I was chosen because of its high activity in the ligation buffer. The ability of *Sma*I to react in a low salt buffer avoided the requirement of microdialysing the preparation after the digest. Hence, the ligation mixture was electroporated directly into *E. coli* strain MC1022 (Method 2.6.6).

Potential recombinants were identified through their resistance to ampicillin and their white colony morphology. The appropriate colonies were picked and their plasmids extracted using the rapid boil method (Method 2.4.2). These mini preparations were later digested with *Eco*RI and *Pst*I to liberate the expected 199bp fragment (Method 2.6.1). One colony that exhibited this band on a 1% agarose gel was designated pSPV3.

4.4.2.7 Sequencing of pSPV3

The pSPV3 plasmid preparation obtained in 4.4.2.6 was cleaned using Method 2.9.1 and later sequenced using both the Universal -40 Primer and the Reverse Primer. The sequence is shown in Figure 4.9.

The sequence determined contradicted that published by Alton and Vapnek (1979) who first elucidated the CAT gene sequence. It appears that an unreported 1bp deletion must have occurred at the +41-43 nucleotide position, where a group of three C nucleotides has been reduced to two. The sequence also differs from that reported by Rossi *et al* (1983). A fusion consisting of the first 56⅔ amino acid codons of the CAT gene was observed here, yet Rossi *et al* (1983), noted the fusion occurring after the first 52⅔ codons. The important issue in both of these investigations was the fact that the fusion occurred two thirds of the way through the codon bordering on the reporter gene. This knowledge can be used to ensure that the reporter gene is in-frame with the rbs in the final construct.

4.4.3 Cloning of *gus*A-rbs-MCS Fragment

The final step in the production of the bidirectional vector was the transfer of the transcriptionally fused *gus*A-rbs-MCS construct into the unidirectional *lacZ* transcriptional expression vector pMP220.

4.4.3.1 Cloning the *Spe*I Fragment of pSPV2 into pMP220

Earlier, the pUC21 MCS was primarily selected for its broad range of enzyme sites, and its flanking *Spe*I sites. This latter property enabled the entire cassette to be excised as a single *Spe*I/*Spe*I fragment. Hence, an *Spe*I digestion was performed on the pSPV2 clone and the products electrophoresed on a 1% Seaplaque gel. The 2kb *gus*A-rbs-MCS module was isolated, purified and then blunt-ended (Methods 2.5.5 and 2.6.2).

Vector DNA was prepared in a similar fashion. The purified preparation of pMP220, made in 4.4.2.3, was digested with *Hind*III (Method 2.6.1). This enzyme cleaved once at each end of the polylinker resulting in the removal of the entire MCS. Importantly, this excision avoids repetition of enzyme sites after the MCS associated with the insert gets cloned into pMP220. To guarantee that the *Hind*III/*Hind*III MCS fragment would be lost during a precipitation step, its size was reduced to an unsalvageable level. This was accomplished through a second digestion using the restriction enzyme *Eco*RI, which cleaved the plasmid at a unique site mid-way through the MCS.

Figure 4.9. Sequence Through the Ribosome Binding Site in pSPV3

```
  T CCA CTA GTT CTA GAG CGG CCG CCA CGC GGA TAT CGG ATC CAT ATG ACG
  A GGT GAT CAA GAT CTC GCC GGC GGT GCG CCT ATA GCC TAG GTA TAC TGC

TCG CGT CTG CAG AAG CTT CGA ATT CGA GCT CCC GGG TAC CAT GGC ATG CAA
AGC GCA GAC GTC TTC GAA GCT TAA GCT CGA GGG CCC ATG GTA CCG TAC GTT

GCT TCG ACG AGA TTT TCA GGA GCT AAG GAA GCT AAA ATG GAG AAA AAA ATC
CGA AGC TGC TCT AAA AGT CCT CGA TTC CTT CGA TTT TAC CTC TTT TTT TAG

ACT GGA TAT ACC ACC GTT GAT ATA TCC AAT GGC ATC GTA AAG AAC ATT TTG
TGA CCT ATA TGG TGG CAA CTA TAT AGG TTA CCG TAG CAT TTC TTG TAA AAC

AGG CAT TTC AGT CAG TTG CTC AAT GTA CCT ATA ACC AGA CCG TTC AGC TGG
TCC GTA AAG TCA GTC AAC GAG TTA CAT GGA TAT TGG TCT GGC AAG TCG ACC

ATA TAA CGG CCT TTT TAA AGA CCG TAA AGA AAA ATA AGC ACA AGT TTT ATC
TAT ATT GCC GGA AAA ATT TCT GGC ATT TCT TTT TAT TCG TGT TCA AAA TAG

CGG ATC CGC CTG AGA ACC CAA
GCC TAG GCG GAC TCT TGG GTT
  ↑
```

Figure 4.9.

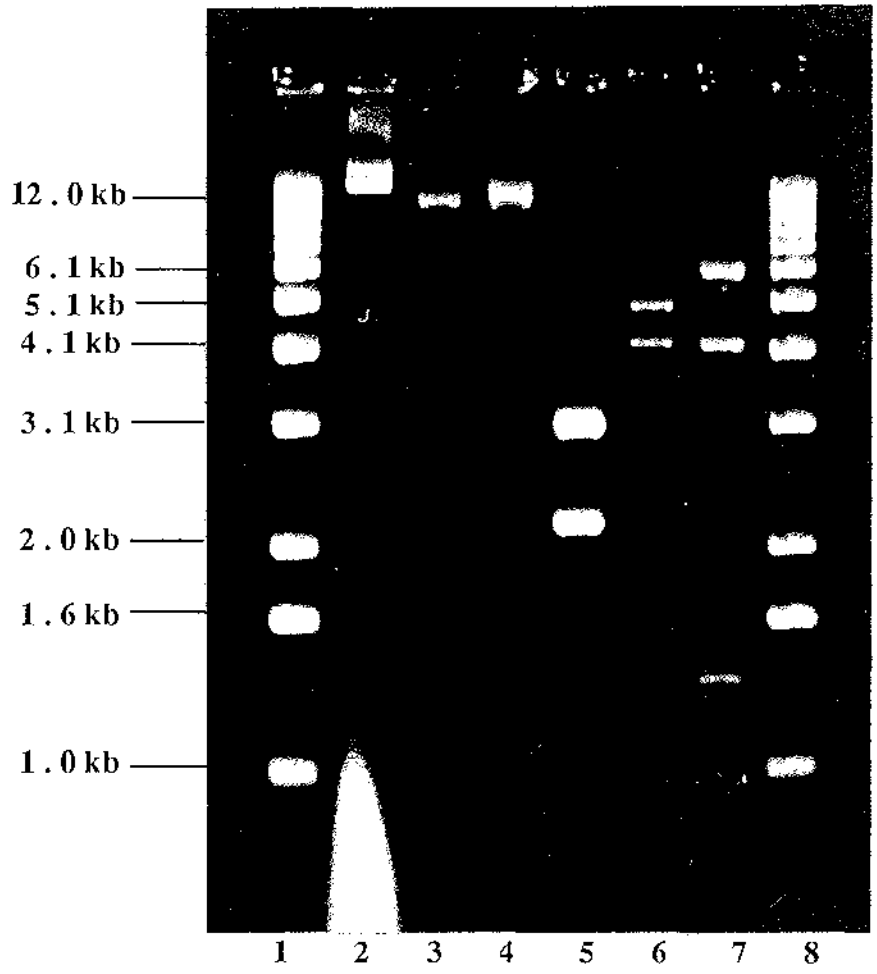
DNA sequence of the portable rbs positioned upstream of the *gusA* gene in pSPV3. The sequence was initially derived from the chloramphenicol acetyl transferase gene found in *E. coli* and utilised in the construction of pMP220. Bold type indicates the rbs. This consists of a Shine-Dalgarno sequence (underlined) and an ATG translational start codon (double underline) separated by a specific spacer region. The nucleotides have been arranged in codon triplets to facilitate identification of the reading frame with respect to the start codon of the rbs. The two and a half lines of sequence preceding the rbs is not translated and therefore plays little role in the reporter gene fusion. Accordingly, the entire sequence of this untranslated region has not been presented. The codons following the rbs are from the *E. coli* CAT gene. The overlined hexamer denotes the *Bam*HI restriction enzyme site which marks the border of the portable rbs. The arrow indicates that the point of *Bam*HI cleavage occurs two-thirds of the way through the 57th codon after the rbs ATG start site. It is at this site that reporter genes may be fused. Sequence following this *Bam*HI site is part of the MCS found in pSPV2 which was cloned into pSPV3.

The remaining *Hind*III/*Hind*III 10kb vector fragment was end-filled with Klenow and then CAP treated to discourage self-ligation events (Methods 2.6.2 and 2.6.3). Appropriate amounts of blunt-ended fragments were later combined in a ligation reaction (Method 2.6.4). Once sufficient time had elapsed for the reaction to go to completion, the ligase was heat inactivated and a 'suicide' digest set up. The only unique enzyme site that was not present in the desired construct was *Bgl*II. The single position recognised by this enzyme in the intact pMP220 had been removed as part of the MCS. Similarly, the only *Bgl*II site on the *Spe*I/*Spe*I fragment had been destroyed by the compatible-end ligation event between *Bam*HI and *Bgl*II (see Section 4.4.2.3). Hence, the ligation mix was digested with *Bgl*II, mainly to eliminate any background caused by religated or uncut vector molecules.

The reaction mixture was later microdialysed and then electroporated into DH1, an *E. coli* strain chosen for its *end*A$^-$, *rec*$^-$ genotype. Tetracycline-resistant colonies were picked and later cultured in preparation for further analysis. The plasmids were isolated from the DH1 cells and digested with diagnostic enzymes to determine whether an insert had been ligated into the vector. In this case, where no suitable enzyme sites flanked the region of interest, the plasmids were merely linearised with *Eco*RI. The restricted DNA was later electrophoresed on a 0.7% agarose gel where a single band of 12kb (10kb vector + 2kb insert) was expected. From the examination of 20 potential clones, nine were found to yield a linear molecule of the correct size.

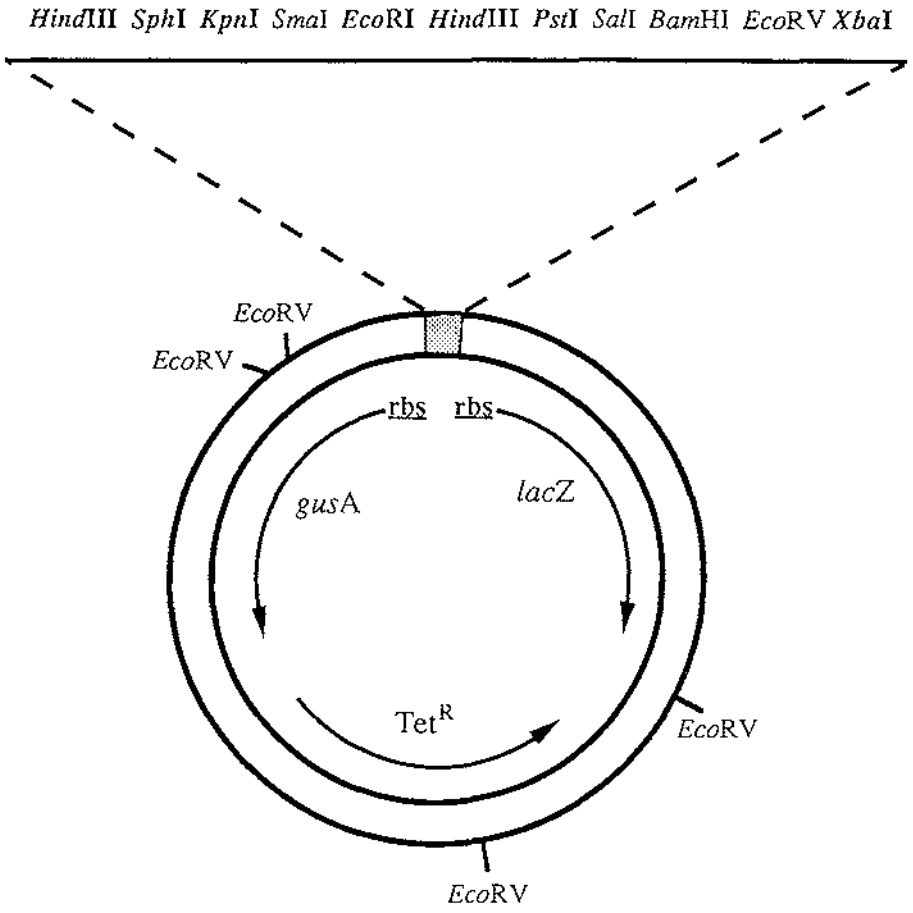
The bidirectional vector would only be functional if the insert was cloned with the *gus*A gene separated from the *lac*Z gene by the MCS. The opposite orientation would result in a unidirectional *gus*A reporter vector, with the *lac*Z gene being redundant. To ascertain the polarity of the insert, fresh plasmid preparations of the nine recombinants were digested with *Eco*RV. The asymmetric distribution of the *Eco*RV sites around the predicted final vector construct, allowed the orientation of the insert to be deduced (see Figures 4.10 and 4.11). Upon digestion with this enzyme, construct containing the correctly oriented insert was expected to give band sizes of 0.24kb + 0.90 + 1.4kb + 4.0kb + 6.3kb while the reversed orientation should have given a 0.24kb + 0.90 + 1.4kb + 5.0kb + 5.3kb pattern. Pleasingly, once the restricted plasmids were electrophoresed, six of the recombinants yielded fragments commensurate with the band pattern expected from a correctly cloned insert. The three remaining recombinants were observed to have inserts cloned in the opposite orientation. The desired recombinant plasmid was labelled pSPV4. This vector represented the final configuration of the bidirectional reporter vector and fulfilled the aim of this part of the project.

Figure 4.10. Restriction Analysis of pSPV4



1. 1kb ladder
2. pSPV4 uncut
3. pMP220 *Eco*RI cut
4. pSPV4 *Eco*RI cut
5. pSPV2 *Spe*I cut
6. pMP220 *Eco*RV cut
7. pSPV4 *Eco*RV cut
8. 1kb ladder

Figure 4.11. Restriction Enzyme Map of Bidirectional Reporter Vector pSPV4



*Eco*RV Cut of Potential Recombinants

Correct Orientation of *gus*A (kb)	Incorrect Orientation of *gus*A (kb)
0.24	0.24
0.90	0.90
1.4	1.4
4.0	5.0
6.3	5.3

Figure 4.11.

Diagrammatic representation of the bidirectional expression vector. The two reporter genes are transcribed divergently from the extensive MCS which has been enlarged. The enzymes which cleave the vector only in the MCS are in bold. To determine the orientation of the *gus*A-rbs-MCS module within pSPV4, the asymmetric *Eco*RV sites were used.

Chapter 5 - REGULATORY ANALYSIS OF PCR-GENERATED REGIONS CONTAINING AN *R. loti nod*-BOX

5.1 AMPLIFICATION OF SPECIFIC *nod*-BOX-CONTAINING REGIONS

The aim of this investigation was to examine the regulatory influence conferred by the flanking regions of the *R. loti nod*-box. These bordering DNA sequences may modulate the putative promoter activity exerted by the *nod*-box. Hence, this project endeavours to monitor the regulatory effect of cloning the *nod*-box sequence plus varying lengths of DNA bracketing it, into a bidirectional expression vector.

The generation of such a diverse range of DNA products can be performed in two ways. Firstly, the method employed by Teo (1990), involved the use of the non-specific exonuclease *Bal*31. Since the *nod*-box cloned by Collins-Emerson (1991), was contained within a 0.5kb fragment of pPN25, a series of *Bal*31 deletions could be performed on the two flanking regions. This produced desirable deletions of the regions bordering the *nod*-box which could be compared to the original undeleted fragment.

The major disadvantage with the *Bal*31 protocol was the non-specificity of the final deletion product. To overcome this randomness, the alternate method of polymerase chain reaction (PCR) amplification was undertaken. This strategy entails the prudent selection of specific primers that constitute the boundaries of the desired fragments. Therefore, specified regions of DNA flanking the *nod*-box could be investigated instead of the randomly-generated deletions implicit in *Bal*31 digestions.

5.1.1 Selection of Primers

The identification and subsequent cloning of DNA regions containing the two *nod*-boxes of interest had previously been accomplished. The *nod*-box upstream of the *nod*ACIJ operon had been cloned as a 0.5kb *SalI*/*Sph*I fragment in pPN25 (Collins-Emerson, 1991). Similarly, Young (pers. comm.) constructed a plasmid, pPN37, that contained a 4.1kb *Sph*I/*Sph*I fragment bearing the *nod*-box located upstream of the *nod*B operon. Both *nod*-boxes were subsequently analysed and their sequences determined.

This sequence information enabled oligonucleotides to be designed such that their nucleotide sequences were complementary to the DNA sequences at the borders of the desired region and such that their 5' to 3' polarity was oriented towards each other. Figures 5.1 and 5.2 show the distribution of primers that were synthesised for this

investigation, along with the oligonucleotides kindly donated by Professor Barry Scott. The arrangement of the primers was intended to represent a cross-section of possible fragments containing intact *nod*-boxes, yet with varying lengths and patterns of their respective flanking regions. Clearly for these criteria to be satisfied, the two primers incorporated into an individual PCR reaction had to be separated by the 47bp putative *nod*-box.

As seen in Figure 5.1 and 5.2, the most extensive DNA flanking regions could be generated using the BS primers. These sets of oligonucleotides were designed to anneal to sequences located towards the extremities of the cloned fragments containing the *nod*-boxes. Hence, in the case of the *nod*-box in pPN25, the DNA fragment amplified between the two primers BS1 and BS2, would be fundamentally equivalent to the undeleted fragment of Teo (1990).

To generate fragments whose flanking regions were not so extensive, the SP series of primers was synthesised. These oligonucleotides were designed so that their point of annealing would be positioned between the distal BS primer binding site and the *nod*-box. Significantly, this provides the potential to produce fragments with flanking regions that have been reduced in length by a predetermined amount. These sites were selected with the arbitrary intention of excluding the ATG start codons of the divergently expressed *nod* genes. In theory the inclusion of an ATG start codon may have compromised the formation of the reporter gene fusion. However, the presence of termination codons in all three reading frames prior to the rbs, meant that no read-through could occur in this construct. Once the SP series of primers had been chemically synthesised, a number of oligonucleotide combinations were used to generate specific PCR products (Table 5.1).

5.1.2 Polymerase Chain Reaction

Since the PCR was first introduced it has revolutionised the analysis of DNA and is increasingly being applied to routine clinical diagnosis of medical disorders as well as basic research science. In both of these areas, PCR has enormous potential because it is rapid, inexpensive and simple to perform. However, PCR is not without its problems. Apart from the significant yet avoidable threat of DNA contamination, there are a number of other factors that affect PCR efficiency. Such variables as annealing temperature and template concentration had to be optimised for each combination of template and primers.

Figure 5.1. DNA Sequence of the *nodA nod*-Box

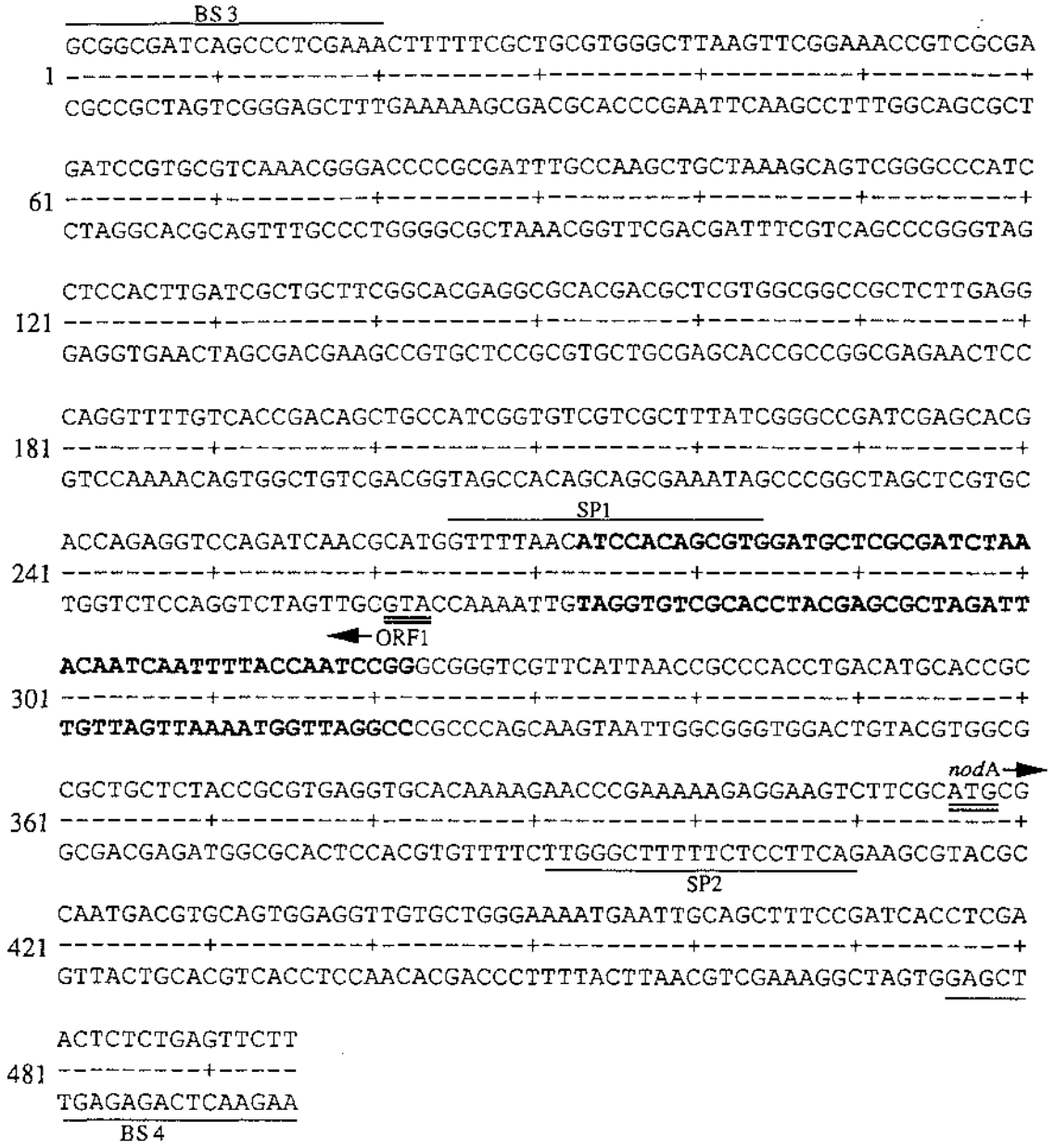
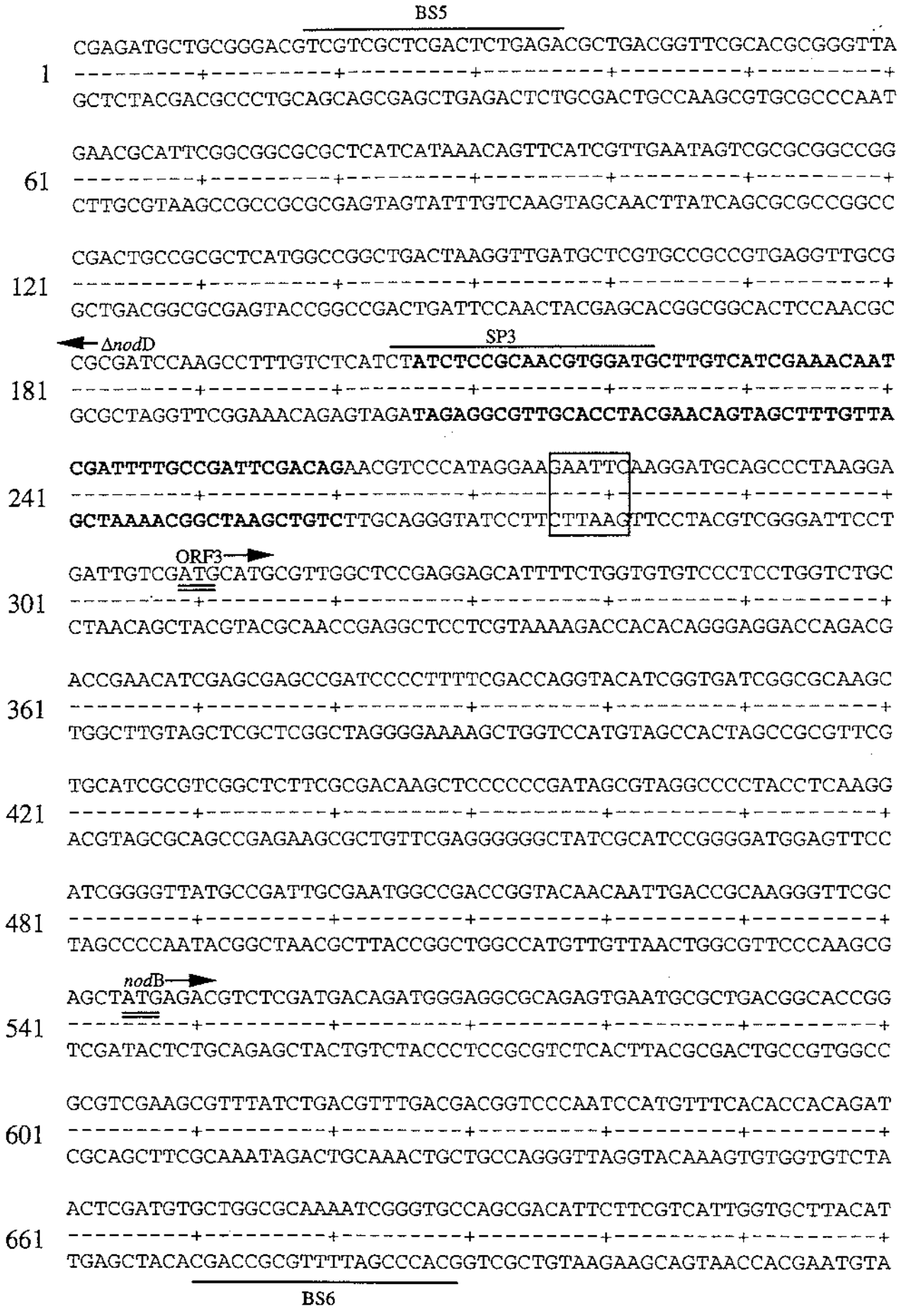


Figure 5.1 DNA sequence of the intercistronic region of the *nod*D gene fragment (ORF1) and the *nod*ACIJ operon in *R. loti*.

Figure 5.2 DNA sequence of the intergenic region of the deleted *nod*D gene and the ORF3/*nod*B fusion.

The putative translational initiation start sites are doubly underlined. The *nod*-box sequences are in bold, while the single underlining denotes the sequences to which the primers bound. Each primer annealing site has been labelled with the name of the primer that binds there.

Figure 5.2 DNA Sequence of the *nod*B *nod*-Box



5.1.2.1 DNA Source and Concentration

The PCR is an enzymatic amplification process which can yield large amounts of a desired DNA region from only a small amount of often very crude starting DNA. In fact, high yields of PCR products A1, B1 and B4 were obtained from using whole cells as the DNA template (Table 5.1). However, for products whose size was under 400bp, a cleaner DNA template was required. In this case, a rapid boil preparation of the plasmid containing the *nod*-box of interest provided acceptable starting DNA.

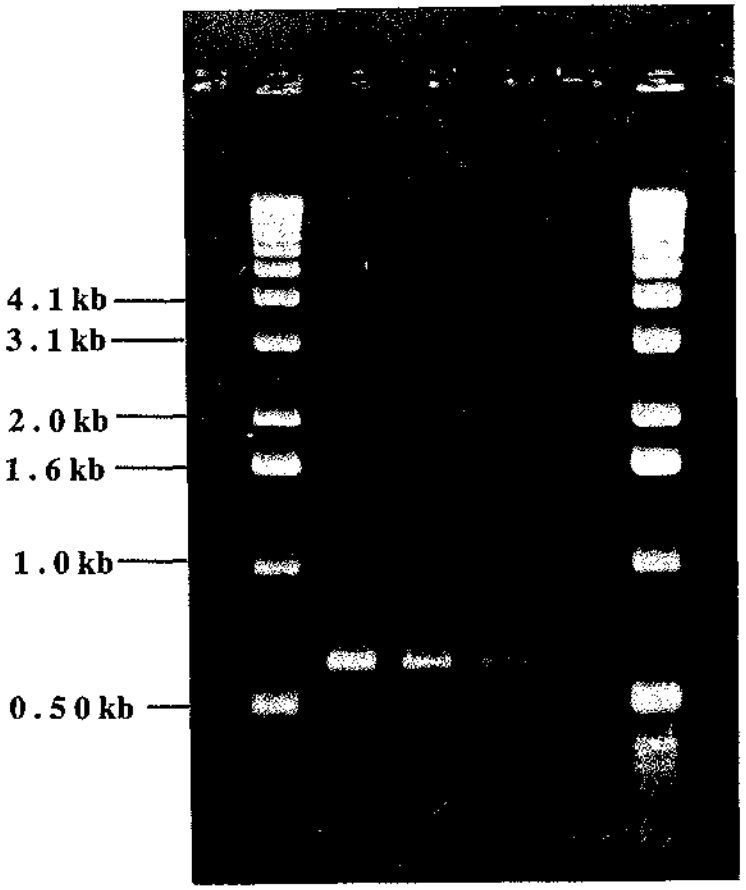
Not only was the purity of the DNA template of vital importance, but so was its concentration in the reaction mixture. The addition of an excess of plasmid DNA into the PCR caused extensive non-specific priming to occur. This could be observed on an agarose mini-gel as a range of DNA bands instead of the single band of interest (see Figure 5.3). To overcome this problem, less template was included in the PCR. Often, non-specific priming was eliminated only after the amount of starting DNA was reduced to 1ng.

5.1.2.2 Annealing Temperature

Increasing the annealing temperature enhances discrimination against incorrectly annealed primers, and reduces misextension of incorrect nucleotides at the 3' end of primers. Hence, to obtain PCR products with a high fidelity, it is essential to provide stringent annealing temperatures. This is particularly important in the early stages of the PCR. The fragments formed in the first few cycles of PCR form the amplification products upon which the massive amplification takes place. Any misprimed PCR product will, therefore, be greatly amplified resulting in a vast quantity of non-specific DNA. The optimum temperature for primer annealing depends on such factors as the base composition, length and concentration of the amplification primers. As there are two distinct primers in each PCR, it follows that the annealing temperature had to be optimised for each individual set of primers.

In this investigation, it was considered unlikely that the optimum annealing temperature would lie outside the 45°C-55°C temperature interval. Therefore, the initial trial run was always done at 50°C. The absence of PCR products at this temperature probably resulted from an overly stringent annealing temperature, and the next trial was performed at 48°C. Conversely, if several PCR-generated bands appeared on an agarose gel, it was clear that non-specific priming had occurred during the PCR. Hence, a more stringent primer annealing temperature was applied in the next trial. This method of trial and error continued until an appropriate yield of specifically primed DNA was routinely generated by the PCR.

Figure 5.3. Effect of DNA Concentration on PCR



1. 1kb ladder
2. PCR containing 200 ng template DNA
3. PCR containing 20 ng template DNA
4. PCR containing 2 ng template DNA
5. PCR containing 0.2 ng template DNA
6. 1kb ladder

Figure 5.3.

The effect of varying the amount of template DNA in a PCR has been demonstrated. An excess of template DNA (lanes 2 and 3) resulted in non-specific DNA products, while an insufficient amount of template resulted in a poor yield of the desired DNA product (lane 5). The optimum amount of template for this reaction was 2ng (lane 4).

5.1.2.3 Primer Concentration

Typically, the concentration of the primers within a PCR are between 0.1 and 0.5μM. At higher concentrations non-specific priming can occur resulting in the accumulation of one or more undesirable product(s). High levels of primers also increase the probability of generating template-independent artifacts termed primer-dimers. If either of these events are initiated, low yields of the desired PCR product usually result. However, this investigation found that increased primer concentrations were occasionally essential to produce acceptable yields of PCR product.

For a summary of the PCR products and the conditions used to obtain these products see Figure 5.7 and Table 5.1.

Table 5.1 **Generation of PCR Products**

	PCR Product	PCR Primers	Fragment Size (bp)	Annealing Temperature (°C)	Template DNA	PCR Conditions (see Table 2.4)
nodA ***nod*** - box	A1	BS3 / BS4	494	45	Whole cells	Programme 1.
	A2	BS3 / SP2	408	48	Plasmid DNA (20ng)	Programme 2.
	A3	SP1 / SP2	145	52	Plasmid DNA (2ng)	Programme 2 (excess primers)
nodB ***nod*** - box	B1	BS5 / BS6	739	45	Whole cells	Programme 1.
	B4	SP3 / BS6	552	50	Whole cells	Programme 1.

For identification of PCR primers see Figures 5.1 and 5.2

5.2 CLONING OF PCR PRODUCTS

Ultimately, the PCR products were destined to be cloned into pSPV4, the bidirectional vector constructed in this project. However, at this stage, the direct cloning of these products into pSPV4 would create insurmountable logistical problems. The most overwhelming drawback would be the lack of a convenient system which could select positive recombinants over background vector-only constructs. The inefficient nature of cloning small PCR fragments would mean the subsequent screening of unacceptably large numbers of potential clones. Notwithstanding this cloning handicap, the sequence confirmation of the PCR products would also present problems. Although direct sequencing of a portion of a plasmid (or genome) is technically possible, it continues to be a technique fraught with difficulty and hence was not seriously contemplated here.

Undoubtedly, an alternate, more manageable cloning vector had to be utilised in the characterisation of the PCR products. The foremost prerequisite in such a vector was the presence of an effective selection system. The blue/white selection system is a common technique for identifying recombinant clones. In this system, insertion of DNA into an MCS located within a *lacZ*α gene results in inactivation of the α-peptide. In the presence of X-gal, colonies containing the disrupted gene are white, while those without an insert are blue. One cloning vector that employs this labour-saving *lacZ* chromogenic substrate screen is the widely used plasmid pUC118. Measuring only 3kb in length, this high copy number plasmid was perfectly suited to act firstly as a cloning vector and later as a sequencing vector.

5.2.1 Cloning PCR Products into pUC118

Despite the broad applicability of PCR, a simple and efficient method of cloning the resultant products has not been universally accepted. A number of methods have emerged and three of these have been employed in this study.

5.2.1.1 T-Tailed Vector

The inherent terminal transferase activity associated with *Taq* polymerase has been found to frequently incorporate an adenosine base onto the 3' ends of a DNA duplex (Clark, 1988). This template-independent activity results in single nucleotide overhangs which renders the direct ligation of these fragments into a blunt-ended vector very inefficient. To overcome this problem, Marchuk *et al.* (1991), reported the use of *Taq* polymerase to add a thymine base onto the 3' termini of a blunt-ended vector molecule. This

provides a complementary single base overhang which is designed to enhance the efficiency of ligation (Figure 5.4a). Although the addition of T-tails is primarily to provide cohesive ends, it also has the secondary function of reducing vector-only background. Linearised vector molecules carrying single nucleotides on their 3' termini cannot religate on themselves. Hence, theoretically, the only residual background should be due to uncut plasmids. However, inefficiencies in *Taq* polymerase terminal transferase resulted in an appreciable proportion of linearised plasmids remaining untailed and subsequently blunt-end religating. The recircularisation of these unmodified vector termini was found to contribute significantly to the undesirable background.

While in theory, the T-tailing method appears to be an effective cloning system, it does have some weaknesses. The major disavantage was found to be the unexpectedly low proportion of PCR products that were originally bearing a 3' adenosine nucleotide. Presumably the weak terminal transferase activity of *Taq* polymerase left a variety of termini ranging from blunt ends to 3'overhangs of variable length. It appeared that termini exhibiting an adenosine overhang were, in fact, a minority of the total population of ends. As most termini were incompatible with the single nucleotide vector overhangs, a low frequency of ligation was observed. Three PCR products (A1, B1 and B4) were cloned using this method, yielding plasmids pSPA1 pSPB1 and pSPB4.

To significantly increase the effectiveness of the T-tailed vector, further manipulation of amplifed products usually had to be undertaken. These PCR products were modified by the *Taq* polymerase, so that an A-tail was introduced onto any fragments bearing a blunt end. While this does not solve the problem of fragments bearing ragged ends, it did result in an improved efficiency of ligation. This improvement, however, was not substantial enough to recommend that the T-tailing method be considered appropriate for routine cloning.

5.2.1.2 End-Filling of PCR Products

The unpredictable nature of the PCR product termini appeared to be the downfall of the T-tailing method outlined above. One way of circumventing this complication is to fill in the overhangs (Method 2.6.2). The fact that no dependence is placed on the identity of the unpaired base(s) is the great advantage conferred by this method. After treatment, every PCR product was expected to possess blunt ends, and therefore have the potential to be ligated into a *Sma*I-cut vector.

The maximisation of fragments available for cloning by blunt-ending, does have drawbacks, however. The additional manipulation step incurs small yet significant losses of insert DNA. This depletion of PCR product is especially evident when the fragment size is less than approximately 300bp. The second disadvantage of this method is associated with the vector termini. In order to prevent the blunt-end religation of the plasmid ends, the removal of the 5' phosphate group must occur. This reaction was mediated by the calf alkaline phosphatase (CAP) enzyme (Method 2.6.3), and had to be done thoroughly to abolish high vector background, yet not so extensively as to inflict other non-specific damage on the vector (Figure 5.4b).

Unquestionably, these minor inconveniences contributed little towards the low frequency of positive recombinants observed in this investigation. The major contributing factor must certainly have been the poor efficiency of cloning small blunt-ended products into much larger vector molecules. Interestingly, when inserts were found in recombinant plasmids, they frequently represented multimers of the intended PCR product. For reasons currently unknown, this multimeric construct was almost invariably in the form of a triplet. One PCR product that was successfully ligated into pUC118 was A3 resulting in the plasmid pSPA3.

The unsatisfactory ligation efficiency exhibited by single end-filled PCR products and blunt-end vectors, precludes this method from being used extensively as a routine protocol.

5.2.1.3 Addition of *Sma*I Into Ligation Mixture

The CAP treatment of vector ends is not infallible, and often a low background of non-recombinants can be observed after a ligation event. With the blue/white selection system employed by pUC118, these vector-only plasmids do not represent a screening hindrance. However, any circularised plasmid not containing an insert engenders a loss in cloning potential.

To remedy this situation, an adaptation of the previously described end-filling protocol was undertaken. This strategy is based on the knowledge that when DNA, cleaved with a blunt-end cutter, religates on itself the enzyme site is reconstituted. Therefore, addition of the restriction enzyme *Sma*I increases the proportion of vector molecules accessible for insert ligation.

Figure 5.4

a) **T-Tailed Vector**

The terminal transferase activity of *Taq* polymerase often adds an adenosine base onto the 3' end of the PCR product. This made ligation into the *Sma*I site of pUC118 very inefficient. Using *Taq* polymerase to incorporate a thymine base onto the 3' end of the blunt-ended vector improved the efficiency of the ligation due to the complementarity of adenosine and thymine.

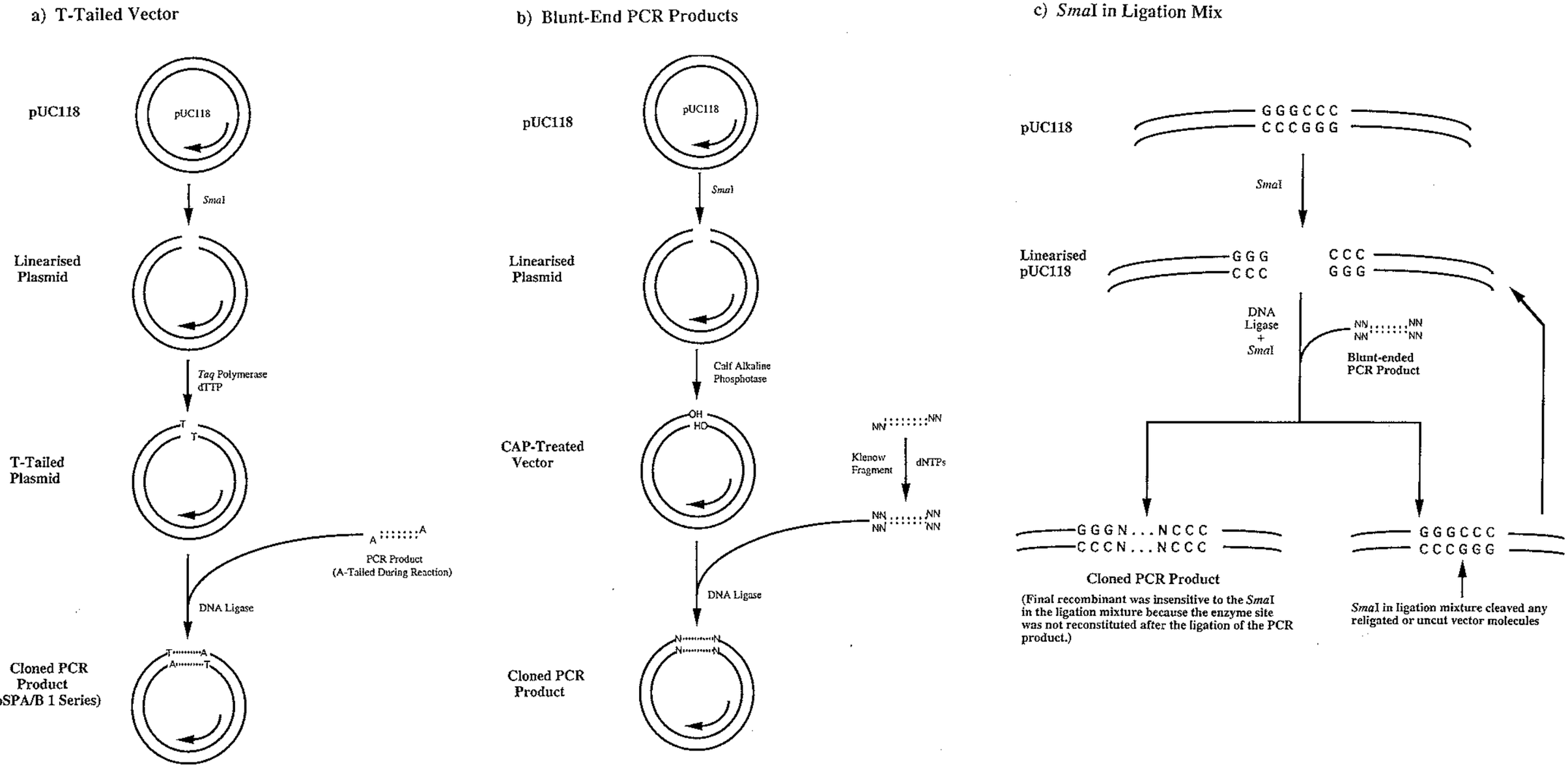
b) **Blunt-End Ligation**

Inefficiencies in the terminal transferase activity often left 3' ends of variable lengths on the PCR products. To overcome the non-uniformity of these termini, end-filling was used to create blunt-ends. This maximised the number of PCR products with the potential to ligate into pUC118.

c) ***Sma*I In Ligation Mixture**

The addition of *Sma*I in the ligation mixture was an extension of the blunt end ligation method (b). The inclusion of this enzyme decreased the proportion of vector molecules that were unavailable for ligation due to self-ligation. Any vector that did ligate to itself, automatically reconstituted the *Sma*I site, and was therefore susceptible to *Sma*I cleavage. These *Sma*I-cleaved vectors were subsequently available for further cloning of PCR products. A recombinant containing a PCR product was insensitive to the *Sma*I in the ligation mixture because the enzyme site was not reconstituted after the ligation.

Figure 5.4. Strategies for Cloning PCR Products



For this procedure to be applied successfully, the nucleotide sequence of the PCR product must be known. This sequence must not contain any *Sma*I cleavage sites, nor must it regenerate a *Sma*I site when it is blunt-end ligated into the vector. Should either of these criteria not be fulfilled, this technique will destroy the desired products as readily as the background vector-only molecules. Clearly, this would be a counterproductive course of events, and would result in an extremely low yield of desirable clones (Figure 5.4c).

Consequently, the additon of *Sma*I into the ligation mixture has a positive effect on the efficiency of ligation in most circumstances. However, this technique does not address the low efficiency of blunt-end ligation experienced in the earlier section. Therefore, it's widespread use may still be limited by the corresponding low efficiencies of ligation. One PCR product that was cloned in this manner was A3. The resulting plasmid was designated pSPA3.

5.3 IDENTIFICATION OF RECOMBINANT PLASMIDS CONTAINING PCR PRODUCTS

MC1022 was transformed with the resulting constructs and then plated on selective media to obtain colonies (Method 2.6.6). White, ampicillin-resistant, single colony isolates were cultured and small-scale rapid boil preparations made (Method 2.4.2). The plasmids were then cut with *Kpn*I and *Hind*III and run on a mini-gel to check for the presence of the correct inserts (Methods 2.6.1 and 2.5.3). See Figure 5.5 and lanes 2, 3, 4, 5 and 8 of Figure 5.6.

5.4 SEQUENCE DETERMINATION OF CLONED PCR PRODUCTS

It was essential to guarantee that no large-scale sequence discrepancies had been introduced by the PCR. Verification of the fidelity of the clone was established through sequence analysis.

The integration of a filamentous phage f1 *ori* and packaging signals allows pUC118 to generate single-stranded plasmid DNA in the presence of a helper phage (Dotto *et al.*, 1981; Dente *et al.*, 1983; Vieira and Messing, 1987). Single-stranded sequencing yields greater amounts of legible data than the comparable double-stranded method. However, as the major consideration of the sequence analysis was to determine the integrity of the PCR product, it was not essential to generate immense stretches of sequence data.

Figure 5.5. PCR Product Cloned into pUC118

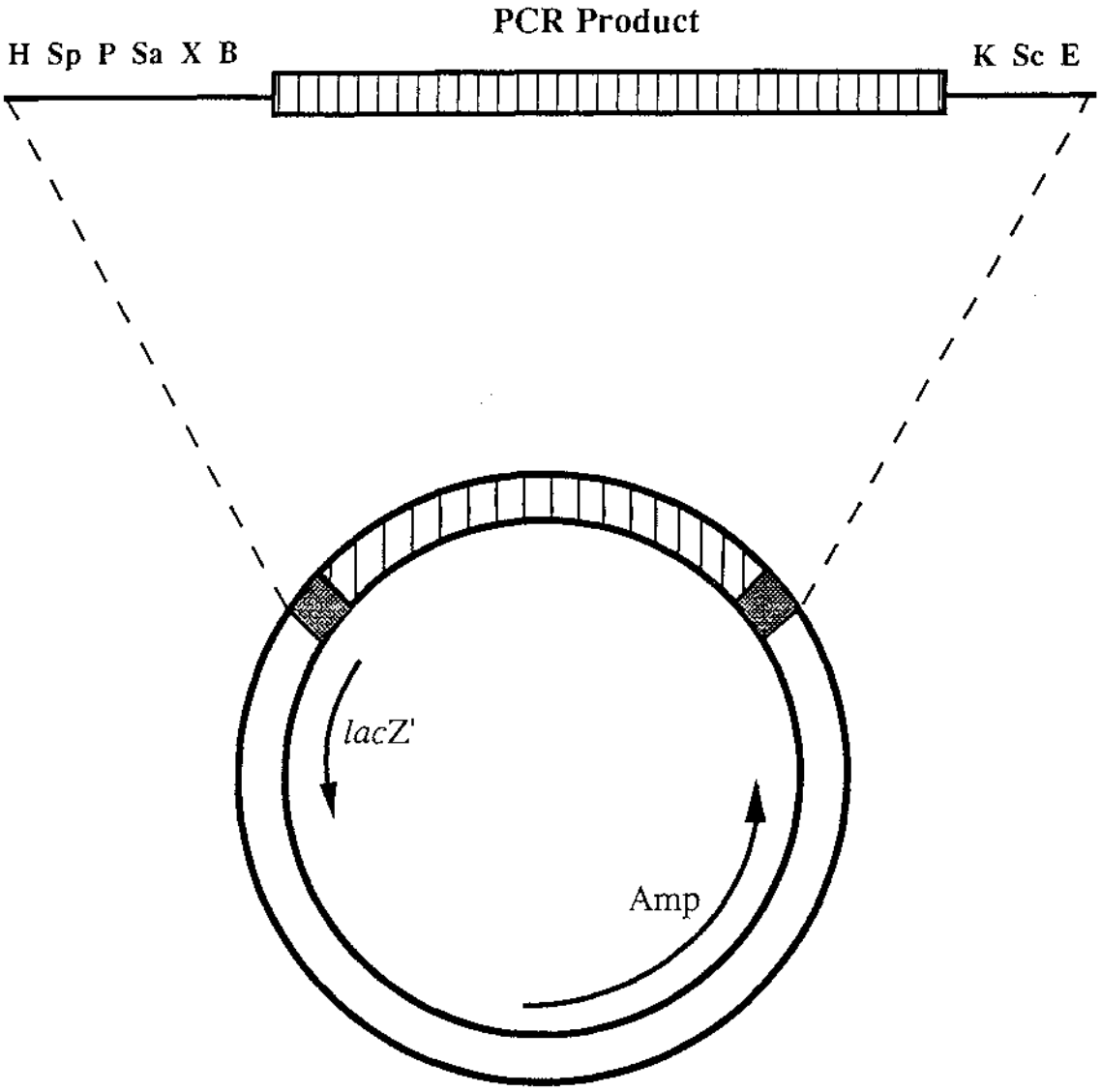
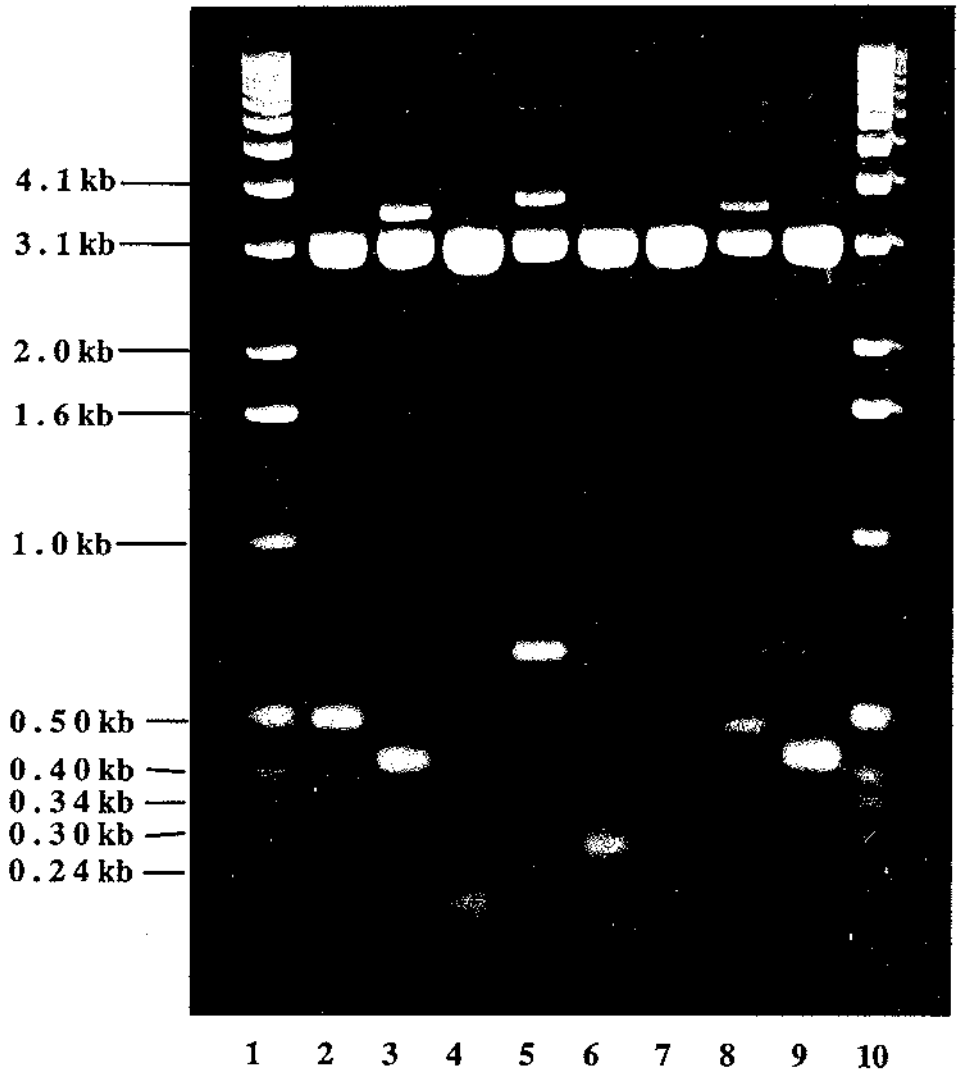


Figure 5.5

Diagrammatic representation of a *nod*-box-containing PCR product cloned into the *Sma*I site of pUC118. The hatched lines denote the cloned PCR product while the darkly hatched regions refer to the MCS. Enzyme sites preserved after the ligation event are shown in the enlarged region: H-*Hind*III; Sp-*Sph*I; P-*Pst*I; Sa-*Sal*I; B-*Bam*HI; K-*Kpn*I; Sc-*Sac*I; E-*Eco*RI;

Figure 5.6. Restriction Analysis of PCR-Generated Inserts in pUC118



1. 1kb ladder
2. pSPA1 *Eco*RI + *Hind*III cut
3. pSPA2 *Eco*RI + *Hind*III cut
4. pSPA4 *Eco*RI + *Hind*III cut
5. pSPB1 *Kpn*I + *Hind*III cut
6. pSPB2 *Eco*RI + *Hind*III cut
7. pSPB5 *Eco*RI + *Hind*III cut
8. pSPB4 *Kpn*I + *Hind*III cut
9. pSPB3 *Eco*RI + *Hind*III cut
10. 1kb ladder

Hence, the simpler double-stranded sequencing protocol was employed (Method 2.9). Using the -40 Universal Primer and the Reverse Primer, this method was used to determine the sequence of the boundaries between the PCR product and the pUC118 polylinker. Not only did this data demonstrate correctness of the insert sequence, but it also confirmed the extent of the regions bracketing the *nod*-box. In all cases, the generated sequences agreed almost precisely with the published data (Collins-Emerson, 1991). The few discrepancies were attributed to the error frequency associated with the *Taq* polymerase. Deduction of the insert orientation with respect to the pUC118 was also facilitated by the sequence determination.

5.5 PRODUCTION OF ADDITIONAL CLONES POSSESSING REDUCED FLANKING REGIONS

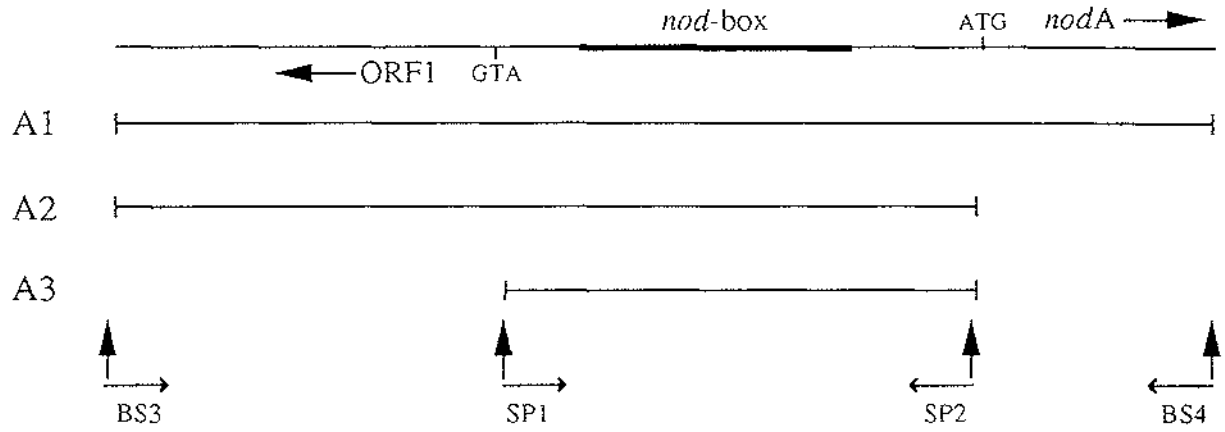
Sequence analysis by Collins-Emerson (1991), established the presence of an *Eco*RI enzyme site to the right of the *nod*-box associated with the *nod*B gene. Utilisation of this site enabled three extra 'deletion' fragments to be cloned and subsequently analysed. Importantly, these clones allowed the regulatory effect conferred by the region to the right of the *nod*B *nod*-box to be investigated. This stretch of DNA was expected to contain promoter elements regulating the *nod*D gene if they were located outside the *nod*-box.

The clones designated pSPB1 and pSPB4 both contained this *Eco*RI enzyme site. Restriction at this site, coupled with cleavage at a second *Eco*RI site within the MCS allowed cohesively-ended fragments to be produced (Figure 5.7). Significantly, these fragments of DNA were bereft of any of the regulatory regions to the right of the *nod*-box and despite their small size, were considerably more efficient to clone than a PCR product.

5.5.1 Construction of pSPB2, pSPB3 and pSPB5

A small-scale rapid boil preparation of pSPB1 was digested with *Eco*RI (Methods 2.4.2 and 2.6.1). This cleavage resulted in the liberation of the region between the *Eco*RI site in the PCR product and the site in the MCS of pUC118. Contained within this excised fragment was the DNA sequence flanking the right-hand side of the *nod*B *nod*-box. Consequently, the putative promoter activity of the *nod*-box was still associated with the initial pSPB1 plasmid.

a) *nod*A *nod*-box



b) *nod*B *nod*-box

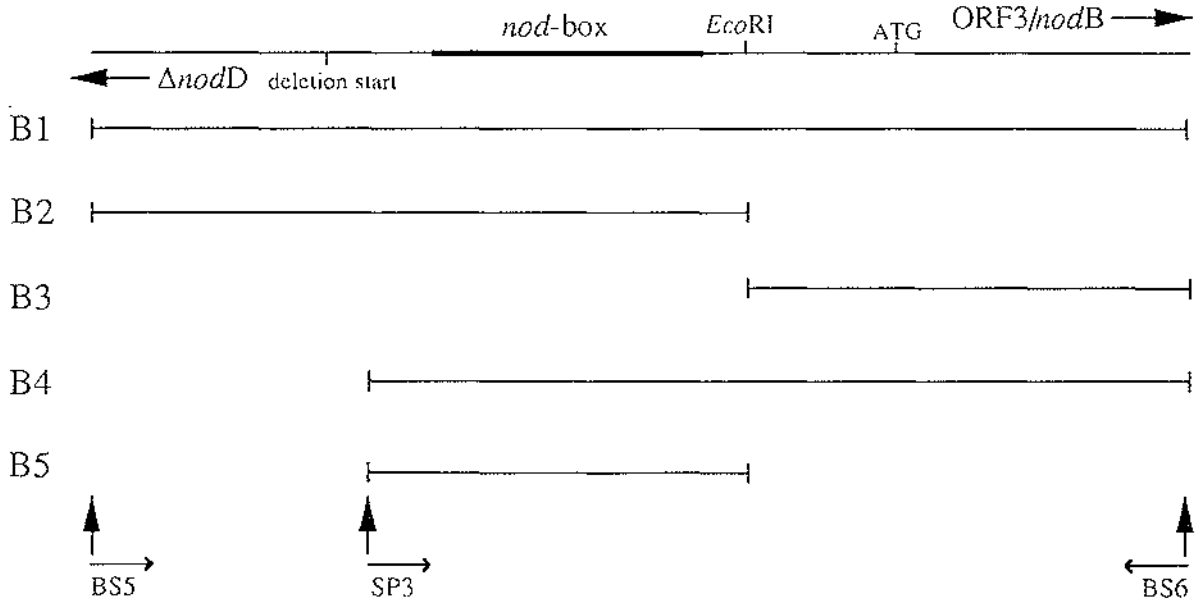


Figure 5.7.

Extent of the flanking regions for each PCR-generated product examined in the course of this investigation. The fragments associated with the *nod*A and *nod*B *nod*-boxes are represented in (a) and (b) respectively. The primers which specify the borders of the fragments are indicated below.

The two restricted fragments were later separated on a 1% Seaplaque gel (Method 2.5.5). After purification, the larger fragment containing the majority of the pSPB1 clone, was intentionally self-ligated. The resulting construct, entitled pSPB2, therefore consisted of the entire regulatory domain of interest, apart from the region located to the right of the *nod*-box.

The *Eco*RI fragment liberated earlier from pSPB1 was also isolated from the Seaplaque gel. The purified 468bp fragment was later cut with *Kpn*I and ligated into an *Eco*RI/*Kpn*I double-digested pUC118 vector (Method 2.6.1). Employment of *Kpn*I, in this instance, was simply to prevent vector-religation and was used in preference to CAP-treating the vector termini. The outcome of this ligation was the clone pSPB3 which was intended to constitute a negative promoter control as it did not include a *nod*-box sequence.

One further clone was produced in a similar manner to pSPB2. The only difference was that the *Eco*RI/*Kpn*I fragment was excised from the plasmid pSPB4 in lieu of pSPB1. Since these two clones had inserts cloned in opposite orientations, the displaced 74bp fragment did not exclusively contain the DNA region flanking the *nod*B *nod*-box. Instead, it carried the *nod*-box itself, along with vastly depleted flanking regions. The reduction in these bordering segments can be attributed to the positions of the *Eco*RI site and the SP3 primer binding site associated with pSPB4. These sites were found on opposite sides of the *nod*-box, yet closely neighbouring this suggested regulatory region. Hence, once the 74bp fragment was ligated into the *Eco*RI/*Kpn*I restricted vector, the resulting construct, pSPB5, was considered to be a minimal *nod*-box-containing clone.

All of the newly generated clones were later electroporated into MC1022, and identified using the selection system outlined in Section 5.3. Similarly, these clones were confirmed (Figure 5.5 and lanes 6, 7 and 9 of Figure 5.6) and sequenced in exactly the same fashion as the earlier PCR clones (Section 5.4). Predictably, the sequence agreed rigorously with the sequence data produced earlier for the pSPB1 and pSPB4 clones.

5.6 CLONING PCR-GENERATED PRODUCTS FROM THE pSPA/B 1 SERIES INTO THE BIDIRECTIONAL VECTOR

The production of eight clones, each defining a different pair of flanking sequences (usually included with the *nod*-box) was sufficient for this investigation. These fragments, preliminarily cloned into pUC118, were subsequently excised and re-cloned into the bidirectional expression vector, pSPV4, constructed as outlined in Chapter 4. As emphasised earlier, divergent reporter genes of pSPV4 enabled the activity initiated from a cloned fragment to be read simultaneously in both directions. Hence, all of the PCR-generated products only needed to be re-cloned once. The inserts contained within pSPB2 and pSPB4 were cloned in both orientations as a control of the reciprocity of the bidirectional strategy and in order to get a measure of the relative activities of the two reporter genes.

5.6.1 Cloning Strategy

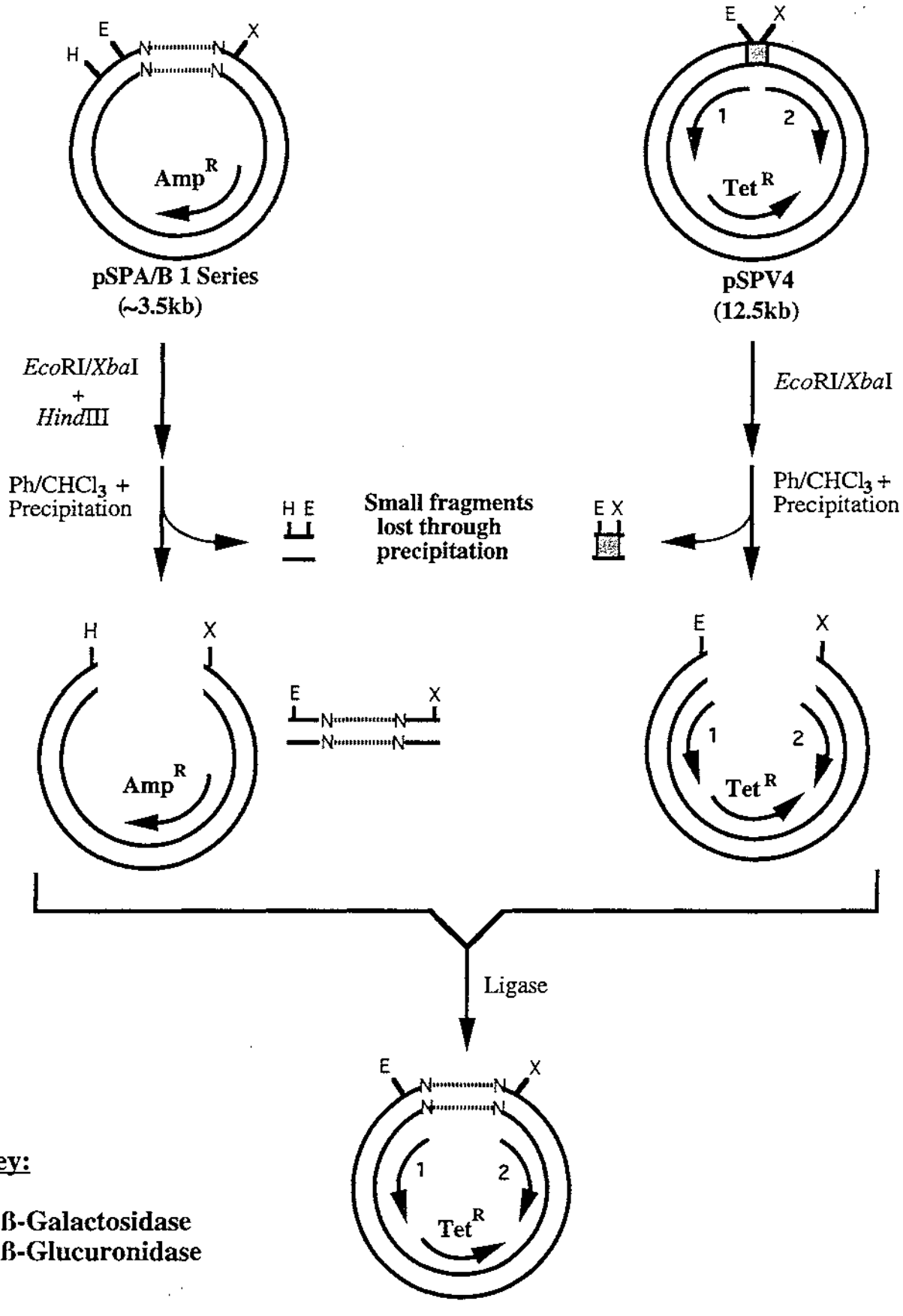
Small-scale plasmid preparations of the pSPA/B 1 series were individually produced using the rapid boil method (Method 2.4.2). These constructs were then digested with *Eco*RI, *Xba*I and *Hind*III (Method 2.6.1) to release the *Eco*RI/*Xba*I DNA fragment of interest. The *Hind*III cut ensured the larger fragment was incompatible with the cloning vector. Consequently, the only pieces of DNA with *Eco*RI/*Xba*I termini were the PCR-generated inserts. (It should be noted that fragments B1 and B4 each contained an *Eco*RI site, rendering this strategy ineffective. The plasmids containing these fragments were cut with *Kpn*I instead of *Eco*RI).

To accommodate these individually-prepared insert regions, a large-scale preparation of pSPV4 was firstly purified on a CsCl gradient, and then digested with *Eco*RI and *Xba*I (Methods 2.4.1 and 2.6.1). Once these enzymes had been inactivated, the restricted DNA was added to the insert preparations which had also been freed from enzyme activity. The *Eco*RI/*Xba*I cohesive ends endowing each of the eight inserts were subsequently ligated into the *Eco*RI/*Xba*I restricted sites of the pSPV4 bidirectional vector (Method 2.6.4). The enzyme mediating this reaction, T4 DNA ligase, was later heat inactivated and a 'suicide cut' employed to reduce the undesirable vector-only background. The enzyme chosen to fulfil this role was *Pst*I.

Figure 5.8.

Strategy for cloning PCR-generated inserts (contained within pUC118) into pSPV4. Cleavage of the *Eco*RI/*Xba*I-cut pSPA/B 1 series with *Hind*III circumvented the need to isolate the fragment of interest as the only fragment bearing *Eco*RI/*Xba*I cohesive ends was the one containing the PCR product. The other products of the triple digestion were not compatible with the *Eco*RI/*Xba*I cut vector molecule, and were therefore not cloned. Final recombinants were termed the pSPA/B 10 series.

Figure 5.8. Cloning PCR Inserts Into Bidirectional Vector pSPV4



Key:

1. ß-Galactosidase
2. ß-Glucuronidase

There is a unique *Pst*I enzyme site within the MCS of the intact pSPV4 vector, but this site gets excised when the plasmid is successfully restricted with both enzymes. As there were no *Pst*I sites in the *Eco*RI/*Xba*I insert fragments, their subsequent incorporation into the enzymatically-prepared vector resulted in the complete absence of *Pst*I enzyme sites. Hence, the *Pst*I digestion vastly reduced the number of potential recombinants that had to be screened. Following the 'suicide' digests, enzyme activities were terminated by heat inactivation and the residual salts were removed by microdialysis (Method 2.6.7).

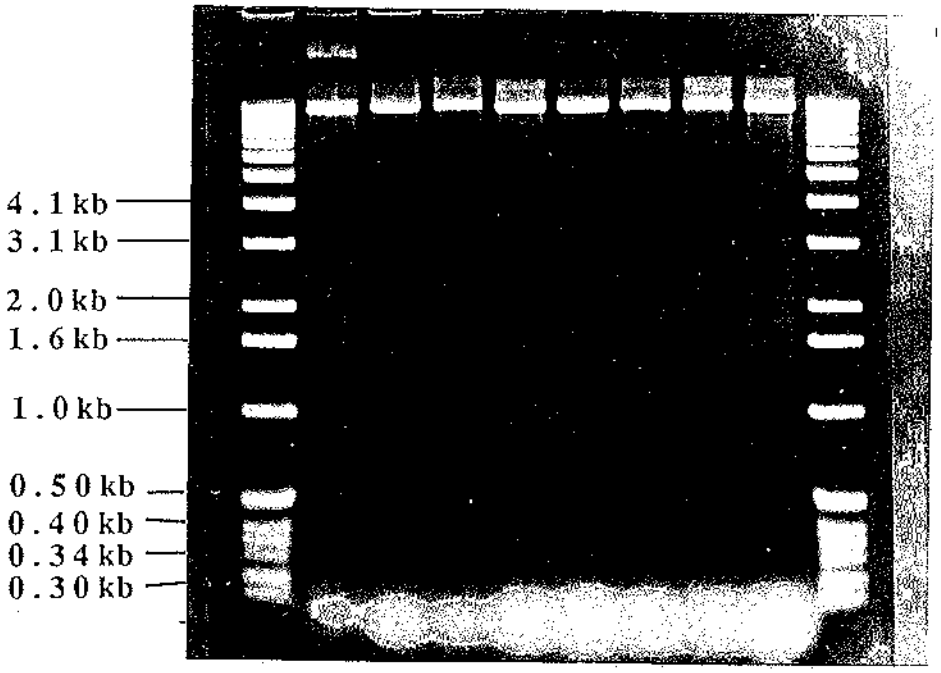
Due to complications inherent in extracting plasmid DNA from *Rhizobium* species, the eight constructs were first electroporated into the *rec*A$^-$, *end*A$^-$ *E. coli* strain DH1 (Method 2.6.6). Tetracycline-resistant single colony isolates were cultured and their plasmids isolated using the rapid boil method (Method 2.4.2). These potential recombinant plasmids were later digested with *Hind*III and *Xba*I to release any ligated insert and the presence of the PCR-generated inserts was checked on a 1.5% agarose mini-gel (Figure 5.9). Recombinants containing the desired inserts were termed the pSPA/B 10 series.

5.6.2 Cloning PCR-Generated Products in pSPV4 in the Opposite Orientation

In order to check the reciprocity of the reporter genes in pSPV4, two PCR-generated fragments were subsequently cloned into pSPV4 in the reverse orientation. To achieve this reversal in polarity, the restriction enzyme sites *Eco*RI and *Sph*I were utilised from the pSPV4 MCS (Figure 4.11). In pUC118, the order of these enzymes with respect to *Xba*I and a cloned PCR product, is *Eco*RI-insert-*Xba*I-*Sph*I. As the MCS in pSPV4 is arranged *Sph*I-*Eco*RI-*Xba*I, the insert can easily be inverted by cloning the *Eco*RI/*Sph*I fragment excised from the pSPA/B 1 series of constructs.

In theory, it should be possible to clone all of the PCR-generated products into pSPV4 in both orientations. However, the presence of an *Sph*I site in A1 ruled out this fragment, while the *Eco*RI site mentioned in the previous section made cloning B1 and B4 much more difficult. In this latter case, substituting *Kpn*I for *Eco*RI was not as effective as it was in Section 5.6.1, because this left no 'suicide enzyme sites' available to reduce the vector-only background. Hence, although cloning these fragments was possible, it presented a problem during the screening of positive recombinants. Despite this disadvantage, the PCR product from pSPB4 was cloned into pSPV4 in the reverse orientation. The resulting plasmid was designated pSPB14a and the plasmid containing the insert in the original orientation was termed pSPB14b. Another PCR-generated product cloned into pSPV4 in the reverse orientation was B2. This insert was cloned

Figure 5.9. Restriction Analysis of PCR-Generated Products Cloned into pSPV4



1. 1kb ladder
2. pSPV4 uncut
3. pSPA11 *Eco*RI + *Xba*I cut
4. pSPA12 *Eco*RI + *Xba*I cut
5. pSPB11 *Kpn*I + *Xba*I cut
6. pSPB12 *Eco*RI + *Xba*I cut
7. pSPB14 *Kpn*I + *Xba*I cut
8. pSPB13*Eco*RI + *Xba*I cut
9. pSP11 *Eco*RI + *Xba*I cut
10. 1kb ladder

Figure 5.9.

Excision of the various PCR-generated fragments from the bidirectional reporter vector pSPV4 (12kb). Only the plasmids containing larger inserts have been included on this gel as they are easier to observe.

as an *Eco*RI/*Sph*I fragment using *Kpn*I as a 'suicide enzyme'. The plasmid bearing the fragment in reverse polarity was designated pSPB12b, while the original B2-containing pSPV4 recombinant was named pSP12a.

5.6.3 Generation of Control Strains

Three different types of control were used to ascertain the effectiveness of the bidirectional reporter vector, pSPV4, in *R. loti*. The first control consisted of untransformed *R. loti* PN184 cells. As this was the strain into which the plasmids constructed in Sections 5.6.1 and 5.6.2 were to be introduced, this first control was used to give an indication of the endogenous β-galactosidase and GUS. For pSPV4 to be serviceable in *R. loti*, the basal levels of each reporter gene product must be relatively low. This eliminates the difficulties associated with detecting low or fluctuating reporter gene activity against high endogenous levels. As mentioned earlier (Section 4.2), neither β-galactosidase nor GUS are thought to be present at high concentrations, thereby circumventing this problem.

A second form of negative control was also employed in the analysis of pSPV4. This control consisted of *R. loti* PN184 cells which had been transformed with pSPV4 containing no insert. The resulting clone, designated SP100, was intended to determine the levels of activity attributable to the vector alone. It is thought that the presence of a functional rbs is enough to ensure that any randomly-primed transcriptional read-through will be translated into potentially-active reporter gene product (Sharma, pers. comm.). The termination codons found in all three reading frames upstream of the portable rbs (Section 5.1.1), were expected to minimise this undesirable activity.

To check the functionality of the divergent reporter genes in pSPV4, the *R. meliloti nod*-box was cloned into this vector. The activity of this *nod*-box has been well-characterised making it an appropriate positive control. Using the strategy outlined in Section 5.6.1, the control sequence was ligated into an *Eco*RI/*Xba*I-cut pSPV4 vector (Method 2.6.4), before being electroporated into *E. coli* DH1 cells (Method 2.6.6). Recombinants containing the 0.48kb insert were identified as in Section 5.6.1. The resulting clone of interest was named SP11. After pSP11 had been transformed into *Rhizobium* cells to yield strain SP101, it was expected that the cloned *nod*-box would promote strong constitutive activity in the *lacZ* direction and highly inducible activity (in the presence of luteolin) in the *gus*A direction. Such results would help to confirm the utility of pSPV4.

5.7 ELECTROPORATION OF BIDIRECTIONAL REPORTER CONSTRUCTS

Once the plasmids of interest had been constructed and identified in *E. coli* (Sections 5.6.1, 5.6.2 and 5.6.3), they were transformed into *R. loti* strain PN184. As determined in Chapter 3, the fastest and most efficient method of transforming *Rhizobium* is the technique of electroporation. Hence, after isolating the plasmid DNA from the appropriate *E. coli* strains (Method 2.4.2), the constructs were electroporated into pSPV4 (Methods 3.3.1 and 3.3.2).

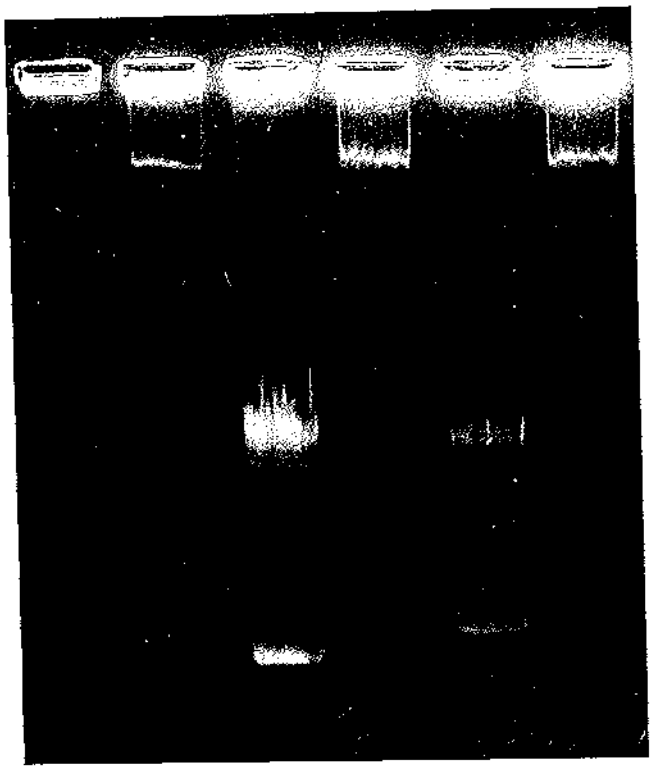
Electroporation of a competent cell sample lacking DNA gave rise to absolutely no colonies when plated on antibiotic-containing media. Therefore, any colonies produced from cells electrotransformed in the presence of plasmid DNA, were presumed to contain the intended Tet^R construct. To verify this assumption, the presence of introduced plasmid DNA was observed on an Eckhardt gel (Method 2.12). This gel is presented as Figure 5.10. Lane 2 of this gel demonstrates the presence of the indigenous megaplasmid of R. loti. The remaining lanes in the Eckhardt gel confirm the successful transfer of the plasmids pSPV4 and pSPB11 from DH1 into *R. loti* PN184 via electroporation. It is relevant to note that the plasmid bands corresponding to pSPB11 are slightly higher on the gel than those of pSPV4. This correlates with the presence of the 0.739kb insert in pSPB11.

5.8 ASSAY OF BIDIRECTIONAL REPORTER CONSTRUCTS

Once the presence of the desired constructs had been confirmed in *Rhizobium*, a number of enzymatic assays were performed. For each recombinant and control, the activity of the *lacZ* and *gusA* transcriptionally-fused genes was determined in the presence and absence of flavonoid inducers (Method 2.10). To obtain statistically meaningful results, the assays were repeated six times over several days (see Appendix). The averages of these bidirectional assays were subsequently calculated and presented in Tables 5.2, 5.3 and 5.4.

The results of the two negative control assays (Table 5.2) were as expected. Both PN184 and SP100 gave relatively low reporter gene activities. The apparent *lacZ* activity was slightly higher than the *gusA* activity due to the endogenous β-galactosidase levels in the cell. An explanation for the high constitutive activity (*lacZ*) and low inducibility (*gusA*) of the positive control SP101 is less easily

Figure 5.10. Eckhardt Gel Confirming the Presence of Plasmids in *R. loti*



1. *E. coli* strain DH1
2. *R. loti* strain PN184
3. DH1 + pSPV4
4. PN184 + pSPV4
5. DH1 + pSPB101
6. PN184 + pSPB101

Figure 5.10.

Lane 1 shows the absence of native plasmids in *E. coli* DH1 cells, while lane 2 shows the presence of the indigenous megaplasmid of *R. loti* PN184 cells. The remaining lanes demonstrate the presence of smaller pSPV4-derived plasmid DNA. Size comparisons of these plasmids in *E. coli* and *R. loti* confirm the successful transfer of this DNA via electroporation.

provided. As outlined fully in the Discussion, this anomalous activity is probably attributable to a low binding constant between the *R. loti* NodD protein and the *R. meliloti nod*-box.

The overall functionality of pSPV4 was confirmed by two sets of important results. Firstly, the longest *nod*-box-containing fragments, SPA101 and SPB101, gave results which were typical of *nod*-box promoter activity (Tables 5.3 and 5.4). This helped to verify that both of the reporter gene fusions were functional and responsive to promoters cloned into the MCS of pSPV4. The second data set that proved the functionality of pSPV4 was that reported by SPB102a and SPB102b (Table 5.4). The results observed from this pair of clones demonstrated the reciprocity of the two reporter genes, and therefore the utility of pSPV4.

As seen in Tables 5.3 and 5.4, the majority of the PCR-generated inserts cloned into pSPV4, have promoted the type of bidirectional activity expected from a conventional *nod*-box (constitutive expression in the *nod*D direction and flavonoid-inducible expression in the opposite direction). Apparent differences in promoter strength between some of the strains (eg SPA101 and SPA102 constitutive activity levels) are discussed in detail in the next chapter. It is thought that these differences may be due to either regulatory elements or variations in spacing between the *nod*-boxes and the reporter genes. Also considered in the Discussion, are the unexpected results observed for SPB103 and SPB104b. It shall be argued that the inordinately high constitutive activities reported for these strains are cloning artifacts and not true reflections of the promoter activity of the fragments.

With the functionality and applicability of pSPV4 demonstated by using the regulatory properties of the *nod*-box, this part of the project is complete. The analysis of the regulatory properties of the regions bracketing the *nod*A and *nod*B *nod*-boxes is presented in Chapter 6.

Table 5.2.

Assay results for the two negative controls, PN184 (no plasmid) and SP100 (pSPV4 containing no insert), and the positive control SP101. The *Rhizobium* strain SP101 carried the plasmid pSP11 which contained the *R. meliloti nod*-box cloned into pSPV4. SP101 was 'induced' by the *R. meliloti* flavonoid, luteolin. Constitutive activity was expected in the *lacZ* direction and luteolin-induced activity was expected in the *gusA* direction.

Table 5.3.

Bidirectional assay results for the SPA 100 series. The plasmids from these strains contain fragments bearing the *nodA nod*-box along with variably-lengthed bracketing regions.

Table 5.4.

Bidirectional assay results for the SPB 100 series. The plasmids from these strains contain fragments carrying the *nodB nod*-box along with variably-lengthed bracketing regions. One plasmid not containing a *nod*-box was pSPB13. The PCR-generated inserts contained within pSPB2 and pSPB4 have been cloned into pSPV4 in both orientations to give SPB102a/b and SPB104a/b respectively. The former pair of clones allowed the reciprocity of pSPV4 to be examined.

Table 5.2 Control Assays

Control	*lacZ* -SE	*lacZ* +SE	*gusA* -SE	*gusA* +SE
PN184	7	8	2	1.5
SP100	10	8.1	2.2	2.1
SP101 *(R. loti)*	241	211	2.4	2.4

Table 5.3 *nodA nod*-Box Bidirectional Assay Results

Construct	Transcription Towards *nodD* -SE	Transcription Towards *nodD* +SE	Transcription Towards *nodA* -SE	Transcription Towards *nodA* +SE
SPA101	48	45	1	29
SPA102	22	17	1.7	26
SPA103	6.3	8.4	4	20

Table 5.4 *nodB nod*-Box Bidirectional Assay Results

Construct	Transcription Towards *nodD* -SE	Transcription Towards *nodD* +SE	Transcription Towards *nodB* -SE	Transcription Towards *nodB* +SE
SPB101	50	46	9	81
SPB102a	18	19	1.9	6.6
SPB102b	7.5	7.1	2.5	19
SPB103	1.3	1.6	247	250
SPB104a	10	13	12	14
SPB104b	6.5	7.6	330	280
SPB105	11	12	4.5	36

Note: Underlined numbers denote activity reported by the *gusA* fusion

Chapter 6 - DISCUSSION

6.1 BIDIRECTIONAL REPORTER VECTOR CONSTRUCTION

The construction of the bidirectional reporter vector, pSPV4, involved a number of intermediary plasmids. It has not gone unnoticed that some of these transitional constructs may be of value to other researchers. This is particularly pertinent given the current trend towards the utilisation of *gus*A as a reporter gene.

The first plasmid intermediate which could be a useful reporter vector was pSPV1 (Figures 4.1 and 4.2). This construct contained the *gus*A coding region from pRAJ260 (Jefferson *et al.*, 1986) cloned in the opposite orientation to the *lac*Zα gene of pUC21 (Vieira and Messing, 1991). Accordingly, a vector with the potential to form translational fusions with the *gus*A coding sequence was formed. For the proposed fusion to occur, a promoter region through to a portion of an exogenous coding region, would have to be cloned into the MCS of the vector such that it was in-frame with the translational start codon of *gus*A. Similar translational fusions could presumably be formed using the *lac*Zα gene resulting in a vector with the potential to form translational fusions in either direction. Unfortunately, the blue/white selection system associated with the *lac*Zα gene was abolished after the *gus*A gene was cloned into the pUC21 MCS. Hence, the identification of recombinants may pose a small problem in this vector. The need to screen large numbers of potential clones could be avoided, however, by digesting the ligation mixture with an appropriate 'suicide enzyme' (see Section 4.4.2.4). Although the proposed *gus*A translational fusion in pSPV1 would have enormous potential as a histochemical reporter in many plant and mammalian systems, the number and variety of unique restriction enzyme sites in this vector also makes it especially useful in the construction of alternative vectors. Clearly, the easily-excisable *gus*A/MCS module from pSPV1 could be used in a variety of experimental contexts if cloned into an appropriate plasmid.

The other important functional intermediate was pSPV2 (Figure 4.1). In terms of its construction, this plasmid differed from its predecessor (pSPV1) in that it had the portable rbs module cloned in-frame with the *gus*A gene translational start site (Figure 4.4). The resulting reporter gene fusion was preferable to the previous construct as it could report on the promoter activity of any fragment cloned into the MCS of the vector. In the other direction, the potential *lac*Zα translational fusion was unaltered with respect to pSPV1. The lack of symmetry between the two reporter gene fusions may restrict the applicability of pSPV2. However, one potential function for this vector would be in the analysis of two aspects of single cloned fragments. Clearly, the DNA of interest would have to be cloned in both orientations for this strategy to be effective.

The final version of the bidirectional reporter vector, pSPV4, was obtained by cloning the *gus*A/rbs/MCS module into the *lac*Z reporter vector pMP220 (Figure 4.1). This cloning event resulted in both the *gus*A and the *lac*Z constructs divergently forming transcriptional fusions with a DNA fragment cloned into the MCS (Figure 4.11). Using this vector, it was possible to monitor the bidirectional activity of a cloned putative promoter element in the same culture of cells. This allowed the relative promoter activity of the fragment to be directly compared with the corresponding activity in the opposite direction. The significance of obtaining the relative bidirectional activity of a promoter was particularly important in this investigation as the precise physiological state of the cells may be crucial for *nod*-box regulation. Not only was this vector useful in terms of *nod*-box analysis, but it could also be used to assay transcription in any other organism in which IncP plasmids are functional, eg *Pseudomonas*, *E. coli* and *Agrobacterium*. In organisms where IncP plasmids can't be used, the *gus*A/rbs/MCS module of pSPV2 could be cloned into an existing *lac*Z reporter vector that was already compatible with the desired host.

6.2 ANALYSIS OF CONTROL ASSAYS (Table 5.2)

Three classes of control were designed to demonstrate that the activity of the pSP 10 plasmids was almost exclusively attributable to the ligated PCR fragment. Initially two forms of negative control were used to check that the background *lac*Z and *gus*A activity was not significant. The first of these examined the basal levels of untransformed *R. loti*. As predicted, the endogenous level of the *gus*A reporter gene was almost undetectable (2U). In contrast, the *lac*Z activity of *R. loti* was convincingly higher (10U). The reason for this low, yet significant, level of activity was due to the presence of a functional β-galactosidase gene in *Rhizobium*. The product of this gene is known to be present constitutively (Sharma, pers. comm.).

The second negative control, SP100, measured the activity of *R. loti* after it had been electrotransformed with the bidirectional reporter vector, pSPV4. With no insert cloned into the MCS of this vector, the activity of the reporter genes was expected to be comparable to the basal levels observed earlier. Indeed, the results indicate only a slight increase in the two activities. This analysis confirmed that the bidirectional reporter vector possessed no endogenous promoter activity. Had the vector been shown to contain a promoter element that operated independently of the cloned DNA fragment, then it could scarcely have been relied upon to yield meaningful results.

To ensure the functionality of the bidirectional reporter vector, it was important to clone and assay a known promoter. Construction of this positive control, termed pSP11 (Section 5.6.), was accomplished by inserting the *R. meliloti nod*-box region into the MCS of the vector. The divergent transcription regulated by this *nod*-box has been well-characterised making it an appropriate promoter control. Once this construct had been transformed into the native *R. meliloti* background it was bidirectionally assayed (Sharma pers. comm.). The results obtained were inconsistent with the results expected from earlier studies (data not presented). Sharma attributes the observed lack of *nod* gene inducibility to the complex interactions between the other *nod*D genes present in *R. meliloti*. Hence, to conclusively show the functionality of the vector in *R. meliloti*, a strain containing mutations in *nod*D_2, *nod*D_3 and *syr*M would have to be used.

The positive control plasmid, pSP11, was also electroporated into *R. loti* host cells (Section 5.6.1). The purpose of this transformation was to determine the functionality of a *R. meliloti nod*-box in a different rhizobial environment. Measurement of the bidirectional activity of this *nod*-box indicated that the *nod*ABC operon was not induced by luteolin yet demonstrated high constitutive expression in the *nod*D direction (Table 5.2). To explain these results, it is proposed that the *R. loti* NodD protein has a very low binding constant for the *R. meliloti nod*-box and that transcription in the *nod*D direction continued to be constitutive (Section 1.5.4.3). Even the presence of flavonoid inducer did not apparently facilitate the functional binding of the RNA polymerase. Hence, transcription of *nod*D continued at high levels, while transcription of *nod*ABC was not induced due to the proposed low binding constant between RNA polymerase and the *nod*ABC promoter. The uninducibility of *nod*ABC observed here supports the notion presented in Section 1.5.4.1, where it was hypothesised that protein-protein interactions between the NodD-flavonoid complex and the RNA polymerase, allow the latter to bind to the *nod*ABC promoter region and transcribe this operon.

The same control was re-assayed using *Lotus* seed exudate (data not presented). Interestingly, similar results were obtained supporting the proposal that the lack of *nod* gene induction was due to an inadequate interaction between the *R. loti* NodD protein and the *R. meliloti nod*-box carried by pSP11. It appears that the proposed protein-protein interactions between the induced *meliloti* NodD protein and the *meliloti* RNA polymerase may not be conserved by the induced *R. loti* NodD protein causing the altered binding constants. The great sensitivity of binding constants to both nucleotide sequence and amino acid sequence has already been highlighted in the λ repressor/*cro* system (Section 1.4). In that system, both minor alterations in the amino acid sequence in the helix-turn-helix region as well as minor alterations in the nucleotide sequence of the tripartite operator sites have a significant effect on the regulatory properties of the phage.

The fragment, B2, was used to test the symmetry of the bidirectional vector (Table 5.4). It was cloned from pSPB2 into pSPV4 in both orientations (Section 5.6.2), and the resulting strains, SPB102a and SPB102b (SPB102a/b), were assayed in both directions. The results clearly demonstrate that the type of bidirectional regulation observed in each clone was not a function of the reporter genes (Table 5.4). However, the magnitude of the reporter gene units were not directly relatable. For some reason, the *lac*Z values were always higher than the *gus*A values. The symmetry of the RNA polymerase binding sites means that this discrepancy probably does not lie at the transcriptional level. It is more likely that the β-galactosidase hybrid proteins are more stable than the GUS hybrids. Hence, the accumulation of functional β-galactosidase hybrids manifests itself in the form of disproportionately high activity when compared to the GUS hybrids.

In order to accurately assess the relative levels of *gus*A and *lac*Z activity in unidirectionally-cloned constructs, the ratio of GUS/β-galactosidase activity must be determined. A first approximation of this ratio was calculated from the assay results of the bidirectionally-cloned fragment B2 (Table 5.4). The comparisons between the *gus*A activity and the *lac*Z activity from SPB102a/b yielded an average ratio of 0.38. It is worth noting that this ratio did not include the data from the 'transcription towards *nod*B (-SE)' assays due to the low activity. As the *lac*Z and *gus*A assays are both known to fluctuate to some extent, these low levels would magnify such inconsistencies out of proportion (Sharma, pers. comm.). The other pair of bidirectionally-oriented fragments, contained in pSPB14a/b, were not considered in the determination of the ratio due to their unusual regulatory properties (see Section 6.4)

As mentioned earlier, the value of 0.38 as the GUS/β-galactosidase ratio is only preliminary. The importance of this parameter will be limited until a series of bidirectionally-cloned fragments can be compared. Therefore, at this stage, while the vector, pSPV4, could be relied upon to report qualitatively on the bidirectional modes of regulation of a randomly cloned fragment, direct comparisons between the activity of the two reporter genes are still of dubious quantitative significance.

6.3 Analysis of the *nod*A *nod*-BOX (see Table 5.3)

The *R. loti nod*-box separating the *nod*ACIJ operon from the truncated *nod*D gene had previously been analysed by Teo (1990). Using the reporter vector pMP220 (Spaink *et al.* 1987), Teo obtained results suggesting that the *nod*A *nod*-box promoted inducible transcription in both directions. This was surprising as the *nod*D gene is transcribed constitutively in all other *Rhizobium* species studied to date.

To investigate these unusual regulatory results further, the bidirectional reporter vector, pSPV4 was utilised (Figure 4.11). Three different PCR-generated fragments each containing the *nod*A *nod*-box were cloned into the bidirectional vector (Section 5.6.1). The longest of these, designated A1, contained extensive 5' and 3' regions flanking the *nod*-box. Accordingly, this clone was thought to possess all the *cis*-acting regulatory elements that may affect the promoter activity of the *nod* genes. The other two plasmid constructs, pSPA12 and pSPA13, contained fragments whose flanking regions were shorter than those of pSPA11 (PCR products A2, A3 and A1 respectively in Figure 5.7). These fragments, specifically delineated by PCR primer binding sites, were used to identify regulatory elements located outside the *nod*-box region. All three of these clones were bidirectionally assayed in the native background of *R. loti* (Section 5.8). The results of these assays have been presented in Table 5.3, and the Appendix.

The investigation of the *nod*A *nod*-box raised several interesting points regarding the regulatory properties of this putative bidirectional promoter region. Firstly, the results of these assays strongly indicate that transcription of the *nod*ACIJ operon is induced by seed exudate, while transcription in the '*nod*D' direction is constitutive. While these observations are typical for most *Rhizobium* species, the latter finding differs from the results obtained previously for *R. loti* (Teo, 1990). Several explanations for this discrepancy may be considered. From a methodological point of view, the requirement of Teo to clone every *nod*-box-containing fragment in both orientations, and assay each construct independently could lead to ambiguous data being obtained. This concern is heightened when it is realised that the constructs were assayed only once. On the other hand, the fragments in this investigation were generated by PCR using *Taq* polymerase. This enzyme, which has an inherently high error rate, may have introduced point mutations that affected the regulatory properties of the *nod*-box in the current study. However, the fact that constitutive transcription was invariably observed in the *nod*D direction for each construct indicates that this is probably not the case. This evidence, coupled with the universally constitutive nature of rhizobial *nod*D genes, supports the results observed in this investigation.

The plasmid pSPA11, which contained the longest *nod*A *nod*-box fragment (A1), displayed the highest level of constitutive activity in the '*nod*D' direction (Table 5.3). Predictably, the same fragment also exhibited relatively high inducibility in the opposite direction in the presence of *L. pedunculatus* seed exudate. The anticipated character of these results confirmed the premise that the cloned fragment in pSPA11 contains all of the known *cis*-acting elements that regulate the *nod* genes. When the remaining strains in the SPA 100 series of clones were compared with SPA101, several differences in activity became evident (Table 5.3). Firstly, it was noticed that the strain SPA102 (containing pSPA12) had considerably lower levels of constitutive activity than that seen in SPA101. Conceivably, the decrease in the activity of SPA102 may have been due to the exclusion of a promoter element located in the region that was omitted from A1 (fragments A1 and A2 in Figure 5.7). However, given the relatively minor two-fold magnitude of the drop, it seems unlikely that this is the case. If a promoter region was involved at all, only a small part of it must have been affected. Indeed, the integrity of most, if not all, of the regulatory DNA-protein and protein-protein interactions must have remained faithful in SPA102, as control of the *nod* genes was not dramatically altered. Therefore, the apparent decrease in constitutive activity observed in SPA102 can probably be attributed to differences in the spacing between the *nod*-box and the *lac*Z gene. Such spacing variations are well-known to affect the regulatory properties of a cloned fragment (Spaink *et al.*, 1987).

The third clone, SPA103, contained a short PCR fragment that was ligated in the opposite orientation to the other *nod*A *nod*-box fragments. The consequence of this reversed orientation is that the divergently-expressed reporter genes in pSPA13 measured activity generated in the opposite direction relative to the other pSPA 10 fragments. Thus, while the *gus*A gene in pSPA13 measured transcriptional activity in the *nod*D direction, the *lac*Z gene of pSPA11 and pSPA12 was responsible for the corresponding measurement. To compare the reporter gene activities between strains SPA101 and SPA102, the GUS/β-galactosidase ratio calculated in Section 6.1 was applied. By multiplying the *lac*Z assay results by 0.38, all of the these numbers become amenable to direct comparison with the *gus*A data. While the constitutive transcription appears to be commensurate with SPA102 (6.3 compared with 8.4 (without seed exudate), and 8.4 compared with 6.5 (with seed exudate)), strain SPA103 was not induced to the levels observed in SPA102 (1.5 compared with 1.7 (without seed exudate), and 7.6 compared with 26 (with seed exudate)). These results support the tentative proposal that a promoter element lies in the region omitted from right of the *nod*-box in fragments A2 and A3. The effect of this deletion is a reduction of the constitutive activity in the '*nod*D' direction. The other interesting feature regarding the assay results from SPA103 is the apparent decrease in inducibility of the *nod*ACIJ operon. As the absence of a region to the left of the *nod*-box correlates with this reduction in activity, this suggests

the presence of an additional regulatory element. Further investigation of this region will be required before the exact nature of this element can be precisely determined. Realistically, at this stage, the only information that can be confidently gleaned from the SPA103 assay results relates to the nature of the *nod* gene regulation. As with the other clones, transcription is constitutive in the '*nod*D' direction and inducible in the *nod*A direction.

In summary, the analysis of the *nod*A *nod*-box yielded several important results. Most significantly, the data demonstrated that transcription in the '*nod*D' direction is constitutive and not inducible as previously thought (Teo, 1990). This indicates that the regulation conferred by the *nod*-box region in *R. loti* is similar to that in other *Rhizobium* species despite the rearrangements in its common *nod* genes. Furthermore, two putative regulatory elements have been identified bracketing the *nod*A *nod*-box. One such element, located to the right of the *nod*-box, appears to have a role in regulating expression in the '*nod*D' direction. Deletion of this region gave consistently low levels of constitutive activity. An earlier investigation by Rossen *et al.* (1986), showed that the 30bp to the right of the *nod*-box were important in the expression of the *nod*D gene in *R. leguminosarum*. The results from this current study are in agreement with this observation. The second proposed regulatory region, located to the left of the *nod*-box, appears to affect the inducibility of the *nod*ACIJ operon. When this region was omitted, a marked decrease in the expression in the *nod*ACIJ direction was observed. It should be pointed out that these results do not comprehensively prove the presence of regulatory elements. Differences in the spacing between the promoter and the reporter gene have been known to cause changes in reporter gene activity. Additional research will have to be undertaken to ascertain the authenticity of both of these regulatory regions.

6.4 Analysis of the *nod*B *nod*-BOX

Essentially, the results generated from the *nod*B *nod*-box were very similar to those observed earlier for the *nod*A *nod*-box (Tables 5.3 and 5.4). While this *nod*-box has been referred to as the *nod*B *nod*-box, it should be pointed out that the transcriptional fusions formed in the *nod*B direction are actually fusions involving ORF3 (Figure 1.6). The significance of this ORF3/*nod*B fusion is not likely to be great given that the *nod*-box is thought to contain all the *nod* gene promoter activity. As expected, most of the bidirectional constructs containing a *nod*B *nod*-box showed constitutive activity in the '*nod*D' direction and inducible activity in the *nod*B direction. Strain SPB101 demonstrated this effect most convincingly (Table 5.4 line 1). As this clone also carries the longest stretches of DNA flanking the *nod*-box (PCR product B1 in Figure 5.7), it is possible that these flanking sequences provide binding sites for various uncharacterised

regulatory proteins. Conversely, these regions may simply place the *nod*-box in a more natural environment in which to be assayed. The juxtaposition of *nod*-box regions with vector DNA in some of the smaller clones may have caused some of the regulatory anomalies that were observed.

One strain that gave unexpected results was SPB102 (Table 5.4 lines 2 and 3). The activity generated by the *nod*-box in this plasmid was markedly lower than that seen for SPB101. This raised the possibility that a regulatory element was present in B1, yet absent in B2 (Figure 5.7). Furthermore, strains SPB105, and SPB104a/b, which each contained the same region flanking the right-hand side of the *nod*-box, also showed a significant reduction in activity in the '*nod*D' direction (PCR products B5 and B4 respectively in Figure 5.7, and Table 5.4). Thus, in conclusion, this decrease in constitutive activity can be attributed to either spacing discrepancies (as mentioned earlier), or the absence of a binding site for a transcriptional activator.

The strain, SPB105, represents a *nod*-box bearing such small flanking regions that it could almost be considered as a stand-alone *nod*-box. While it has been tentatively suggested that this clone is missing a regulatory binding site affecting '*nod*D' transcription, its high inducibility in the *nod*B direction indicates that there are no important regulatory elements to the left of the *nod*-box.

The most unexpected results were obtained by assaying strains SPB103 and SPB104b. These strains gave constitutive expression of *lacZ*(247-SE / 250+SE) and *lacZ*(330SE / 280+SE) respectively, which was many fold greater than those observed within the other constructs. Especially puzzling was the fact that this high activity was in the direction of the inducible *nod*B gene. Transcription in the '*nod*D' direction was considerably lower than that reported for the *nod*B 'control strain' SPB101. [*lacZ*(50-SE / 46+SE) in SPB101 compared to *lacZ*(1.3-SE / 1.6+SE) in SPB103 and *gusA*(10-SE / 13+SE) in SPB104b]. Hence, the unexpectedly high assay results were not bidirectional. Significantly, the clone, SPB103, originally intended to act as a negative control, did not contain a *nod*-box. Therefore, at first glance, it appears that these unexpected observations are probably due to contamination of the assays with an organism that has high levels of endogenous *lacZ* activity. However, this possibility was discounted through repeated independent assays. Moreover, the same results were obtained when the constructs were assayed by an independent researcher (Sharma, pers. comm.).

To explain the surprising results of SPB104b and SPB103 in terms of regulatory elements lying outside the *nod*-box region is tempting. However, several lines of evidence exist which cast doubt on this appealing hypothesis. The most crucial of these was the observation by Spaink *et al.* (1987), that cloning artifacts sometimes cause the activity of the *lacZ* reporter gene in pMP220 to give unusual results. In this case, the term 'cloning artifact' refers to the placement of a non-regulatory region of DNA next to a reporter gene, which results in high levels of expression. The unexpected transcription is thought to be caused by the accidental production of a functional RNA polymerase binding site. Unfortunately, the presence of this site yields results which do not reflect the promoter activity of the cloned DNA fragment in its natural environment. In order to demonstrate that the unusual results observed in pMP220 were due to cloning artifacts, and not genuine promoter activity, Spaink *et al.* (1987), deleted an internal fragment from the putative promoter element. The deleted fragment of interest still contained constitutive promoter activity. Furthermore, when the unmodified fragment was re-cloned into different enzyme sites within the reporter vector, the promoter activity was halved. Therefore, it was more likely that the irregular promoter activity observed in this case was a cloning artifact rather than a *nod* promoter (Spaink *et al.*, 1987).

In the current investigation, the validity of the SPB104b data was checked by cloning the insert in the opposite orientation (Section 5.6.2). If the cloned fragment was a promoter element, the high activity reported by the *lacZ* gene should be reciprocated by the *gusA* gene. However, if a cloning artifact was responsible for the high expression, lower activity should be reported by the *gusA* gene in SPB104a. As the results show (Table 5.4), the latter case was observed, supporting the suggestion of a cloning artifact. In the *nodD* direction, the integrity of regulation appeared to be restored. Strains SPB104a/b both showed constitutive levels of activity comparable with those found in SPB102a/b (Table 5.4). As all of these fragments contain the same sequence to the right of the *nod*-box, this result was expected (fragments B2 and B4 in Figure 5.7). Similarly, the lack of activity in the *nodD* direction in strain SPB103 (containing cloned fragment B3) was also expected due to the lack of a *nod*-box (Table 5.4 and Figure 5.7).

In conclusion, the analysis of the *nodB nod*-box revealed several interesting insights into gene regulation. The most important of these was the tentative identification of a binding site of a transcriptional activator. Located between the BS6 primer binding site and the *Eco*RI site immediately to the right of the *nod*-box (Figure 5.7), this site is apparently involved in the control of the *nodD* gene. However, it should not be overlooked that the data implying the presence of a regulatory element may simply be due to the effects of spacing on reporter gene activity. As with the putative regulatory elements flanking the *nodA nod*-box, a great deal of further work will be required to confirm or deny the presence of this site. The ongoing problem of artifactual results in

reporter gene systems has also been highlighted in the current investigation. Initially reported in pMP220 by Spaink *et al.* (1987), these spurious results were occasionally observed in pSPV4. On a more positive note, the symmetry of the pSPV4 was confirmed by cloning the fragment B2 in both orientations and measuring the bidirectional activity (Table 5.4). This result was important as it verified the assumption that the observed mode of *nod*-box regulation was not dependent on the orientation of the cloned fragment.

6.5 FACTORS AFFECTING THE LEVEL OF INDUCIBILTY OF THE *nod* GENES

The inducibility of the *nod* genes in *R. loti* appears to be very low in comparison to many other *Rhizobium* species. Indeed, Rossen *et al.* (1985), using *R. leguminosarum*, observed constitutive levels in the *nod*D direction of over 600 β-galactosidase units, and inducible levels in the *nod*ABC direction of over 2000 β-galactosidase units. Excluding the apparently anamolous results of SPB103 and SPB104b, the highest constitutive level observed in *R. loti* was 50 β-galactosidase units (SPB101), and the highest inducible activity was 81 GUS units (SPB101). This latter figure converts to 213 relative β-galactosidase units after applying the GUS/β-galactosidase ratio. The low *R. loti* assay results are probably attributable to the crude preparation of seed exudate used in these assays. Such seed extracts may not contain the correct concentration of flavonoid inducers, or conversely, may contain inhibitory compounds that affect *nod* gene expression. Until the flavonoid inducer for *R. loti* has been purified, it will not be possible to accurately compare the activity of the *R. loti nod* genes with those from other species of *Rhizobium*. If the inducibility of the *R. loti nod* genes was still relatively low after using purified flavonoids, the presence of a NolR repressor protein would have to be considered. It was pointed out earlier that strains containing this repressor exhibited a decrease in the expression of the *nod* genes.

Although *R. loti nod* gene inducibility was invariably low in comparison to other *Rhizobium* species, there was variability between each clone. This was probably due to either spacing differences between the *nod*-box and the reporter genes, or the presence of *cis*-acting regulatory elements. One other alternative is that the copy number of the recombinant expression vector was slightly different in each case. Moore *et al.* (1990), working with yeast, demonstrated that variation in gene expression can be due to this dosage effect. Any difference in copy number would reflect greatly on the observed reporter gene activity, as pSPV4 is a low copy number plasmid. To determine whether the variation in gene expression was attributable to plasmid copy number, the copy number of the plasmid would have to be measured. This can be achieved in individual transformants by dot-blotting, using a single-copy gene as a control (Moore *et al.*, 1990).

6.6 FURTHER RESEARCH

While the current study has resolved a number of important questions regarding *nod* gene regulation, it has also raised some new ones. Firstly, the presence of the putative regulatory elements to the left of the *nod*A *nod*-box and to the right of the *nod*B *nod*-box must be investigated in greater depth. Linker analysis may have to be used to eliminate the effect of spacing in the reporter system. In the short term, a series of different, appropriately defined fragments should be cloned into pSPV4 and bidirectionally assayed. This would help to elucidate the presence of the regulatory elements. If a regulatory binding site was found in this region, further research would focus on purifying the regulatory protein thought to be responsible for this binding. It would then be interesting to check for the presence of this protein (and its corresponding gene(s)) in other rhizobial systems.

Secondly, the regulatory properties of the third *nod*-box (associated with the intact *nod*D_3 gene) should be elucidated. This *nod*-box is particularly interesting as it bears only limited homology to conventional *nod*-boxes. Surprisingly, the conserved regions lie almost completely in one half of the *nod*-box. The other half however, is very poorly conserved leading to the speculation that the bidirectional regulatory properties of the *nod*-box may be affected (Figure 6.1). There is, in fact, no apparent requirement for this *nod*-box to promote transcription in both directions as no *nod* genes have been identified to the right of this region. In the long term, footprinting analyses and gel retardation studies should be performed on this *nod*-box to reveal the pattern of NodD binding. These studies may provide answers to the question of whether a NodD complex binds to a NodD box (Goethals *et al.*, 1992), or whether the whole *nod*-box is required for NodD binding.

In the future, further research on *R. loti* may be focused on the formal demonstration of NodD binding to the *nod*A and *nod*B *nod*-boxes using gel retardation and footprinting analyses. As these *nod*-boxes are well-conserved with respect to *R. meliloti*, it is unlikely that anything new will emerge from these studies (Scott *et al.*, unpublished). Similarly, the precise definition of the transcriptional start sites of the divergent *nod* genes will also probably only confirm what has already been elucidated from comparisons with the *R. meliloti* system. One investigatory avenue that would yield interesting results, is the identification of a *nol*R homologue in *R. loti*. As mentioned in Section 1.5.3.2, many strains of *Rhizobium* contain this gene, though it is not always a functional copy. If *R. loti* was shown to possess a functional copy of this gene, footprinting could be used to determine its binding site.

Figure 6.1. Comparison Between the *nod*D$_3$ *nod*-box From *R. loti* and the Classical *nod*-box Sequence

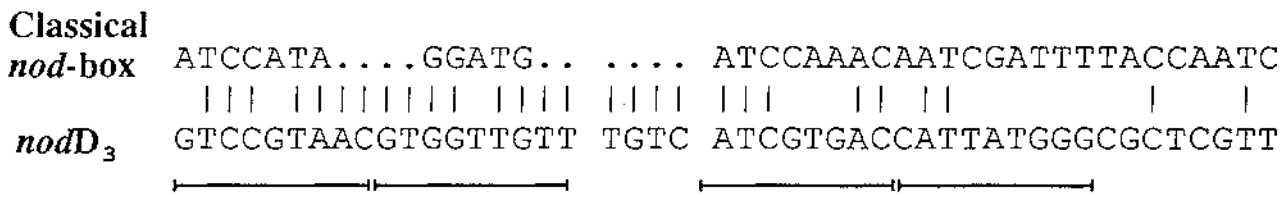


Figure 6.1.

A comparison between the divergent *nod*D$_3$ *nod*-box of *R. loti* (Scott *et al.*, unpublished), and the classical *nod*-box consensus sequence (Wang and Stacey, 1991). Matching nucleotides are indicated.

Application of the three bidirectional expression vectors offer interesting potential for further research. The two potential translational fusion vectors offer enormous potential in the bidirectional analysis of many promoter regions. Furthermore, the *Spe*I enzyme sites bracketing the functional *gus*A gene fusion, facilitates the construction of new vectors. The vector, pSPV4, also provides a convenient vehicle for which to study a variety of promoter activities in *Rhizobium* species. While it has only been used in the analysis of the *R. loti nod*-box to date, further uses for this vector appear to be unlimited.

Appendix

Construct	Transcription Towards *lacZ* -SE	+SE	Transcription Towards *gusA* -SE	+SE
PN184	8.0	11	3.1	2.1
	8.6	9.2	2.0	2.3
	6.2	6.4	1.9	0.6
	7.1	8.2	2.6	1.7
	5.9	5.9	0.0	0.8
	6.2	7.3	2.4	1.5
SP100	12	9.6	2.6	3.1
	10	7.5	1.1	1.4
	9.2	7.7	1.7	1.8
	8.5	7.2	2.8	2.2
	12	9.4	2.9	2.6
	8.3	7.2	2.1	1.5
SP101	309	262	4.1	3.1
(R.loti)	206	179	2.2	1.7
	262	219	2.6	2.4
	185	151	1.7	2.5
	276	279	0.9	1.6
	208	176	2.9	3.1

Construct	Transcription Towards *nodD* -SE	+SE	Transcription Towards *nodA* -SE	+SE
SPA101	44	39	0.0	22
	56	51	2.6	37
	53	45	1.5	29
	50	57	0.8	30
	39	32	0.0	24
	43	46	1.1	33
SPA102	22	17	1.2	21
	19	16	0.0	16
	26	24	1.5	26
	15	12	2.5	35
	2*	10	2.1	23
	28	21	2.9	32

* Anomalous result not included in averages.

Continued

Construct	Transcription Towards *nodD* -SE	+SE	Transcription Towards *nodA* -SE	+SE
SPA103	4.6	8.1	6.4	22
	6.3	8.9	2.7	17
	2.5	8.7	4.4	19
	8.1	6.1	6.2	21
	7.7	9.5	0.0	15
	8.6	9.1	4.3	23

Construct	Transcription Towards *nodD* -SE	+SE	Transcription Towards *nodB* -SE	+SE
SPB101	49	43	6.2	78
	59	54	12	97
	50	46	3.8	69
	46	45	9.9	88
	43	38	13	91
	50	50	9.1	65
SPB102a	19	16	3.1	7.8
	16	21	1.1	3.4
	21	20	1.5	3.7
	16	17	2.1	10
	16	22	2.9	9.6
	17	16	0.7	5.2
SPB102b	8.5	9.1	5.6	27
	7.9	6.1	2.2	16
	6.1	6.1	0.0	15
	5.8	7.8	1.9	18
	7.2	5.9	3.6	21
	9.5	7.7	1.7	19
SPB103	1.1	1.2	189	215
	0.9	2.1	253	242
	1.9	1.5	241	258
	1.6	2.0	267	253
	1.2	1.6	271	272
	1.1	1.2	261	260

Continued

Construct	Transcription Towards *nodD* -SE	+SE	Transcription Towards *nodB* -SE	+SE
SPB104a	12	13	16	19
	7.4	8.8	10	10
	12	15	9.6	14
	7.9	12	12	15
	13	19	8.6	9.6
	7.7	12	14	11
SPB104b	6.6	6.2	301	240
	4.5	5.2	333	326
	7.1	12	366	260
	6.5	6.5	355	300
	8.4	8.2	328	282
	5.9	7.5	297	272
SPB105	6.9	9.7	4.0	34
	12	14	3.2	14
	14	13	6.1	55
	9.5	12	3.5	42
	7.8	8.8	2.5	29
	14	15	7.7	40

BIBLIOGRAPHY

Albersheim, P. and Darvill, A.G. (1985).
Oligosaccharins. Scientific American, September: 44-50.

Alton, N.K. and Vapnek, D. (1979).
Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. Nature 282: 864-869.

Ames, P. and Bergman, K. (1981).
Competitive advantage provided by bacterial motility in the formation of nodules by *Rhizobium meliloti*. J. Bacteriol. 148: 728-729.

Banfalvi, Z., Nieuwkoop, A.J., Schell, M.G., Besl, L. and Stacey, G. (1988).
Regulation of *nod* gene expression in *Bradyrhizobium japonicum*. Mol. Gen. Genet. 214: 420-424.

Barnet, Y.M. (1991).
Distribution and characteristics of root-nodule bacteria isolated from Australian *Acacia*-spp. Plant Soil 135: 109-120

Bashan, Y. and Levanony, H. (1987).
Horizontal and vertical movement of *Azospirillum brasilense* Cd in the soil and along the rhizosphere of wheat and weeds in controlled and field environments. J. Gen. Microbiol. 133: 3473-3480.

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

Bergman, K., Gulash-Hoffee, M., Hovestadt, R.E., Larosiliere, R.C., Ronco II, P.G and Su, L. (1988).
Physiology of behavioral mutants of *Rhizobium meliloti*: evidence for a dual chemotaxis pathway. J. Bacteriol. 170: 3249-3254.

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

Beringer, J.E., Brewin, N., Johnston, A.W.B., Schulman, H.M. and Hopwood, D.A. (1979).
The *Rhizobium*-legume symbiosis. Proc. R. Soc. Lond. B204: 219-233.

Bhuvaneswari, T.V., Bhagwat, A.A. and Bauer, W.D. (1981).
Transient susceptibility of root cells in four common legumes to nodulation by rhizobia. Plant Physiol. 68: 1144-1149.

BioRad: Gene Pulser Apparatus Methods in Electroporation.

Boivin, C., Camut, S., Malpico, C.A., Truchet, G. and Rosenberg, C. (1990).
Rhizobium meliloti genes encoding catabolism of trigonelline are induced under symbiotic conditions. The Plant Cell 2: 1157-1170.

Bowra, B.J. and Dilworth, M.J. (1981).
Motility and chemotaxis towards sugars in *Rhizobium leguminosarum*. J. Gen. Microbiol. 126: 231-235.

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.

Bulawa, C.E. and Wasco, W. (1991).
Chitin and nodulation. Nature 353: 710.

Burn, J.E., Rossen, L. and Johnston, A.W.B. (1987).
Four classes of mutations in the *nod*D gene of *Rhizobium leguminosarum* biovar *viciae* that affect its ability to autoregulate and/or activate other *nod* genes in the presence of flavonoid inducers. Genes Dev. 1: 456-464.

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 *nod*D of *Rhizobium*. Mol. Microbiol. 3: 1567-1577.

Caetano-Anolles, G., Crist-Estes, D.K. and Bauer, W.D. (1988).
Chemotaxis of *Rhizobium meliloti* to the plant flavone luteolin requires functional nodulation genes. J. Bacteriol. 170: 3164-3169.

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

Casadaban, M.J. and Cohen, S.N. (1980).
Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138: 179-207.

Chassey, B.M. and Flickinger, J.L. (1987).
Transformation of *Lactobacillus casei* by electroporation. FEMS Microbiol. 44: 173-177.

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.

Clark, J.M. (1988).
Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucl. Acids Res. 16: 9677-9686.

Collins-Emerson, J.M. (1991).
Characterisation of a *Rhizobium loti* nodulation mutant. Ph.D. thesis, Massey University, New Zealand.

Collins-Emerson, J.M., Terzaghi, E.A., Young, C.A. and Scott, D.B.
Traces of sequence rearrangements among the common *nod* genes of *Rhizobium loti*. In preparation.

Dart, P.J. (1977).
Infection and Development of Leguminous Nodules. A Treatise on Dinitrogen Fixation. In: Section 3, Biology (editors: Hardy, R.W.F. and Silver, W.S.). Wiley (Interscience), New York, pp. 367-472.

Davis, E.O. and Johnston, A.W.B. (1990).
Regulatory functions of the three *nod*D genes of *Rhizobium leguminosarum* biovar *phaseoli*. Mol. Microbiol. 4: 933-941.

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

de Boer, H.A., Hui, A., Comstock, L.J., Wong, E. and Vasser, M. (1983).
Portable Shine-Dalgarno regions: a system for a systematic study of defined alterations of nucleotide sequences within *E. coli* ribosome binding sites. DNA 2: 231-235.

Denarie, J., Debelle, F. and Rosenberg, C. (1992).
Signal exchange in nodulation. Ann. Rev. Microbiol. 295: 497-531.

Dente, L., Cesareni, G. and Cortese, R. (1983).
pEMBL: a new family of single stranded plasmids. Nucl. Acids Res. 11: 1645-1655.

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

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

Djordjevic, M.A., Gabriel, D.W. and Rolfe, B.G. (1987).
Rhizobium - the refined parasite of legumes. Ann. Rev. Phytopathol. 25: 145-168.

Dotto, G.P., Enea, V. and Zinder, N.D. (1981).
Functional analysis of the bacteriophage f1 intergenic region. Virology 114: 463-473.

Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988).
High efficiency transformation of *E. coli* by high voltage electroporation. Nucl. Acids Res. 16: 6127-6145.

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

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.

Dusha, I., Bakos, A., Kondorosi, A., de Bruijn, F.J. and Schell, J. (1989).
The *Rhizobium meliloti* early nodulation genes (*nodABC*) are nitrogen-regulated: isolation of a mutant strain with efficient nodulation capacity on alfalfa in the presence of ammonium. Mol. Gen. Genet. 219: 86-96.

Eckhardt, T. (1978).
A rapid method for the identification of plasmid deoxyribonucleic acid in bacteria. Plasmid 1: 584-588.

Ehrhardt, D.W., Atkinson, E.M. and Long, S.R. (1992).
Depolarization of alfalfa root hair membrane potential by Rhizobium meliloti Nod factors. Science 256: 998-1000.

Engelke, T.H., Jagadish, M.N. and Puhler, A. (1987).
Biochemical and genetical analysis of *Rhizobium meliloti* mutants defective in C_4-dicarboxylate transport. J. Gen. Microbiol. 133: 3019-3029.

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

Fisher, R.F. and Long, S.R. (1989).
DNA footprint analysis of the transcriptional activator proteins NodD1 and NodD3 on inducible *nod* gene promoters. J. Bacteriol. 171: 5492-5502.

Fisher, R.F. and Long, S.R. (1992).
Rhizobium-plant signal exchange. Nature 357: 655-660.

Fisher, R.F. and Long, S.R.
Interactions of NodD at the *nod* box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. J. Mol. Biol., in press.

Fisher, R.F., Brierley, H.L., Mulligan, J.T. and Long, S.R. (1987b).
Transciption of *Rhizobium meliloti* nodulation genes: Identification of a *nodD* 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. (1987a).
Extended region of nodulation genes in *Rhizobium meliloti* 1021. II. Nucleotide sequence, transcription start sites and protein products. Genetics 117: 191-201.

Franssen, H.J., Thompson, D.V., Idler, K., Kormelink, R., van Kammen, A. and Bisseling, T. (1989).
Nucleotide sequence of two soybean ENOD2 early nodulin genes encoding Ngm-75*. Plant Mol. Biol. 14: 103-106.

Gillman, M.A. (1990).
The use of *lac* fusions to analyse the regulation of a *nod* gene region of *Rhizobium loti*. M.Sc. thesis, Massey University, New Zealand.

Gitte, R.R., Rai, P.V. and Patil, R.B. (1978).
Chemotaxis of *Rhizobium sp.* towards root exudate of *Cicer arietinum* L. Plant Soil 50: 553-566.

Goethals, K., Van Montagu, M. and Holsters, M. (1992).
Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. Proc. Natl. Acad. Sci. USA 89: 1646-1650.

Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981).
Translational initiation in prokaryotes. Ann. Rev. Microbiol. 35: 365-403.

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

Guerinot, M.L., Morisseau, B.A. and Klapatch, T. (1990).
Electroporation of *Bradyrhizobium japonicum*. Mol. Gen. Genet. 221: 287-290.

Guinel, F.C. and LaRue, T.A. (1991).
Light microscopy study of nodule initiation in *Pisum sativum L.* cv sparkle and its low-nodulating mutant E2 (*sym* 5). Plant Physiol. 97: 1206-1211.

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.

Hall, M.N., Gabay, J., Debarbouille, M. and Schwartz, M. (1982).
A role for mRNA secondary structure in the control of translation initiation. Nature 295: 616-618.

Hanahan, D. (1983).
Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166: 557-580.

Harte, D.J.G. (1991).
Transposon Tn-5-induced plasmidic rearrangement in the soil isolate OR168. Masters Thesis, Massey University, New Zealand.

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.

Hirsch, A.M., Bhuvaneswari, T.V., Torrey, J.G. and Bisseling, T. (1989).
Early nodulin genes are induced in alfalfa root outgrowths elicited by auxin transport inhibitors. Proc. Natl. Acad. Sci. USA 86: 1244-1248.

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. Nucl. 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 gene. Proc. Natl. Acad. Sci. USA 84: 8558-8562.

Honma, M.A., Asomaning, M. and Ausubel, F.M. (1990).
*Rhizobium meliloti nod*D genes mediate host-specific activation of *nod*ABC. J. Bacteriol. 172: 901-911.

Hubac, C., Guerrier, D., Ferran, J., Tremolieres, A. and Kondorosi, A. (1992).
Lipid and protein composition of outer and inner membranes in wild-type strains and *nod* mutants of *Rhizobium meliloti*. J. Gen. Microbiol. 138: 1973-1983.

Hui, A., Hayflick, J., Dinkelspiel, K. and de Boer, H.A. (1984).
Mutagenesis of the three bases preceding the start codon of the β-galactosidase mRNA and its effect on translation in *Escherichia coli*. EMBO J. 3: 623-629.

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

Jefferson, R.A., Burgess, S.M. and Hirsh, D. (1986).
β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. Proc. Natl. Acad. Sci. USA 83: 8447-8451.

Jefferson, R.A. (1987).
Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol. Biol. Reporter 5: 387-405.

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.

Kape, R., Parniske, M. and Werner, D. (1991).
Chemotaxis and *nod* gene activity of *Bradyrhizobium japonicum* in response to hydroxycinnamic acids and isoflavonoids. Appl. Envir. Micro. 57: 316-319.

Kondorosi, E. and Kondorosi, A. (1986).
Nodule induction on plant roots by *Rhizobium*. TIBS 11: 296-299.

Kondorosi, E., Buire, M., Cren, M., Iyer, N., Hoffmann, B. and Kondorosi, A. (1991a).
Involvement of the *syr*M and *nod*D3 genes of *Rhizobium meliloti* in *nod* gene activation and in optimal nodulation of the plant host. Mol. Microbiol. 5: 3035-3048.

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.

Kondorosi, E., Pierre, M., Cren, M., Haumann, U., Buire, M., Hoffmann, B., Schell, J and Kondorosi, A. (1991b).
Identification of NolR, a negative transacting factor controlling the *nod* regulon in *Rhizobium meliloti*. J. Mol. Biol. 222: 885-896.

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.

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

Low, B. (1968).
Formation of merodiploids in matings with a class of rec⁻ recipient strains of *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA 60: 160-167.

McSweeney, D.G. (1987).
The use of lac fusions to analyse the regulation of a symbiotic region of *Rhizobium loti*. Honours thesis, Massey University, New Zealand.

Maillet, F., Debelle, F. and Denarie, J. (1990).
Role of the *nod*D and *syr*M genes in the activation of the regulatory gene *nod*D3, and of the common and host-specific *nod* genes of *Rhizobium meliloti*. Mol. Microbiol. 4: 1975-1984.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989).
Molecular Cloning: A Laboratory Manual.(2nd Edition) Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. (1991).
Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucl. Acids Res. 19: 1154.

Marvel, D.J., Torrey, J.G. and Ausubel, F.M. (1987).
Rhizobium symbiotic genes required for nodulation of legume and nonlegume hosts. Proc. Natl. Acad. Sci. USA 84: 1319-1323.

Mattanovich, D., Ruker, F., Machado, A.D.C., Laimer, M., Regner, F., Steinkellner, H., Himmler, G. and Katinger, H. (1989).
Efficient transformation of *Agrobacterium* spp. by electroporation. Nucl. Acids Res. 17: 6747.

Mellor, H.Y., Glenn, A.R., Arwas, R. and Dilworth, M.J. (1987).
Symbiotic and competitive properties of motility mutants of *Rhizobium trifolii* TA1. Arch. Microbiol. 148: 34-39.

Mersereau, M., Pazour, G.J. and Das, A. (1990).
Efficient transformation of *Agrobacterium tumefaciens* by electroporation. Gene 90: 149-151.

Miller, J.F., Dower, W.J. and Tompkins, L.S. (1988).
High-voltage electroporation of bacteria: genetic transformation of *Campylobacter jejuni* with plasmid DNA. Proc. Natl. Acad. Sci. USA 85: 856-860.

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

Mole, S.E., Iggo, R.D. and Lane, D.P. (1989).
Using the polymerase chain reaction to modify expression plasmids for epitope mapping. Nucl. Acids Res. 17: 3319.

Moore, P.A., Bettany, A.J.E. and Brown, A.J.P. (1990).
Expression of a yeast glycolytic gene is subject to dosage limitation. Gene 89: 85-92.

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.

Mulligan, J.T. and Long, S.R. (1989).
A family of activator genes regulate expression of *Rhizobium-meliloti* nodulation genes. Genetics 122: 7-18.

Murphy, P.J., Heycke, N., Banfalvi, Z., Tate, M.E., de Bruijn, F., Kondorosi, A., Tempe, J. and Schell, J. (1987).
Genes for the catabolism and synthesis of an opine-like compound in *Rhizobium meliloti* are closely linked and on the Sym plasmid. Proc. Natl. Acad. Sci. USA 84: 493-497.

Nagel, R., Elliott, A., Masel, A., Birch, R.G. and Manners, J.M. (1990).
Electroporation of binary Ti plasmid vector into *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. FEMS Microbiol. Letters 67: 325-328.

Newcomb, W. (1981).
Nodule morphogenesis and differentiation. Int. Rev. Cytol. Supp. 13: 247-298.

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.

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

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

Pankhurst, C.E., Hopcroft, D.H. and Jones, W.T. (1987).
Comparative morphology and flavolan content of *Rhizobium loti* induced effective and ineffective root nodules on *Lotus* species, *Leuceana leucocephala*, *Carmichaelia flagelliformis*, *Ornithopus sativus*, and *Clianthus puniceus*. Can. J. Bot. 65: 2676-2685.

Perotto, S., Vandenbosch, K.A., Butcher, G.W. and Brewin, N.J. (1991).
Molecular composition and development of the plant glycocalyx associated with the peribacteroid membrane of pea root nodules. Development 112: 763-773.

Peters, N.K. and Long, S.R. (1988).
Alfalfa root exudates and compounds which promote or inhibit induction of *Rhizobium meliloti* nodulation genes. Plant Physiol. 88: 396-400.

Peters, K.N., 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.

Powell, I.B., Achen, M.G., Hillier, A.J. and Davidson, B.E. (1988).
A simple and rapid method for genetic transformation of lactic streptococci by electroporation. Appl. Environ. Microbiol. 54: 655-660.

Ptashne, M. (1987).
A genetic switch: Gene control and phage lambda. Blackwell Scientific Publications and Cell Press: Australia.

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

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 peribacteriod membranes. J. Cell Sci. 30: 129-149.

Roche, P., Debelle, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Denarie, J. and Prome, J-C. (1991).
Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: *nod*H and *nod*PQ genes encode the sulfation of lipo-oligosaccharide signals. Cell 67: 1131-1143.

Ronson, C.W., Lyttleton, P. and Robertson, J.G. (1981).
C_4-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. Proc. Natl. Acad. Sci. USA 78: 4284-4288.

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*A,B,C genes. EMBO J. 4: 3369-3373.

Rossi, J.J., Soberon, X., Marumoto, Y., McMahon, J. and Itakura, K. (1983).
Biological expression of an *Escherichia coli* consensus sequence promoter and some mutant derivatives. Proc. Natl. Acad. Sci. USA 80: 3203-3207.

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.

Salminen, S.O. and Streeter, J.G. (1992).
Labelling of carbon pools in *Bradyrhizobium japonicum* bv *viciae* bacteroids following incubation of intact nodules with $^{14}CO_2$. Plant Physiol. 100: 597-604.

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

Saunders, S.E. and Burke, J.F. (1990).
Rapid isolation of miniprep DNA for double strand sequencing. Nuc. Acids Res. 18: 4948.

Schlaman, H.R.M., Okker, R.J.H. and Lugtenberg, B.J.J. (1992).
Regulation of nodulation gene expression by NodD in rhizobia. J. Bacteriol. 174: 5177-5182.

Schlaman, H.R.M., Horvath, B., Vijgenboom, E., Okker, R.J.H. and Lugtenberg, B.J.J. (1991).
Suppression of nodulation gene expression in bacteroids of *Rhizobium leguminosarum* biovar *viciae*. J. Bacteriol. 173: 4277-4287.

Schlaman, H.R.M., Spaink, H.P., Okker, R.J.H. and Lugtenberg, B.J.J. (1989).
Subcellular localisation of the *nod*D gene product in *Rhizobium leguminosarum*. J. Bacteriol. 171: 4686-4693.

Scott, D.B., pers. comm., Massey University, Palm. Nth., New Zealand

Scott, D.B., Young, C.A., Collins-Emerson, J.M., Terzaghi, E.A., Rockman, E.S., Lewis, P.E. and Pankhurst, C.E.
Mutational and structural analysis of *Rhizobium loti* nodulation genes. In preparation.

Sharma, S.B., pers. comm., AgResearch Grasslands, CRI, Palm. Nth., New Zealand.

Sharma, S.B. and Signer, E.R. (1990).
Temporal and spatial regulation of the symbiotic genes of *Rhizobium meliloti* in planta revealed by Tn-5-*gus*A. Genes Dev. 4: 344-356.

Shepard, H.M., Yelverton, E. and Goeddel, D. (1982).
Increased synthesis in *E. coli* of fibroblast and leukocyte interferons through alterations in ribosome binding sites. DNA 1: 125-131.

Shigekawa, K. and Dower, W.J. (1988).
Electroporation of eukaryotes and prokaryotes: a general approach to the introduction of macromolecules into cells. BioTechniques 6: 742-751.

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

Smit, G., Logman, T.J.J., Boerrigter, M.E.T.I., Kijne, J.W. and Lugtenberg, B.J.J. (1989).
Purification and partial characterisation of the *Rhizobium leguminosarum* biovar *viciae* Ca^{2+}-dependent adhesin, which mediates the first step in attachment of cells of the family *Rhizobiaceae* to plant root hair tips. J. Bacteriol. 171: 4054-4062.

Spaink, H.P., Okker, R.J.H., Wijffelman, C.A., Pees, E. and Lugtenberg, B.J.J. (1987).
Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Plant Mol. Biol. 9: 27-39.

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

Spaink, H.P., Sheeley, D.M., van Brussel, A.A.N., Glushka, J., York, W.S., Tak, T., Geiger, O., Kennedy, E.P., Reinhold, V.N. and Lugtenberg, B.J.J. (1991).
A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. Nature 354: 125-130.

Stormo, G.D., Schneider, T.D., Gold, L. and Ehrenfeucht, A. (1982).
Characterization of translational initiation sites in *E. coli*. Nucl. Acids Res. 10: 2997-3011.

Teo, J. (1990).
Analysis of the role of the *nod*-box in the regulation of the *nodACIJ* gene cluster of *Rhizobium loti* NZP2037. Honours thesis, Massey University, New Zealand.

Terzaghi, E.A., pers. comm., Massey University, Palm. Nth., New Zealand.

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.

Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., de Billy, F., Prome, J-C. and Denarie, J. (1991).
Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenesis in alfalfa. Nature 351: 670-673.

van Batenburg, F.H.D., Jonker, R. and Kijne, J.W. (1986).
Rhizobium induces marked root hair curling by redirection of tip growth: a computer simulation. Physiol. Plant 66: 476-480.

Vazquez, M., Davalos, A., De Las Penas, A., Sanchez, F. and Quinto, C. (1991).
Novel organization of the common nodulation genes in *Rhizobium leguminosarum* bv. *phaseoli* strains. J. Bacteriol. 173: 1250-1258.

Verma, D.P.S. and Long, S. (1983).
The molecular biology of *Rhizobium*-legume symbiosis. Int. Rev. Cytol. 14 (Suppl.): 211-245.

Vieira, J. and Messing, J. (1987).
Production of single-stranded plasmid DNA. Methods Enzymol. 153: 3-11.

Vieira, J. and Messing, J. (1991).
New pUC-derived cloning vectors with different selectable markers and DNA replication origins. Gene 100: 189-194.

Wang, S-P. and Stacey, G. (1991).
Studies of the *Bradyrhizobium japonicum nod*D_1 promoter: a repeated structure for the *nod* box. J. Bacteriol. 173: 3356-3365.

Winsor, B.A.T. (1989).
A *nod* at differentiation: the *nod*D gene product and initiation of *Rhizobium* nodulation. TIG 5: 199-201.

Yarosh, O.K., Charles, T.C. and Finan, T.M. (1989).
Analysis of C_4-dicarboxylate transport genes in *Rhizobium meliloti*. Mol. Microbiol. 3: 813-823.

Young, C.A., pers. comm., Massey University, Palm. Nth., New Zealand.

Zeugin, J.A. and Hartley, J.L. (1985).
Ethanol precipitation of DNA. Focus 7: 1-2.