IDENTIFICATION AND CHARACTERIZATION OF NOVEL LATE-NODULIN GENES FROM A MODEL LEGUME *LOTUS JAPONICUS*

By

Philipp V. Kapranov

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Program in Genetics

2000

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF NOVEL LATE-NODULIN GENES FROM A MODEL LEGUME *LOTUS JAPONICUS*

By

Philipp Kapranov

The formation of nitrogen-fixing nodules represents an unusual example of externally induced organogenesis that unites the plant host and the symbiotic bacteria in a microenvironment appropriate for the support of bacterial nitrogen fixation and plant-mediated assimilation of nitrogen. The process of nodule development can be subdivided into early and late stages. To study the complexity of genes expressed during in a fully functioning, nitrogen-fixing nodule, the mRNA expression profile was examined in determinate nodules of the diploid legume *Lotus japonicus* using the differential display method. A range of novel expressed sequence tags (ESTs) associated with late developmental events during *Lotus japonicus* nodule organogenesis was identified and three genes, *LjNOD16*, *LjNPP2C1* and *LjNOD70* were characterized in detail. The *LjPLP* gene family was identified based on nucleotide sequence homology to a cDNA, *LjNOD16*, encoding *L. japonicus* late nodulin, Nlj16. We show that LjPLP proteins consist of an N-terminal domain that shares ~ 40% primary sequence identity with yeast

PITP, and is also capable of complementing a temperature-sensitive phenotype of a yeast *sec14* mutant strain, and a C-terminal domain consisting of either Nlj16, or highly related amino-acid sequence. The Nlj16 C-terminal domain was found to be involved in targeting PLPs to the cell plasma membrane. Furthermore, nodule-specific *LjNOD16* mRNA is shown to be the result of an unusual transcriptional event, mediated by a nodule-specific, bi-directional promoter present in an intron of a member of the *LjPLP* gene family. The results obtained suggest that nodulin Nlj16 may exert a dominant negative effect and function as a component of a mechanism that inactivates the function of LjPLPs specifically in nodules. Expression of the *LjNPP2C1* gene, encoding a protein phosphatase type 2C enzyme, was found to increase significantly in the mature *L. japonicus* nodules. Expression of the *LjNPP2C1* gene was found to be drastically altered in specific *L. japonicus* lines carrying monogenic recessive mutations in symbiosis-related loci, suggesting that the product of the *LjNPP2C1* gene may function at both early and late stages of nodule development. Two major mRNA species corresponding to the *LjNOD70* gene were induced in nodules and shown to be the result of a mechanism resembling alternative splicing. The longer, presumably unspliced, mRNA species was shown to contain a single open reading frame (ORF), encoding a polytopic hydrophobic protein, LjN70, with a predicted molecular mass of 70 kD. The predicted amino-acid sequence of nodulin LjN70 revealed structural features characteristic of transport proteins, and was found to share similarity with the oxalate/formate exchange protein of *Oxalobacter formigenes*. Therefore, the *L. japonicus LjNOD70* gene family is postulated to encode nodule-specific transport proteins, which may have evolved as a result of exon-intron shuffling.

To my parents and my friends who believed in me

ACKNOWLEDGEMENTS

I want to thank my advisor Frans de Bruijn for giving me the opportunity to work in his lab on several independent and often esoteric projects. I wish to extend my special gratitude and appreciation, which is hard to express in words, to Krzysztof Szczyglowski, who always gave me his friendship, support and guidance, who taught me a lot about what I currently know about science and molecular biology, in particular, and without whom this thesis would not have been possible. I want to thank my committee members Lee McIntosh, Lee Kroos and Doug Gage for always finding time to discuss science and life, for their encouragement and suggestions along the way. I owe a lot to all of the past and present members of the de Bruijn lab, who made it a genuinely pleasant place to work. I really enjoyed working with this group of people. Especially, I want to thank Mary Ellen Davey for her support and numerous helpful advices about life and science; David Silver and Susan Fujimoto, for those many discussions we had about everything; Jodi Trzebiatowski, Chris Vriezen, Judith Wopereis and Ann Milcamps, without whom the last few years would have been much less enjoyable; and everybody for being fantastic people to work with. Thanks to the past and present members of PRL, especially to Sridhar Venkataraman and Denis Maxwell, who made it a special place to work. To Xiaoqiong Qin for introducing me to raquetball and sushi, and also helping us to solve a nasty technical problem. To Ryan Bushey for excellent technical assistance with generation of transgenic plants. To Marlene and Kurt for excellent artwork.

PREFACE

The experiments in the Chapter 2 were initiated and guided by Dr. Krzysztof Szczyglowski and represent a truly collective effort of a small team also including Dr. Dirk Hamburger and the author of the thesis. As to the Chapters 3, 4 and 7, all of the experiments were done by the author of the thesis.

In the chapter 5, the original differential display product was cloned by the author of this thesis and the spliced cDNA clone 48-23 was isolated by Drs. Krzysztof Szczyglowski and Dirk Hamburger. The author of the thesis also contributed the following: the RT-PCR analysis; the discovery of the spliced, presumably functional, form of *LjNOD70* mRNA; the northern blot analysis of *LjNOD70* expression in the different tissues; cloning of a related gene from *L. japonicus* and the generation of the anti-LjN70 antibody. Dr. Krzysztof Szczyglowski determined the pattern of *LjNOD70* expression during nodule developmental and the complexity of the gene family represented by *LjNOD70* in the *L. japonicus* genome. The experiments to localize the LjN70 protein to the peribacteroid membrane in *L. japonicus* nodules were performed by Jim Guenther and Dr. Dan Roberts, University of Tennessee, Knoxville, using the anti-LjN70 antibodies generated by the author of the thesis. . The experiments in the Chapter 6 were performed equally by the author and Dr. Krzysztof Szczyglowski.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

SYMBIOTIC NITROGEN FIXATION

1.1. Legumes.

Plants are an important source of nourishment for the world's rapidly growing population. Plant growth and development depends on the availability of nitrogen and other nutrient compounds in the soil, which are rapidly depleted in modern agricultural practices. After photosynthesis, the acquisition of nitrogen is considered to be the second most important process for plant growth (Vance, 1998). To replenish soil nitrogen content, chemical fertilizers are extensively used. However, this approach suffers from serious drawbacks, including the large amount of energy consumed during fertilizer production and its severe negative impact on the environment, manifested by nitrate contamination of the groundwater. Biological nitrogen fixation by free-living and symbiotic prokaryotic organisms contributes approximately 120 million tons of fixed nitrogen per year to the biosphere, almost twice the amount produced by nitrogen fertilizer industry (Vance, 1998). The ability to fix nitrogen is limited to prokaryotes. However, a number of families of flowering plants have developed the ability to form a symbiotic relationship with nitrogen-fixing soil microorganisms. The latter belong to two distinct groups, gram-negative bacteria of the genera *Rhizobium*, *Bradyrhizobium*, *Mezorhizobium*, *Azorhizobium* and *Sinorhizobium* (collectively referred to as rhizobia) and gram-positive bacteria of the genus *Frankia*. The bacterial partner fixes dinitrogen by

converting it to ammonia, which is utilized by the plant host for growth. In return, the plant host provides reduced carbon products as a source of energy for the bacterial endosymbiont and provides an ecological niche.

The legume-*Rhizobia* symbiosis is the most prominent example of this relationship, and the most members of this plant family are capable of establishing symbiotic interactions (Gualtieri and Bisseling, 2000). This symbiosis also has the highest economical importance, since it has been estimated that that 80% of biologically fixed nitrogen is generated by the legume-*Rhizobia* symbiosis (Vance, 1998). Therefore it is not surprising, that this symbiosis has been the subject of scientific investigations since the late 19th century, when Hellriegel and Willfarth conclusively showed that root nodules allow the assimilation of nitrogen from the air (Hellriegel and Willfarth, 1888). Since then, the physiological, biochemical and, in past two decades, the molecular aspects of our understanding of the legume-rhizobia symbiosis have greatly advanced (reviewed in Gualtieri and Bisseling, 2000; Kaminski et al., 1998; Mylona et al., 1995; Pueppke, 1996; Hirsch, 1992; Nap and Bisseling, 1990). In particular, significant progress has been made in the development of the molecular genetics of the microsymbiont. Many of the rhizobial genes and the biochemical pathways required for the symbiosis have been identified and well characterized, in particular the nature of the primary rhizobial signaling molecules, the Nod factors have been elucidated in detail (reviewed in Schultze and Kondorosi, 1998; Pueppke, 1996; Long, 1996). Moreover, the symbiotic plasmid of the broad-host range symbiont *Rhizobium* sp. NGR234, containing many of the genes required for symbiosis, has been completely sequenced (Freiberg et al., 1997), the sequencing of the entire genome

of *Rhizobium meliloti* is well advanced (http://cmgm.stanford.edu/~mbarnett/genome.htm), and the entire DNA sequence of the symbiosis island of *Mesorhizobium loti* has been completed (Frans de Bruijn, personal communication). This information is ushering in a new era in systematic genome-wide analyses of the intricacies of the *Rhizobial* biology, both in the free-living and symbiotic stages (for example see Perret et al., 1999).

On the other hand, the understanding of the molecular basis of the symbiotic interactions from the plant side is lagging far behind. Despite the availability of a variety of plant mutants defective in various stages of symbiosis, only a single gene underlying a symbiotic phenotype has been cloned thus far (Schauser et al., 1999). However, a vast number of plant genes, specifically induced during the interaction with rhizobia have been identified, but the biochemical and biological function of many of these genes is still unknown (see below). Below, I will first summarize what is currently known about the physiology, biochemistry and (molecular) genetics of the legume-rhizobia symbiosis.

2. Nodule development and functioning

2.1. Early stages of nodule development

2.1.1. Signal exchange between free-living rhizobia and the plant host. Symbiotic nitrogen fixation represents a unique example of interactions between two very distant species, and takes place in specialized organ-like structures called root- or stem nodules. The development of this organ is a complex multistep process which can be subdivided into early and late stages (Sprent, 1989). The early stages of nodule development start

with mutual recognition by the symbiotic partners. This, in turn, sets in motion a cascade of events leading to the initiation of root cortical cell divisions eventually giving rise to a nodule primordium. Subsequently, the cells of the nodule primordium are infected and colonized by rhizobia. At this stage, the nitrogen fixation commences and nodule development enters into its late stage.

Different stages of nodule organogenesis are hallmarked by the expression of specific sets of plant genes, termed nodulin genes, to reflect their specific association with the processes of nodule development and/or functioning (van Kammen, 1984). Nodulin genes have been traditionally categorized as "early or late", depending on the time point of their induction (Nap and Bisseling, 1990). Here I will present a brief overview of the events taking place during early stages of nodule development. For comprehensive reviews on this topic see Gualtieri and Bisseling, (2000); Schultze and Kondorosi, (1998); Geurts and Franssen, (1996); Long, (1996); Mylona et al, (1995); Hirsh, (1992); Sprent, (1989).

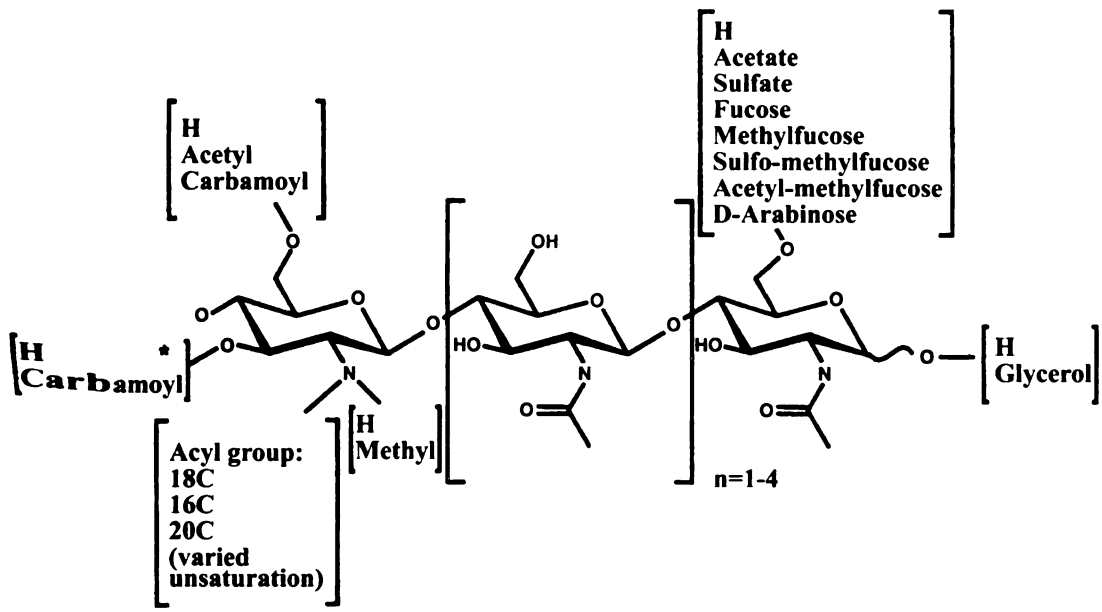
Nodule morphogenesis is normally initiated by the mutual recognition of specific signal molecules produced by both plant and bacterial symbiotic partners (Fig. 1.1; Pueppke, 1996; Long, 1996; Geurts and Franssen, 1996). Flavonoid compounds secreted by the plant host induce in rhizobia the synthesis of a specific set of morphogenic lipochitooligosaccharide molecules, the Nod factors (Fig. 1.1; Spaink, 1992; Mylona et al., 1995; Long, 1996; Pueppke, 1996). Nod factor molecules, in turn, trigger a cascade of events resulting in penetration of the plant root by rhizobia and the initiation of nodule organogenesis (Fig. 1.1; Mylona et al., 1995; Geurts and Franssen, 1996). A typical Nod

factor resembles a short chitin molecule and consists of 3-6 moieties of β-1,4-N-acetylglucosamine, with a fatty acid linked to the non-reducing end (Fig. 1.2; Schultze and Kondorosi, 1998; Pueppke, 1996). Different rhizobia species produce one or more distinct Nod factors, which differ primarily in the spectrum of side groups attached to the chitin backbone (Fig. 1.2; Pueppke, 1996). The host range of a bacterial endosymbiont is primarily determined by the repertoire of the Nod factors it can produce, due to the high specificity of recognition of a particular Nod factor by a particular plant symbiotic partner (Spaink, 1992; Spaink and Lugtenberg, 1994; Long, 1996; Pueppke, 1996; Niebel et al., 1999). Exogeneous application in the absence of rhizobia of the purified Nod factors at picomolar concentrations initiates several cellular responses characteristic of early stages of symbiotic interactions. In addition, it can lead to the induction of expression of certain nodulin genes and, in some cases, even cause the formation of nodule-like structures (Spaink et al., 1991; van Brussel et al., 1992; Spaink and Lugtenberg, 1994; Long, 1996).

Figure 1.1. Schematic representation of the early events taking place during the symbiotic interactions. Plant cells secrete flavonoid compounds which in turn stimulate production in rhizobia of specific signaling compounds, the Nod factors. The latter interact with yet unidentified plant receptor and set in a motion a cascade of cellular events depicted on the diagram which eventually lead to the formation of nodule primordium. From Hirsch, (1992).

Rhizobium meliloti
nod genes induced
Nod factor
(flavonoids) inducers
Root hair deformation
Infection thread formation
Altered hormone levels
Increase in flavonoids
Cytoskeletal rearrangements
Ca2+

Figure1.2. A generic structure of the Nod factors. A typical Nod factor molecule is based on a oligomeric backbone of 3-6 β-1,4-linked N-acetylglucosamine moieties containing a spectrum of substitutions at the reducing and non-reducing ends. The substituting groups identified in the Nod factors produced by various species of rhizobia are shown. From Long, (1996).


H
Acetyl
Carbamoyl
H
Acetate
Sulfate
Fucose
Methylfucose
Sulfo-methylfucose
Acetyl-methylfucose
D-Arabinose
H
Carbamoyl
Acyl group:
18C
16C
20C
(varied
unsaturation)
H
Methyl
n=1-4
H
Glycerol
Non-reducing end
Reducing end

2.1.2. Perception of the Nod factors and subsequent events. The interactions of Nod factors with yet hypothetical plant receptor(s) lead to a number of rapid responses in the host plants, such as the depolarization of the membrane of root epidermis cells (Felle et al., 1995; Kurkdjian, 1995), rapid intracellular alkalization (Felle et al., 1996), calcium spiking in root hairs (Ehrhardt et al., 1996), and root hair deformation and curling (Heidstra et al., 1994). There is mounting evidence that significant rearrangements of the root hair cytoskeleton occur within minutes after Nod factor application (Crdenas et al., 1998). The Nod factors signal transduction pathway(s) *in planta* and the molecular mechanisms of their recognition by plant are presently unknown. However, evidence for the involvement of G-proteins in the Nod factor signaling cascades has recently been presented based on pharmacological studies (Pingret et al., 1999). In addition, the early nodulin gene *rip1*, which encodes a peroxidase, is rapidly but transiently induced soon after plant-*Rhizobium* interaction, suggesting that oxidative processes may take place during early stages of symbiotic interactions (Cook et al., 1995). Root hair curling results in the formation of a structure referred to as a "shepherd hook", which physically entraps the bacteria. Local hydrolysis of the plant cell wall occurs at the region of rhizobial attachment to the root hair curl, and invagination of the plant plasma membrane allows the bacteria to penetrate the cells of the root epidermis (Van Spronsen et al., 1994). New plant cell wall material is deposited around the bacteria as it enters the plant cell, leading to the formation of a tubular structure, the infection thread, in which the bacteria divide and migrate to the nodule primordium located in the root cortex (Kijne, 1992). Two well studied early nodulins, *ENOD5* and

ENOD12, are expressed during this process. *ENOD5* transcripts are present in cells containing growing infection threads (Scheres, 1990a). *ENOD12* is expressed in root hairs and root cells which contain growing infection threads, as well as in cells positioned several cell layers ahead of the growing infection threads (Scheres, 1990b). Both these nodulin genes encode proline-rich proteins, which are thought to be components of the infection thread, and, in case of the ENOD12 protein, the cell wall of cortical cells located in path of the extending infection thread (Schultze et al., 1994). However, it appears that *ENOD12* is not required for successful nodulation in at least some alfalfa varieties (Csanadi et al., 1994).

Concomitant with root hair deformation and infection thread formation, fully differentiated cells of the root cortex are induced to divide (Nap and Bisseling, 1990; Verma, 1992; Mylona et al., 1995; Hirsch, 1992). Generally, temperate legumes such as alfalfa, pea, and clover will form indeterminate type of nodules, which are characterized by the presence of a persistent apical meristem (Newcomb et al., 1979). In contrast, tropical legumes such as soybean, *Sesbania rostrata*, and *Lotus japonicus* develop determinate type of nodules, which do not have meristematic activity but are rather assumed to grow by a cell expansion mechanism (Newcomb et al., 1979). The position of the root cortical cells which are induced to divide first depends on the type of nodules a particular legume species forms (Hirsch, 1992). In the case of indeterminate nodules, these are the cells of the inner cortex, while the cells of the outer cortex initiate the formation of the determinate type of nodules. Division of root cortical cells results in the formation of a rudimentary nodule structure, the nodule primordium, which, as pointed

above, can also be induced by the application of purified Nod factors (Spaink et al., 1991; Spaink and Lugtenberg, 1994).

2.1.3. Regulation of the early stages of nodule development. The program of nodule development is determined primarily by the plant host. The most convincing evidence for it comes from the existence of a plethora of plant mutants affected at different stages of nodule development. The mutant phenotypes range from a total lack of nodules (non-nodulation; Nod^-), to the formation of ineffective nodules (Nod^+Fix^-) with varying amounts of residual nitrogenase activity, to the production of excessive amounts of nodules (supernodulation; Nod^{++}, reviewed in Caetano-Anolles and Gresshoff, 1991; also see Szczyglowski et al., 1998; Schauser et al., 1998). Most of the genes underlying the symbiotic phenotypes in these plant mutants have not yet been identified. In fact, the first and, so far, the only plant gene in which a mutation causes a Nod^- phenotype has been cloned from *Lotus japonicus nin* mutant (Schauser et. al., 1999). *Nin* mutant plants do not show the formation of infection threads or cortical cell divisions, however they do undergo root hair curling and deformation, suggesting that at least some elements of the Nod factor perception are functional in this mutant line (Schauser et. al., 1999). The protein product of the *nin* gene has regional similarity to transcription factors and a predicted DNA binding/ dimerization domain, suggesting that the nin protein may regulate the expression of genes involved in early stages of the symbiotic interaction (Schauser et. al., 1999). Interestingly, the *nin* mRNA is consitutively expressed in *L. japonicus* roots and increases in mature nitrogen-fixing nodules (Schauser et. al., 1999).

Not surprisingly, increasing evidence suggests that plant hormones play significant roles in regulating various aspects of the symbiotic interactions (Hirsch and Fang, 1994). Application of exogenous auxin and cytokinin induces root cortical cell divisions (Libbenga et al., 1973). Treatment of alfalfa with auxin transport inhibitors (NPA or TIBA), induces the formation of nodule-like structures or "pseudonodules", in which early nodulin genes, such as *ENOD2* and *Nms 30*, are expressed (Hirsch et al., 1989). Exogenous cytokinin has been shown to induce expression of the early nodulin gene *ENOD2* (Dehio and de Druijn, 1992) and *ENOD40* (Fang and Hirsch, 1998). A role of ethylene as an inhibitor of nodule formation has been extensively documented at least in some legumes. The application of exogenous ethylene, or its precursor 1-aminocyclopropane-1-carboxylate (ACC), has been shown to be detrimental to nodule formation in pea (Lee and Larue, 1992). On the other hand, inhibitors of ethylene synthesis, such as aminoethoxyvinylglycine (AVG), have been shown to have a stimulatory effect on nodule formation in alfalfa (Peters and Crist-Estes., 1989). A direct genetic link between ethylene signaling and regulation of nodulation has been established by the isolation of the *sickle* mutant of *Medicago truncatula*, which forms excessive number of nodules comparing to wild-type plants and also exhibits insensitivity to ethylene (Penmetsa and Cook, 1997). However, the role and/or mechanisms of action of ethylene during the symbiotic interactions may vary among different legume species, since an ethylene insensitive soybean mutant has been shown to have a normal number of nodules (Schmidt et al., 1999). Ethylene has also been implicated in providing positional

information for the formation of the nodule primordia opposite of protoxylem poles (Heidstra et al., 1997).

2.2 Late stages of nodule development

2.2.1. Formation of the mature nitrogen-fixing nodule. The late stages of nodule development commence with the release of bacteria from the infection thread into the cells of the central region of nodule primordium, via a process resembling endocytosis (Basset et al., 1977). The bacterial cells replicate and enlarge, filling up most of the cytoplasm of the infected cells, and subsequently differentiate into a distinct form, called the bacteroid. A fraction of the cells in the central region in the nodule primordium remain uninfected and are typically much smaller than infected cells. Infected and uninfected cells form the central tissue of nodule, and are surrounded by peripheral tissues. The organization of the latter varies among the legumes. In the case of indeterminate-type nodules, closest to the central zone usually lies the nodule vascular system fully embedded in nodule parenchyma, followed by cell layers of the nodule endodermis and cortex. In the determinate type of nodules, outside of the parenchyma, cell layers of the sclerenchyma, nodule endodermis, nodule cortex and periderm are found to be located (Hirsch, 1992). Nodule peripheral tissues are believed to constitute a barrier for oxygen diffusion (see below). In indeterminate-type nodules, several developmental zones can be distinguished in the central tissue, from the nodule apex to the root. Zone I, the meristematic zone, is located in the apex of the nodule and it contains actively dividing cells which give rise to

the remainder of cells. The next zone is invasion zone II, where plant cells are infected by rhizobia. The N_2 fixation commences in the adjacent region, interzone II-III, and continues throughout zone III, the nitrogen fixation zone. Zone IV is the senescence zone where the cytoplasm of plant cells and bacteroids are degraded. Determinate type of nodules lack a persistent meristem and, at any given point, the plant cells within a single nodule of this type are at more or less similar developmental stage. The comparison of the structures of the determinate- and indeterminate-type nodules is shown in Figure 1.3.

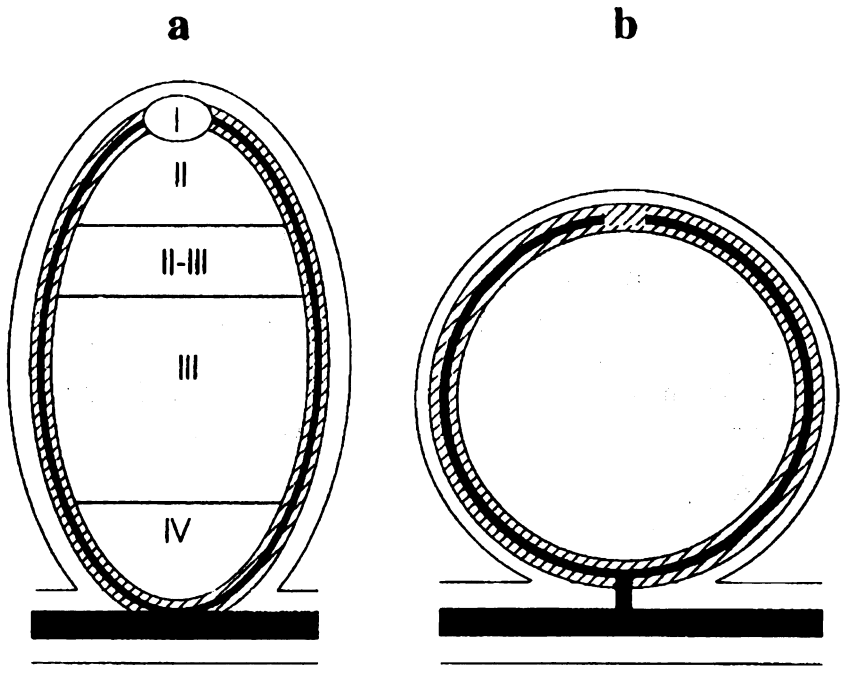


Figure 1.3. Structure of the indeterminate- and determinate-type legume nodules. (a) Structure of indeterminate-type nodule. This type of nodules contains 5 distinctive zones

in the central region: zone I-the meristematic zone; zone II-the pre-fixation zone; zone II-III- the interzone; zone III-the nitrogen fixation zone and zone IV-the senescence zone. (b) Structure of determinate-type nodule. The central region contains cells in a similar stage of development. Adapted from Hadri et al., (1998).

Bacteria released into plant cells undergo a multistage process of "terminal" differentiation ultimately resulting in the formation of their nitrogen-fixing forms, called bacteroids. The process of differentiation is accompanied by drastic changes in bacterial morphology and physiology, and has been extensively reviewed by Kaminski et al., (1998) and Kahn et al., (1998). Essentially, the metabolism and overall physiology of bacteroids is adapted to fixing and providing nitrogen for the plant host and utilizing the carbon compounds provided by the host in a low-oxygen environment.

2.2.2. Nitrogenase and nitrogen fixation. The actual process of symbiotic nitrogen fixation is catalyzed by a multimeric rhizobial enzyme, called nitrogenase, and can be summarized as follows:

$$N_2 + 8H^+ + 8e^- + 16\text{Mg-ATP} \rightarrow 2NH_3 + H_2 + 16\text{Mg-ADP} + 16P_I$$

A typical nitrogenase enzyme complex is composed of two protein components, usually referred to as the Fe protein and FeMo proteins (reviewed in Howard and Rees, 1996; Burgess and Lowe, 1996). The Fe protein is a homodimer, containing one Fe_4S_4 cluster and two ATP binding sites along the interface of the two subunits, which are encoded by rhizobial *nifH* gene. The FeMo protein is the site of actual nitrogen reduction. It consists

of two subunits (α and β), encoded by the bacterial *nif D* and *nifK* genes, which interact with a single homodimer molecule of the Fe protein. The FeMo protein contains two novel types of FeS clusters: a P cluster and a FeMo cluster, containing molybdenum in a traditional nitrogenase: in some other variants the Mo atom is replaced with either Fe or V (Eady, 1996). The Fe protein receives electrons from an electron donor, such as ferredoxin, and passes them to the FeMo component in a process requiring the hydrolysis of 2 ATP molecules per single transferred electron (Seefeld and Dean, 1997). The crystal structure of a complex between the Fe and FeMo proteins with bound ADP-AlF_4^- (a compound resembling a possible intermediate of ATP hydrolysis) has been solved (Schindelin et al., 1997). The association of Fe and FeMo proteins requires the binding of 2 ATP molecules by the Fe protein. Nucleotide binding induces drastic conformational changes in the Fe protein, allowing it to interact with and transfer electrons to the FeMo protein (Schindelin et al., 1997). These changes are reversed by the hydrolysis of ATP by the Fe-FeMo protein complex, since Fe protein alone can not hydrolyze ATP. Nucleotide hydrolysis is required for the dissociation of the two parts of the nitrogenase complex. X-ray crystallographic data suggest that the binding and hydrolysis of ATP by the Fe protein act as a molecular switch in a mode similar to that of GTP-binding proteins, such as Ras (Schindelin et al., 1997).

2.2.3. The oxygen paradox. A unique and perplexing feature of nitrogenase is its high sensitivity to molecular oxygen: contact with O_2 irreversibly denatures the enzyme. This property of nitrogenase conflicts with the high energetic cost of nitrogen fixation.

The latter requires a high rate of bacteroid respiration which, in turn, relies on a constant supply of oxygen. The oxygen dilemma is solved by several means. 1. The concentration of oxygen in the central zone within a nodule is maintained at a very low level, 5-30 nM, comparing to the equilibrium O_2 concentration in aerobic cells (230 μM). Moreover, the existence of an oxygen diffusion barrier has been demonstrated by directly measuring the oxygen concentrations inside different regions of a nodule. The concentration of oxygen decreases significantly in the peripheral tissues, suggesting that these tissues provide a major obstacle for oxygen diffusion (Tjepkema and Yokum, 1974). 2. Specialized plant proteins, leghemoglobins, bind and transport oxygen to the respiring bacteroids, decreasing the free O_2 concentration in the cytoplasm (Appleby, 1984). Leghemoglobins are late nodulins that constitute the most abundant group of proteins in nodules (Brisson et al., 1982). 3. The rate of bacterial respiration is maintained at a high rate to maximize the consumption of oxygen. Specialized bacterial terminal oxidases, with a high affinity for oxygen, are induced during the nitrogen fixation process (reviewed in Kaminski et al., 1998). In summary, free oxygen does pose a serious problem for nitrogen fixation but both symbiotic partners have evolved a set of complex mechanisms to deal with the problem.

In this respect, it is quite interesting that a free-living nitrogen fixing bacteria, *Streptomyces thermoautotrophicus* UBT1, has been recently discovered which contains a nitrogenase which is totally insensitive to molecular oxygen (Ribbe et al., 1997). Moreover, this nitrogenase requires superoxide anion radicals ($O_2^{\cdot-}$) as a source of

electrons for nitrogen fixation and therefore depends on $O_2^{\cdot-}$ and O_2 (Ribbe et al., 1997). Another unusual properties of the *S. thermoautotrophicus* enzyme is a lower MgATP requirement, only 4-12 MgATPs per molecule of N_2 fixed, comparing to 16 MgATPs in case of typical nitrogenases. Moreover, it is unable to reduce acetylene (Ribbe et al., 1997). The full amino acid sequence of any of the protein components of this novel nitrogenase is not yet available in the literature. This information together with a detailed molecular and structural characterization of this enzyme will be invaluable for a better understanding of the enzymology of nitrogen fixation and perhaps, for engineering more efficient and less oxygen sensitive nitrogenases in rhizobia and economically important free-living nitrogen fixing microorganisms. However, by far the most exciting experiment would be to express this or a similar nitrogenase enzyme in plants This would circumvent one of the most severe limitations of expressing a typical nitrogenase in plants: the sensitivity to oxygen, which is ubiquitous in plant tissues. If successful, this attempt may allow plants to fix free nitrogen, which has always been a major long-term goal of nitrogen-fixation research.

2.2.4. Assimilation of fixed nitrogen. It has been long believed that bacteroids provide fixed nitrogen to the plant in the form of ammonia (Udvardi and Day, 1997). First the fixed nitrogen in the form of NH_3 diffuses passively through the bacterial membranes into the peribacteroid space (PBS), where it is converted to NH_4^+, due to the relatively low pH of this compartment. The ammonium ion is then transported through the peribacteroid membrane (PBM) via a channel or a transporter. A recently identified monovalent cation channel GmSAT1 from soybean nodules is located in the PBM and is

thought to be involved in the transport of NH_4^+ from the PBS into the plant cytosol (Kaiser et al., 1998). However, recent evidence suggests that alanine, not ammonia, is transported across bacterial membranes and the PBM in soybean nodules, since alanine was found to be the major ^{15}N-labeled compound excreted by the purified soybean bacteroids during incubation in ^{15}N atmosphere in the presence of D,L-malate (Waters et al., 1998). The reason for the discrepancies within the previously mentioned studies was the inclusion of a sucrose density gradient purification step by Waters et al., (1998) which removed the contaminating enzymes of the plant cytosol from the bacteroid fraction. Addition of plant cytosolic fraction to the bacteroids resulted in a quick production of ammonia from alanine (Waters et al., 1998). Incorporation of ammonia in alanine is most likely mediated by bacterial alanine dehydrogenase, whose activity increases in the symbiotic state in contrast to the other rhizobial enzymes involved in ammonia assimilation (Waters et al, 1998; Werner et al., 1980)

The first step in the utilization of the fixed nitrogen by plant is the synthesis of glutamate via the glutamine synthetase/glutamate synthetase (GS/GOGAT) cycle (Fig. 1.4; Lam et al., 1996). Glutamate is subsequently used for the synthesis of nitrogen transport compounds which, depending on the species of legumes, can be either amides, predominantly asparagine, or ureides, mostly allantoin or allantoic acid (Fig. 1.4; Verma and Fortin, 1989). Most, if not all, key plant enzymes involved in nitrogen assimilation in nodules have been purified and the corresponding cDNAs and genes have been cloned and characterized. Nodule specific GS isoforms has been isolated from a number of legumes (Bennett et al., 1989; Boron et al., 1989; Boron and Legocki, 1993; Stanford et al., 1993;

Temple et al., 1995; Forde et al., 1989). The GOGAT from alfalfa nodules has been extensively investigated (Gregerson et al., 1993; Vance et al., 1995). The predominant form of GOGAT in this tissue is NADH-GOGAT, whose activity increases substantially during nodule development in parallel with an increase in its mRNA and protein level (Gregerson et al., 1993; Vance et al., 1995). Plants contain at least one more GOGAT, the Fd-GOGAT, which is most abundant in green tissues, however its activity, mRNA and protein levels in nodules are undetectable or low comparing to that of NADH-GOGAT (Vance et al., 1995 and references therein). NADH-GOGAT activity, mRNA and protein levels are significantly downregulated in mutant ineffective nodules which do not fix nitrogen (Gregerson et al., 1993; Vance et al., 1995). NADH-GOGAT contains a targeting peptide, however it is not clear to which organelle the protein is targeted (Vance et al., 1995).

The nitrogen from glutamate is incorporated into aspartate and asparagine via the consecutive action of aspartate aminotransferase (AAT) and asparagine synthase (AS; Fig. 1.4; Lam et al., 1995). Asparagine is the major nitrogen transport compound in many plant species (Lea et al., 1990), and the final product of nitrogen fixation in temperate legumes, such as alfalfa (Vance, 1990). Alfalfa nodule AAT and AS have been purified and extensively characterized at the molecular level (Gantt et al., 1992; Shi et al., 1997; Vance and Gantt, 1992; Vance et al., 1994 and references therein). Alfalfa contains two AAT isoforms: AAT-1 appears to be constitutive while the expression of AAT-2 isoform is enhanced 15-20 folds in the mature nodules comparing to other tissues (Gantt et al., 1992). Analysis of the AAT-2 amino acid sequence and immunogold labeling

suggest that it is localized to amyloplast (Robinson et al., 1994). AS is dramatically induced on the level of mRNA and protein in mature nitrogen-fixing nodules of alfalfa (Shi et al., 1997). AS appears to be encoded by a single gene in alfalfa and the corresponding gene is also induced in leaves, after a dark treatment (Shi et al., 1997). In alfalfa, the transcripts of NADH-GOGAT, AAT, AS and phosphoenolpyruvate carboxilase (PEPC, see below) have been detected by *in situ* hybridization in both infected and uninfected cells of the alfalfa nodules (Vance et al., 1995; Robinson et al., 1994; Shi et al., 1997). This suggests that uninfected cells harbor the necessary enzymes for the incorporation of fixed nitrogen into glutamate and asparagine (Shi et al., 1997; also see below). The reason for the presence of GS in uninfected cells is still an open question (Temple et al., 1995; Forde et al., 1989; Shi et al, 1997).

In tropical legumes, such as soybean, the nitrogen assimilated via the GS/GOGAT pathway is channeled into the purine biosynthetic pathway, leading to the synthesis of uric acid in the infected cells (Verma and Fortin, 1989). The enzymology of this pathway is highly conserved (Zalkin and Dixon, 1992). Several enzymes involved in purine biosynthesis have been identified in nodules of tropical legumes (Atkins, 1991; Boland and Schubert, 1983; Schnorr et al., 1996). Uric acid is converted to allantoin and allantoic acid via the action of uricase and allantoinase, which are localized in the peroxisomes of uninfected cells (Bergmann et al., 1983; Hanks et al., 1983; Nguyen et al., 1985). Nodules contain a nodule-specific form of uricase- uricase II, which consists of four subunits of nodulin 35 (Bergmann et al., 1983).

2.2.5. Carbon metabolism in nodules. Fixation of 1 gram of nitrogen requires estimated 5-10 grams of carbon (Phillips, 1980), and the cost of N_2 fixation can amount up to 1/3 of the total photosynthate in perennial legumes (Maxwell et al., 1984). In nodules, the metabolism of carbon is adapted to the specific requirements of the nitrogen fixation process. 1. The plant partner provides the appropriate carbon compounds to the bacteroids for the production of ATP, rapidly consumed during nitrogenase activity, and for other biosynthetic processes. 2. Assimilation of fixed nitrogen into glutamine, glutamate and nitrogen transport compounds requires a constant supply of carbon skeletons. 3. Carbon metabolism in infected cells has to operate under anoxic conditions.

It has been well established that nodules derive carbon from photosynthetic tissues primarily in the form of sucrose (Reibach and Streeter, 1983), which is broken to UDP-glucose and fructose by the enzymatic activity of sucrose synthase (SS). Nodules contain a specific isoform of SS, initially identified as nodulin 100 (Thummler and Verma, 1987). UDP-glucose is believed to be metabolized to phosphoenolpyruvate (PEP) via the action of glycolytic enzymes in nodules (Kahn et al., 1998 and references therein). However, sucrose and other sugars do not appear to be major carbon metabolites in nodules (reviewed in Kahn et al., 1998). Rather, C_4 dicarboxylic organic acids (DAs), such as succinate, fumarate and primarily malate, are the most important players in the nodule carbon metabolism (Fig. 1.4). There is a substantial amount of evidence that DAs are the actual source of plant carbon for nitrogen-fixing bacteroids. For example, DAs are efficient substrates for *in vitro* nitrogen fixation by purified bacteroids (Bergersen, 1977). Moreover, isolated symbiosomes can readily take up (Udvardi et al., 1988) and

metabolize DAs (Salminen and Streeter, 1987). On the other hand, sugars are not efficiently transported into bacteroids (Udvardi et al., 1990).

The rhizobial DA transport system (*dctAB*) has been found to be absolutely required for the establishment of effective nodules (Ronson et al., 1981). A plant DA transporter activity has been identified in the peribacteroid membrane (PBM), but the corresponding gene has not yet been cloned (Udvardi et al., 1988). Bacteroids utilize DAs via the TCA cycle to generate ATP, reducing equivalents and carbon skeletons for biosynthetic processes. The TCA cycle requires a constant supply of acetyl-CoA, which is most likely generated from malate, via the actions of malic enzymes and pyruvate dehydrogenase (reviewed in Kahn et al., 1988).

The primary route for the generation of malate from oxaloacetate in nodule cells appears to be via the malate dehydrogenase (MDH) pathway (Fig. 1.4; Miller et al., 1998). It is believed that the TCA cycle does not contribute significantly to DA biosynthesis in plant cells (Kahn et al., 1988). A cDNA corresponding to a nodule-enhanced isoform of MDH, neMDH, has been recently cloned from alfalfa (Miller et al., 1998). The neMDH enzyme has a higher capacity for malate production than any of the four other MDH isoforms present in alfalfa (Miller et al., 1998). Taken together with the specific induction of neMDH levels in nodules, this observation suggests that neMDH may be responsible for the dramatic increase of malate content in this organ (Miller et al., 1998).

In nodules, oxaloacetate (OA) is mainly produced by the PEP carboxylase (PEPC) from H_2CO_3 and PEP (Fig. 1.4; Vance et al., 1994). PEPC occurs as multiple isoforms in

many plant species, and its major role in photosynthetic processes in C4 and CAM plants is well known (Ting, 1985). However, PEPC is also one of the central enzymes of nodule carbon- and nitrogen metabolism, primarily because OA is a direct precursor of the two key metabolites, malate and aspartate. PEPC activity, mRNA and protein levels are significantly upregulated in the nodules (Pathirana et al., 1997). The structure and expression of the alfalfa *PEPC-7* gene corresponding to the nodule-enhanced isoform has been well characterized (Pathirana et al., 1992, 1997). The *PEPC-7* gene is expressed throughout the nodule in both infected and uninfected cells, nodule parenchyma, in the pericycle of the nodule vascular system, as well as in a number of non-symbiotic tissues (Pathirana et al., 1997). Immunolocalization studies have shown that PEPC is a cytosolic protein (Robinson et al., 1996). The presence of PEPC in the uninfected cells along with the enzymes of nitrogen metabolism such as NADH-GOGAT, AS, AAT (see above) suggests that this cell type may also be involved in the assimilation of the fixed nitrogen into asparagine (Shi et al., 1997)

2.2.6. Synthesis and functioning of the peribacteroid membrane. The rhizobial endosymbionts are engulfed by a host plasma membrane as they enter plant cells from the infection thread in a process resembling endocytosis (Basset et al., 1977). The peribacteroid membrane (PBM), initially derived from the host plasma membrane, always separates the endosymbiont from the host cell cytosol. A long standing dogma has been that a major function of PBM is to separate the prokaryotic endosymbiont from the environment of the host cytoplasm, thus possibly preventing the induction of defense responses by plant cells. The second function proposed for the PBM is to control the

flow of metabolites to and from the bacteria. For example, the PBM contains transporters for C_4 dicarboxylic acids, such as malate, and ammonia (see above). On the other hand, PBM is almost impermeable to sugars (Udvardi et al., 1990). The bacterial cytoplasm is also enclosed in inner and outer membranes The milieu between the bacterial outer membrane and the PBM constitutes the peribacteroid space (PBS), where a number of enzymes, common to the plant vacuole, have been found (Mellor, 1989). Even though the PBM is derived from the plasmalemma, it also develops its own characteristic features during bacteroid differentiation. In general, the composition of the PBM resembles both plasmalemma and tonoplast membranes (Verma and Hong, 1996). A unique feature of the PBM is the presence of a set of nodulins (Verma, 1992; Cheon et al., 1993). The best studied protein from this group is nodulin-26 (Miao and Verma, 1993), which is related to the a family of water channels (aquaporins) from the major intrinsic protein family (MIP). Nodulin-26 has been shown to be able to transport water and glycerol, suggesting that it may be involved in osmoregulation (Dean et al., 1999).

A characteristic feature accompanying the process of bacterial multiplication in infected cells is a massive synthesis of peribacteroid membranes, the total area of which have been estimated to be 30 times that of the plasma membrane (Verma, 1992). Biogenesis of the PBM and targeting of proteins and other compounds to and from it is not well understood (Verma and Hong, 1996). The PBM is associated with smooth and coated vesicles, however their role in its biogenesis and/or functioning is unclear (Mellor and Werner, 1987). PBM-associated nodulins such as nodulin-26 and nodulin-24 are not found in the normal plant plasma membrane, suggesting that they must contain PBM

specific targeting sequences (Fortin et al., 1985), but the identity of the latter is presently unknown (Fortin et al., 1987). Several small GTP-binding proteins belonging to the Rab family, known to be involved in regulating the vesicle traffic in eukaryotic cells, have been shown to be up-regulated in the nodules (Cheon et al., 1993; Borg et al., 1997). These proteins may, in part, be involved in regulating the vesicle-mediated transport to and/or from PBM (Cheon et al., 1993).

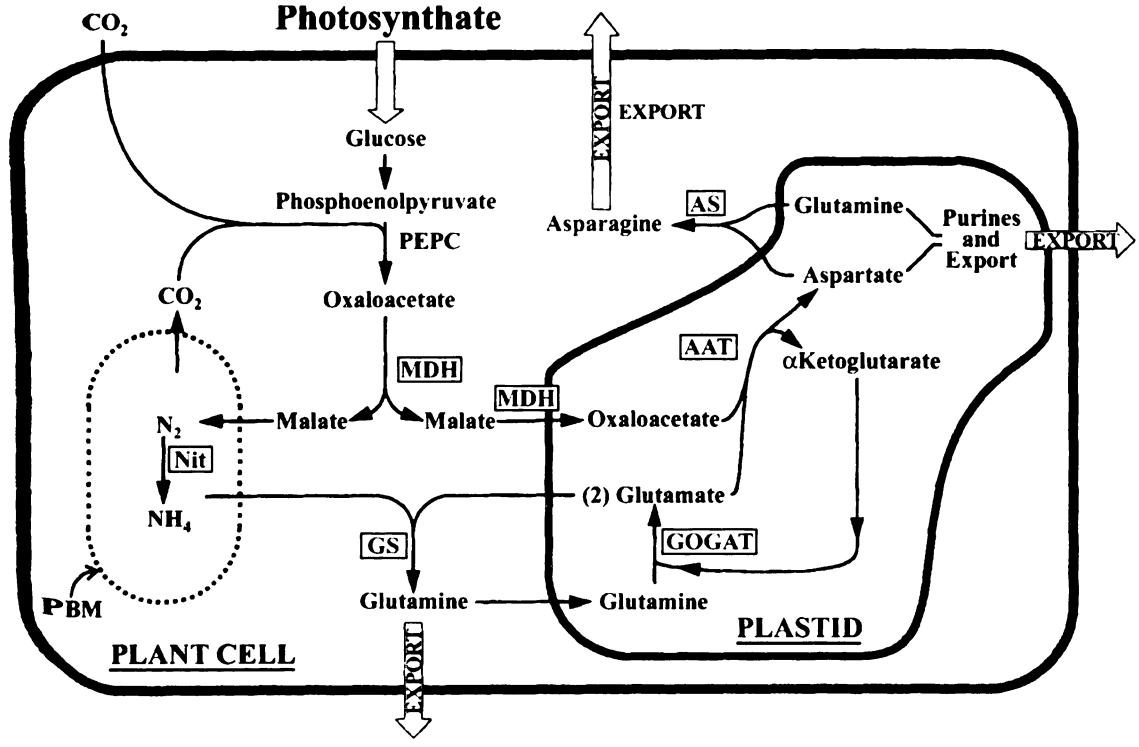


Figure 1.4. Schematic representation of the major pathways of nitrogen and carbon metabolism in nodules. From Vance et al., (1994).

2.2.7. Regulation of the functioning of mature nitrogen-fixing nodules. Given the high complexity of the cytological, physiological and biochemical events taking place in plant host cells at late stages of nodule development and functioning, we know surprisingly little about how these events are regulated. Very few candidates for regulatory molecules specifically involved in late stages of nodule development have been isolated. The *nmh7* gene, encoding a MADS-box containing protein, is expressed specifically in infected cells of nodules and also in flowers, but not in uninfected root or leaf tissues (Heard and Dunn, 1995). However, it is not known at what stages of nodule development the *nmh7* gene, and a related nodule-specific gene *nmhC5*, are induced (Heard and Dunn, 1995; Heard et al., 1997). The nodule-enhanced gene *GmNdx1* from soybean encodes a homeobox-containing protein (Jorgensen et al., 1999). The expression of the *GmNdx1* gene is induced early during symbiotic interactions and gradually increases during the late stages of nodule development (Jorgensen et al., 1999). *GmNdx1* mRNA is also present in dividing cells of other tissues, suggesting that the *GmNdx1* product may be involved in regulatory events at both early and late stages of symbiosis, as well as throughout plant development (Jorgensen et al., 1999). Recently, a number of nodule-enhanced Expressed Sequence Tags (ESTs) with putative regulatory functions have been isolated from young *Medicago truncatula* nodules (Gyorgyey et al., 2000). However, the temporary and spatial expression profiles of these cDNAs are not known (Gyorgyey et al., 2000).

Reversible phosphorylation has also been implicated in the regulation of nodulin activity. Nodulin-26 has been shown to be phosphorylated on the residue 262 by a calcium-dependent protein kinase, which changes its voltage-sensitive channel activity (Weaver and Roberts, 1992; Lee et al., 1995). Phosphorylation of soybean nodule PEPC has been found to result in a decrease of the inhibition of enzyme activity by L-malate (Zhang et al., 1995). The phosphorylation status of PEPC has been directly linked to the photosynthetic status of the plant (Zhang et al., 1995). PEPC kinase has been partially purified from soybean nodules (Zhang and Chollet, 1997). Nodule sucrose synthase has also been shown to be phosphorylated and the corresponding kinase has been purified (Zhang and Chollet, 1997).

Regulation of expression of late nodulin genes has been investigated (reviewed in de Bruijn and Schell, 1993). In particular, the regulation of transcription of the leghemoglobin genes has been studied in great detail (Stougaard et al., 1987; Szabados et al., 1990; de Bruijn and Schell, 1993; Szczyglowski et al., 1994). *Cis*-acting elements required for the general and nodule-infected-cell-specific expression have been identified and well characterized in the 5' flanking regions of leghemoglobin genes from different legumes (Stougaard et al., 1987; Szabados et al., 1990; de Bruijn and Schell, 1993; Szczyglowski et al., 1994;). The search for the transcripton factor(s) involved in the regulation of expression of leghemoglobin or any other late nodulin genes is under way.

3. Molecular biology and classical genetics as tools to understand nodule biology.

3.1. Plant symbiotic mutants and model legumes. Molecular and classical genetics have been widely used to gain basic understanding of the processes taking place during nodule development. A large number of plant mutants affected at different stages of nodule development have been identified in several species of legumes (Caetano-Annoles and Gresshoff, 1991, also see Szczyglowski et al., 1998; Schauser et al., 1999). However, as mentioned above, with the exception of the *nin* mutant of *L. japonicus* (see above), none of the genes underlying these mutant phenotypes have been cloned. The major problem to clone a gene in legumes by a chromosome walking is the huge size and high complexity of their genomes. The most extensively studied legumes such as soybean, pea and bean are relatively recalcitrant to *Agrobacterium*-mediated transformation, and regeneration from tissue culture. This, in turn, makes it extremely challenging, if not impossible, to use transgenic plant approaches to study gene function, complement mutant phenotypes and establish mutagenesis programs based on transposon or T-DNA gene-tagging approaches. To circumvent these shortcomings in the legume research in the early 90s, two legume species have been proposed as model organisms: *Lotus japonicus* and *Medicago truncatula* (Barker et al., 1990; Handberg and Stougaard, 1992; Cook et al., 1997; Jiang and Gresshoff, 1993). Both legumes have a number of features making them well suited for molecular and classical genetic research, and the most important ones are listed below:

1) Both species are diploid, self-fertile, amenable to manual cross-pollination and have small genomes: ~300-400 Mbp (only 3-4 times that of *Arabidopsis*), comparing to the estimated ~3000-4000 Mbp for pea;

2) Protocols for transformation with *Agrobacterium tumefaciens* and *rhizogenes* and regeneration in tissues culture exist for both legumes (Handberg et al., 1994; Stiller et al., 1997; Barker et al., 1990)

3) Plants are small and have short generation times (3 months seed to seed), making it possible to grow large numbers of them in a limited space for mutant screens and evaluate multiple generations in a relatively short time;

4) Polymorphic ecotypes exist for both species, making genetic mapping of the locus underlying a particular phenotype possible;

5) Genetic and physical maps and BAC libraries are being actively developed.

The ability to efficiently generate large numbers of transgenic *Lotus japonicus* lines has prompted the initiation of transposon and T-DNA mediated gene tagging efforts in this legume (Schauser et al., 1998). In total, 1112 transgenic lines have been generated, among which 16 symbiotic mutants have been found and 2 could be potentially tagged (Schauser et al., 1998). One mutant line, *nin* (see above), was subsequently shown to harbor a transposon insertion in a gene involved in nodulation (Schauser et al., 1999). The cloning of this gene does not only represent the first successful example of a gene tagging approach in legumes, a common and well established procedure in *Arabidopsis* and maize,

but also the first isolation of a gene responsible for a symbiotic phenotype. The isolation of *nin* locus clearly demonstrates the full potential of *Lotus japonicus* as a model legume.

In our laboratory, a screen for novel symbiotic mutants of *Lotus japonicus* has been conducted using ethanemethylsulfonate (EMS) as a mutagenic agent. As a result, 20 mutants, comprising at least 14 complementation groups, were found (Szczyglowski et al., 1998). The mutant phenotypes observed could be arranged in several classes: (1) total absence of any visible signs of nodules and normal root phenotype (Nod^-); (2) Nod^- with altered root morphology; (3) formation of small white structures resembling nodules (Nod^{-*}); (4) presence of white ineffective nodules (Nod^+Fix^-); (5) mixed phenotype with wild-type and ineffective nodules present on the same plant ($Nod^+Fix^{-/+}$); (6) supernodulation (Nod^{++}) (Szczyglowski et al., 1998). Presently, some of these mutants are being subjected to more detailed microscopical, physiological and molecular analyses. The most extensively characterized mutant line, *har-1*, forms an excessive number of nodules in response to rhizobia and displays an aberrant "bushy" root phenotype in the non-symbiotic state. (Szczyglowski et al., 1998; Wopereis et al., 2000).

3.2 Identification and characterization of late nodulins. Traditionally, the molecular biological approach has been the isolation and characterization of nodule-specific genes and cDNAs, with the underlying assumption that genes specifically induced in nodules may encode proteins (nodulins) important for symbiosis. A number of late nodulins have been identified in this fashion and have been arbitrarily divided in three groups, based on the level of understanding of their functions in symbiosis. The first

group comprises nodulins with relatively well understood biochemical and symbiotic functions. This group mostly represents the proteins involved in maintaining nodule "housekeeping" functions, such as various enzymes of nitrogen and carbon metabolism, mentioned above (GS, NADH-GOGAT, PEPC, SS, etc), leghemoglobins and nodulin-26. The second group comprises nodulins for which the biochemical functions are either known or a good prediction can be made, but for which exact roles in symbiosis are either obscure or not firmly established. This is a very diverse group, with such proteins as cytochrome P-450 (Szczyglowski et al., 1997), protein phosphatase 2C (Kapranov et al., 1999), peptide transporter LjNOD65 (Szczyglowski et al., 1997), small GTP-binding proteins (Borg et al., 1997), and others. The last group includes nodulins with unknown functions (Gamas et al., 1996; Szczyglowski et al., 1997; Kapranov et al. 1996; Szczyglowski et al., 1998; see also chapters 2-7 of this thesis).

Given the overall complexity of events taking place in mature, fully functioning nodules, it is strongly anticipated that additional genes are likely to be discovered. In general, the identification of novel late nodulin genes is crucial for a thorough molecular understanding of the details of nodule functioning. This is especially true for the most interesting and under-represented class of late nodulins with regulatory functions. Therefore, a systematic search aimed at the identification and characterization of novel late mRNA species associated with the late stages of nodule morphogenesis and functioning in a model legume *L. japonicus* has been initiated using the RNA differential display methodology (Liang and Pardee, 1992). Our major goal was to identify putative regulators of nodule development and functioning. Since mRNAs representing such

proteins are often low-abundant, the differential display is particularly suitable for such endeavor, given the high sensitivity of this PCR-based technology.

In total, we isolated 88 unique differential display products, at least 19 of which represented transcripts specifically associated with the late stages of nodule development as shown by northern blot analysis (Szczyglowski et al., 1997; Kapranov et al., 1997; Kapranov et al., 1999; Szczyglowski et al., 1998). The remaining ESTs did not yield a hybridization signal with root and nodule mRNA derived probes, suggesting that these ESTs may represent genes which are expresses at very low levels (Szczyglowski et al., 1997). A comprehensive description of the differential display screen and the characterization of the expressed sequence tags (ESTs) and cDNAs obtained using this approach also constitutes Chapter 2 of this thesis.

The detailed characterization of three nodulin genes identified in our late-nodulin EST screen, *LjNPP2C1*, *LjNOD70*, and *LjNOD16* constitutes the remaining Chapters of this thesis and has been published in Szczyglowski et al., (1997); Kapranov et al., (1997); Szczyglowski et al., (1998); Kapranov et al., (1999). Chapter 3 describes the molecular and biochemical analysis of a protein phosphatase 2C gene, *LjNPP2C1*, expression of which is induced at the late stages of nodule development. Chapter 4 summarizes the two main approaches undertaken to further investigate the function of the LjNPP2C1 enzyme: the generation of transgenic *L. japonicus* plants (over)expressing sense, antisense or dominant-negative form of *LjNPP2C1* mRNA and the screen for the proteins interacting with LjNPP2C1 using the yeast two-hybrid system. Chapter 5 describes the isolation and characterization of a late-nodulin cDNA, *LjNOD70*, encoding a putative transporter

protein localized to the PBM compartment. The weak but significant amino acid sequence similarity observed between LjNOD70 and the bacterial proteins capable of transporting organic acid, such as oxalate and fumarate, makes this late nodulin an attractive candidate for role of the yet unidentified DA transporter present in the PBM and mentioned above.

Chapter 6 describes the initial cloning and characterization *LjNOD16* cDNA encoding a late nodulin protein Nlj16. The following Chapter 7 shows that *LjNOD16* is a part of larger gene, *LjPLP-IV*, and that *LjNOD16* transcripts originate in *L. japonicus* nodule tissues as a result of unusual transcription events governed by an intron-localized promoter sequence in the *LjPLP-IV* gene. Furthermore, it is shown that the *LjPLP-IV* gene is a member of a gene family encoding a novel class of plant proteins similar to phosphatidylinositol transfer proteins (PITPs) and distinguished from previously described PITPs by the presence of a C-terminal extension comprised of either Nlj16 or other, highly related to Nlj16, amino-acid sequences. The roles of Nlj16 either as a domain within the PITP-like proteins or as a separately expressed protein are also analyzed and discussed in this Chapter.

REFERENCES

Appleby CA (1984) Leghemoglobin and *Rhizobium* interaction. Annu Rev Plant Physiol 35: 443-478

Atkins CA (1991) Ammonia assimilation and export of nitrogen from the legume nodule. In: Dilworth M, Glenn A (eds) Biology and Biochemistry of Nitrogen Fixation, pp293-319, Elsevier Amsterdam

Barker DG, Bianchi S, London F, Dattee Y, Duc G, Essad S, Flament P, Gallusci P, Genier G, Guy P, Muel X, Tourneur J, Denarie J, Huguet T (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. Plant Mol Biol 8: 40-49

Basset B, Goodman RN, Novacky A (1977) Ultrastructure of soybean nodules: Release of *rhizobia* from the infection thread. Can J Microbiol 23: 573-582

Bennett MJ, Lighftoot DA, Cullimore JV (1989) cDNA sequence and differential expression of the gene encoding the glutamine synthetase - polypeptide of *Phaseolus vulgaris* L. Plant Mol Biol 12: 553-565

Bergmann H, Preddie E, Verma DPS (1983) Nodulin-35: a subunit of specific uricase (uricase II) induced and localized in the uninfected cells of soybean nodules. EMBO J 2: 2333-2339

Boland MJ, Schubert KR (1983) Biosynthesis of purines by a proplastid fraction from soybean nodules. Arch Biochem Biophys 220: 179-187

Borg S, Brandstrup B, Jensen TJ, Poulsen C (1997) Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from *Lotus japonicus*, and expression of corresponding mRNAs in developing root nodules. Plant J 11: 237-250

Boron L, Szczyglowski K, Konieczny A, Legocki AB (1989) Glutamine synthetase in *Lupinus luteus*. Identification and preliminary characterization of nodule-specific cDNA clone. Acta Biochim Polon 36: 295-301

Boron L, Legocki AB (1993) Cloning and characterization of a nodule-enhanced glutamine synthetase-encoding gene from *Lupinus luteus*. Gene 136: 95-102

Brisson NA, Pombo-Gentile A, Verma DPS (1982) Organization and expression of leghemoglobin genes. Can J Bot 60: 272-278

Burgess BK, Lowe DJ (1996) Mechanism of molybdenum nitrogenase. Chem Rev 96: 2983-3011

Caetano-Annoles G, Gresshoff P (1991) Plant genetic control of nodulation. Annu Rev Microbiol 45: 345-382

Cheon C-I, Lee N-G, Siddique A-BM, Bal AK, Verma DPS (1993) Roles of plant homologs of Rablp and Rab7p in the biogenesis of the peribacteroid membrane, a subcellular compartment formed *de novo* during root nodule symbiosis. EMBO J 12: 4125-4135

Cook D, Dreyer D, Bonnet D, Howell M, Nony N, VandenBosch K (1995) Transient induction of a peroxidase gene in *Medicago truncatula* precedes infection by *Rhizobium meliloti*. Plant Cell 7: 43-55

Cook DR, VandenBosch K, de Bruijn F, Huguet T (1997) Model legumes get the nod. Plant Cell 3: 275-281

Cooper JB, Long SR (1994) Morphogenic rescue of *Rhizobium meliloti* nodulation mutants by *trans*-zeatin secretion. Plant Cell 6: 215-225

Csanadi G, Szecsi J, Kalo P. Kiss P, Endre G. Kondorosi A, Kondorosi E, and Kiss GB (1994) *ENOD12*, an early nodulin gene, is not required for nodule formation and efficient nitrogen fixation in alfalfa. Plant Cell 6: 201-213

Dean RM, Rivers RL, Zeidel ML, Roberts DM (1999) Purification and functional reconstitution of soybean nodulin 26. An aquaporin with water and glycerol transport properties. Biochemistry 38: 347-353

de Bruijn FJ, Schell J (1993) Regulation of plant genes specifically induced in developing and mature nitrogen-fixing nodules: *cis*-acting elements and *trans*-acting factors. In Verma DPS (ed), Control of plant gene expression. CRC Press, Boca Raton, pp. 241-258

Eady RR (1996) Structure-function relationships of alternative nitrogenases. Chem Rev 96: 3013-3030

Ehrhardt DW, Wais R, Long SR (1996) Calcium spiking in plant root hairs responding to *Rhizobium* nodulation signals. Cell 85: 673-681

Fang Y, Hirsch AM (1998) Studying early nodulin gene *ENOD40* expression and induction by nodulation factor and cytokinin in transgenic alfalfa. Plant Physiol 116: 53-68

Felle H, Kondorosi E, Kondorosi A, Schultze M (1995) Nod signal-induced plasma membrane potential changes in alfalfa root hairs are differentially sensitive to structural modifications of the lipo-chitooligosaccharide. Plant J 7: 939-947

Forde BG, Day HM, Turton JF, Shen W-J, Cullimore JV, Oliver JE (1989) Two glutamine synthetase genes from *Phaseolus vulgaris* l. display contrasting developmental and spatial patterns of expression in transgenic *Lotus corniculatus* plants. Plant Cell 1: 391-401

Fortin MG, Morison NA, Verma DPS (1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. Nucleic Acids Res 15: 813-824

Fortin MG, Zelechowska M, Verma DPS (1985) Specific targeting of membrane nodulins to the bacteroid-enclosing compartment in soybean nodules. EMBO J 4: 3041-3046

Freiberg C, Fellay C, Bairoch A, Broughton WJ, Rosenthal, Perret X (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. Nature 387: 394-401

Gamas P, Niebel FdC, Lescure N, Cullimore J (1996) Use of a substractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol Plant Microbe Interact 9: 233-242

Geurts R Franssen H (1996) Signal transduction in *Rhizobium*-induced nodule formation. Plant Physiol 112: 447-453

Gregerson RG, Miller SS, Twary SN, Gantt JS, Vance CP (1993) Molecular characterization of NADH-dependent glutamate synthase from alfalfa nodules. Plant Cell 5: 215-226

Gualtieri G, Bisseling T (2000) The evolution of nodulation. Plant Mol Biol 42: 181-194

Gyorgyey J, Vaubert D, Jimenez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E (2000) Analysis of *Medicago truncatula* nodule expressed sequence tags. Mol Plant-Microbe Interact 13: 62-71

Handberg K, Stougaard J (1992) *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. Plant J 2: 487-496

Handberg K, Stiller J, Thykjear T, Stougaard J (1994) Transgenic plants: *Agrobacterium* mediated transformation of the diploid legume *Lotus japonicus*. In: Cell Biology: A Laboratory Handbook, Celis JE, ed, 1: 119-127

Hanks JF, Schubert KR, Tolbert NE (1983) Isolation and characterization of infected and uninfected cells from soybean nodules. Role of unifected cells in ureide synthesis. Plant Physiol 71: 869-873

Hadri A-E, Spaink HP, Bisseling T, Brewin NJ (1998) Diversity of root nodulation and rhizobial infection processes. In Spaink HP, Kondorosi A, Hooykaas PJJ (eds), The *Rhizobiaceae*. Kluwer Academic Publishers, Dordrecht, pp.347-360.

Heard J, Dunn K (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc Natl Acad Sci USA 92: 5273-5277

Heard J, Caspi M, Dunn K (1997) Evolutionary diversity of symbiotically induced nodule MADS box genes: characterization of *nmhC5*, a member of a novel subfamily. Mol Plant-Microbe Interact 10: 665-676

Heidstra R, Geurts R, Franssen H, Spaink HP, van Kammen A, Bisseling T (1994) Root hair deformation activity of nodulation factors and their fate on *Vicia sativa*. Plant Physiol 105: 787-797

Heidstra R, Yang WC, Yalcin Y, Peck S, Emons A, van Kammen A, Bisseling T (1997) Ethylene provides positional information on cortical cell division but is not involved in Nod factor-induced root hair tip growth in *Rhizobium*-legume intercation. Development 124: 1781-1787

Hirsch AM (1992) Tansley Review No. 40. Developmental biology of legume nodulation, New Phytol 122: 211-237

Hirsch AM, Bhuvaneswari TV, Torrey JG, 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

Hirsch AM, Fang Y (1994) Plant hormones and nodulation: what's the connection? Plant Mol Biol 26: 5-9

Hong Z, Verma DPS (1994) A phosphatidylinositol 3-kinase is induced during soybean nodule organogenesis and is associated with membrane proliferation. Proc Natl Acad Sci USA 91:9617-9621

Howard JB, Rees DC (1996) Structural basis of biological nitrogen fixation. Chem Rev 96: 2965-2982

Jiang Q, Gresshoff PM (1993) *Lotus japonicus*: a model plant for structure-function analysis in nodulation and nitrogen fixation. Curr Topics Plant Mol Biol 2: 97-110

Jorgensen J-E, Gronlund M, Pallisgaard N, Larsen K, Marcker KA, Jensen EO (1999) A new class of plant homeobox genes is expressed in specific regions of determinate symbiotic root nodules. Plant Mol Biol 40: 65-77

Kahn ML, McDermott TR, Udvardi MK (1998) Carbon and nitrogen metabolism in *Rhizobia*. In Spaink HP, Kondorosi A, Hooykaas PJJ (eds), The *Rhizobiaceae*, Molecular Biology of Model Plant-Associated Bacteria. Kluwer Academic Publishers, Dordrecht, pp. 461-485

Kaiser BN, Finnegan PM, Tyerman SD, Whitehead LF, Bergersen FJ, Day DA, Udvardi MK (1998) Characterization of an ammonium transport protein from he peribacteroid membrane of soybean nodules. Science 281: 1202-1206

Kaminski PA, Batut J, Boistard P (1998) A survey of symbiotic nitrogen fixation by *Rhizobia*. In Spaink HP, Kondorosi A, Hooykaas PJJ (eds), The *Rhizobiaceae*, Molecular Biology of Model Plant-Associated Bacteria. Kluwer Academic Publishers, Dordrecht, pp. 431-460

Kapranov P, de Bruijn FJ, Szczyglowski K (1997) A novel, highly expressed late nodulin gene *LjNOD16* from *Lotus japonicus*. Plant Physiol 113: 1081-1090

Kapranov P, Jensen TJ, Poulsen C, de Bruijn FJ, Szczyglowski K (1999) A protein phosphatase 2C gene, *LjNPP2C1*, from *Lotus japonicus* induced during root nodule development. Proc Natl Acad Sci USA 96: 1738-1743

Kijne, J.W. (1992) The *Rhizobium* infection process. In Stacey G, Burris RH, Evans HJ, (eds), Biological Nitrogen Fixation. Chapman and Hall, New York, pp. 349-398

Lam H-M, Coschigano KT, Oliveira IC, Melo-Oliveira R, Coruzzi GM (1996) The molecular-genetics of nitrogen assimilation into amino acids in higher plants. Annu Rev Plant Physiol Plant Mol Biol 47:569-593

Lea PJ, Robinson SA, Stewart GR (1990) The enzymology and metabolism of glutamine, glutamate and asparagine. In Miflin BJ, Lea PJ (eds), The Biochemistry of Plants, Vol. 16, Intermediary Nitrogen Metabolism. Academic Press, San Diego, pp. 121-159

Lee JW, Zhang Y, Weaver CD, Shomer NH, Louis CF, Roberts DM (1995) Phosphorylation of nodulin 26 on serine 262 affects its voltage-sensitive channel activity in planar lipid bilayers. J Biol Chem 270: 27051-27057

Lee KH, Larue TA (1992) Exogenous ethylene inhibits nodulation of *Pisum sativum* L. cv. Sparkle. Plant Physiol 100: 1759-1763

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257: 967-971

Libbenga KR, van Iren F, Bogers RJ, Schraag-Lamers MF (1973) The role of hormones and gradients in the initiation of cortex proliferation and nodule formation in *Pisum sativum* L. Planta 114: 29-39

Long SR (1996) *Rhizobium* symbiosis: Nod factors in perspective. Plant Cell 8: 1885-1898

Maxwell CA, Vance CP, Heichel GH, Stade S (1984) CO_2 fixation in alfalfa and birdsfoot trefoil root nodules and partitioning of ^{14}C to the plant. Crop Sci 24: 257-264

Mellor RB (1989) Bacteroids in the *Rhizobium*-legume symbiosis inhabit a plant internal lytic compartment: implications for other microbial endosymbioses. J Exp Bot 40: 831-839

Mellor RB, Werner D (1987) Peribacteroid membrane biogenesis in mature legume root nodules. Symbiosis 3: 75-100

Miao G-H, Verma DPS (1993) Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterologous plants. Plant Cell 5: 781-794

Miller SS, Driscoll BT, Gregerson RG, Gantt JS, Vance CP (1998) Alfalfa malate dehydrogenase (MDH): molecular cloning and characterization of five different forms reveals a unique nodule-enhanced MDH. Plant J 15: 173-184

Mylona P, Pawlowski K, Bisseling T (1995) Symbiotic nitrogen fixation. Plant Cell 7: 869-885

Nap J-P, Bisseling T (1990) Developmental biology of a plant-prokaryote symbiosis: The legume root nodule. Science **250,** 948-954

Newcomb W, Sippel D, Peterson RL (1979) The early morphogenesis of *Glycine max* and *Pisum sativum* root nodules. Can J Bot 57: 2603-2616

Nguyen T, Zelechowska MG, Foster V, Bergmann H, Verma DPS (1985) Primary structure of the soybean nodulin-35 gene encoding a nodule-specific uricase localized in peroxisomes of uninfected cells of soybean. Proc Natl Acad Sci USA 82: 5040-5044

Niebel A, Gressent F, Bono JJ, Ranjeva R, Cullimore J (1999) Recent advances in the study of Nod factor perception and signal transduction. Biochimie 81: 669-674

Pathirana SM, Vance CP, Miller SS, Gantt JS (1992) Alfalfa root nodule phosphoenolpyruvate carboxylase: Characterization of the cDNA and expression in effective and ineffective nodules. Plant Mol Biol 20: 437-450

Perret X, Freiberg C, Rosenthal A, Broughton WJ, Fellay R (1999) High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234. Mol Microbiol 32: 415-425

Peters NK, Crist-Estes DK (1989) Nodule formation is stimulated by the ethylene inhibitor aminoethoxyvinylglycine. Plant Physiol 91: 690-693

Phillips DA (1980) Efficiency of symbiotic nitrogen fixation in legumes. Annu Rev Plant Physiol Plant Mol Biol 31: 29-49

Pingret J-L, Journet E-P, Barker DG (1998) *Rhizobium* Nod factor signaling: evidence for a G protein-mediated transduction mechanism. Plant Cell 10: 659-671

Pueppke SG (1996) The genetic and biochemical basis for nodulation of legumes by *Rhizobia*. Crit Rev Biotechnology 16: 1-51

Reibach PH, Streeter JG (1983) Metabolism of ^{14}C-labeled photosynthate and distribution of enzymes of glucose metabolism in soybean nodules. Plant Physiol 72: 634-640

Ribbe M, Gadkari D, Meyer O (1997) N_2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N_2 reduction to the oxidation of superoxide produced from O_2 by a molybdenum-CO dehydrogenase. J. Biol. Chem. 272: 26627-26633

Robinson DL, Kahn ML, Vance CP (1994) Cellular localization of nodule enhanced aspartate amonitransferase in *Medicago sativa* L. Planta 92:202-210

Robinson DL, Pathirana SM, Gantt JS, Vance CP (1996) Immunogold localization of nodule-enhnanced phosphoenolpyruvate carboxylase in alfalfa. Plant Cell Environ 19: 602-608

Ronson CW, Lyttleton P, Robertson, JG (1981) C_4-dicarboxylate transport mutants of *Rhizobium trifolii* from ineffective nodules on *Trifolium repens*. Proc Natl Acad Sci USA 78: 4284-4288

Salminen SO, Streeter JG (1987) Involvement of glutamate in the respiratory metabolism of *Bradyrhizobium japonicum* bacteroids. J Bacteriol 169: 495-499

Schauser L, Handberg K, Sandal N, Stiller J, Thykjaer T, Pajuelo E, Nielsen A, Stougaard J (1998) Symbiotic mutants deficient in nodule establishment identified after T-DNA transformation of *Lotus japonicus*. Mol Gen Genet 259: 414-423

Schauser L, Roussis A, Stiller J, Stougaard J (1999) A plant regulator controlling development of symbiotic root nodules. Nature 402: 191-195

Scheres B, van Engelen F, van den Knaa, E, van de Wiel C, van Kammen A, and Bisseling T (1990a) Sequential induction of nodulin gene expression in the developing pea nodule. Plant Cell 2: 687-700

Scheres B, Van de Wiel C., Zalensky A, Horvath B, Spaink H, Van Eck H, Zwartkruis F, Wolters AM, Gloudemans T, Van Kammen A, and Bisseling T (1990b) The *ENOD12* gene product is involved in the infection process during the pea-*Rhizobium* interaction. Cell 60: 281-294

Schindelin H, Kisker C, Schlessman JL, Howard JB, Rees DC (1997) Structure of ADP-AlF_4^--stabilized nitrogenase complex and its implications for signal transduction. Nature 387: 370-376

Schmidt JS, Harper JE, Hoffman TK, Bent AF (1999) Regulation of soybean nodulation independent of ethylene signaling. Plant Physiol 119: 951-960

Schnorr KM, Laloue M, Hirel B (1996) Isolation of cDNAs encoding two purine biosynthetic enzymes of soybean and expression of the corresponding transcripts in roots and root nodules. Plant Mol Biol 32: 751-757

Schubert KR (1986) Products of biological nitrogen fixation in higher plants: Synthesis, transport, and metabolism. Annu Rev Plant Physiol 37: 539-574

Schultze M, Kondorosi E (1998) Regulation of symbiotic root nodule development. Annu. Rev. Genet. 1998: 33-57

Schultze M, Kondorosi E, Ratet P, Buire M, Kondorosi A (1994) Cell and molecular biology of *Rhizobium*-plant interactions. Int Rev Cytol 156: 1-75

Shi L, Twary SN, Yoshioka H, Gregerson RG, Miller SS, Samac DA, Gantt JS, Unkefer PJ, Vance CP (1997) Nitrogen assimilation in alfalfa: isolation and characterization of an asparagine synthetase gene showing enhanced expression in root nodules and dark-adapted leaves. Plant Cell 9: 1339-1356

Spaink HP (1992) Rhizobial lipooligosaccharides: answers and questions. Plant Mol Biol 20: 977-986

Spaink HP, Lugtenberg BJJ (1994) Role of rhizobial lipo-chitin oligosaccharide signal molecules in root nodule organogenesis. Plant Mol Biol 26: 1413-1422

Spaink HP, Sheeley DM, van Brussel AA, Glushka J, York WS, Tak T, Geiger O, Kennedy EP, Reinhold VN, Lugtenberg BJ (1991) A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. Nature 354: 125-130

Sprent JI (1989) Which steps are essential for the formation of functional legume nodules? New Phytol. 111: 129-153

Stanford AC, Larsen K, Barker DG, Cullimore JV (1993) Differential expression within the glutamine synthetase gene family of the model legume *Medicago truncatula*. Plant Physiol 103: 73-81

Stiller J, Martirani L, Tuppale S, Chian R-J, Chiurazzi M, Gresshoff PM (1997) High frequency transformation and regeneration of transgenic plants in the model legume *Lotus japonicus*. J Exp Bot 48: 1357-1365

Stougaard J, Sandal NN, Gron A, Kuhle A, Marcker KA (1987) 5' Analysis of the soybean leghaemoglobin lc3 gene: regulatory elemnts required for promoter activity and organ specificity. EMBO J 6: 3565-3569

Szabados L, Ratet P, Grunenberg B, de Bruijn FJ (1990) Functional anlysis of the *Sesbania rostrata* leghemoglobin *glb3* gene 5'-upstream region in transgenic *Lotus corniculatus* and *Nicotiana tabacum* plants. Plant Cell 2: 973-986

Szczyglowski K, Szabados L, Fujimoto S, Silver D, de Bruijn FJ (1994) Site-specific mutagenesis of the nodule-infected cell expression (NICE) element and the AT-rich element ATRE-BS2* of the *Sesbania rostrata* leghemoglobin *glb3* promoter. Plant Cell 6: 317-332

Szczyglowski K, Hamburger D, Kapranov P, de Bruijn FJ (1997) Construction of a *Lotus japonicus* late nodulin EST library and identification of novel nodule-specific genes. Plant Physiol. 114: 1335-1346.

Szczyglowski K, Shaw SR, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB, de Bruijn FJ (1998a) Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. Mol Plant-Microbe Interact 11: 684-697

Szczyglowski K, Kapranov P, Hamburger D, de Bruijn FJ (1998b) The *Lotus japonicus LjNOD70* gene encodes a protein with similarities to transporters. Plant Mol Biol 37: 651-661

Temple SJ, Heard J, Ganter G, Dunn K, Sengupta-Gopalan C (1995) Characterization of a nodule enhanced glutamine synthetase from alfalfa: nucleotide sequence, *in situ* localization, and transcript analysis. Mol Plant-Microbe Interact 8:218-227

Thummler, F., and Verma, D.P.S. (1987) Nodulin-100 of soybean is the subunit of sucrose synthetase regulated by the availability of free heme in nodules. J Biol Chem 262: 14730-14736

Ting IP (1985) Crassulacean acid metabolism. Annu Rev Plant Physiol 36: 595-622

Tjepkema JD, Yokum SC (1974) Measurement of oxygen partial pressure within soybean nodules by oxygen microelectrodes. Planta 119: 351-360

Udvardi MK, Day DA (1997) Metabolite transport across symbiotic membranes of legume nodules. Annu Rev Plant Physiol Plant Mol Biol 48: 493-523

Udvardi MK, Price GD, Gresshoff PM, Day DA (1988) A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. FEBS Lett 231: 36-40

Udvardi MK, Ou Yang L-J, Young S, Day DA (1990) Sugar and amino acid transport across symbiotic membranes from soybean nodules. Mol Plant-Microbe Interact 3: 334-340

van Kammen A (1984) Suggested nomenclature for plant genes involved in nodulation and symbiosis. Plant Mol Biol Rep 2: 43-45

van Spronsen PC, Bakhuizen R, van Brussel AAN, Kijne JW (1994) Cell wall degradation during infection thread formation by the root nodule bacterium *Rhizobium leguminosarum* is a two-step process. Eur J Cell Biol 64: 88-94

van Brussel AAN, Bakhuizen R, van Spronsen PC, Spaink HP, Tak T, Lugtenberg BJJ, Kijne JW (1992) Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of *Rhizobium*. Science 257: 70-72

Vance CP (1990) Symbiotic nitrogen fixation: recent genetic advances. In Miflin BJ, Lea PJ (eds), The Biochemistry of Plants, Vol. 16, Intermediary Nitrogen Metabolism. Academic Press, San Diego, pp. 43-88

Vance CP (1998) Legume symbiotic nitrogen fixation: agronomic aspects. In Spaink HP, Kondorosi A, Hooykaas PJJ (eds), The *Rhizobiaceae*, Molecular Biology of Model Plant-Associated Bacteria. Kluwer Academic Publishers, Dordrecht, pp. 509-530.

Vance CP, Miller SS, Gregerson RG, Samac DA, Robinson DL, Gantt JS (1995) Alfalfa NADH-dependent glutamate synthase: structure of the gene and importance in symbiotic N_2 fixation. Plant J 8: 345-358

Vance CP, Gantt JS (1992) Control of nitrogen and carbon metabolism in root nodules, Physiol Plant 85: 266-274

Vance CP, Gregerson RG, Robinson DL, Miller SS, Gantt JS (1994) Primary assimilation of nitrogen in alfalfa nodules: molecular features of the enzymes involved. Plant Sci 101: 51-64

Verma DPS, Fortin MG (1989) Nodule development and formation of the endosymbiotic compartment. In Schell J, Vasil IK, eds, Cell Culture and Somatic Cell Genetics of Plants, Vol. 6, Molecular Biology of Plant Nuclear Genes. Academic Press, San Diego, pp. 329-353

Verma DPS, Hong Z (1996) Biogenesis of the peribacteroid membrane in root nodules. Trends Microbiology 4: 364-368

Waters JK, Hughes BL, Purcell LC, Gerhardt KO, Mawhinney TP, Emerich DW (1998) Alanine, not ammonia, is excreted from N_2-fixing soybean nodule bacteroids. Proc Natl Acad Sci USA 95: 12038-12042

Weaver CD, Roberts DM (1992) Determination of the site of phosphorylation of nodulin 26 by the calcium-dependent protein kinase from soybean nodules. Biochemistry 31: 8954-8959

Weaver CD, Shomer NH, Louis CF, Roberts DM (1994) Nodulin 26, a nodule-specific symbiosome membrane protein from soybean, is an ion channel. J Biol Chem 269: 17858-17862

Werner D, Morschel E, Stripf R, Winchenbach B (1980) Development of nodules of *Glycine max* infected with an ineffective strain of *Rhizobium japonicum*.

Zalkin H, Dixon JE (1992) De novo purine biosynthesis. Prog Nucl Acids Res Mol Biol 42: 259-287

Zhang X-Q, Chollet R (1997) Seryl-phosphorylation of soybean nodule sucrose synthase (nodulin-100) by a Ca^{2+}-dependent protein kinase. FEBS Lett 410: 126-130

Zhang X-Q, Chollet R (1997) Phosphoenolpyruvate carboxylase protein kinase from soybean root nodules: partial purification, characterization, and up/down-regulation by photosynthate supply from the shoots. Arch Biochem Biophys 343: 260-268

CHAPTER 2

Construction of a *Lotus japonicus* late nodulin expressed sequence tag library and identification of novel nodule-specific genes[1]

2.1. ABSTRACT

A range of novel expressed sequence tags (ESTs) associated with late developmental events during nodule organogenesis in the legume *Lotus japonicus* were identified using mRNA differential display. One hundred ten differentially displayed PCR products were cloned and analyzed. Of 88 unique cDNAs obtained, 25 shared significant homology to DNA/protein sequences in the respective databases. This group comprises, among others, a nodule-specific homologue of protein phosphatase 2C, a peptide transporter protein, and a nodule-specific form of cytochrome P450. RNA gel blot analysis of 18 differentially displayed ESTs confirmed their nodule specific expression pattern. The kinetics of mRNA accumulation of the majority of the ESTs analyzed were found to resemble the expression pattern observed for the *L. japonicus* leghemoglobin gene. These results indicate that the newly isolated molecular markers indeed correspond to genes induced during late developmental stages of *L. japonicus* nodule organogenesis, and provide important, novel tools for the study of nodulation.

---

[1] This chapter was published in: Szczyglowski K, Hamburger D, Kapranov P, de Bruijn F (1997) Plant Physiology 114: 1335-1346.

2.2. INTRODUCTION

The formation of nitrogen-fixing nodules represents an unusual example of externally induced organogenesis that unites the plant host and the symbiotic bacteria in a microenvironment appropriate for the support of bacterial nitrogen fixation and plant-mediated assimilation of nitrogen. This complex and highly regulated process begins with the specific recognition between the plant and bacterial partners, which leads to the synthesis of morphogenic lipochito-oligosaccharide molecules of bacterial origin (Nod factors; for a recent review see, Carlson et al., 1995; Spaink, 1996). The Nod factor signal molecules appear to exert their role in the plant root by inducing root hair curling and deformation, as well as redirecting the fate of root cortical cells toward the initiation of a nodule primordium (Spaink, 1996, and references therein). The initial stages of nodule organogenesis also include infection thread formation and penetration of root cells by symbiotic bacteria (for a recent review, see Schultze et al., 1994). Specific plant genes, called early nodulin genes, are activated during early nodule morphogenetic events (Schulze et al., 1994; Mylona et al., 1995). Although the exact functions of the majority of the early nodulin genes identified to date have yet to be determined, the expression patterns of some of these genes have been correlated with processes such as pre-infection (*Mtrip1*; Cook et al., 1995), infection (ENOD5, ENOD12, Scheres et al., 1990a, 1990b), and nodule meristem initiation or nodule structure formation (*cdc2-S5*, ENOD40, ENOD2; Franssen et al., 1987; Miao et al., 1993; Crespi et al., 1994; Matvienko et al., 1994; van de Sande et al., 1996;).

The successful invasion of plant root cells by symbiotic bacteria leads to the final stage of nodule formation, which culminates in the establishment of a fully developed, nitrogen-fixing nodule. However, the developmental cues and molecular events underlying the late steps in nodule organogenesis are largely unknown. Several events that occur in both symbiotic partners at this late stage have been correlated with bacterial release from the infection thread, and plant cell colonization (Sprent, 1989). These events include processes such as proliferation of the membrane system (PBMs; symbiosome membrane), bacteroid differentiation, and molecular and biochemical alterations that create and support the physiological environment required for nitrogen fixation and ammonia assimilation (Verma, 1992; Mylona et al., 1995).

Immediately prior to, or concomitantly with, the initiation of nitrogen fixation, a group of specific plant genes, called late nodulin genes, is activated (Verma, 1992; de Bruijn and Schell, 1992). Since these genes are not expressed in nodules lacking infected cells, it has been postulated that late nodulin genes may be coordinately expressed as a result of a single signal related to the release of bacteria from the infection thread (Nap and Bisseling, 1990; Welters et al., 1993). Typical members of this group include abundantly expressed genes encoding enzymes, or subunits of enzymes, that function in nitrogen or carbon metabolism; proteins associated with the peribacteroid membrane; a family of oxyhemoproteins (leghemoglobins); and a number of proteins the function of which remains to be elucidated (Delauney and Verma, 1988; de Bruijn and Schell, 1992; Mylona et al., 1995). A role for small GTP-binding proteins (Rab1p and Rab7p homologues) in PBM development has also been proposed (Cheon et al., 1993), and a putative plant

transcription factor (NMH7), present specifically in the infected bacteroid-containing cells, has been identified in alfalfa root nodules (Heard and Dunn, 1995). It has been suggested that NMH7 may be involved in cellular activities specific to the differentiation of the infected cells (Heard and Dunn, 1995).

To study the complexity of genes expressed during the transition period between the development of the nodule structure and formation of a functional nodule, we examined the mRNA expression profiles in determinate nodules of the diploid legume *Lotus japonicus* (Handberg and Stougaard, 1992; Jiang and Gresshoff, 1997), using the differential display method (Liang and Pardee, 1992). Here we describe the isolation of numerous ESTs corresponding to novel late-nodulin genes, many of which appear to encode functions that have not previously been implicated in the nodulation process.

2.3. MATERIALS AND METHODS

Plant Material

Lotus japonicus GIFU B-129-S9 seeds (kindly provided by Dr. Jens Stougaard, Aarhus University, Denmark) were surface sterilized and germinated as described by Handberg et al. (1994). *Rhizobium loti* strain NZP2235 (Jarvis et al., 1982) was grown for two days at 28°C in TY medium and used to inoculate seven days old *L. japonicus* seedlings immediately before potting. For the initial experiments, a 3:1 mixture of vermiculite and sand was used as plant growth medium. Later we refined the *L. japonicus* growing conditions, resulting in faster and more efficient levels of nodulation (data not shown). Therefore, for the developmental Northern slot blot experiments shown in Fig. 2.7, a 3:3:1 mixture of grade 2 vermiculite, grade 3 vermiculite, and sand was used. All plants were grown in a controlled environment in growth chambers (18h/6h day/night cycle, 250 μE $sec^{-1}m^{-2}$ light intensity, 22°C/18°C day/night temperature, and 70% humidity). B&D solution (Broughton and Dilworth, 1971) supplemented with 0.5 mM KNO_3 was used to water the plants. The relatively low concentration of combined nitrogen in B&D solution supports growth of the uninoculated control plants, but does not affect nodule formation on roots of infected *L. japonicus* plants (data not shown). Root segments were harvested at various time points after infection, and fully developed

nodules were collected at the 21-day time point. Plant tissues were immediately frozen in liquid nitrogen and stored at -70°C until use.

Nucleic Acid Isolation

R. loti genomic DNA was isolated following the procedure of Marmur (1961), except that the bacterial cells were washed with 1 M NaCl prior to Pronase E digestion. For the isolation of plant genomic DNA the method described by Rogers and Bendich (1988) was used.

Total plant RNA was isolated using the procedure of Verwoerd et al. (1989), except that the extraction buffer used was as described by Hall et al. (1978). The poly $(A)^+$ fraction of mRNA was isolated using a mini-spin column kit (5 Prime-3 Prime, Inc.; Boulder, CO), following the manufacturer's instructions.

Developmental Differential Display of mRNA

The RNA differential display procedure was carried out using RNAmap™ kits from the GenHunter Corporation (Brookline, MA), following the manufacturer's instructions (see also Goormachtig et al., 1995). The cDNA synthesis was performed using 0.5 μg of total RNA isolated from root segments (7, 11, and 13 days after infection), and nodules harvested 21 days after infection. For control experiments, total

RNA isolated from 7- and 21-day-old uninfected roots was used. Selected bands were reamplified and then blunt ended by treatment with the Klenow fragment of DNA polymerase I (Boehringer Mannheim, Indianapolis, IN), phosphorylated at the 5' ends using T4 polynucleotide kinase, and cloned into the *Sma*I digested plasmid pK18 (Pridmore, 1987). Recombinant plasmids were mobilized into the *Escherichia coli* strain InvαF' (Novagen, Madison, WI), using electrotransformation. For each differentially displayed PCR-fragment, 48 recombinant colonies were collected and stored individually in microtiter plates at -70°C, in 30% glycerol.

Southern and Northern Analyses

For Southern blot analyses, 10 μg of plant genomic DNA, or 5 μg of *R. loti* total DNA, were digested to completion using *Eco*RI endonuclease. The digested DNA was separated on a 0.8% agarose gel, transferred onto Hybond-N nitrocellulose filters (Amersham, Arlington Heights, IL), and UV cross-linked, following standard procedures (Sambrook et al., 1989). Hybridizations and subsequent washes were carried out at 65°C, using high stringency conditions (Sambrook et al., 1989).

For northern analyses, 10 μg of total RNA, or 4 μg of poly $(A)^+$ RNA, were separated on a 1.2% denaturing agarose gel (6% formaldehyde, 1xMOPS buffer: 20 mM MOPS, 1.0 mM EDTA, 5.0 mM sodium acetate, pH 7.0) and transferred onto Nitro-Plus membranes (Fisher Scientific, Pittsburgh, PA), as described (Sambrook et al., 1989). Ten

micrograms of total RNA were used for the slot blot RNA hybridization. Prehybridizations and hybridizations were performed according to the procedure described by Church and Gilbert (1984). Filters were washed twice for 15 min in 2X SSC, 0.1% SDS; once for 15 min in 0.3X SSC, 0.1% SDS; and once for 10 min in 0.1x SSC, at 65°C. DNA probes were labeled with α^{32}P-dATP, using the random prime kit from Boehringer-Mannheim (Indianapolis, IN), following the manufacturer's instructions.

Differential Colony Hybridization

For differential colony hybridization, microtiter "combo" plates were developed. A single 96-well microtiter plate combined representatives of eight (A to H) differentially displayed PCR products, each represented by 12 recombinant colonies (8x12). The bacteria were transferred from the "combo" plates onto nitrocellulose filter disks (Schleicher and Schuell, Keene, NH) and grown over night at 28°C. The colonies on the filters were lysed and the liberated DNA was fixed to the filters (Sambrook et al., 1989). Two replica filters were prepared for each microtiter "combo" plate, and hybridized with radiolabeled cDNA probes derived from poly(A)$^+$ mRNA fractions from uninfected control roots, or nodules (see below).

Radiolabeled single-stranded cDNA probes were synthesized using 1 μg of poly (A)$^+$ RNA isolated from 21-day-old uninfected *L. japonicus* roots, or root nodules, respectively. The reaction mixtures contained 50 mM Tris HCl pH 8.5, 8 mM $MgCl_2$, 30

mM KCl, 1 mM DTT, 50 μM of each, dCTP, dGTP and dTTP, 100 μCi α^{32}P-dATP, 10 μM of each of the T_{12}MN primers (RNAmap™ kits; GenHunter Corporation, Brooklyn, MA), 20 units RNasin, and 50 units of AMV reverse transcriptase. After 20 min at 37°C, 2 μl of a mixture containing 10 mM dATP, dCTP, dGTP, and dTTP was added. The reaction was continued for 40 min and terminated by the addition of 1 μl of 0.5 M EDTA (pH 8.0). The RNA was hydrolyzed by the addition of 12 μl of 150 mM NaOH, incubated at 65°C for 1 h, and neutralized with 12 μl 1 M Tris HCl (pH 8.0) and 12 μl 1 N HCl. The radiolabeled, single-stranded cDNAs were separated from unincorporated dNTPs on Sephadex G50 columns and used directly for differential hybridization.

cDNA Library Screening

The cDNA library from mature nodules of *L. japonicus* was kindly provided by Dr. Jens Stougaard, Aarhus, Denmark. The library was constructed with oligo-dT primer in the lambda UniZAP vector from Stratagene (La Jolla, CA). Screening for full-copy cDNAs was performed following standard procedures (Sambrook et al., 1989; Stratagene manual).

DNA Sequencing and Computer Analysis

Manual DNA sequencing was performed using the Sequenase 2.0 kit (USB, Inc., Cleveland OH), following the manufacturer's instructions. Computer analyses of DNA and protein sequences were carried out using the SeqEd (Applied Biosystems, Inc., Foster City, CA), BLAST (Altschul et al., 1990), and GCG (Genetics Computer Group, Madison, WI) software packages. For automated fluorescent sequencing, Taq polymerase-mediated cycle-sequencing reactions were performed, according to Newman et al. (1994). The plasmid templates were prepared by growing bacterial cultures in MR2001 broth (MacConnell Research, San Diego, CA) and extracting of DNA using the Wizard Miniprep Kits (Promega, Madison, WI). One strand of the DNA templates was sequenced (by the MSU Sequencing Facility; MSU-DOE-Plant Research Laboratory, East Lansing, MI).

2.4. RESULTS

Molecular Characterization of *L. japonicus* Root Nodule Development

Upon infection with *R. loti*, *L. japonicus* plants form determinate nodules (Handberg and Stougaard, 1992; Jiang and Gresshoff, 1993). However, a considerable amount of variation in nodulation efficiency has been observed with different *R. loti* strains (data not shown). Therefore, to develop an efficient system, we examined the nodulation characteristics of twenty-one different *R. loti* strains (provided by Dr. D.B. Scott, Massey University, New Zealand). Out of this collection, a highly efficient *R. loti* strain, NZP2235, was selected and used for further nodulation of *L. japonicus* ecotype Gifu plants (K. Szczyglowski and F.J. de Bruijn, unpublished data). Under the plant growth condition used in this study, visible signs of nodule formation could be detected on the roots of *L. japonicus* plants between 3 and 5 d after inoculation. Over the next few days, usually between 7 and 11 days after inoculation, nodule structures enlarged and turned pink, indicative of leghemoglobin synthesis. To further define the early and late stages of *L. japonicus* nodule development, RNA gel blot analyses were performed with two well characterized "marker" genes, namely the *L. japonicus Enod2* gene (R. Chen, K. Szczyglowski, F.J. de Bruijn, unpublished data), and a cDNA corresponding to a *L. japonicus* leghemoglobin gene (J. Stoltzfus, K. Szczyglowski, F.J. de Bruijn, unpublished data). In a number of legume species, the expression of the early nodulin gene *Enod2* and

the leghemoglobin genes has been correlated with early and late stages in nodule development, respectively (Nap and Bisseling, 1990; Mylona et al., 1995).

Developmental northern blot analysis using total RNA isolated from roots and nodules at different time points after *L. japonicus* inoculation revealed that the *Enod2* gene was expressed at a very low level in the uninfected control roots (Fig. 2.1). The level of *Enod2* mRNA gradually increased between 7 and 21 d after inoculation. In comparison, the expression of the *L. japonicus* leghemoglobin gene was first detected at 11 d after inoculation. Based on these observations we concluded that the late developmental events in the *L. japonicus* nodule morphogenesis process are likely to occur between 7 and 11 d after inoculation. Therefore, this time period became the focus of our further molecular analysis.

Construction of a *L. japonicus* Nodule-Specific EST Library

The strategy used to construct the *L. japonicus* nodule-specific expression sequence tag (EST) library was based on the mRNA differential display fingerprinting (Liang and Pardee, 1992). We applied this procedure to detect and clone transcripts that are specifically expressed during the transition period between the formation of the nodule structure and the onset of nitrogen fixation. To minimize false positives (Liang et al., 1993), and maximize the probability of isolating nodule-specific ESTs, we compared the RNA profiles derived from four relatively late phases of nodule development. Using the four degenerate $T_{12}MN$ primers, in combination with 20 arbitrary decamers, RNA

species from the *L. japonicus* roots harvested at 7, 11, and 13 d after infection and from 21-day-old nodules were displayed and compared to the RNA profiles derived from 7 and 21 d old uninfected control roots. Each primer set was found to generate 80 to 150 bands. Since a total of eighty primer combinations were used, approximately 10,000 PCR products were displayed for every time point analyzed. A representative example of these experiments is shown in Fig. 2.2. A visual inspection of all RNA profiles obtained revealed that approximately 1.4% of the bands (137 out of 10,000) appeared to be present in infected *L. japonicus* roots and/or nodules, but not in control, uninfected roots (see Fig. 2.2). Out of 137 bands recovered from the polyacrylamide gels, 110 bands were successfully reamplified and cloned into the pK18 vector (Pridmore, 1987). Given the likelihood that the reamplified PCR products represented a mixture of different mRNA species (Bauer et al., 1993), 48 recombinant colonies were stored per individual PCR product in a single microtiter plate. In addition, 8 PCR products, each represented by 12 recombinant colonies were combined in the 96-well microtiter "combo" plate and used for differential hybridization.

Differential Hybridization and Identification of Nodule-Specific ESTs

Replica filters from each of the 14 microtiter "combo" plates, representing a total of 110 PCR bands, were differentially hybridized with radiolabeled cDNAs derived from nodule or root poly(A)$^+$ RNA. This screening procedure permitted the selection of a group of recombinant colonies, representing 39 differentially displayed PCR products,

which hybridized specifically with the nodule-derived cDNA probe, but not with the probe from uninfected root. Twenty six colonies, representing 7 different PCR bands, hybridized with probes from both uninfected roots and nodules, whereas 11 colonies corresponding to 6 PCR products hybridized only to root cDNAs. The recombinant colonies associated with the remaining 58 PCR bands did not show detectable signals with either probe. An example of this differential hybridization analysis is shown in Fig. 2.3. In this example, several different outcomes of the differential hybridization procedure are shown. The top two rows of nodule-specific strong hybridization signals in Fig. 2.3 (Panel B, upper filter), representing a group of *LjN50* cDNAs, illustrate an example of most (if not all) individually picked transformants containing the same cDNA, that most likely corresponds to a highly expressed nodulin genes. The next two rows and the last two rows on the upper filter illustrate cases where no signal was found with either nodule- or root-specific RNA. This suggests that the cDNAs contained in these colonies either represent false positives or correspond to rare mRNA species. A more sensitive approach will be used to establish their tissue specificity (see also Discussion). The third set of rows on the upper filter illustrate a case of having isolated cDNAs that do not appear to be nodule specific. The examples shown on the lower filter illustrate a case of having only a limited number of positive cDNAs among the transformants analyzed (*LjN77,LjN50*). These results directly show the need to characterize several independent colonies for each putative cDNA. Based on the differential hybridization analysis, we selected 39 nodule-specific *L. japonicus* PCR products (ESTs).

Sequence Analysis of Nodule-Specific ESTs

The entire cDNA inserts of representative members of the 39 nodule-specific EST groups were sequenced. A comparative DNA sequence analysis revealed that the 39 sequences represented 18 unique cDNA species. We refer to these as *L. japonicus* nodulin (*LjN*) cDNAs (Table 2.1). All *LjN* cDNAs were found to contain DNA sequences of the specific primer pairs used during the differential display procedure (data not shown). The majority of the *LjN* cDNAs were represented by single isolates, whereas in a few cases (e.g. *LjN50, LjN132*) the same mRNA species was amplified more then once, using different primer combinations (Table 2.1; see also Discussion). Homology searches using the BLAST algorithm (Altschul et al., 1990) indicated that 6 of the *L. japonicus LjN* cDNAs shared significant homologies to previously described nodulin genes. The *LjN36*, *LjN77*, and *LjN132* cDNAs showed homology to different leghemoglobin genes and appeared to represent three different *L. japonicus* leghemoglobin loci (Table 2.1, Fig. 2.4A). In fact, both the DNA and predicted amino-acid sequence of *LjN77* were identical to the *L. japonicus* leghemoglobin cDNA sequence isolated from a nodule-specific cDNA library and used for the initial northern blot analysis (data not shown; see Fig. 2.1).

The DNA sequence of *LjN13* was similar to *Enod40* genes of various legume species, and *LjN101* shared significant similarity with the *Glycine max* coproporphyrinogen oxidase gene (Yang et al., 1993; Madsen et al., 1993). The DNA or protein alignments for *LjN13* and *LjN101* are shown in Figs. 2.4B and 2.4C, respectively. In addition, 7 different *LjN* clones revealed significant homology to other DNA/protein

sequences stored in the data bases. The translated sequence of *LjN3* showed strong similarity to the *Arabidopsis thaliana* protein phosphatase 2C (Kuromoni and Yamamoto, 1994; Fig. 2.4D; Table 2.1), whereas the deduced *LjN73* protein sequence was similar to a portion of the *Zea mays* cytochrome P-450 protein (Frey et al., 1995; Table 2.1). In the latter case, the 1660bp cDNA fragment was isolated from the *L. japonicus* nodule-specific cDNA library and shown to encode a polypeptide (LjNP450) with significant similarity to different cytochrome P-450 proteins. The heme binding domain, also known as a "signature sequence" characteristic for cytochrome P450 proteins, was detected at the C-terminal end of the LjNP450 protein. (Nelson et al., 1993; see Fig. 2.5). An example of the alignment between LjNP450 and the ripening-related cytochrome P-450 CYP71A1 is shown in Fig. 2.5A (Bozak et al., 1990).

The partial amino-acid sequences deduced from *LjN53* and *LjN63* were short (60 aa and 70 aa respectively), and did not show statistically significant matches in database searches. However, the deduced *LjN53* product was found to share limited similarity with peptide/amino acid transport proteins from plants. The deduced partial amino acid sequence of the *LjN63* showed a high content (34%) of glutamic acid residues (Table 2.1). Full-copy cDNAs corresponding to *LjN53* and *LjN63* were isolated and predicted to encode a 65 kD protein (Nlj65) and a glutamate-rich (27%) protein of 192 amino acids (Nlj21), respectively. The deduced amino-acid sequence of Nlj65 showed significant similarity with the peptide transporter *At*PTR2-B from *Arabidopsis* and other species (Frommer et al., 1994; Fig. 2.5B). The deduced amino-acid sequence of nodulin Nlj21 shared substantial similarity with the Ag13 protein from *Alnus glutinosa* nodules (Guan

et al., 1997), and the pKIWI501 gene specifically expressed during fruit development of kiwifruit (Ledger and Gardner, 1994; Fig. 2.6). The expression of the *ag13* gene has been localized to the pericycle of the vascular bundle of *A. glutinosa* nodules, and in infected cells that exhibit degradation of the endosymbiont (Guan et al., 1997).

The full-copy cDNA corresponding to the *L. japonicus LjN5* cDNA was isolated and shown to encode a 15.6-kD protein (data not shown). The expression of the corresponding gene was induced in infected cells of *L. japonicus* around the time of initiation of nitrogen fixation, and the deduced protein was found to share significant similarity with predicted α-helical domains of two related anonymous *Arabidopsis* ESTs (Kapranov et al., 1997).

The remaining 7 nodule-specific *LjN* cDNAs did not show any significant similarities to DNA/ protein sequences stored in the databases. To gain more insights into the functions of the genes represented by these ESTs, larger cDNA clones were isolated for each of the 7 *LjN* clones (Table 2.1). Out of those, six cDNAs showed significant similarities to the sequences in the databases (Table 2.1). Only two cDNAs, *LjN93* and *LjN71*, were similar to the sequences with known or predicted biochemical functions, such as tyramine hydroxycinnamoyltransferase (Farmer et al., 1999) and putative UDP-glucose 4-epimerase (Table 2.1). The cDNAs *LjN22*, *LjN112* and *LjN80* were similar to the genes previously identified in the screens for cDNAs expressed in a particular tissue or under certain environmental conditions (Table 2.1). For example, *LjN112* represented a homolog of nodulin gene *MtN24* (Gamas et al., 1996); *LjN22* showed similarity to the

gene encoding anoxia-inducible protein (aie) from rice (Huq and Hodges, 1999) and *LjN80* was similar to the embryo-specific genes *ATS1* and *ATS3* (Nuccio and Thomas, 1999). The product of *LjN81* cDNA was highly similar to a putative Arabidopsis protein of unknown function (Table 2.1). Finally, LjN50 cDNA encoded a novel protein. Thus, out of 18 unique *LjN* clones isolated, 14 are likely to represent novel nodulin genes.

Table 2.1. Sequence similarities detected for the nodule specific *LjN* clones.

Clone name	Accession number	Number of isolates	Size range (bp)	Best homology	Significance[a] DNA*/protein
LjN3	AF000382	1	359	*A. thaliana* protein phosphatase 2C[1]	1.4×10^{-19}
LjN13	AF000383	3	220-462	*G. max* mRNA ENOD40-2[2]	$6.9 \times 10^{-27*}$
LjN53	AF000392[b]	1	372	transporter	
LjN63	AF000402[b]	1	342	Glu-rich proteins	
LjN73	AF000403[b]	1	361	*Z. mays* cytochrome P450[3]	6.2×10^{-25}
LjN77	AF000405	2	352-511	*M. sativa* leghemoglobin[4]	3.7×10^{-55}
LjN132	AF000390	5	167-170	*M. sativa* leghemoglobin[4]	$3.0 \times 10^{-6*}$
LjN36	AF000406	1	238	*C. lineata* leghemoglobin mRNA[5]	$3.0 \times 10^{-11*}$
LjN101	AF000407	2	297-300	*G. max* coproporphyrinogen oxidase[6]	2.1×10^{-16}

LjN5	U64964[b]	1	607	Plant PITP-like proteins[7]	
LjN48	AF000404	2	483	*O. formigenes* oxalate/formate exchange protein[8]	
LjN93	AF000391	1	414	NS	
cDNA			1129	*N. tabacum* tyramine hydroxycinnamoyl-transferase[9]	2×10^{-38}
LjN112	AF000389	3	228-415	NS	
cDNA			1014	*M. truncatula* nodulin MtN24[10]	5×10^{-42}
LjN22	AF000388	1	85	NS	
cDNA			570	*O. sativa* anaerobically inducible early gene 2[11]	3×10^{-4}
LjN50	AF000408	9	162-570	NS	
cDNA			670	NS, proline-rich protein	
LjN71	AF000387	1	428	NS	
cDNA			1480	*P. horikoshii* probable UDP-glucose 4-epimerase[12]	4×10^{-24}
LjN80	AF000386	1	115	NS	
cDNA			747	*A. thaliana* embryo-specific protein 3[13]	1×10^{-16}
LjN81	AF000385	3	477-530	NS	
cDNA			1404	*A. thaliana*	1×10^{-127}

hypothetical protein[14]

a = probability (P-value) that such match would occur merely by chance as given by BLAST algorithm; b = full copy or almost full copy cDNA corresponding to the indicated EST has been deposited to GenBank under this accession number; NS = no significant match for the corresponding EST. The information for the full-length or partial cDNA clones representing such ESTs is shown below each EST. (1) Kuromoni and Yamamoto, 1994; (2) Yang et al., 1993; (3) Frey et al., 1995; (4) Lobler and Hirsch, 1992; (5) Acc. No. U09671; (6) Madsen et al., 1993; (7) Kapranov et al. 1997, Chapters 3&4 of this thesis (8) Szczyglowski et al., 1998, Chapter 7 of this thesis; (9) Farmer et al., 1999; (10) Gamas et al., 1996; (11) Huq and Hodges, 1999; (12) Acc. No. H71145; (13) Nuccio and Thomas, 1999; (14) Acc. No. AC013483_38.

Expression Analysis of the *LjN* ESTs

In order to study the temporal expression pattern of selected *L. japonicus* ESTs, developmental slot blot northern analyses were performed. Total RNA isolated from uninfected *L. japonicus* control roots and root segments or nodules harvested 3, 7, 11, and 21 days after infection, was hybridized with a representative selection of radiolabeled cDNA inserts (Fig. 2.7). Since the plant material used in this experiment was generated using slightly modified growth conditions (see Materials and Methods), the *LjN77*-derived insert, encoding one of the *L. japonicus* leghemoglobins, was used as a marker gene for late developmental events (Table 2.1; Fig. 2.7). The mRNA species

corresponding to the *L. japonicus* leghemoglobin genes were first detectable at 7 d after infection, which was slightly earlier then the leghemoglobin gene expression pattern obtained during our initial experiments (compare Figs. 1 and 7). The corresponding mRNA accumulated gradually to a high level in 21-day-old nodules. All *LjN* cDNA inserts analyzed hybridized in a nodule-specific or enhanced manner, except for *LjN3*, which did not give a clearly detectable signal with either root- or nodule-derived total RNA (see below). The expression of some of the genes analyzed could also be detected at a very low level in the uninfected roots (e.g. *LjN 101*, *LjN71*, *LjN93*, *LjN73*). However, a significant increase in the level of their corresponding mRNAs was apparent around 7 d after infection, which was similar to the expression patterns of the other *L. japonicus* ESTs analyzed (Fig. 2.7).

The expression pattern of the majority of genes analyzed resembled that of the leghemoglobin gene (*LjN77*), indicating that they are likely to represent *L. japonicus* late nodulin genes. Interestingly, mRNA of the coproporphyrinogen oxidase homolog *LjN101* accumulated in a slightly different fashion than the mRNA of the leghemoglobin gene *LjN77*. The gene corresponding to *LjN101* appeared to be induced to a high level around 7 d after inoculation, with no significant changes in the steady-state level of mRNA accumulation during the later stages of nodule development (Fig. 2.7).

The mRNA corresponding to the *LjN13*, an *ENOD40* homolog (Table 2.1 and Fig. 2.4B), was weakly detectable in uninfected roots, but the hybridization signals were significantly enhanced in infected roots and fully developed 21-day-old nodules. The poly $(A)^+$ fraction of total RNA from uninfected roots and nodules was used to further

analyze tissue-specific expression of *LjN3*, the putative protein phosphatase 2C homologue. The *LjN3* insert hybridized with both root and nodule mRNA species of approximately 1600 nt in length (Fig. 2.8). However, the level of the corresponding mRNA in *L. japonicus* nodules was found to be approximately six times higher than in uninfected roots, confirming the nodule-specific/enhanced pattern of *LjN3*gene expression.

Since all EST sequences analyzed were generated using PCR-based procedures, their plant (*L. japonicus*) origin needed to be confirmed using Southern blot analysis. All *LjN* cDNA inserts hybridized specifically with *L. japonicus* genomic DNA, but not with total DNA isolated from *R. loti* strain NZP2235 (data not shown).

Further Characterization of the *L. japonicus* EST Library

To further characterize the *L. japonicus* EST library, we employed automated DNA sequencing. Two randomly selected recombinant colonies, both corresponding to a single differentially displayed PCR product, were used for DNA sequencing analysis. The recombinant colonies were selected from the collection of *L. japonicus* EST "combo" plates, based on their lack of hybridization with the radiolabeled root and nodule cDNA probes used during the differential hybridization procedure. The nucleotide sequence of 142 cDNA inserts, representing a total of 71 PCR products, was established. DNA sequence comparisons showed that they corresponded to 69 unique cDNA sequences (data not shown). The EST nucleotide sequences obtained were compared to the nucleotide and protein sequences in the data bases by BLASTN and BLASTX searches,

respectively (Altschul et al., 1990). Nine of 69 EST sequences analyzed showed a moderate to high level of similarity to DNA/protein sequences stored in the data bases. A summary of the results of this analysis is shown in Table 2.2. We refer to these clones as *L. japonicus LjEST* cDNAs. All cDNAs, except *LjEST59*, appeared to encode unique enzymatic functions, including subtilisin-like protease (Ribeiro et al., 1995), adenosylosuccinate synthetase (Zeidler et al., 1993), tyrosine decarboxylase (Maldonado-Mendoza et al., 1996), dehydroquinate dehydratase/shikimate dehydrogenase (Booner and Jensen, 1994), heme oxygenase (Shibahara et al., 1985), chalcone synthase (Goormachtig et al., 1995), and glutamine phosphoribosylpyrophosphate amidotransferase (Kim et al., 1995). However, their specific involvement in the nodulation process needs to be confirmed using RNA blot hybridization or other more sensitive protocols, such as RT-PCR. The remaining 60 EST sequences obtained via the random sequencing approach did not show any significant homologies or similarities and were therefore classified as anonymous *L. japonicus* ESTs.

Table 2.2. Sequence similarities detected for randomly sequenced EST clones.

Clone name	Accession number	Number of isolates	Size (bp)	Best homology	Significance DNA*/protein
LjEST38	AF000393	1	411	*Arabidopsis thaliana*	3.8×10^{-18}

				subtilisin-like protease[1]	
LjEST58	AF000394	1	527	yeast adenylosuccinate synthetase[2]	2.7×10^{-7}
LjEST59	AF000395	1	258	unknown *Trypanosoma brucei* protein[3]	5.4×10^{-8}
LjEST66	AF000396	1	245	*Papaver somniferum* tyrosine decarboxylase[4]	1.5×10^{-13}
LjEST103	AF000398	1	131	*Nicotiana tabacum* dehydroquinase/shikimate dehydrogenase mRNA 3'end[5]	$1.8 \times 10^{-12*}$
LjEST105	AF000399	1	238	Rat heme oxygenase[6]	3.4×10^{-5}
LjEST118	AF000400	1	296	*Sesbania rostrata* chalcone reductase[7]	1.7×10^{-11}
LjEST120	AF000401	1	192	*Glycine max* glutamine phosphoribosylpyrophosphate amidotransferase[8]	8.0×10^{-8}

(1) Ribeiro et al., 1995; (2) Zeidler et al., 1993; (3) Acc. No. U05313; (4) Maldonado-Mendoza et al., 1996; (5) Booner and Jensen, 1994; (6) Shibahara et al., 1985; (7) Goormachtig et al., 1995; (8) Kim et al., 1995.

2.5. DISCUSSION

We have previously reported the successful application of the mRNA differential display technique for the identification and isolation of early nodulin genes from *Sesbania rostrata* stem nodules (Goormachtig et al., 1995). Here, we employed the RT-PCR-based differential display procedure to identify a range of molecular markers associated with relatively late developmental events during *L. japonicus* nodule organogenesis. A well-characterized nodulin gene, namely the leghemoglobin gene, was used to define the late stages in *L. japonicus* nodule morphogenesis. Using 80 primer combinations, the profiles of approximately 10,000 PCR products were analyzed and 110 differentially displayed nodule-specific or enhanced bands were successfully reamplified and cloned. Thus, a library of differentially displayed *L. japonicus* ESTs was established. The differential hybridization procedure revealed relatively abundant mRNA species and allowed the selection of 39 nodule-specific PCR products, representing 18 unique cDNA sequences. For the purpose of this paper, the term "specific" was used merely to reflect the significant differences in the expression levels observed between the uninfected roots, and infected roots or nodules.

The apparent redundancy found among some of the isolated sequences (between 2 and 9, see Table 2.1) was in part due to the different combinations of primers used during the differential display procedure. Interestingly, in most cases analyzed, different

positions of T_{12}MN primers at the 3' end of the mRNA, in combination with a single arbitrary decamer primer, gave rise to multiple products derived from the same mRNA species (e.g. *LjN13*, *LjN101*, *LjN81*, Table1). This finding may be explained by the presence of multiple poly(A) sites in a given plant mRNA species (Wu et al., 1995). In the case of the *LjN50* cDNA group, in addition to differences in the position of the T_{12}MN primers, three different arbitrary decamers contributed to multiple independent isolations of the respective cDNA species (data not shown). The sequence analysis of nine partial cDNAs belonging to the *LjN50* group revealed that all of them shared highly conserved 3'-terminal sequences of approximately 220 bp. However, significant differences in the corresponding DNA sequences at the 5' ends were found, indicating that they may, in fact, correspond to related, but not identical, genes in the *L. japonicus* genome (data not shown).

The comparative sequence analysis of representatives of all 18 cDNA groups established that the majority of them are likely to correspond to novel nodule-specific genes. This group of *L. japonicus* ESTs includes putative homologues of protein phosphatase 2C and cytochrome P450, both candidates for proteins with regulatory functions. The protein phosphatase activity in plants, of which the least well-characterized member is PP2C protein serine/threonine phosphatase, has been implicated in such processes as signal transduction, hormonal regulation, mitosis, and control of carbon and nitrogen metabolism (Smith and Walker, 1996). Cytochrome P450 enzymes, on the other hand, are membrane-bound, heme-containing enzymes, implicated in a variety of biosynthetic reactions (Nelson et al., 1993; Frey et al., 1995). In plants they

are typically involved in the synthesis of chemically diverse secondary metabolites, often involved in defense mechanism, synthesis of plant hormones, or cell-wall related substances (Holton et al., 1993; Winkler and Helentjaris, 1995; Frey et al., 1995; Szekeres et al., 1996). The specific functions of these interesting new nodulin genes (*LjN3* and *LjN73*) in *L. japonicus* nodules remain to be elucidated.

cDNA clones that shared significant similarity to already characterized late-nodulin genes from different legume species were also identified. The latter group includes *L. japonicus* leghemoglobin genes and a putative homologue of soybean coproporphyrinogen oxidase, an enzyme catalyzing the oxidative decarboxylation of coproporphyrinogen III to protoporphyrinogen IX, in the heme and chlorophyll biosynthetic pathway(s) (Madsen et al., 1993). The latter finding supports the notion that specific plant host functions participate in the synthesis of heme, which is needed in nodules for increased hemoprotein biosynthesis (Madsen et al., 1993).

The kinetics of mRNA accumulation of the *L. japonicus* ESTs analyzed closely resemble the pattern observed for leghemoglobin gene expression in *L. japonicus* nodules, indicating that a majority of them are almost certainly related to late stages in nodule development. Since the goal of our analysis was to identify molecular markers associated specifically with late developmental stages in nodule organogenesis, our results illustrate the advantage of using nodules with a determinate developmental pattern for late-nodulin expression analysis, as opposed to indeterminate nodules, where successive early and late developmental stages coexists.

Since a majority of the *L. japonicus* EST clones did not hybridize with the root- or nodule-specific cDNA probes used during differential hybridization screening, we employed a random sequencing approach to further characterize the remaining differentially displayed *L. japonicus* ESTs. This resulted in the generation of 69 unique cDNA sequences. Only nine of the cDNAs shared similarity with protein sequences from the data bases (see Table 2.2). Since selection of all of these sequences was based exclusively upon the observed differential display patterns, it is not certain whether they indeed correspond to nodule-specific or nodule-enhanced genes. Clearly, more in-depth analyses are needed to unambiguously establish their tissue and or cell specificity. However, it is important to note that at least three of them, namely subtilisin-like protease, chalcone reductase, and glutamine phosphoribosylpyrophosphate amidotransferase have previously been reported to be induced during nodulation (Ribeiro et al., 1995; Goormachtig et al., 1995; Kim et al., 1995). Several roles for lipoxygenases (LOX) in plant-microbe interactions including symbiotic nodule formation have also been suggested (Croft et al., 1993; Veronesi et al., 1996; Gardner et al., 1996; Perlick et al., 1996)

In summary, of 110 *L. japonicus* differential amplification products, 88 unique partial cDNA sequences were obtained. Nineteen of these were further analyzed and shown to correspond to nodule-specific genes, with the majority of them likely to encode novel nodule-specific functions. Twenty-five *L. japonicus* ESTs showed varying degrees of similarity to different DNA/protein sequences stored in the data bases. For most of these, solid predictions about their activities could be made. However, their specific roles

in the *L. japonicus* nodules need to be further analyzed. A group of 59 EST sequences failed to reveal significant homology to any sequences in the data bases, and their tissue specificity has not been firmly established. We are currently analyzing these ESTs with regard to their nodule-specific expression. We assume that some of them may, in fact, correspond to a group of weakly expressed genes with specific functions during late stages of *L. japonicus* nodule development. In fact, a preliminary analysis of 16 different LjEST clones revealed that 8 of them corresponded to low-abundant mRNAs induced in *L. japonicus* nodules (Kasiborski et al., unpublished). The novelty and diversity of the isolated nodule-specific genes should greatly facilitate further molecular analyses of the late stages of nodule development. The collection of ESTs reported here will also be an indispensible tool for the development of a *L. japonicus* genetic map, an essential step toward future map-based cloning of symbiosis-specific genes in this model legume plant.

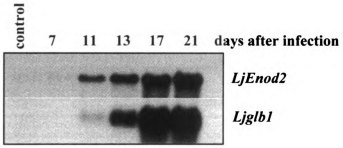


Figure 2.1. RNA gel blot analysis of *L. japonicus LjEnod2* and leghemoglobin (*Ljglb1*) gene expression. 10 μg of total RNA isolated from 7-day-old uninfected roots (control), and root segments or nodules harvested 7, 11, 13, and 21 d after infection were analyzed. α-^{32}P-radiolabeled cDNAs encoding *L. japonicus LjEnod2* (upper panel), and *Ljlb1* (lower panel) were used as molecular probes.

Figure 2.2. Developmental mRNA differential display. An example is shown of mRNA amplification profiles from 21 (lane 1) and 7 (lane 2) days old uninfected control roots, and root segments and nodules harvested 7 (lane 3), 11 (lane 4), 13 (lane 5), and 21 (lane 6) days after infection. The decamer primer AP9 (CGTGGCAATA), in combination with four 3'- anchoring primers (T_{12}MG, T_{12}MA, T_{12}MT, and T_{12}MC, respectively), was used to generate the RNA profiles shown. The dots (•) indicate putative nodule-specific bands. The differentially displayed band generated using AP9/T_{12}MC primer combination (see lane 6) corresponds to the *LjN77* EST and encodes a *L. japonicus* leghemoglobin (see Table 2.1).

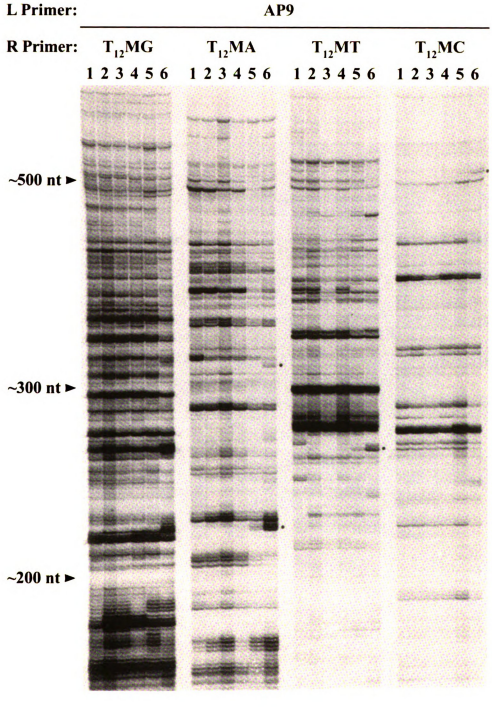

L Primer:
AP9
R Primer:
T12MG
T12MA
T12MT
T12MC
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
~500 nt ►
~300 nt ►
~200 nt ►

Figure 2.3. Differential colony hybridization. The result of differential hybridization, using recombinant colonies derived from two "combo" microtiter plates is shown. The filters shown in Panels A and B were hybridized using radiolabeled cDNA derived from control 21-day-old uninfected roots and 21-day-old *L. japonicus* nodules, respectively. For further details see text.

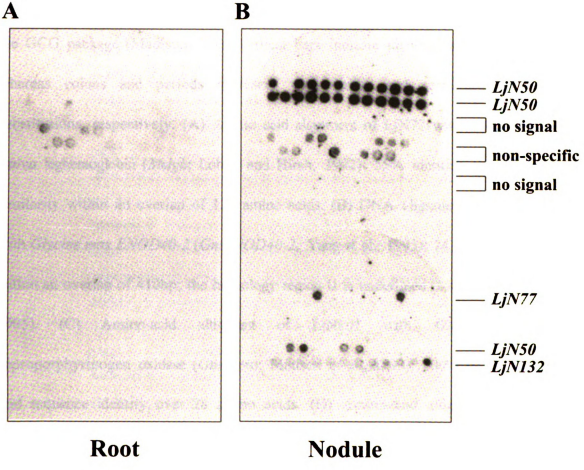

A
B
LjN50
LjN50
no signal
non-specific
no signal
LjN77
LjN50
LjN132
Root
Nodule

Figure 2.4. DNA/protein sequence alignments of LjN77, *LjN13*, LjN101, and LjN3. The alignments shown were performed using the BESTFIT program from the GCG package (Madison, WI). Vertical bars indicate identical amino acids, whereas colons and periods represent conservative and semiconservative substitutions, respectively. (A) Amino acid alignment of LjN77 with *Medicago sativa* leghemoglobin (*Mslgb*; Lobler and Hirsh, 1992): 55% identity and 68% similarity within an overlap of 122 amino acids. (B) DNA alignment of *LjN13* with *Glycine max ENOD40-2* (*GmENOD40-2;* Yang et al., 1993*)*: 74.5% identity within an overlap of 410bp; the homology region II is underlined (see Vijn et al., 1995). (C) Amino-acid alligment of LjN101 with *Glycine max* coproporphyrinogen oxidase (*Gm*copro; Madsen et al., 1993): 100% of amino-acid sequence identity over 28 amino acids. (D) Amino-acid alligment of *L. japonicus* LjN3 with *Arabidopsis thaliana* protein phosphatase 2C (*At*pp2C; Kuromoni and Yamamoto, 1994): 50% identity and 63% similarity within an overlap of 79 amino acids.


A
LjN77 AIVFCSTPOYWKKPQ.....LLKTCSPSKGFWTHAQSSTP...SPML 40
Mslgb ALVNSSWESFKQNPGNSVLFYTIILEKAPAAKGMFSFLKDSAGVQDSPKL 58
LjN77 .....KRFLDJHAMAAQLLAKGEVTLADASLGAVHVQKAVADPHFAVVKE 85
Mslgb QSHAEKVFGMVRDSAAQLRATGGVVLGDATLGAIHIQKGVVDPHFAVVKE 108
LjN77 ALLKTVQAAVGDKWSEELSTAWGVAYDGLAAAIKKAM 122
Mslgb ALLKTIKEVSGDKWSEELNTAWEVAYDALATAIKKAM 145
B
LjN13 gagagaagctttggctacagcctggcgaaaccggcaagtcac.agaaagg 73
GmENOD40-2 gaggagaggcttggctacagcctggc.aaaccggcaagtcacaaaaaagg 202
LjN13 caatggaccccattaggtttcttatggctatgtatgaatgactcttttgt 123
GmENOD40-2 caatggactccattggggtctctatggctatgta.....gtgctcatgt 246
LjN13 agttcttcttcaagtagaatgtaataaacaaaaatgttcttcttccttt. 172
GmENOD40-2 agttcttcttgctgtagaatgtaataataaacaaagttggtcttcctttt 296
LjN13 ....gaggtggtgttcattcatatcacccaattttgcagctgac 212
GmENOD40-2 gagaagttaccagcttttgctgtccaaaattactcaa.tttgcagctgac 345
LjN13 tagactcccagtttgctcttcagtttcttcgaagatgagtaggtaggtaa 262
GmENOD40-2 tagaattcc.tttctctcttcagtttct..gcagatgagtaggtaggcaa 392
LjN13 .ttatgatcaccttgtctcttctttttctgtg...tctgttttccttttc 308
GmENOD40-2 tttgtgatcac....tcccttcccttttcatgtcttctgtgttccctttt 438
LjN13 tcatgtgtgtatggttgttgtcttgtggcctatatgcaataatagtattt 358
GmENOD40-2 ccatgcttgtttgtgttgttagttatgacct.tatgaggaaataaaagaa 487
LjN13 gagtgatgtttttgttccttcaagtgtaggattgtgttttgtagaagttg 408
GmENOD40-2 tagtacaattctagtccct..cagtttaggattgtattctattgaacttt 535
LjN13 at 410
GmENOD40-2 at 537
C
LjN101 RWEYDHKPEEGSEEWKLLDACINPKEWI 28
Gmcopro RWEYDHKPEEGSEEWKLLDACINPKEWI 384
D
LjN3 SVTKRSSKDEFLILASDGLWDVISSEMACQVVRKCLNG.QIRRICNENQS 49
Atpp2C TVTDRTDEDECLILASDGLWDVVPNETACGVARMCLRGAGAGDDSDAAHN 360
LjN3 RASEAATLLAEIALAKGSRDNTSVIVIELR 79
Atpp2C ACSDAALLLTKLALARQSSDNVSVVVVDLR 390

Figure 2.5. Protein sequence alignments of LjNP450 and LjN65. The alignments shown were performed as describe in Figure 4. (A) *L. japonicus* LjNP450 protein with cytochrome P-450 CYP71A1 from avocado fruit (Bozak at al., 1990): 46% identity and 68% similarity; the heme-binding domain for cytochrome P-450 (the signature sequence FXXGXXXCXG; Nelson et al., 1993) is underlined. (B) *L. japonicus* Nlj65 with *Arabidopsis thaliana* A*t*PTR2-B transporting protein; 38% identity and 59% similarity (Frommer et al., 1994).

A


CYP71A1 MAILVSLLF 9
LjNP450 VALMILRKNLKKPDSIPNIPPGPWKLPIIGSIPHLVGSPPHRKLRD 46
CYP71A1 LAIALTFFLLKLN.EKREKKPNLPPSPPNLPIIGNLHQL.GNLPHRSLRS 57
LjNP450 LAKKYGPLMHLQLGEVIFIIVSSAEYAKEVMKTHDVTFASRPRSLFTDIV 96
CYP71A1 LANELGPLILLHLGHIPTLIVSTAEIAEEILKTHDLIFASRPSTTAARRI 107
LjNP450 FYGSTDIGFSPYGDYWRQVRKICNVELLSMKRVQSLWPIREEEVKNLIQR 146
CYP71A1 FYDCTDVAFSPYGEYWRQVRKICVLELLSIKRVNSYRSIREEEVGLMMER 157
LjNP450 I..ASEEGSVVNLSQAIDSLIFTITSRSAFGKRYMEQEE....FISCVRE 190
CYP71A1 ISQSCSTGEAVNLSELLLLLSSGTITRVAFGKKYEGEEERKNKFADLATE 207
LjNP450 VMKLAGGFNIADLFPSAKWLENLTRMRSKFEYLHQKMDRILETIIDDHKA 240
CYP71A1 LTTLMGAFFVGDYFPSFAWVDVLTGMDARLKRNHGELDAFVDHVIDDHLL 257
LjNP450 NSRTKEGQVEGGEEDLIDVLLKYENSSTDQDFHLTIRNIKAILFDIFIAG 290
CYP71A1 .SRKANGSDGVEQKDLVDVLLHLQKDSS.LGVHLNRNNLKAVILDMFSGG 305
LjNP450 SETSATTINWTMAEMMKDPILLKKAQDEVREIFQRRGKVDETCIYELKYL 340
CYP71A1 TDTTAVTLEWAMAELIKHPDVMEKAQQEVRRVVGKKAKVEEEDLHQLHYL 355
LjNP450 KAFINEVLRLHPPGPLVF.RECRQACEINGYHIPAKSTVLVNTFAIGTDS 389
CYP71A1 KLIIKETLRLHPVAPLLVPRESTRDVVIRGYHIPAKTRVFINAWAIGRDP 405
LjNP450 KYWAEPERFCPERFIDSSIDYKGTNFEHLPFGAGRRICPGINYGMANVEL 439
CYP71A1 KSWENAEEFLPERFVNNSVDFKGQDFQLIPFGAGRRGCPGIAFGISSVEI 455
LjNP450 VLALLLYHFDWTLPKGIKNEDLDLTEEFGVTVSKKEDLCLIPSISHPLPST 490
CYP71A1 SLANLLYWFNWELPGDLTKEDLDMSEAVGITVHMKFPLQLVAKRHLS 502

B

```
Nlj65                                                          M 1

AtPTR2-B                                          MGSIEEEARPLIEEGLILQ 19

Nlj65      EGKGYTLDGTVDLAGRPVLSSLTGKQKACTYILVYRVLERFAYYGVGANL 51
AtPTR2-B   EVKLYAEDGSVDFNGNPPLKEKTGNWKACPFILGNECCERLAYYGIAGNL 69

Nlj65      VNFMTTQLNKDVVSSITSFNNWSGLATLTPILGAYIADTYTGRFWTITFS 101
AtPTR2-B   ITYLTTKLHQGNVSAATNVTTWQGTCYLTPLIGAVLADAYWGRYWTIACF 119

Nlj65      LLIYAIGLVLLVLTTTLKSLRPA.CENGICREASNLQVALFYTSLYTIAV 150
AtPTR2-B   SGIYFIGMSALTLSASVPALKPAECIGDFCPSATPAQYAMFFGGLYLIAL 169

Nlj65      GSGAVKPNMSTFGADQFDDFRHEEKEQKVSFFNWWAFNGACGSLMATLFV 200
AtPTR2-B   GTGGIKPCVSSFGADQFDDTDSRERVRKASFFNWFYFSINIGALVSSSLL 219

Nlj65      VYIQEKNGWGLAYSLSAIGFLLSSIIFFWGSPVYRHKSRQARSPSMNFIR 250
AtPTR2-B   VWIQENRGWGLGFGIPTVFMGLAIASFFFGTPLYRFQ.KPGGSPITRISQ 268

Nlj65      VPLVAFRNRKLQLPCNPSELHEFQLNYYISSGARKIHHTSHFSFLDRAAI 300
AtPTR2-B   VVVASFRKSSVKVPEDATLLYETQDKNSAIAGSRKIEHTDDCQYLDKAAV 318

Nlj65      ......RESNTDLSNPPCTVTQVEGTKLVLGMFQIWLLMLIPTNCWALES 344
AtPTR2-B   ISEEESKSGDYSNSWRLCTVTQVEELKILIRMFPIWASGIIFSAVYAQMS 368

Nlj65      TIFVRQGTTMDRTLGPKFRLPAASLWCFIVLTTLICLPIYDHYFIPFMRR 394
AtPTR2-B   TMFVQQGRAMNCKIGS.FQLPPAALGTFDTASVIIWVPLYDRFIVPLARK 417

Nlj65      RTGNHRGIKLLQRVGIGMAIQVIAMAVTYAVETQRMSVIKKHHIAGPEET 444
AtPTR2-B   FTGVDKGFTEIQRMGIGLFVSVLCMAAAAIVEIIRLHMANDLGLVESGAP 467

Nlj65      VPMSIFWLLPQNIILGVSFAFLATGMLEFFYDQSPEEMKGLGTTLCTSCV 494
AtPTR2-B   VPISVLWQIPQYFILGAAEVFYFIGQLEFFYDQSPDAMRSLCSALALLTN 517

Nlj65      AAGCYINTFLVTMI.......DKLNWIGNNLNDSQSRITIMPFFSVISAL 503
AtPTR2-B   ALGNYLSSLILTLVTYFTTRNGQEGWISDNLNSGHLDY.FFWLLAGLSLV 567

Nlj65      NFGVFLWVSSGYIYKKENTSTTEVHDIEMSAEKTVKY              574
AtPTR2-B   NMAVYFFSAARYKQKKASS                                585
```

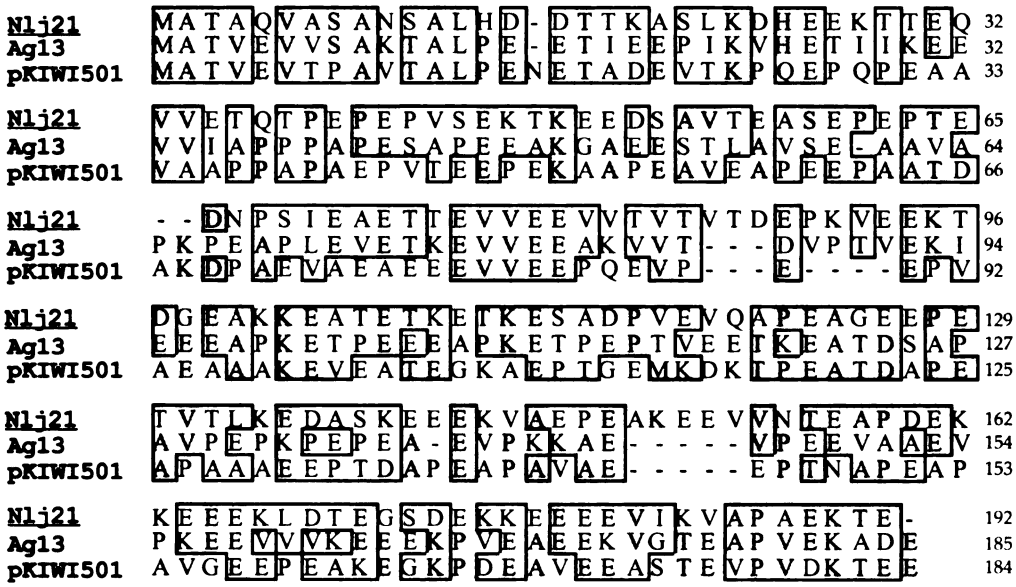
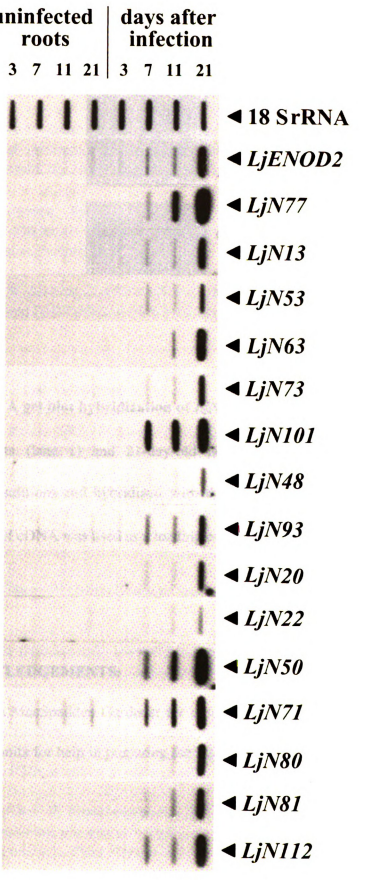


Figure 2.6. Amino-acid sequences comparison of Nlj21, Ag13, and PKIWI501 . The deduced amino-acid sequence of the *L. japonicus* nodulin Nlj21 is compared with *A.glutinosa* Ag13 and kiwifruit pKIWI501 (Guan et al.,1997; Ledger and Gardner, 1994). The identical residues, as well as conservative substitutions, are boxed.

Figure 2.7. Developmental slot blot northern analysis. 10 μg of total RNA isolated from uninfected (control), and infected root segments and nodules at the time points indicated were transferred onto nitrocellulose filters using a manifold slot blot vacuum apparatus and hybridized with radiolabeled cDNA inserts corresponding to 14 unique *LjN* ESTs. The cDNAs *LjN112* and *LjN20* were later found to be derived from the same mRNA species (data not shown). The *LjENOD2* and leghemoglobin (*LjN77*) cDNAs were included as nodule-developmental markers, whereas the 18S cDNA was used as a loading control


uninfected roots
days after infection
3 7 11 21 3 7 11 21
18 SrRNA
LjENOD2
LjN77
LjN13
LjN53
LjN63
LjN73
LjN101
LjN48
LjN93
LjN20
LjN22
LjN50
LjN71
LjN80
LjN81
LjN112

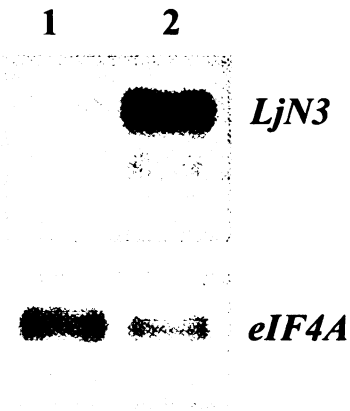


Figure 2.8. RNA gel blot hybridization of *LjN3*. 4 μg of poly(A)$^+$ mRNA from control uninfected roots (lane 1) and 21-day-old nodules (lane 2) were separated under denaturating conditions and hybridized with the α^{32}P- labeled cDNA insert from EST *LjN3*. The *eIF4A* cDNA was used as a loading control (Taylor et al., 1993).

2.6. ACKNOWLEDGEMENTS:

We thank Brian MacSpadden Gardener for comments and helpful discussions. We also thank Kurt Stepnitz for help in preparing the Figures.

2.7. LITERATURE CITED:

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410.

Bauer D, Muller H, Reich J, Riedel H, Ahrenkiel V, Warthoe P, Strauss M (1993) Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). Nucleic Acids Res **21:** 4272-4280

Booner CA, Jensen RA (1994) Cloning of cDNA encoding the bifunctional dehydroquinase-shikimate dehydrogenase of aromatic-amino-acid biosynthesis in *Nicotiana tabacum*. Biochem J **302:** 11-14

Bozak KR, Yu H, Sirevag R, Christoffersen, RE (1990) Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proc Natl Acad Sci USA **87:** 3904-3908.

Broughton WJ, Dilworth MY (1971) Control of leghemoglobin synthesis in snake beans. Biochem J **125:** 1075-1080

Carlson RW, Price NPJ, Stacey G (1995) The biosynthesis of rhizobial lipo-oligosaccharide nodulation signal molecules. Mol. Plant-Microbe Interact **7:** 684-695

Cheon C-III, Lee N-G, Siddique A-BM, Bal AK, Verma DPS (1993) Roles of plant homologs of Rablp and Rab7p in the biogenesis of the peribacteroid membrane, a subcellular compartment formed *de novo* during root nodule symbiosis. EMBO J **12:** 4125-4135

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA **81**: 1991-1995

Cook D, Dreyer D, Bonnet D, Howell M, Nony N, VandenBosch K (1995) Transient induction of a peroxidase gene in *Medicago truncatula* precedes infection by *Rhizobium meliloti*. Plant Cell **7:** 43-55

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) *enod40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. EMBO J **13:** 5099-5112

Croft KPC, Juttner F, Slusarenko AJ (1993) Volatile products of the lipoxygenase pathway evolved from *Phaseolus vulgaris* (L.) leaves inoculated with *Pseudomonas syringae* pv *phaseolicola*. Plant Physiol **101:** 13-24

de Bruijn FJ, Schell J (1992) Regulation of plant genes specifically induced in developing and mature nitrogen-fixing nodules: *cis*-acting elements and *trans*-acting factors. *In* DPS Verma, ed, Control of Plant Gene Expression, CRC Press, Boca Raton, pp 241-258

Delauney AJ, Verma DPS (1988) Cloned nodulin genes for symbiotic nitrogen fixation. Plant Mol Biol Rep **6:** 279-285

Farmer MJ, Czernic P, Michael A, Negrel J (1999) Identification and characterization of cDNA clones encoding hydroxycinnamoyl-CoA:tyramine N-hydroxycinnamoyltransferase from tobacco. Eur J Biochem **263:** 686-694

Franssen HJ, Nap J-P, Gloudemans T, Stiekema W, van Dam H, Govers F, Louwerse J, van Kammen A, Bisseling T (1987) Characterization of cDNA for nodulin-75 of soybean: a gene product involved in early stages of root nodule development. Proc Natl Acad Sci USA **84:** 4495-4499

Frey M, Kliem R, Saedler H, Gierl A (1995) Expression of a cytochrome P450 gene family in maize. Mol Gen Genet **246:** 100-109

Frommer WB, Hummel S, Rentsch D (1994) Cloning of an *Arabidopsis* histidine transporting protein related to nitrate and peptide transporters. FEBS Lett **347:**185-189

Gardner CD, Sherrier DJ, Kardailsky IV, Brewin NJ (1996) Localization of lipoxygenase in the lumen of infection threads. Mol Plant-Microbe Interact **9:** 282-289

Goormachtig S, Valerio-Lepieniec M, Szczyglowski K, Van Montagu M, Holsters M, de Bruijn FJ (1995) Use of differential display to identify novel *Sesbania rostrata* genes enhanced by *Azorhizobium caulinodans* infection. Mol Plant-Microbe Interact **8:** 816-824

Guan Ch, Akkermans AD, van Kammen A, Bisseling T, Pawlowski K (1997) *ag13* is expressed in *Alnus glutinosa* nodules in infected cells during endosymbiont degradation and in the nodule pericycle. Physiol Plant (in press)

Hall TC, Ma Y, Buchbinder BU, Pyne JW, Sun SM, Bliss FA (1978) Messenger RNA for G1 protein of French bean seeds: cell-free translation and product characterization. Proc Natl Acad Sci USA **75:** 3196-3200.

Handberg K, Stiller J, Thykjear T, Stougaard J (1994) Transgenic plants: *Agrobacterium* mediated transformation of the diploid legume *Lotus japonicus*. In: Cell Biology: A Laboratory Handbook, Celis JE, ed, **1**: 119-127

Handberg K, Stougaard J (1992) *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. Plant J **2:** 487-496

Heard J, Dunn K (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc Natl Acad Sci USA **92:** 5273-5277

Holton TA, Brugliera F, Lester DR, Tanaka Y, Hyland CD, Menting JG, Lu CY, Farcy E, Stevenson TW, Cornish EC (1993) Cloning and expression of cytochrome P450 genes controlling flower colour. Nature **366:** 276-269

Huq E, Hodges TK (1999) An anaerobically inducible early (aie) gene family from rice. Plant Mol Biol **40:** 591-601

Jiang Q, Gresshoff PM (1993) *Lotus japonicus*: a model plant for structure-function analysis in nodulation and nitrogen fixation. Curr Topics Plant Mol Biol **2:** 97-110

Jiang Q, Gresshoff PM (1997) Classical and molecular genetics of the model legume *Lotus japonicus*. Mol Plant-Microbe Interact **1**: 59-68

Jarvis BDW, Pankhurst CE, Patel JJ (1982) *Rhizobium loti*, a new species of legume root nodule bacteria. Inter J Syst Bacteriol **32:** 378-380

Kapranov P, de Bruijn FJ, Szczyglowski K (1997) A novel, highly expressed late nodulin gene *LjNOD16* from *Lotus japonicus*. Plant Physiol 113: 1081-1090

Kim JH, Delauney AJ, Verma DPS (1995) Control of *de novo* purine biosynthesis genes in ureide-producing legumes: induction of the glutamine phosphoribosylpyrophosphate amidotransferase gene and characterization of its cDNA from soybean and *Vigna*. Plant J **7:** 77-86

Kuromoni T, Yamamoto M (1994) Cloning of cDNAs from *Arabidopsis thaliana* that encode putative protein phosphatase 2C and a human Dr-1 like protein by transformation of the fission yeast mutant. Nucleic Acids Res **22:** 5296-52301

Ledger SE, Gardner RC (1994) Cloning and characterization of five cDNAs for genes differentially expressed during fruit development of kiwifruit (*Actinidia deliciosa* var. deliciosa). Plant Mol Biol **25:** 877-886

Liang P, Averboukh L, Pardee AB (1993) Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. Nucleic Acids Res **21:** 3269-3275

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **257:** 967-971

Lobler M, Hirsch AM (1992) An alfalfa (*Medicago sativa* L.) cDNA encoding acidic leghemoglobin (MsLb3). Plant Mol Biol **20:** 733-736

Madsen O, Sandal L, Sandal NN, Marcker K.A. (1993) A soybean coproporphyrinogen oxidase gene is highly expressed in root nodules. Plant Mol Biol **23:** 35-43

Maldonado-Mendoza IE, Lopez-Meyer M, Galef JR, Burnett, RJ, Nessler CL (1996) Molecular analysis of a new member of the opium poppy tyrosine/3,4-dihydroxyphenylalanine decarboxylase gene family. Plant Physiol **110:** 43-49

Marmur J (1961) A procedure for the isolation of deoxiribonucleic acid from microorganisms. J Mol Biol **3:** 208-218

Matvienko M, Van De Sande K, Yang W-C, Van Kammen A, Bisseling T, Franssen H (1994) Comparison of soybean and pea ENOD40 cDNA clones representing genes expressed during both early and late stages of nodule development. Plant Mol Biol **26:** 487-493

Miao G-H, Hong Z, Verma DPS (1993) Two functional soybean genes encoding $p34^{cdc2}$ protein kinases are regulated by different plant developmental pathways. Proc Natl Acad Sci USA **90:** 943-947

Mylona P, Pawlowski K, Bisseling T (1995) Symbiotic nitrogen fixation. Plant Cell **7:** 869-885

Nap J-P, Bisseling T (1990) Developmental biology of a plant-prokaryote symbiosis: The legume root nodule. Science **250:** 948-954

Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzales FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, Nebert DW (1993) The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol **12:** 1-51

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C (1994) Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol **106:** 1241-1255

Nuccio, ML, Thomas TL (1999) *ATS1* and *ATS3*: two novel embryo-specific genes in *Arabidopsis thaliana*. Plant Mol Biol **39:** 1153-1163

Perlick AM, Albus U, Stavridis T, Frühling M, Küster H, Pühler A (1996) The *Vicia faba* lipoxygenase gene VfLOX1 is expressed in the root nodule parenchyma. Mol Plant-Microbe Interact **9:** 860-863

Pridmore RD (1987) New and versatile cloning vectors with kanamycin resistance marker. Gene **56:** 309-312

Ribeiro A, Akkermans ADL, van Kammen A, Bisseling T, Pawlowski K (1995) A nodule-specific gene encoding a subtilisin-like protease is expressed in early stages of Actinorhizal nodule development. Plant Cell **7:** 785-794

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. *In*: SB Gelvin, RA Schilperoort, eds, Plant Molecular Biology Manual. Kluwer Academic Publishers, Dordrecht, **A6** 1-10

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A Laboratory Manual, ED2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Scheres B, Van de Wiel C, Zalensky A, Horvath B, Spaink H, Van Eck H, Zwartkruis F, Wolters AM, Gloudemans T, Van Kammen A, Bisseling T (1990a) The ENOD12 gene product is involved in the infection process during the pea-*Rhizobium* interaction. Cell **60:** 281-294

Scheres B, van Engelen F, van den Knaap E, van de Wiel C, van Kammen A, Bisseling T (1990b) Sequential induction of nodulin gene expression in the developing pea nodule. Plant Cell **2:** 687-700

Schultze M, Kondorosi E, Ratet P, Buire M, Kondorosi A (1994) Cell and molecular biology of *Rhizobium*-plant interaction. Int Rev Cytol **156:** 1-75

Schibahara S, Muller R, Taguchi H, Yoshida T (1985) Cloning and expression of cDNA for rat heme oxigenase. Proc Natl Acad Sci USA **82:** 7865-7869

Smith RD, and WalkerJC (1996) Plant protein phosphatases. Annu Rev Plant Physiol Plant Mol Biol **47:** 101-125

Spaink HP (1996) Regulation of plant morphogenesis by lipo-chitin oligonucleotides. Crit Rev Plant Sci **15:** 559-582

Sprent JI (1989) Which steps are essential for the formation of functional legume nodules? New Phytol **111:**129-153

Szekeres M, Nemeth K, Koncz-Kalman Z, Mathur J, Kauchmann A, Altman T, Redei GP, Nagy F, Schell J, Koncz C (1996) Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in *Arabidopsis*. Cell **85:** 171-182

Szczyglowski K, Kapranov P, Hamburger D, de Bruijn FJ (1998b) The *Lotus japonicus LjNOD70* gene encodes a protein with similarities to transporters. Plant Mol Biol 37: 651-661

Taylor CB, Bariola PA, DelCardayre SB, Raines RT, Green PJ (1993) RNS2: A senescence-associated RNase of *Arabidopsis* that diverged from the S-RNases before speciation. Proc Natl Acad Sci USA **90:** 5118-5122

van de Loo FJ, Turner S, Somerville C (1995) Expressed sequence tags from developing castor seeds. Plant Physiol **108:** 1141-1150

van de Sande K, Pawlowski K, Czaja I, Wieneke U, Schell J, Schmidt J, Walden R, Matvienko M, Wellink J, van Kammen A, Franssen H, Bisseling T (1996) Modification of phytohormone response by a peptide encoded by ENOD40 of legumes and nonlegumes. Science **273:** 370-373

Verma DPS (1992) Signals in root nodule organogenesis and endocytosis of *Rhizobium*. Plant Cell **4:** 373-382

Veronesi C., Rickauer M, Fournier J, Pouenat M-L, Esquerre-Tugaye M-T (1996) Lipoxygenase gene expression in the tobacco-*phytophtora parasitica nicotianae* interaction. Plant Physiol **112:** 997-1004

Verwoerd TC, Dekker BMM, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res **17:** 2362

Vijn I, Yang W-C, Pallisgard N, Jensen EO, van Kammen A, Bisseling T (1995) *VsENOD5, VsENOD12* and *VsENOD40* expression during *Rhizobium*-induced nodule formation on *Vicia sativa* roots. Plant Mol Biol **28:** 1111-1119

Welters P, Metz B, Palme K, Szczyglowski K, de Bruijn FJ (1993) Interaction of a rhizobial DNA-binding protein with the promoter of a plant leghemoglobin gene. Plant Physiol **102:** 1095-1107

Winkler RG, Helentjaris T (1995) The maize Dwarf3 gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis. Plant Cell **7:** 1307-1317
Wu L, Ueda T, Messing J (1995) The formation of mRNA 3'-ends in plants. Plant J **8:** 323-329

Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, Spee J, van Kammen A, Bisseling T, Franssen H (1993) Characterization of the GmENOD40, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J **3:**573-585

Zeidler R, Hobert O, Johannes L, Faulhammer H, Krauss G (1993) Characterization of two novel single stranded DNA-specific autonomously replicating sequence-binding proteins from *Saccharomyces cerevisiae*, one of which is adenylosuccinate synthetase. J Biol Chem **268:** 20191-20197

CHAPTER 3

A protein phosphatase 2C gene, *LjNPP2C1*, from *Lotus japonicus* induced during root nodule development[2]

3.1. ABSTRACT

Symbiotic interactions between legumes and compatible strains of rhizobia result in root nodule formation. This new plant organ provides the unique physiological environment required for symbiotic nitrogen fixation by the bacterial endosymbiont and assimilation of this nitrogen by the plant partner. We have isolated two related genes (*LjNPP2C1* and *LjPP2C2*) from the model legume *Lotus japonicus*, encoding protein phosphatases type 2C (PP2C). Expression of the *LjNPP2C1* gene was found to be specifically enhanced in *L. japonicus* nodules, while the *LjPP2C2* gene was expressed at a similar level in nodules and roots. A GST-LjNPP2C1 fusion protein was shown to have a Mg^{2+} or Mn^{2+} dependent, and okadaic acid insensitive PP2C activity *in vitro*. A chimeric construct containing the full length *LjNPP2C1* cDNA, under the control of the *Saccharomyces cerevisiae* alcohol dehydrogenase promoter, was found to be able to complement a yeast PP2C deficient mutant (*pct1Δ*) The transcript level of the *LjNPP2C1* gene was found to increase significantly in mature nodules, and its highest expression level

---

[2] This chapter was published in Kapranov P, Jensen TJ, Poulsen C, de Bruijn FJ, Szczyglowski K (1999) Proc. Natl. Acad. Sci. USA 96: 1738-1743.

occurred after leghemoglobin (*lb*) gene induction, a molecular marker for late developmental events in nodule organogenesis. Expression of the *LjNPP2C1* gene was found to be drastically altered in specific *L. japonicus* lines carrying monogenic recessive mutations in symbiosis-related loci, suggesting that the product of the *LjNPP2C1* gene may function at both early and late stages of nodule development.

3.2. INTRODUCTION

Interaction of legume plants with bacteria belonging to the genera *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium*, *Mesorhizobium*, and *Azorhizobium*, triggers a developmental program leading to the formation of a highly specialized organ, the nodule. Nodules provide a unique environment for symbiotic fixation and assimilation of dinitrogen. The early events during nodule organogenesis are set in motion by molecular cross-talk between the two symbiotic partners culminating in the synthesis of bacterially encoded lipochito-oligosaccharide signal molecules, the Nod factors (for recent reviews see 1, 2). Morphogenic activity of the Nod factor induces diverse cellular responses in the root of a compatible host plant, resulting in the redirection of a subset of the root cortical cells towards the formation of a nodule primordium (2). Invasion and subsequent intracellular colonization of the nodule primordium by the microsymbiont initiate the final differentiation process of plant and bacterial cells, which leads to the formation of a functional nitrogen-fixing nodule (3, 4). The induction of a number of plant host genes, named nodulins (5, 6), accompanies nodule organogenesis and functioning. We have been interested in studying the complexity and regulation of plant genes expressed during relatively late stages of the symbiotic interaction (late nodulin genes; 7, 8). The components of the signal transduction pathways involved in the late developmental events of nodule morphogenesis and functioning remain mostly uncharacterized, although nodule-specific MADS-box genes have been identified (9, 10). In order to identify novel genes with potential regulatory functions during late stages of nodule organogenesis and/or

nodule functioning, we have constructed a *Lotus japonicus* late nodulin expressed sequence tag (EST) library using the mRNA differential display procedure (8). One of the ESTs identified during our initial screening, *LjN3*, showed strong similarity to protein phosphatase 2C genes from different organisms (8). Here we describe the molecular characterization of two homologous *L. japonicus* genes, *LjNPP2C1* and *LjPP2C2*. The *LjNPP2C1* gene corresponds to EST *LjN3* and was found to encode a functional protein phosphatase 2C (LjNPP2C1). During *L. japonicus* nodule development, the expression of the *LjNPP2C1* gene was found to correlate with relatively late stages of nodule organogenesis. In contrast, the *LjPP2C2* gene, a member of a small family of related genes in the *L. japonicus* genome, was found to be expressed at a similar level in nodules and uninfected roots. Therefore, we postulate that the *LjNPP2C1* gene product may play a specific role in a signaling cascade at a late stage of nodule development or functioning.

In addition, two experimental approaches, such as generation of transgenic *L. japonicus* plants and a yeast 2-hybrid screen, were used to test this hypothesis and to further address the biological functions of LjNPP2C1. The results of these experiments are described in the following chapter.

3.3. MATERIALS AND METHODS

Plant Material. The *Lotus japonicus* GIFU B-129-S9 wild-type line and mutant lines derived thereof were grown as previously described (11, 12). *Mesorhizobium loti* wild-type strain NZP2235 was used for all nodulation experiments (12). Uninoculated control roots were harvested 7,11, 21, and 35 days after sowing. Roots of wild-type plants inoculated with *M. loti* NZP2235 were harvested 7 and 11 days after inoculation (dai), and fully developed nodules were collected 21 and 35 dai. Inoculated roots, and nodule-like structures from mutant lines were harvested 35 dai. Leaves and stems were harvested from 21 day old plants, whereas flowers were obtained from mature *L. japonicus* plants.

Northern Blot Analysis. Northern blot hybridization was carried out in 0.5 M phosphate buffer (pH 7.2) containing 7% SDS and 1% BSA at 65°C, essentially as described (8, 11). Filters were washed either at high stringency: 25 min in 2X SSC, 0.1% SDS; 25 min in 1X SSC, 0.1% SDS at 65°C and 15 min in 0.1X SSC, 0.1% SDS at 65°C. Poly(A)$^+$ mRNA was isolated using the PolyATtract system (Promega) following the manufacturer instructions. Signals from northern blot assays were quantified using PhosphorImager analysis (Molecular Dynamics) and were standardized to ubiquitin RNA levels.

cDNA Library Screening and 5' RACE Analysis. A ^{32}P-radiolabeled *LjN3* cDNA fragment and a partial cDNA corresponding to the *LjPP2C2* gene were used to screen a *L. japonicus* nodule-specific cDNA library (kind gift from Dr. Jens Stougaard, Aarhus University, Denmark), following standard protocols (13). The 5' RACE procedure (14) was performed using the 5' RACE system (Gibco/BRL) following the manufacturer's instructions. For the reverse transcription reaction 1 μg of poly(A)$^+$ mRNA derived from *L. japonicus* nodules and a *LjNPP2C1* gene specific primer (5'-CAACCCCGTCACGGTCTCCTC-3') were used.

***In vitro* Protein Phosphatase Assay.** A portion of the *LjNPP2C1* cDNA, corresponding to amino acids 96-362 of the LjNPP2C1 protein, was inserted into the pGEX-5X-1 vector (Pharmacia) to express a glutathione S-transferase (GST)-LjNPP2C1 fusion protein in *Escherichia coli*. Affinity purification of the resulting fusion protein was carried out using glutathione-agarose beads (Sigma, product G-4510), following the procedure described by Smith and Johnson (15). Purified protein was aliquoted and stored in 10% glycerol at -20°C.

Protein phosphatase activity was assayed using phosphorylated casein as a substrate, essentially as described by McGowan and Cohen (16). Briefly, 5 mg of dephosphorylated bovine casein (Sigma, product C-4765) was phosphorylated using 50 μCi of [γ-^{32}P]-ATP (3000 Ci/mmol) and 25 units of the catalytic subunit of cAMP-dependent protein kinase (Sigma, product P-2645) for 4 hr at 30°C, as described (16). The

radiolabelled casein was precipitated with 15% trichloroacetic acid (TCA), washed three times with cold 20% TCA, twice with cold acetone, air-dried, dissolved in 0.2 M Tris-HCl (pH 8.0) and stored at 4°C. The purified GST-LjNPP2C1 fusion protein was assayed for phosphatase activity in a time-course experiment in 50 µl of incubation mixture consisting of 50 mM Tris-HCl (pH 7.0), 0.1 mM EGTA, 0.1% 2-mercaptoethanol, 20 mM magnesium acetate, 1 µM okadaic acid, ~ 0.2 µg of the fusion protein, and $2x10^4$ cpm (Cerenkov counts) of ^{32}P-labelled casein at 30°C. In the control reactions the fusion protein was replaced with 1-2 µg of purified GST protein. Divalent cation dependence was analyzed using similar conditions, except that magnesium acetate was substituted with either 20 mM $MnSO_4$, 20 mM $CaCl_2$, or 10 mM EDTA. All reactions were terminated by the addition of 50 µl of cold 30% TCA. After 5 min of incubation on ice, the samples were centrifuged for 10 min at 14,000 g, and the radioactivity in the supernatant and pellet was measured by Cerenkov counting. All assays were performed in triplicate.

Site-Directed Mutagenesis of the *LjNPP2C1* cDNA. An *Arabidopsis abi1-1*-type amino acid substitution (see Results section) was created by site-directed mutagenesis of the *LjNPP2C1* cDNA following the method described by Kunkel (17). An oligonucleotide (5'-TTTGACGGTCATGACGGAGCTCAGG-3'), containing the *abi1-1* type mutation (G→A), was used to change the Gly133 residue to Asp. In addition, a silent C→T substitution was introduced two base-pairs upstream from the G→A mutation in order to

create a unique *Bsp*HI restriction site. The presence of the base pair substitutions was confirmed by DNA sequence analysis.

Functional Complementation Assay in *S. cerevisiae*. The *LjNPP2C1* cDNA fragment was cloned into the yeast expression vector pDBL2 in the sense (construct pDBL3) and antisense (construct pDBL5) orientation, with respect to the yeast alcohol-dehydrogenase promoter ADH1 (18). In addition, a mutant *LjNPP2C1* construct was prepared by inserting the *LjNPP2C1* cDNA, containing the *abi1-1* type substitution, into the pDBL2 vector in the sense orientation to generate plasmid pDBL3M.

Plasmids were introduced into yeast strain TM126 (19), carrying a disruption in the *PTC1* locus encoding yeast protein phosphatase 2C, using a modified lithium acetate/PEG method (20). The pDBC1 construct, containing the wild-type yeast *ptc1* gene under the control of the ADH1 promoter (19), as well as the pDBL2 vector alone, were also introduced into strain TM126 and used as positive and negative controls, respectively. Complementation assays were performed essentially as described by Bertauche *et al.* (21). Each transformant was grown to saturation at 28°C in liquid synthetic SD medium (22), lacking uracil and leucine. For the complementation assay, 3 µl of the yeast suspension culture, containing approximately 10^4 cells/µl, were replica plated onto YPD media (22). Replica plates were incubated at either 28°C (permissive-), or at 37°C (non-permissive) temperatures, for 30-36 hours. Subsequently, the phenotype of the individual strains was carefully analyzed to examine the ability of the recombinant

plasmids to complement the temperature-sensitive growth phenotype of strain TM126 (19).

3.4. RESULTS

The *L. japonicus LjNPP2C1* Gene Shares Significant Similarity with Protein Phosphatase 2C Genes. Using the mRNA differential display procedure (23), we previously isolated a range of *L. japonicus* ESTs associated with relatively late stages of nodule development and/or nodule functioning (8). One of these ESTs, *LjN3*, was found to share significant similarity with a number of protein phosphatase 2C genes from different organisms (8). Subsequent screening of a nodule specific cDNA library, using radiolabelled *LjN3* cDNA as a probe, facilitated the cloning of the corresponding full length cDNA. The nucleotide sequence of the longest cDNA clone (*LjNPP2C1-16*) was shown to be 1235 bp in length, and was found to contain a 1086 bp long open reading frame (ORF), starting with an ATG codon at base pair 25. Since the ATG_{25} codon was not preceded by an in-frame stop codon, the 5' Rapid Amplification of cDNA Ends (RACE) procedure was used to further characterize the 5'-terminal sequence of the *LjNPP2C1* mRNA. The sequence of the longest 5'-RACE cDNA product extended the 5' end of the *LjNPP2C1-16* cDNA by 83 bp (data not shown). However, no additional in-frame ATG codons were found in the 5' sequence. Moreover, the length of the *LjNPP2C1* cDNA correlates well with the observed mRNA size on northern blots (~1300 nt). Based on these observations, we propose that the ATG_{25} of *LjNPP2C1-16* cDNA (*(*ATG_{108} in the extended sequence) represents the initiation codon of the *LjNPP2C1* gene, and that the *LjNPP2C1-16* cDNA contains the entire coding region for the LjNPP2C1 protein.

The amino acid sequence of the deduced LjNPP2C1 protein was found to correspond to a polypeptide of 362 residues and a predicted molecular mass of ~ 39.5 kD. In agreement with our earlier observations (8), the deduced protein was found to share a high level of similarity with a number of PP2C proteins from different eukaryotes, including the ABI1 and ABI2 proteins from *Arabidopsis thaliana* (Fig. 3.1).

The *L. japonicus LjPP2C2* gene. In the course of an independent research project, a distinct partial cDNA clone, encoding a putative *L. japonicus* protein phosphatase type 2C, was fortuitously identified (data not shown). A corresponding full copy cDNA clone (*LjPP2C2*) was isolated from the *L. japonicus* cDNA library and was found to encode a polypeptide (LjPP2C2) of 282 residues and a predicted molecular mass of ~ 30.8 kD. Similar to the LjNPP2C1 protein, the predicted amino-acid sequence of the LjPP2C2 was found to share a high level of similarity with a number of PP2C proteins from different eukaryotes, including *L. japonicus* LjNPP2C1, and *A. thaliana* ABI1 (Fig. 3.1). Based on these results, we conclude that the *LjNPP2C1* and *LjPP2C2* genes are likely to encode two similar, but distinct *L. japonicus* protein phosphatases type 2C.

***LjNPP2C1*, but not *LjPP2C2* transcripts, accumulate preferentialy in *L. japonicus* nodules.** We previously reported that *LjNPP2C1* mRNA levels were enhanced in *L. japonicus* nodules versus uninoculated control roots (8). In order to address the question whether the *LjPP2C2* gene was also preferentialy expressed in nodules, a northern blot analysis was performed. Poly(A)$^+$ mRNA derived from uninfected *L. japonicus* control

roots and nodules was hybridized sequentially with [α-^{32}P]-dATP labeled *LjNPP2C1*, *LjPP2C2*, and *Sesbania rostrata* ubiquitin cDNA probes (Fig. 3.2). A significantly higher level of *LjNPP2C1* mRNA was detected in nodules than in uninfected *L. japonicus* control roots, confirming and extending our earlier observations (8). In contrast, the level of *LjPP2C2* mRNA was found to be equal in both tissues. Based on these results, we conclude that the *LjNPP2C1* is likely to have a unique function(s) associated with nitrogen fixing nodules. Therefore, our further analysis concentrated on a more detailed characterization of the *LjNPP2C1* gene.

The *LjNPP2C1* gene encodes a functional protein phosphatase type 2C. To investigate whether the *LjNPP2C1* gene encoded a functional protein phosphatase 2C, an *in vitro* phosphatase assay was carried out. For this purpose, a recombinant protein was created containing the C-terminal catalytic domain of LjNPP2C1 fused to the glutathione S-transferase (GST) protein. The GST-LjNPP2C1 fusion protein was expressed in *Escherichia coli*, purified, and found to display a time-dependent phosphatase activity using phosphorylated casein as a substrate (Fig. 3.3A). The phosphatase activity of the GST-LjNPP2C1 fusion protein was found to be insensitive to 1 μM okadaic acid, a potent inhibitor of protein phosphatases other than type 2C (24), and to be dependent on the presence of the divalent cations Mg^{2+} or Mn^{2+} (Fig. 3.3B). The GST protein alone, expressed and purified under similar conditions, was found to be inactive in the phosphatase assay (Fig. 3.3B).

The biochemical evidence for protein phosphatase 2C activity of the *LjNPP2C1* gene product was further supported by heterologous genetic complementation experiments in yeast. *Saccharomyces cerevisiae* strain TM126 carries a disrupted PP2C gene (*ptc1Δ)*, resulting in a temperature sensitive growth phenotype. Cells of the *ptc1Δ* mutant strain grow slower at 37°C than at 28°C (19). The *LjNPP2C1* cDNA, fused to the yeast ADH1 promoter was introduced into *S. cerevisiae* strain TM126 on plasmid pDBL3 and was indeed found to complement the yeast *ptc1* mutant phenotype (Fig. 3.4). A plasmid carrying the wild-type *LjNPP2C1* cDNA in the antisense orientation (pDBL5), failed to complement the temperature sensitive growth phenotype of *ptc1Δ* yeast cells (Fig. 3.4).

In order to provide further evidence for the designation of the *LjNPP2C1* gene product as a protein phosphatase 2C, we made use of the phenotype of an *Arabidopsis thaliana* mutant PP2C protein. The *A. thaliana abi1-1* mutation is caused by a single base-pair transition that substitutes amino-acid residue Gly180 with an Asp residue, resulting in a significant decrease of PP2C activity *in vitro*, as well as the generation of a dominant negative mutation leading to abscisic acid insensitivity (21, 25, 26). An *abi1-1*-type mutation (Gly133→Asp), was created in the wild-type *LjNPP2C1* gene by site directed mutagenesis. The resultant mutant LjNPP2C1 protein was indeed found to be unable to complement the *ptc1Δ* mutant phenotype of yeast strain TM126 (Fig. 3.4).

***LjNPP2C1* mRNA Accumulates During Relatively Late Stages of Nodule Development**. To correlate the expression pattern of the *LjNPP2C1* gene with (a) specific phase(s) in nodule development, a time course northern hybridization analysis was performed. *LjNPP2C1* mRNA was found to be present at a low level in uninoculated *L. japonicus* roots, and to be slightly increased at 7 and 11 days after inoculation with *M. loti*. However a ~ 20 fold increase of *LjNPP2C1* mRNA was observed in nodules harvested 21 days after inoculation (Fig. 3.5). Interestingly, the most dramatic enhancement of *LjNPP2C1* mRNA level was found to occur after the induction of *L. japonicus* leghemoglobin (*lb*) gene expression (a molecular marker for late stages of nodule development; 7), which was found to occur in nodules 11 days after inoculation (Fig. 3.5; see also 12).

Tissue-specific expression of the *LjNPP2C1* gene. To gain further insight into the tissue specificity of *LjNPP2C1* gene expression, poly(A)$^+$ mRNA isolated from different *L. japonicus* tissues was probed with a radiolabeled *LjNPP2C1* cDNA. *LjNPP2C1* mRNA was found to accumulate to the highest level in nodules and flowers (Fig. 3.6). A low level of *LjNPP2C1* mRNA could also be found in uninfected root and leaf tissues (Fig. 3.6), whereas no hybridization signal, even upon prolonged exposure period, was detected in the *L. japonicus* stems (data not shown). Therefore, we conclude that *LjNPP2C1* gene expression is not nodule-specific, but that expression of this gene is substantially enhanced in fully developed nodules and flowers.

***LjNPP2C1* Gene Expression is Altered in Symbiotic mutants of *L. japonicus*.** The expression pattern of the *LjNPP2C1* gene was further analyzed by examining six non-allelic *L. japonicus* symbiotic mutants. Four non-nodulating lines (Nod^-; LjEMS34, LjEMS46, LjEMS70, and LjEMS76), which fail to display macroscopically visible signs of nodulation, and two distinct mutant lines (LjEMS88 and LjEMS75), forming white, mostly ineffective ($Nod^+Fix^{-/+}$) nodules (12), were examined for *LjNPP2C1* expression.

Mutant lines LjEMS88 and LjEMS75 differ with respect to their ability to support late stages of nodule development. Mutant line LjEMS88 forms only small white nodule-like structure which do not fix nitrogen, and accumulate only background levels of leghemoglobin mRNA (≤50 fold less than wild type level; 12). In contrast, the LjEMS75 line produces a mixed population of well developed white nodules, in addition to pink wild-type-like nodules, that fix nitrogen, albeit at a significantly diminished level (12). Both types of the LjEMS75 nodules were harvested and analyzed separately. White nodules of line LjEMS75 were found to contain a significant level of *L. japonicus lb* mRNA (only ~ 3 fold lower than the wild type; Fig. 3.7A), which clearly distinguishes them from white nodules formed on line LjEMS88 (see above).

35-days-old wild-type nodules, and pink nodules of mutant line LjEMS75 showed similar, 15-20 times elevated, levels of *LjNPP2C1* mRNA, as compared with uninfected control roots. White nodules of LjEMS75 were found to have a two-fold lower level of *LjNPP2C1* transcripts than wild-type nodules, but an approximately 7 times higher transcript level than uninfected control roots (Fig. 3.7A). Analysis of *LjNPP2C1*

gene expression in white nodule-like structures formed on the LjEMS88 mutant line showed only a background level of the corressponding mRNA (Fig. 3.7B). A background level of *LjNPP2C1* mRNA was also detected in wild-type uninfected roots and infected roots of lines LjEMS70 and LjEMS76. In contrast, the infected roots of the non-nodulating mutant line LjEMS34 showed a relatively modest increase (two times higher than the wild-type uninoculated roots) of *LjNPP2C1* transcript levels, whereas LjEMS46 was found to display a significant levels of *LjNPP2C1* mRNA: ~50% of that present in nitrogen fixing nodules.

3.5. DISCUSSION

Reversible phosphorylation of proteins has been implicated in the regulation of cellular processes as diverse as metabolism, transcription and translation, cell division, membrane transport and secretion, stress response, fertilization, and memory (27). The role of protein kinases in controlling the level of phosphorylation of proteins has been well documented (28). The finding that the activity of protein phosphatases is regulated in specific cases, has led to the hypothesis that cellular responses to external stimuli might result from direct activation or inhibition of protein phosphatases, rather then through the action of protein kinases (24, 28). So far, three out of the four known major classes of Ser/Thr protein phosphatases have been described in plants (PP1, PP2A and PP2C), while additional protein phosphatases unrelated to these major groups are thought to exist (24). In general, PP2C is the least well-characterized class of protein phosphatases. However, recent studies have suggested an intriguing connection between PP2C enzyme activity and several cellular processes in different eukaryotes (for recent reviews see 24, 28, 29). Plant members of this group of protein phosphatases include four different PP2Cs from *A. thaliana*: KAPP (30), AtPP2C (31), ABI1 (25, 26), ABI2 (32), and a protein phosphatase 2C (MP2C) from alfalfa (33).

Here we report the identification and characterization of a plant gene, *LjNPP2C1*, encoding a protein with amino acid similarity to protein phosphatases type 2C, the expression of which is significantly enhanced during *L. japonicus* root nodule organogenesis. The following biochemical and genetic experiments have led us to the

conclusion that the *LjNPP2C1* gene encodes a functional protein phosphatase type 2C. 1. A recombinant GST-LjNPP2C1 protein is capable of dephosphorylating phosphorylated casein, a commonly used artificial substrate for measuring PP2C activity (16). The activity of the LjNPP2C1 catalytic domain present in the GST-LjNPP2C1 fusion is insensitive to inhibition by okadaic acid, and dependent on the presence of divalent cations (Mg^{2+} or Mn^{2+}). 2. The *LjNPP2C1* gene is able to complement a yeast PP2C deficient mutant (*pct1Δ*.) 3. The *LjNPP2C1* cDNA, carrying an *abi1* type single amino-acid substitution (Gly→Asp), is unable to complement the temperature sensitive phenotype of yeast strain TM126, thereby mimicking the behavior of the *A. thaliana* ABI1 mutant protein used in similar complementation experiments (21). We also describe a second gene, *LjPP2C2,* which encodes a protein sharing significant amino-acid similarity with LjNPP2C1, as well as other PP2C proteins from different eukaryotes. Unlike in the case of the *LjNPP2C1* gene, the *LjPP2C2* transcript appears to accumulate to a similar level in both uninoculated *L. japonicus* roots and nodules, suggesting that two different regulatory mechanisms are involved in the expression of these otherwise similar genes. ***LjNPP2C1*** gene is represented by a single- or low-copy gene number in *L. japonicus* genome, while *LjPP2C2* gene is a member of a small family of related genes, as evidenced by Southern blot hybridization (data not shown).

LjNPP2C1 appears to be the only plant protein phosphatase 2C reported thus far, whose expression is enhanced during plant-microbe interactions in general, and symbiotic root nodule formation in particular. Moreover, the temporal expression pattern

of the *LjNPP2C1* gene during rhizobial infection and nodule formation appears to be unusual in comparison to that of other nodulin genes. Only a limited level of *LjNPP2C1* mRNA accumulation is observed at relatively early stages of infection and nodule morphogenesis. In contrast, a 20-fold higher level of *LjNPP2C1* transcripts, as compared with the control uninoculated roots, is observed in 21 day-old nodules. The prominent increase in *LjNPP2C1* mRNA levels observed is preceded by the developmental transition between nodule ontogeny and the establishment of a functional, nitrogen fixing nodule, as indicated by the expression pattern of the *L. japonicus* leghemoglobin gene, which serves as a molecular marker for this transition (Fig. 3.5; see also 12). This observation is supported by results obtained from studies of *LjNPP2C1* gene expression in symbiotic mutants of *L. japonicus*. The highest levels of *LjNPP2C1* transcripts were found in nodules, or nodule-like structure, already containing a substantial level of leghemoglobin mRNA. For example, the white-, and pink nodules formed on mutant line LjEMS75 accumulate a high level of both *lb* and *LjNPP2C1* gene transcripts, whereas only a low level of these mRNA species are found in white, ineffective nodule-like structures formed on line LjEMS88 (Fig. 3.7; see also 12).

On the basis of these results, it appears that the LjNPP2C1 protein has a specific function(s) in nodules that are in the process of initiating nitrogen fixation, and have already passed the developmental time-point at which expression of a late nodulin molecular marker gene *lb* is activated. The intriguing observation that the roots of two particular non-nodulating *L. japonicus* mutant lines, LjEMS34 and LjEMS46, contain elevated level of *LjNPP2C1* mRNA is difficult to interpret. However, it is conceivable

that the enhanced expression of *LjNPP2C1* gene in these mutant lines is somehow related to their Nod$^-$ phenotypes: it might at least partially be responsible for the observed phenotypes, especially in the case of LjEMS46 mutant line. It is possible that inappropriate timing or level of *LjNPP2C1* induction may have quite different phenotypic effects at the early vs late stages of nodule development. At present we are investigating whether the LjEMS46 mutant allele maps at or near the *LjNPP2C1* locus. Further analysis of *LjNPP2C1* gene expression, the phenotype(s) of transgenic plants (over)expressing the *LjNPP2C1* transcript in sense- and antisense orientation, and characterization of the relevant *in vivo* substrate(s) of LjNPP2C1 will be essential to resolve the function of the *LjNPP2C1* gene during nodulation or symbiotic nitrogen fixation.

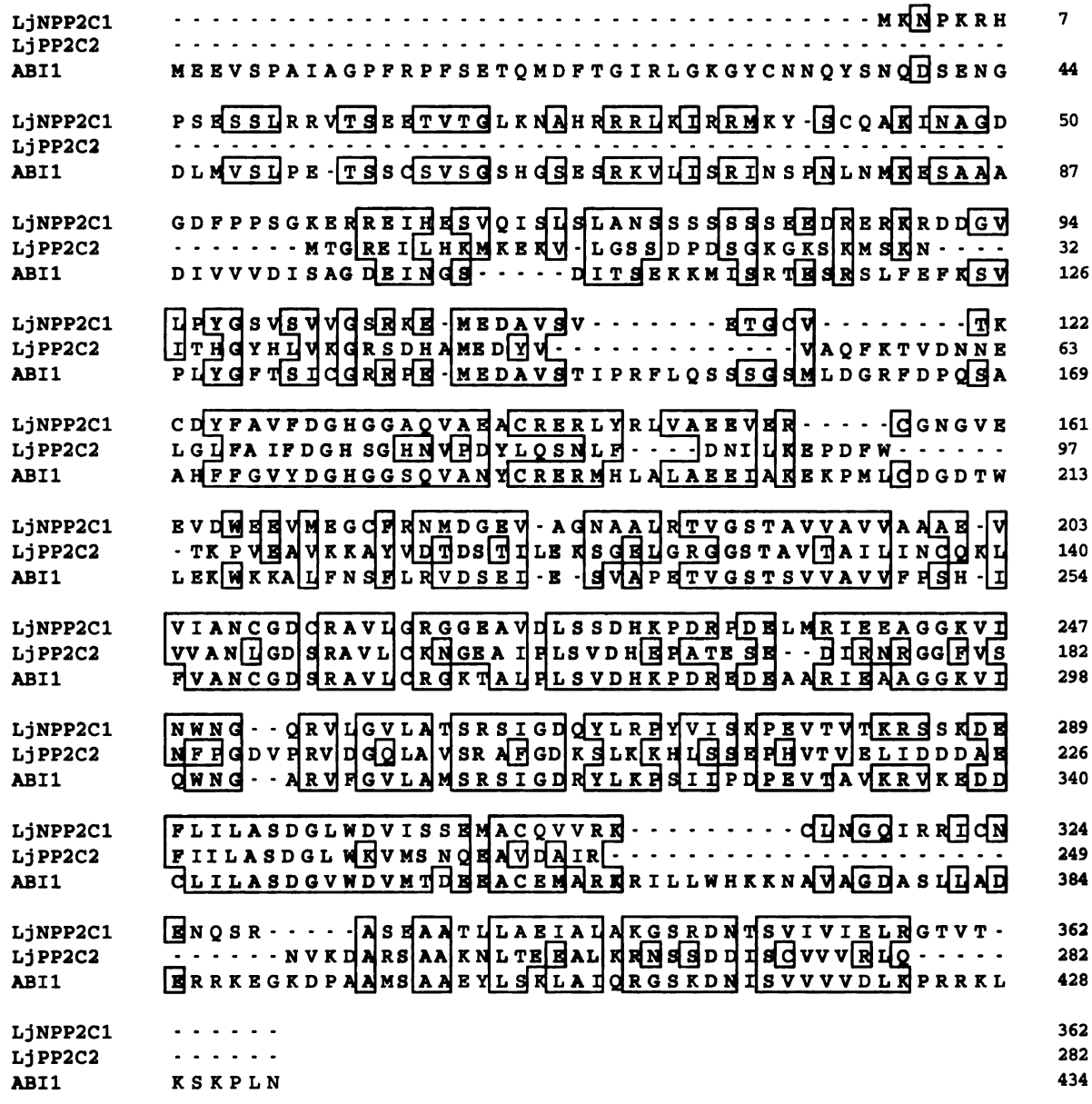


Figure 3.1. Amino acid sequence alignment of *L. japonicus* LjNPP2C1 and LjPP2C2 and *A. thaliana* ABI1 (25, 26) proteins. The alignment was performed using the PileUp algorithm from the GCG software package (Genetics Computer Group, Madison, WI). Identical amino acids are shaded and conservative substitutions are boxed using the SeqVu 1.1 program. Gaps introduced to allow an optimal alignment are represented by dashes.

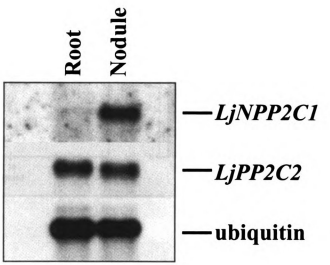


Figure 3.2. Northern blot analysis of *LjNPP2C1* and *LjPP2C2* expression. Two micrograms of poly $(A)^+$ mRNA from uninfected roots and mature nodules were separated under denaturing conditions, and hybridized sequentially with α-^{32}P-labelled probes corresponding to the *L. japonicus LjNPP2C1*, *LjPP2C2*, and *Sesbania rostrata* ubiquitin cDNAs.

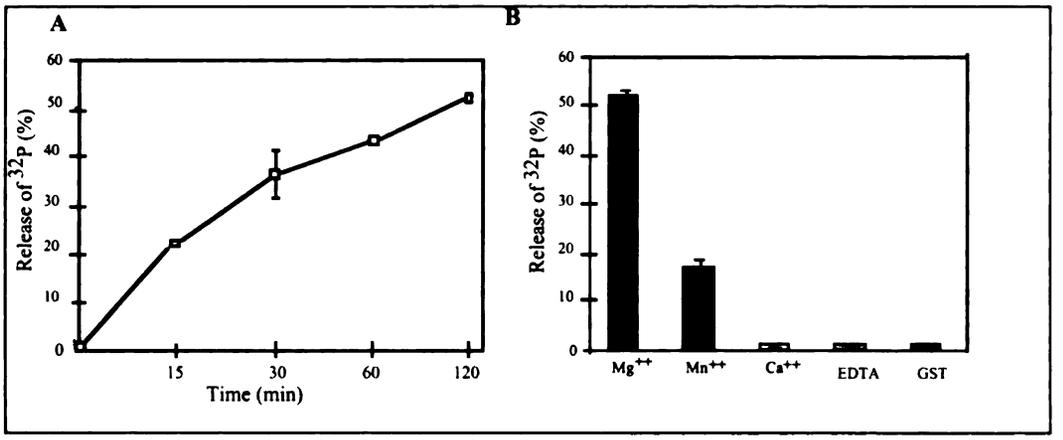
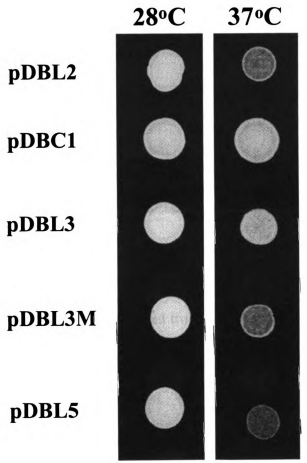


Figure 3.3. Protein phosphatase type 2C activity of a GST-LjNPP2C1 protein. (**A**) Experiments showing a time-dependent dephosphorylation of [^{32}P]-labelled casein by a recombinant GST-LjNPP2C1 protein in the presence of 20 mM magnesium acetate and 1 μM okadaic acid. (**B**) Divalent cation requirement for the protein phosphatase activity of LjNPP2C1. The recombinant GST-LjNPP2C1 protein was assayed for its PP2C activity in the presence of Mg^{2+}, Mn^{2+} or Ca^{2+}, or in the absence of a divalent cation and in the presence of EDTA. "GST" denotes the assay performed in the presence of the GST protein alone.

Figure 3.4. Complementation of the PP2C deficient yeast mutant strain TM126 (*ptc1Δ*). For each strain, 3 µl of cell suspension of each yeast strain, containing approximately $3x10^4$ cells, was spotted onto agar plates (see Material and Methods) and incubated at either 28°C or 37°C for two days. Strain TM126 was transformed with the yeast expression vector pDBL2 as a control, plasmid pDBC1 containing the wild-type yeast *PTC1* gene (19), plasmids pDBL3 and pDBL3M containing either wild-type, or mutant *L. japonicus LjNPP2C1* genes, and plasmid pDBL5 containing the *LjNPP2C1* gene in the anti-sense orientation, with respect to the yeast ADH promoter. The poor growth phenotype of strain TM126 at the non-permissive temperature (37°C) is evidenced by a less dense (partially translucent) spot; complementation of the temperature sensitive growth phenotype is evidenced by a denser (non-translucent) spot.


28°C
37°C
pDBL2
pDBC1
pDBL3
pDBL3M
pDBL5

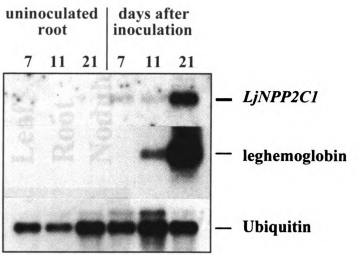


Figure 3.5. Northern blot analysis of developemntal *LjNPP2C1* gene expression. Two micrograms of poly(A)$^+$ mRNA isolated from uninoculated roots, harvested 7, 11 and 21 days after sowing, and roots harvested from *L. japonicus* plants 7 and 11 dai, as well as 21 days old nodules, were analysed. The blot was sequentially probed with the radiolabelled *LjNPP2C1* cDNA, the *LjN77* EST corresponding to a *L. japonicus* leghemoglobin gene (8), and the *S. rostrata* ubiqutin cDNA as a loading control.

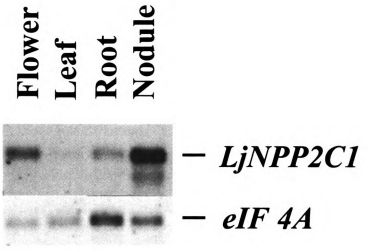
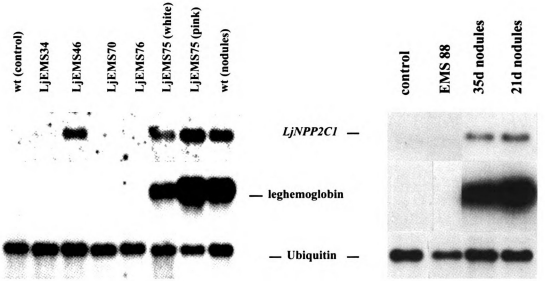


Figure 3.6. Northern blot analysis of *LjNPP2C1* expression in different tissues of *L. japonicus*. Four micrograms of poly(A)$^+$ RNA isolated from flowers, leaves, uninoculated roots and mature nodules of *L. japonicus* were separated under denaturing conditions, and probed with radiolabeled probes corresponding to the *L. japonicus LjNPP2C1* and *Arabidopsis eIF4A* cDNAs.

Figure 3.7. Comparison of *LjNPP2C1* transcript levels in: (**A**) wild-type uninfected roots (control), inoculated roots of non-nodulating mutant lines LjEMS34, LjEMS46, LjEMS70, and LjEMS76, white (LjEMS75 white) and pink (LjEMS75 pink) nodules of mutant line LjEMS75, and 35 day-old wild-type nodules. **(B)** wild-type uninfected roots (control), white nodule-like structures induced on the mutant line LjEMS88, and 35- and 21-days-old wild-type nodules. The blots shown in panels A and B were sequentially hybridized with [α-^{32}P]dATP labeled probes corresponding to the *LjNPP2C1* cDNA, the *L. japonicus* leghemoglobin gene (EST *LjN77*; 8) and a *S. rostrata* ubiquitin cDNA, as loading control.


wt (control)
LjEMS34
LjEMS46
LjEMS70
LjEMS76
LjEMS75 (white)
LjEMS75 (pink)
wt (nodules)
LjNPP2C1
leghemoglobin
Ubiquitin
control
EMS 88
35d nodules
21d nodules

3.6. ACKNOWLEDGMENTS

We are grateful to Dr. Haruo Saito (Harvard University) for kindly providing the yeast strain TM126, vectors and advice on the yeast complementation experiments. We wish to thank Kurt Stepnitz for expert photographic assistance. This work was supported by grants from the Department of Energy (DE-FG02-91ER20021), the National Science Foundation (NSF-09630189) and the U.S. Department of Agriculture.

3.7. LITERATURE

1. Long, S. R. (1996) *Plant Cell* **8**, 1885-1898.
2. Spaink, H.P. (1996) *Crit. Rev. Plant Sci.* **15**, 559-582.
3. Verma D. P. S. (1992) *Plant Cell* **4**, 373-382.
4. Mylona, P., Pawlowski, K. & Bisseling, T. (1995) *Plant Cell* **7**, 869-885.
5. Legocki, R.P. & Verma, D.P.S. (1980) *Cell* **20**, 153-163.
6. van Kammen, A. (1984) *Plant Mol. Biol. Rep.* **2**, 43-45.
7. de Bruijn, F. J. & Schell, J. (1992) in *Control of Plant Gene Expression*, ed. Verma D. P. S. (CRC Press, Boca Raton), pp. 241-258.
8. Szczyglowski, K., Hamburger, D., Kapranov, P. & de Bruijn, F. J. (1997) *Plant Physiol.* **114**, 1335-1346
9. Heard, J., & Dunn, K. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 5273-5277.
10. Heard, J., Caspi, M. & Dunn, K. (1997) *Mol. Plant-Microbe Interact.* **10**, 665-676.
11. Kapranov, P., de Bruijn, F. J. & Szczyglowski, K. (1997) *Plant Physiol.* **113**, 1081-1090.
12. Szczyglowski, K., Shaw, R.S., Wopereis, J., Copeland, S., Hamburger, D., Kasiborski, B., Dazzo, F. & de Bruijn, F.J. (1998) *Mol. Plant-Microbe Interact.* **11**, 684-697.
13. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.
15. Smith, D.B. & Johnson, K.S. (1988) *Gene* **67**, 31-40.

16. McGowan, C.H. & Cohen, P. (1988) *Methods Enzymol.* **159**, 416-426.

17. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 488-492.

18. Milne, G. T. & Weaver, D. T. (1993) *Genes and Dev.* **7**, 1755-1765.

19. Maeda, T., Tsai, A. Y. & Saito, H. (1993) *Mol. Cell. Biol.* **13**, 5408-5417.

20. Elble, R. (1992) *Biotechniques* **13**, 18-20.

21. Bertauche, N., Leung, J. & Giraudat, J. (1996) *Eur. J. Biochem.* **241**, 193-200.

22. Sherman, F., Fink, G. R. & Hicks, J. B. (1983) *Methods in Yeast Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

23. Liang, P. & Pardee, A. (1992) *Science* **257**, 967-971.

24. Smith R. D. & Walker, J. C. (1996) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 101-125.

25. Leung, J., Bouvier-Durand, M., Morris, P.-C., Guerrier, D., Chefdor, F. & Giraudat, J. (1994) *Science* **264**, 1448-1452.

26. Meyer, K., Leube, M.P. & Grill, E. (1994) *Science* **264**, 1452-1455.

27. Cohen, P. (1989) *Annu. Rev. Biochem.* **58**, 453-508.

28. Hunter, T. (1995) *Cell* **80**, 225-236.

29. Luan S. (1998) *Trends Plant Sci.* **3**, 271-275.

30. Stone, J.M., Collinge, M.A., Smith, R.D., Horn, M.A., & Walker, J.C. (1994) *Science* **266**, 793-795.

31. Kuromori, T., & Yamamoto, M. (1994) *Nucl. Acids Res.* **22**, 5296-5301.

32. Leung, J., Merlot, S. & Giraudat, J. (1997) *Plant Cell* **9**, 759-771.

33. Meskiene, I., Bögre, L., Glaser, W., Balog, J., Brandstötter, M., Zwerger, K., Ammerer, G. & Hirt, H. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1938-1943.

CHAPTER 4

Analysis of the biological functions of LjNPP2C1

4.1. ABSTRACT

Protein phosphatases type 2C (PP2Cs) play key roles in regulating multiple biological processes in eukaryotes. In plants, PP2Cs have been implicated in regulating abscisic acid signaling, the stress induced MAP kinase pathway and signal transduction via receptor-like protein kinases. In addition, a novel PP2C from *L. japonicus*, LjNPP2C1, has been implicated in regulating the events taking place at late stages of nodule development. This chapter describes two independent approaches taken to further address the possible role and significance of LjNPP2C1 in *L. japonicus* nodule development and functioning, as well as in other aspects of plant physiology. First, transgenic *L. japonicus* plants were generated containing chimeric constructs designed to either increase or decrease endogenous levels of *LjNPP2C1* mRNA. In addition, transgenic plants expressing a dominant-negative isoform of LjNPP2C1 were generated. Second, a yeast two-hybrid screen was used to identify *L. japonicus* cDNAs encoding proteins interacting with LjNPP2C1. One such cDNA, *LBP*, was identified and characterized.

4.2. INTRODUCTION

Protein Phosphatases type 2C (PP2Cs) belong to the least well-characterized class of the protein serine/threonine phosphatases, sharing no apparent amino acid sequence similarity with other major classes of to protein phosphatases, such as PP1, PP2A and PP2B (Cohen, 1989; Cohen and Cohen, 1989). PP2Cs require Mg^{2+} or Mn^{2+} for enzyme activity and exhibit an insensitivity to okadaic acid, a potent inhibitor of protein phosphatases belonging to classes PP1 and PP2A (Cohen, 1989). PP2Cs are believed to function as monomeric proteins, and no additional regulatory subunits have been identified (Cohen, 1989; Smith and Walker, 1996). The distribution of PP2C enzymes in plants appears to be more restrictive than that of other protein phosphatases (Smith and Walker, 1996). The lack of genetic approaches and the absence of specific inhibitors have greatly hindered the analysis of PP2C functions. However, recent studies have suggested intriguing connections between PP2C enzymes and several cellular signal transduction systems. PP2Cs have been implicated in the regulation of mammalian AMP-activated protein kinases (Davis et al., 1995) and in male sexual development in the nematode *Caenorhabditis elegans* (Pilgrim et al., 1995; Chin-Sang and Spence, 1996). In *Saccharomyces cerevisiae*, the *PTC1* and *PTC3* genes, encoding PP2C enzymes, have been shown to be suppressors of the *sln1* mutation (Maeda et al., 1994). The *SLN1* gene encodes a transmembrane protein homologous to histidine kinases and the response regulator proteins of bacterial two-component systems, which have been identified in a wide variety of signaling pathways (Ota and Varshavsky, 1993). In the fission yeast

Sacharomyces pombe , analysis of the three *ptc* (phosphatase two C) genes (*ptc1*, *ptc2* and *ptc3*), has led to the hypothesis that they play important roles in the osmo-sensing MAP kinase signal transduction pathway (Shiozaki and Russell, 1995). Interestingly, the Ser protein phosphatase SpoIIE from *Bacillus subtilis*, which acts on SpoIIA triggering the specific activation of a sporulation transcription factor, has recently been reported to have sequence similarity to the PP2C family of eukaryotic Ser/Thr protein phosphatases (Adler et al., 1997).

Besides *LjNPP2C1*, several PP2C genes from from plants have been cloned and characterized (Rodriguez, 1998), including five from *Arabidopsis* (*PP2C-At*, *KAPP*, *ABI1*, *ABI2* and *AtP2C-HA*; Kuromori and Yamamoto, 1994; Stone et al., 1994; Leung et al.,1994; Leung et al.,1997; Meyer et al.,1994; Rodriguez et al., 1998) and one from *Medicago sativa*, *MP2C* (Meskiene et al., 1998). The common feature of plant PP2C proteins is presence of the N-terminal extensions of variable lengths in addition to the PP2C catalytic region (Rodriguez, 1998; Kapranov et al., 1999). The N-terminal regions of plant PP2Cs typically do not share any sequence similarity with each other or with any other protein in the databases and have not been found in any of the known fungal or mammalian PP2Cs, except for FEM-2 of *C. elegans* (Smith and Walker, 1996, Pilgrim et al., 1995).

The protein phosphatase *PP2C-At* gene was identified via a complementation screen of the *S. pombe pde1* mutant, which is defective in cAMP phosphodiesterase, a component of the cAMP-dependent protein kinase cascade (Kuromoni and Yamamoto, 1994). Its role in *Arabidopsis*, however, remains unknown. The second PP2C cloned,

termed "KAPP", for kinase-associated protein phosphatase, was isolated based on its ability to interact *in vitro* with a serine/threonine receptor-like kinase, RLK5, from *Arabidopsis* (Stone et al., 1994). RLK5 is a member of a family of related receptor-like protein kinases (RLKs) that participate in diverse biological functions, including defense responses, self incompatibility, and plant development (Stone et al., 1994). Moreover, KAPP interacts *in vitro* with several RLK catalytic domains, suggesting that this PP2C may regulate a number of signaling pathways mediated by RLKs (Braun et al., 1997). In fact, KAPP has been implicated in the negative regulation of CLAVATA1, a receptor-like kinase from *Arabidopsis* involved in controlling the size of floral and apical meristems (Williams et al., 1997).

The *MP2C* cDNA was isolated in a screen for negative regulators of the yeast mitogen-activated protein kinase (MAPK) pathway (Meskiene et al., 1998). Purified MP2C protein was found to be able to inactivate the stress-activated MAPK (SAMK) in plant extracts (Meskiene et al., 1998). The SAMK pathway is normally induced by a variety of abiotic stresses, such as drought, cold, wounding, etc. (Bogre et al., 1997; Jonak et al., 1996). Moreover, the *MP2C* mRNA was transiently induced by wounding and its expression pattern correlated with the inactivation of the SAMK pathway (Meskiene et al., 1998). These results suggested that MP2C is a negative regulator of the SAMK pathway in plants (Meskiene et al., 1998).

Additional evidence that type 2C protein phosphatases are involved in plant signaling pathways comes from the isolation of the *ABI1* and *ABI2* genes. Originally, a dominant mutation (*abi1-1*) conferring abscisic acid (ABA) insensitivity in *Arabidopsis*

was identified and the corresponding *ABI1* gene shown to encode a PP2C enzyme (Meyer et al., 1994; Leung et al., 1994; Bertauche et al., 1996). The *abi1* mutant is severely impaired in a wide spectrum of ABA responses, including reduced seed dormancy, excessive water loss and abnormal drought rhizogenesis (Meyer et al., 1994; Leung et al., 1994, and references therein). Thus, the control of the phosphorylation state of cell signaling components by the *ABI1* product may be mediating the observed pleiotropic hormone responses (Meyer et al., 1994; Leung et al., 1994). The *abi1-1* mutation converts the Gly residue at position 180, which is embedded in a sizable block of contiguous sequence identities among different PP2Cs, to an Asp residue and is dominant over the wild-type *ABI1* allele (Meyer et al., 1994; Leung et al., 1994). The *abi1-1* dominant mutant allele of the *ABI1* gene has been introduced into *Nicotiana benthamiana*, and the resulting transgenic plants showed a complementary subset of mutant phenotypes, including the tendency to wilt and reduced seed dormancy (Amstrong et al., 1995). These results suggest the presence of a significant degree of conservation between elements in the ABI1 signaling cascade(s) across species.

A novel PP2C gene encoding a protein highly related to *ABI1* gne has been isolated from *Arabidopsis* and was mapped to the *abi2* locus. A mutation in the latter confers ABA insensitivity similarly with the *abi1* mutation and has dominant properties at least under certain conditions (Leung et al., 1997 and references therein). Interestingly, the sequence of the ABI1 homolog from the *abi2* mutant revealed that it had the same Gly→ Asp substitution identical to that found in the *abi1-1* allele (Leung et al., 1997). Further

sets of experiments confirmed that the identified ABI1 homolog was indeed the *ABI2* gene as well as a PP2C (Leung et al., 1997). It has been proposed that ABI1 and ABI2 function in overlapping but not identical pathways of ABA signaling (Leung et al., 1997).

The above examples illustrate the important roles that PP2Cs play in plant signaling systems. The *LjNPP2C1* cDNA was shown to encode a PP2C enzyme by biochemical and genetic criteria (Kapranov et al., 1999; chapter 3, this thesis). The expression pattern of *LjNPP2C1* mRNA resembled that of a marker gene for late stages of nodule development, leghemoglobin (Kapranov et al., 1999; previous chapter in this thesis). Therefore, *LjNPP2C1* was classified as a late-nodulin gene and was hypothesized to perform specific regulatory function(s) during the late stages of nodule development and/or nodule function. (Kapranov et al, 1999; chapter 3, this thesis). Two independent approaches were taken to address the function or role of this gene in symbiotic interactions. 1. Generation of transgenic *L. japonicus* plants with altered expression of *LjNPP2C1* mRNA or expressing a dominant-negative form of this enzyme. 2. Screen for proteins interacting with LjNPP2C1 using yeast two-hybrid system. In this chapter, I will summarize the results of these experiments.

4.3. MATERIALS AND METHODS

Plant Growth Conditions

Wild-type *L. japonicus* ecotype Gifu plants were grown as described (Kapranov et al., 1997; Szczyglowski et al., 1997). The transgenic *L. japonicus* plants were inoculated with *Rhizobia loti* strain 2235 (Jarvis et al., 1982) before moving from tissue culture to soil and grown under the same conditions. Transgenic plants were watered with B&D nutrient solution (Broughton and Dilworth, 1971) supplemented with 1 mM KNO_3. The nodulation and general phenotype of transgenic plants were evaluated starting from 2-3 weeks after inoculation. Subsequently the transgenic plants were moved to a rich soil and grown for seed production.

Isolation of a λ phage clone containing the *LjNPP2C1* locus

The genomic library from *L. japonicus* ecotype Gifu was constructed in the FIX II λ phage vector by Stratagene (La Jolla, CA) and was kindly provided by Dr. Jens Stougaard, Aarhus, Denmark. The library was screened with a *LjNPP2C1* cDNA as a probe following the manufacturer's protocols.

Plasmid Construction for the Transgenic Plants

The *LjNPP2C1* cDNA used for the anti-sense and over-expression experiments contains the entire 1086 bp LjNPP2C1 coding region with additional 23 bp of 5'UTR and

129 bp 3'UTR regions. The cDNA clone was originally cloned in pBluescript SK^- vector in the anti-sense orientation with respect to *LacZ* promoter.

The construct **pBIPCR3anti,** containing the *LjNPP2C1* cDNA in the anti-sense orientation with respect to 35S CAMV promoter, the cDNA was isolated as a 1.2 kb BamHI-KpnI fragment and the KpnI site was made blunt by treatment with T4 DNA polymerase. The DNA fragment was then cloned into pBI121 binary vector (Clontech). The vector was digested with BamHI-SacI to remove the GUS gene and the SacI site was made blunt by treatment with T4 DNA polymerase. For over-expression experiments, the *LjNPP2C1* cDNA was placed behind the 35S CAMV promoter in the sense orientation in the construct **pBIPCR3sense,** engineered as follows. The 1.2 kb *LjNPP2C1* cDNA was isolated as a EcoRI-KpnI fragment, the KpnI site was made blunt by T4 DNA polymerase and the fragment was cloned into pBI121 digested with EcoRI-SmaI.

The *abi-1*-type substitution was introduced in the LjNPP2C1 cDNA as previously described (Kapranov et al., 1999; previous chapter in this thesis). The mutant cDNA was isolated as a 1.2 kb BamHI-KpnI fragment and the KpnI site was made blunt by treatment with T4 DNA Polymarase. The DNA fragment was subsequently subcloned in pBluescript II SK^-, digested with EcoRV-SacI, with the SacI site also blunt-ended by T4 DNA Polymerase. As, a result, the plasmid **pBSPCR3M,** containing the mutant *LjNPP2C1* cDNA in the anti-sense orientation with respect to the *lacZ* promoter, was selected to facilitate the future cloning steps. The 1.2 kb cDNA fragment was excised from pBSPCR3M by digestion with EcoRI-HindIII and the HindIII site was filled-in with

Klenow DNA Polymerase. The cDNA was then subcloned in pBI121 previously digested with EcoRI-SmaI to generate construct **pBIPCR3M**.

A 7.5 kb EcoRI fragment from the *L. japonicus* genome containing the *LjNPP2C1* gene with 4.3 kb of 5' and 0.5 kb 3' flanking regions was used in all the constructs designed to over-express *LjNPP2C1* mRNA under the control of the cognate regulatory elements. The 7.5 kb EcoRI fragment was derived from a λ phage isolated with the LjNPP2C1 cDNA probe. The pBI101 binary vector was digested with EcoRI-XbaI to remove the GUS gene, filled-in with Klenow DNA Polymerase and re-ligated to generate **pBI101ΔEX**. The 7.5 kb EcoRI fragment was subcloned into pBI101ΔEX digested with EcoRI and de-phosphorylated to generate construct **pBIGen2**. The 35S CAMV promoter was excised as HindIII-BamHI fragment from pBI121 and cloned into pBluescript SK^-, digested with HindIII-BamHI, to generate the construct **pBS35S**. The 7.5 kb EcoRI fragment representing the *LjNPP2C1* gene was filled-in with Klenow and ligated into pBS35S, digested with XbaI-EcoRV. The XbaI site was filled-in with Klenow DNA Polymerase and the vector was dephosphorylated. The resulting construct **pBSGen235E** contains the 35S CAMV enhancer sequence (-90...-800) fused to the 5' flanking regions of the *LjNPP2C1* gene. The 7.8 kb DNA fragment containing this fusion was excised with EcoRI-ClaI and the ClaI site was filled-in with Klenow DNA Polymerase. The fragment was then cloned into pBI101 binary vector digested with SmaI-EcoRI to generate construct **pBIGen235E**.

The 7.5 kb EcoRI fragment representing the *LjNPP2C1* gene was subcloned into pBluescript SK$^-$ vector and the *abi-1*-type substitution was incorporated in it to generate plasmid **pBSGen2SDM**. The 7.5 kb region was excised from it by EcoRI digestion and subcloned into pBI101ΔEX, digested with EcoRI and dephosphorylated, to generate construct pBISDM. The **pBISDM35E** construct, containing the 35S CAMV enhancer fusion to the mutant *LjNPP2C1* gene in pBI101, was generated in the same way as the pBIGen235E construct.

Plasmid Construction for the Yeast Two-Hybrid Screen

The 1.2 kb *LjNPP2C1* cDNA was excised as a BamHI-KpnI fragment and treated with Klenow and T4 DNA Polymerases to blunt both sites. The fragment was then cloned into the pAS2-1 vector (Clontech) digested with BamHI-PstI, also treated with Klenow and T4 DNA Polymerases and dephosphorylated. The resulting construct **pASPCR3** contains the GAL4 DNA Binding Domain (BD) fused to the complete LjNPP2C1 coding sequence (**BD-LjNPP2C1**). The same BamHI-KpnI fragment was also cloned in the pACT2 vector (Clontech), digested with BamHI, filled-in with Klenow DNA Polymerase and dephosphorylated. The resulting construct **pACTPCR3** contains the GAL4 Activation Domain (AD)-LjNPP2C1 fusion (**AD-LjNPP2C1**).

The pASPCR3 construct was digested with NdeI, filled-in with Klenow DNA Polymerase and re-ligated to generate a frame-shift between the GAL4-BD and LjNPP2C1 domains (**BD-LjNPP2C1(Fr)**). The plasmid **pASPCR3ΔN** was engineered to

contain the GAL4-BD fused to the PP2C catalytic domain of LjNPP2C1, lacking the first 96 amino acids of the protein (**BD-LjNPP2C1ΔN**). To construct it, the 0.9 kb BsiWI-XhoI region of the *LjNPP2C1* cDNA was isolated, filled-in with Klenow DNA Polymerase and cloned into pAS2-1, digested with SmaI and dephosphorylated.

The mutant *LjNPP2C1* cDNA with the *abi-1*-type substitution was isolated as a 1.2 kb HindIII-EcoRI fragment from pBSPCR3M (see above) and both sites were filled-in with Klenow DNA Polymerase. The DNA fragment was then cloned blunt-end into the pAS2-1 vector digested with NdeI, filled-in with Klenow DNA Polymerase and dephosphorylated to generate **pASPCR3M** construct containing **BD-LjNPP2C1-M** fusion.

The construct **pASLjPP2C2**, containing the fusion between the GAL4BD and LjPP2C2 (Kapranov et al., 1999; previous chapter), was generated as follows. The coding region of *LjPP2C2* cDNA was amplified using primers DB623 (5'-ATGACTGGCAGAGAGATTCTC-3') and DB624 (5'-TCACTGAAGTCTCACAACAAC-3'). The PCR product, treated with Klenow DNA Polymerase and T4 polynucleotide kinase, was cloned blunt-end into the pAS2-1 vector digested with BamHI-PstI, treated with Klenow and T4 DNA Polymerases and dephosphorylated. To generate the **GAL4BD-Nlj16** fusion, the 0.5 kb region of *LjNOD16* cDNA (Kapranov et al., 1997) was excised with BglII-HindII and treated with Klenow DNA Polymerase. The DNA fragment was then subcloned blunt-end into the pAS2-1 vector digested with BamHI-PstI, treated with Klenow and T4 DNA

Polymerases and dephosphorylated, to generate construct **pASPCR5**. The plasmids containing the BD-laminin and BD-p53 fusions were provided with the MATCHMAKERII system.

The pAD-LBP plasmid was digested with EcoRI, filled-in with Klenow DNA Polymerase and re-ligated to generate a construct containing a frame-shift between the GAL4AD and the LjNPP2C1 Binding Protein (LBP) domains (**AD-LBP(Fr)**). The construct **pASLBP**, containing the GAL4BD-LBP fusion (**BD-LBP**), was created as follows. The 1.1 kb LBP cDNA was excised from pLBP with EcoRI-XhoI and ligated into pAS2-1 digested with EcoRI-SalI.

All constructs were verified by DNA sequencing.

Northern blot Analysis

Total RNA was extracted from different tissues of *L. japonicus* as previously described (Kapranov et al., 1997; Szczyglowski et al., 1997). Ten micrograms of total RNA was separated on agarose-formaldehyde gel, transferred to nitrocellulose and hybridized with radiolabelled probes under the conditions described previously (Kapranov et al., 1997; Szczyglowski et al., 1997). The signal intensities were quantified using PhosphoImager(Molecular Dynamics, Boulder, CO).

Generation of transgenic *L. japonicus* plants

The constructs described above were conjugated into the *Agrobacterium tumefaciens* strain LBA4404 using tri-parental mating. The transgenic plants were generated using the modification of the original hypocotyl-transformation method (Handberg et al., 1994) essentially as described by Stiller et al., 1997.

Yeast Two-Hybrid Screen

The *L. japonicus* nodule cDNA library, used for the screen, was kindly provided by Dr. Carsten Poulsen, Aarhus, Denmark. The library was constructed by Stratagene in the HybriZAP vector and than converted into the pAD-GAL4 library by mass *in vivo* excision. The two-hybrid screen was performed using the CG-1945 yeast strain from MATCHMAKER II system (Clontech, Palo Alto, CA) essentially following the manufacturer's instructions. The quantitative β-galactosidase assays were performed on liquid yeast cultures using o-nitrophenyl β-D-galactopyranoside (ONPG) as a substrate following the manufacturer's protocol. One unit of β-galactosidase was defined by hydrolysis of 1 μmol of ONPG per minute per cell. All other manipulations with yeasts were also performed essentially following the manufacturer's protocols.

4.4 RESULTS

Transgenic *L. japonicus* plants do not have an obvious phenotype.

The T-DNA regions of the constructs introduced into transgenic *L. japonicus* plants are schematically shown in Figure 1. Table 1 summarizes the number of independent transgenic lines obtained for each construct.

Table 1

Construct	Number of lines
pBIPCR3Anti	69
pBIPCR3Sense	34
pBIPCR3M	45
pBIGen2	15
pBIGen235E	10
pBISDM	14
pBISDM35E	8

None of lines displayed any obvious alterations in the symbiotic phenotype compared to the wild-type *L. japonicus* that could be correlated with the presence of transgenes. Several transgenic lines showed morphological abnormalities commonly observed in the tissue culture-derived plants: partial sterility, dwarfism, abnormal flower

morphology etc. Similar abnormalities were also seen in transgenic *L. japonicus* plants containing unrelated constructs and most likely can be attributed to somaclonal variation.

Yeast two-hybrid screen

For unknown reasons, the CG-1945 strain harboring pASPCR3 had a much lower transformation efficiency then expected. Therefore, only 150,000 yeast transformants were screened in total. Unexpectedly, most of the primary transformants could grow and form variable-size colonies on a media lacking histidine even in the presence of the inhibitor of HIS3 enzyme, 3-amino-1,2,4-triazole (3-AT; 5-15 mM). A similar phenomenon was observed when a different PP2C, ABI1, was used as bait in a two-hybrid screen (Dr. Jeff Leung, personal communication). Therefore, the selection based on the activation of *HIS3* gene could not be reliably used. To circumvent this problem, the activation of the second reporter gene *lacZ* was used as a sole indication of an interaction at the first round of screening. The β-galactosidase activity was determined directly on the plates with the primary transformants using the filter-lift assay. Two colonies exhibiting the most intense blue staining where chosen for further analysis. The pAD-GAL4 plasmids were isolated from these colonies and shown to contain an identical cDNA insert, designated *LBP* (LjNPP2C1 Binding Protein). The original pAD-GAl4 plasmid containing this insert will be referred to as pAD-LBP.

Since the yeast two-hybrid screen is prone to false-positives, a number of additional controls were performed. Strain CG-1945 was co-transformed with the pairs of plasmids containing different GAL4BD and AD fusions and the activation of *HIS3* and *lacZ* genes reporter genes was evaluated as indicated in Fig. 4.2A&B. For each plasmid pair, four independent transformants were streaked out on the media without histidine supplemented with 5-15mM 3-AT to score the histidine-independent growth. The representative results for each GAL4BD and AD fusion pair are shown in Fig. 4.2A. As an independent measure of interaction, the β-galactosidase acitivity was measured quantitatively for each of the four independent transformants and the averages for each pair of GAL4Bd and AD fusions are shown in Fig. 4.2B. The expression of *HIS3* and *lacZ* genes could be activated when in the strain expressing both the AD-LBP fusion and the fusion between BD and the catalytical PP2C domain of LjNPP2C1 lacking the N-terminal extension. LBP and LjNPP2C1 could activate the expression of the reporter genes when switched to another GAL4 domain, i.e. BD-LBP and AD-LjNPP2C1. LBP could not activate the expression of *HIS3* or *lacZ* gene by itself in the presence of the pAS2-1 vector expressing only the GAL4BD. As expected, a frame-shift engineered between a GAL4 domain and either LBP or LjNPP2C1 abolished the activation.

The *abi-1*-type mutation is known to significantly decrease the catalytical activity of both ABI1 and ABI2 PP2Cs (Bertauche et al., 1996; Leung et al., 1997). Interestingly, the *abi-1*-type substitution, introduced into the LjNPP2C1 domain of the BD-LjNPP2C1 fusion, abolished the activation of *HIS3* and *lacZ* genes when co-

expressed with the AD-LBP fusion. The result shows that this mutation may prevent the interaction between the LjNPP2C1 and LBP proteins.

No activation was observed when the AD-LBP fusion was co-expressed with BD fused to related PP2C from *L. japonicus*, LjPP2C2 (Kapranov et al., 1999; previous chapter) or to any other unrelated protein, such as Nlj16 (Kapranov et al., 1997), laminin and p53. Taken together, these results strongly suggest that LBP specifically interacts with LjNPP2C1 in the yeast two-hybrid system.

LBP is a novel protein

The pAD-LBP plasmid contains a 1.1 kb cDNA with a major ORF of 183 amino acids fused in-frame with the GAL4-AD. The BLAST search showed that LBP is related to a number of hypothetical *Arabidopsis* proteins (40-70% identity and 59-88% similarity, data not shown). Based on the sequence similarity, it is likely that *LBP* cDNA contains almost the entire LBP coding region and probably lacks several N-terminal amino acids (data not shown). LBP did not show a significant similarity to proteins with known functions. It had a weak similarity (E-values 0.095-3.0, Fig. 4.3B) to a number of ribonucleases and pathogen induced proteins from plants, such ginseng ribonuclease 2 (Moiseyev et al., 1997; Fig. 4.3B). The similarity was restricted to a region of approximately 90 amino acid including a conserved amino acid stretch GDLGIGSVR (Fig. 4.3B). The latter motif may constitute a P-loop region implicated in binding nucleotides, such as ATP or GTP (Saraste et al., 1990).

LBP mRNA is up-regulated in uninfected roots of *L. japonicus*

LBP mRNA is expressed in flowers, leaves, stems and nodules of *L. japonicus*, with the highest levels in uninfected roots (data not shown). The level of *LBP* mRNA actually declines as the nodules mature (Fig. 4.4A&B). Interestingly, the level of *LBP* mRNA in the uninfected roots of *L. japonicus* non-nodulating Nod^- mutant LjEMS45 (Szczyglowski et al., 1999) is lower than in the wild-type roots or roots of other Nod^- mutants (Fig. 4.4 A&B).

4.5. DISCUSSION

The absence of a mutant phenotype in the transgenic plants containing various *LjNPP2C1* constructs may be due to a variety of different reasons. First of all, the major alterations in the expression of this gene may be lethal. On the other hand, the LjNPP2C1 enzyme may not be essential for nodulation and other biological processes in *L. japonicus*, for example a different PP2C enzyme can fulfil the function(s) of LjNPP2C1. It is also possible that the anti-sense approach was not sufficient to decrease the expression of *LjNPP2C1* gene below a threshold required to maintain the normal levels of phosphorylation of LjNPP2C1 substrate(s). Whatever the reasons are, it is clear that in order to address the biological function and significance of LjNPP2C1, it is necessary to generate a complete knock-out mutation in this gene. Only then would it be possible to make direct, unambiguous conclusions as to the role of this PP2C in nodulation and other plant processes.

A parallel and complementary approach to address the biological function(s) of LjNPP2C1 is to identify the substrates of this PP2C. The yeast two-hybrid system (Fields and Song, 1989) was used to screen for cDNAs encoding proteins that may interact with LjNPP2C1. As a result, one candidate cDNA clone was isolated. The protein product of this cDNA was tentatively designated LBP. The latter can interact with the PP2C catalytic domain of LjNPP2C1 with or without the N-terminal extension. LBP and LjNPP2C1 can also interact when their corresponding GAL4 domains are exchanged. A strong supportive evidence for the specificity of LjNPP2C1-LBP interaction comes from observations that LBP can not interact with a related PP2C

protein, LjPP2C2, or with any other protein tested in the two-hybrid system. Interestingly, the *abi-1*-type substitution can abolish the interaction, suggesting that correctly folded and active PP2C domain is required for the interaction. Finally, as expected, the interaction is dependent on the production of the GAL4 fusions with LjNPP2C1 and LBP, since the frame-shift constructs are not active. All of the above results strongly argue that LBP represents a *bona fide* LjNPP2C1-interacting protein, at least in the yeast two-hybrid system.

Unfortunately, the biological significance of this interaction remains unknown. LBP is a novel protein, with a weak similarity to ginseng ribonuclease 1 and 2 (Moiseyev et al., 1997) and a family of proteins including pathogenesis-induced PR-10 proteins (Wang et al., 1999) and related plant allergens (Acc. No. O49065). The region of similarity includes a conserved amino acid stretch of LBP (see above) which is similar to the P-loop regions of proteins binding ATP or GTP nucleotides (Saraste et al., 1990). The importance of this motif for the biological function of a protein was suggested by the X-ray crystal structure of a birch pollen allergen Bet v 1, related to the PR-10 proteins (Gajhede et al., 1996).

The *LBP* and *LjNPP2C1* genes have a different, even somewhat opposite patterns of expression. The level of *LBP* mRNA is highest in the uninfected roots of *L. japonicus* and decreases significantly in the mature *L. japonicus* nodules, while the level of *LjNPP2C1* mRNA goes up ~20-fold (Kapranov et al., 1999). The activity of LBP may be controlled by several independent mechanisms, including the regulation at the level of *LBP* mRNA abundance and also post-translationally, by protein

phosphorylation/dephosphorylation. The optimal functioning of the *L. japonicus* nodules may require the decrease in the level of LBP mRNA and concomitant dephosphorylation of this protein via the increase in the LjNPP2C1 levels. On the other hand, the *LjNPP2C1* mRNA is present, albeit at low levels, in *L. japonicus* wild-type roots and induced in the roots of *L. japonicus* Nod$^-$ mutant LjEMS46 (Kapranov et al., 1999). It is therefore possible, that LjNPP2C1 interacts with LBP in the roots and perhaps, other *L. japonicus* tissues as well. The LBP mRNA is also present in the roots of LjEMS45 mutant line, albeit at slightly lower levels than in the roots of the wild-type *L. japonicus* (Fig 4.4A&B). Finally, it is important to state that even if LjNPP2C1 interacts with LBP, the latter may not be a substrate of LjNPP2C1. Alternative possibilities exist, for example, LBP may bind and regulate in some way the activity of LjNPP2C1.

At present, it is difficult to address the biological significance of LBP-LjNPP2C1 interaction. **First and foremost, the loss-of-function mutations in *LBP* and *LjNPP2C1* genes are needed in order to address the biological functions of both genes.** The transgenic *L. japonicus* plants expressing *LBP* anti-sense transcripts behind the 35S CAMV promoter were generated (data not shown). No alterations in the symbiotic or other plant phenotypes have been observed so far. Second, the biochemical function of LBP has to be established, for example, whether it has an RNAase activity or whether it can binds ATP or GTP nucleotides.

Figure 4.1. Schematic representation of the T-DNA regions of the constructs used for transformation of *Lotus japonicus*. A, Constructs designed to (over)express the wild type *LjNPP2C1* mRNA under control of the cognate regulatory elements with and without 35S CaMV enhancer (pBIGen235E and pBIGen2, respectively), and under control of a strong constitutive 35S CaMV promoter (pBIPCR3Sense). B, Construct with the wild type *LjNPP2C1* cDNA cloned in the anti-sense orientation behind the 35S CaMV promoter. C, Constructs designed to (over)express the mutant form of *LjNPP2C1* mRNA under control of the cognate regulatory elements with and without 35S CaMV enhancer (pBISDM35E and pBISDM, respectively), and under control of 35S CaMV promoter (pBIPCR3M). NOSp- nopaline synthetase promoter, NOSt-nopaline synthetase polyadenylation signal sequence, NPTII-neomycin phosphotransferase gene, 35Sp- 35S CaMV promoter, 35Se- 35S CaMV enhancer. *LjNPP2C*-Gen represents the 7.5 kb fragment containing the *LjNPP2C1* coding region (represented by the arrow) and the 5' and 3' flanking regions. Asterisk denotes the presence of the Gly133 to Asp substitution (also indicated by the arrow).

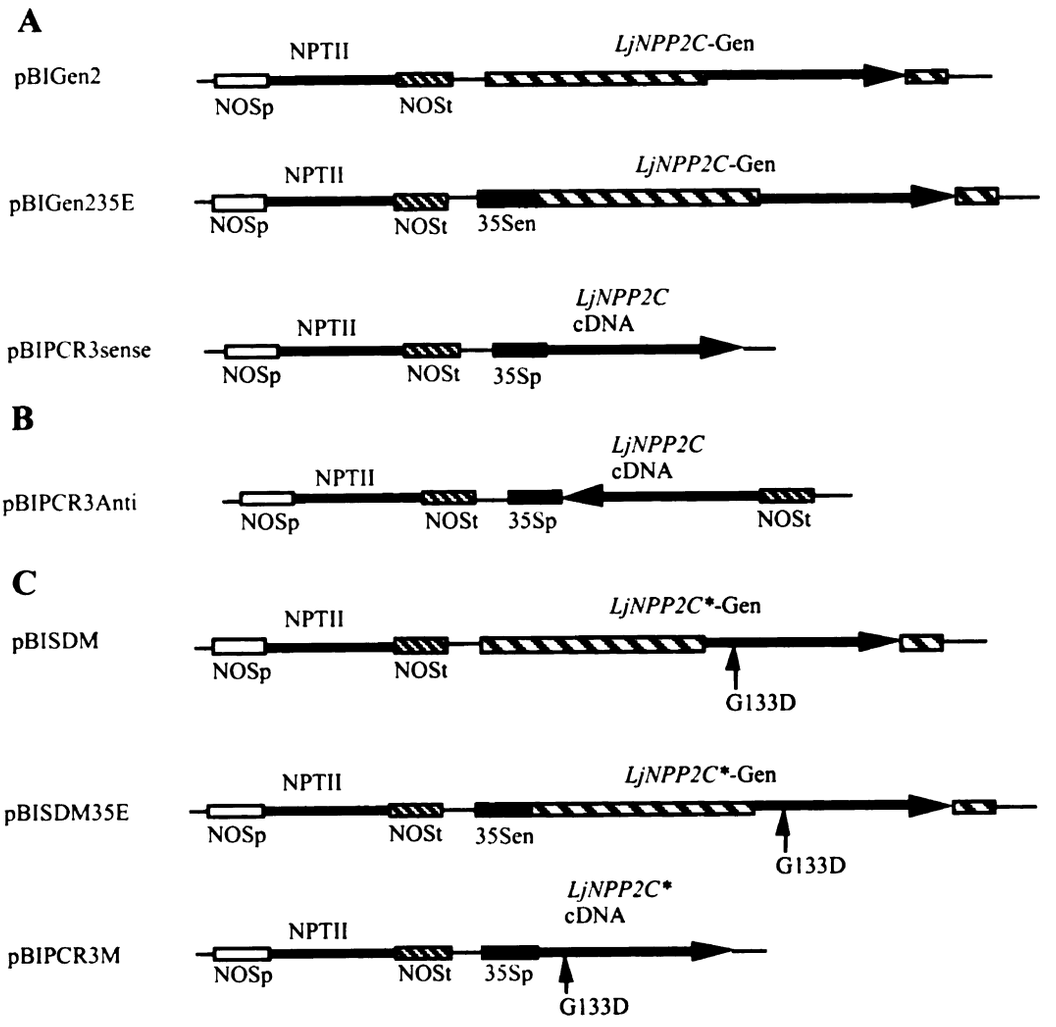

A
pBIGen2
NPTII
LjNPP2C-Gen
NOSp
NOSt
pBIGen235E
NPTII
LjNPP2C-Gen
NOSp
NOSt
35Sen
pBIPCR3sense
NPTII
LjNPP2C
cDNA
NOSp
NOSt
35Sp
B
pBIPCR3Anti
NPTII
LjNPP2C
cDNA
NOSp
NOSt
35Sp
NOSt
C
pBISDM
NPTII
LjNPP2C*-Gen
NOSp
NOSt
G133D
pBISDM35E
NPTII
LjNPP2C*-Gen
NOSp
NOSt
35Sen
G133D
pBIPCR3M
NPTII
LjNPP2C*
cDNA
NOSp
NOSt
35Sp
G133D

Figure 4.2. LBP interacts specifically with LjNPP2C1 in the yeast two-hybrid system. (**A**), The growth with and without histidine of the yeast strains expressing the indicated GAL4 BD and AD fusions.

A
+HIS
-HIS+3-AT
BD-LjNPP2C1 + AD-LBP
BD-LjNPP2C1ΔN + AD-LBP
BD-LjNPP2C1 + AD-LBP(Fr)
BD-LjPP2C2 + AD-LBP
BD-lamin + AD-LBP
BD + AD-LBP
BD-LBP + AD-LjNPP2C1
BD-LjNPP2C1(Fr) + AD-LBP
BD-Nij16 + AD-LBP
BD-LjNPP2C1-M + AD-LBP
BD-p53 + AD-LBP

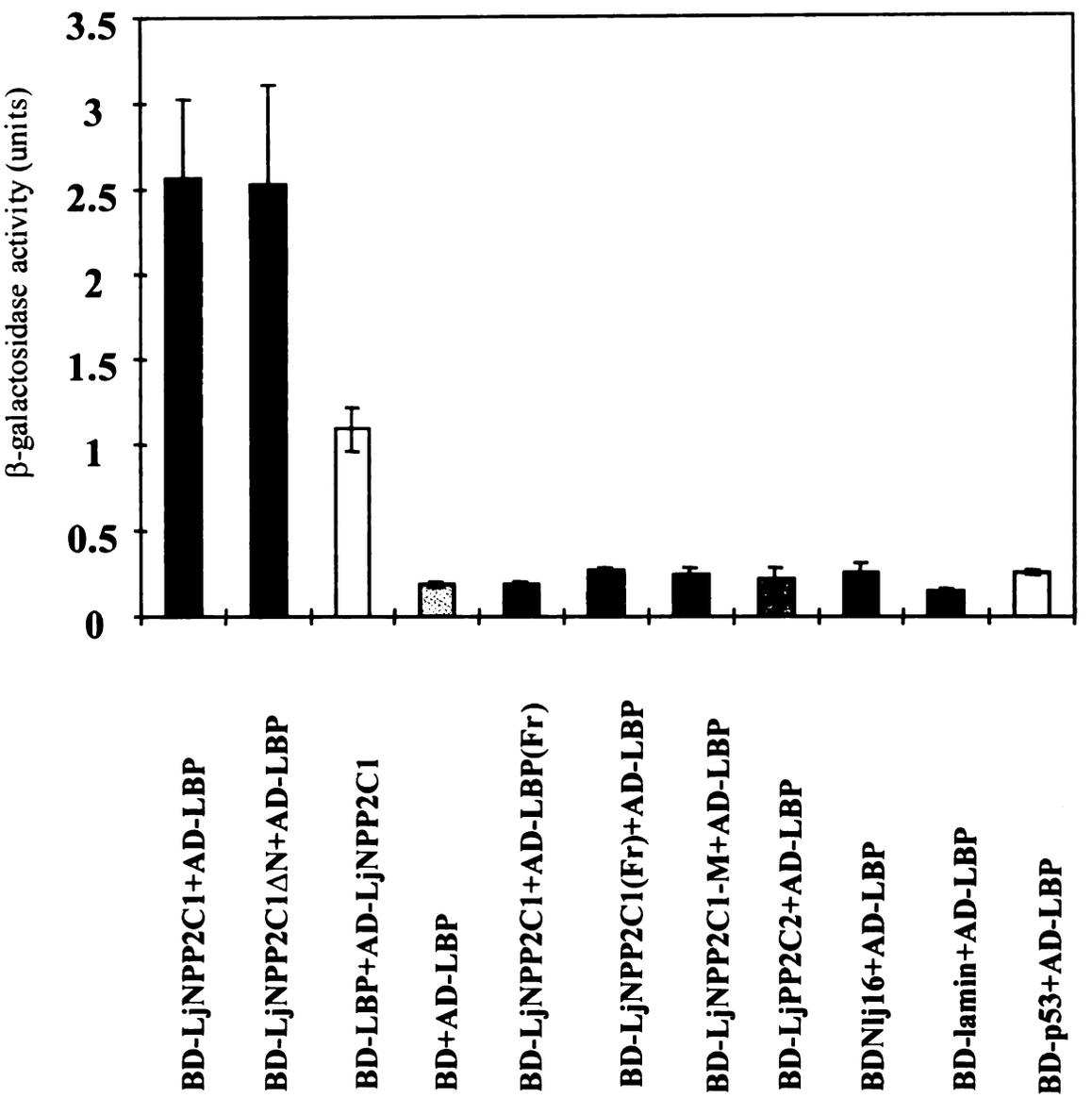


Figure 4.2B. LBP interacts specifically with LjNPP2C1 in the yeast two-hybrid system. The average β-galactosidase activity measured for each pair of the indicated GAL4 BD and AD fusions. The bars represent confidence intervals ($p<0.05$).

Figure 4.3. The LBP cDNA encodes a protein with a weak similarity to ginseng ribonuclease 2. (**A**) Nucleotide and deduced amino acid sequences of LBP mRNA. The DNA sequence shown lacks the 5' linker (GAATTCGGCACGAG) used to clone the cDNA in-frame with the GAL4-AD region. The asterisk indicates a stop codon.

A

```
1    TGGTGATGAA CCTTACAGTG CTATTGAGTC ACAGTACATT AGGAGACATC ACAAGCATGA ACTCAGGGAC
       G  D  E  P  Y  S  A  I  E  S  Q  Y  I  R  R  H  H  K  H  E  L  R  D
71   AATCAGTGCA CTTCTGCACT TGTCAAACAC ATAAAGGCCC CTGTTCATCT TGTGTGGTCT CTGGTTAGAA
       N  Q  C  T  S  A  L  V  K  H  I  K  A  P  V  H  L  V  W  S  L  V  R
141  GATTTGATCA GCCTCAGAAG TATAAACCAT TTGTGAGCAG GTGTATCATG CAAGGAGACC TTGGCATTGG
       R  F  D  Q  P  Q  K  Y  K  P  F  V  S  R  C  I  M  Q  G  D  L  G  I  G
211  AAGTGTTAGA GAAGTGAATG TTAAATCTGG TCTTCCAGCA ACAACTAGCA CTGAGAGGTT GGAACAGCTT
       S  V  R  E  V  N  V  K  S  G  L  P  A  T  T  S  T  E  R  L  E  Q  L
281  GATGATGAGG AACACATCCT TGGTATCAGG ATTGTTGGAG GTGATCACAG GCTCAGGAAC TATTGTTCCA
       D  D  E  E  H  I  L  G  I  R  I  V  G  G  D  H  R  L  R  N  Y  C  S
351  TAATCACTGT CCATCCGGAG GTCATCGATG GAAGACCCGG TACCATGGTG ATTGAGTCAT TTGTGGTGGA
       I  I  T  V  H  P  E  V  I  D  G  R  P  G  T  M  V  I  E  S  F  V  V  D
421  TGTGCCTGAA GGGAACACCC GAGATGAAAC TTGTTACTTC GTGGAGGCTC TGATCAGGTG CAACCTAAGC
       V  P  E  G  N  T  R  D  E  T  C  Y  F  V  E  A  L  I  R  C  N  L  S
491  TCTTTGGCTG ATGTCTCAGA GAGGATGGCC GTGCAGGGTC GAACCGATCC GCTCAACCAT TAAGCATGTC
       S  L  A  D  V  S  E  R  M  A  V  Q  G  R  T  D  P  L  N  H  *
561  TAAGGGAGAG TGAAGTGTTG GGTATGTGAT CCATTGCATT CATTCTGCCC AGTTCAAGTG TGGAGATTGG

631  TTGGATTAGT CCTTCCGTGT TCCCGAAAGT TTTGGATGTT CATTTTCCTC TTATCGTGGA CACCTTCCCT

701  GCAACCATCC TCCTACAATT TTAGCTGTGT ATTTCTCTTT TGAACAGCTG AGTCTCTCTT CTACAGTTTT

771  TGTAAGATGA GATTCTGGAT TTCAAAAGCA CACAGGGATG GGAGTGGGGG AAGTTGTTTT TTACTGCTAA

841  ATTCTTCCTG GTGGTGTGTT ACTACAGTTA TTTGTAAATA TTTTTGTTGG AAAAAAAAGA CATTTCTGTA

911  GTTTTTCATG TTTAGTGGCA ACTAGAAACT ATTATGTGCT GGTGGGAGAG GTCCCCAAAC AGTAGATAGT

981  ACTGCCTTGT TACTTTGGTT TACTTTAAGT AATGTTGTAC TAAANTGTAA GTTATCTATT CCAAATTCTT

1051 CTTGTTTTAG AACAAAAAAG AAAAAAAAAA AAAAA
```

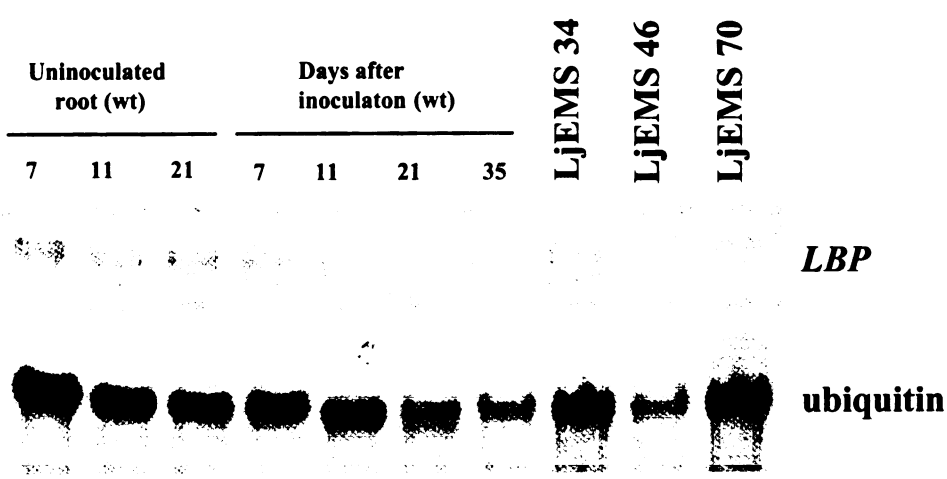
B

```
sp|P80890|RNS2_PANGI RIBONUCLEASE 2
          Length = 153

 Score = 38.7 bits (88), Expect = 0.095
 Identities = 21/89 (23%), Positives = 47/89 (52%), Gaps = 6/89 (6%)
 Frame = +2

Query: 185 IMQGDLGIGSVREVNVKSGLPATTSTERLEQLDDEEHILGIRIVGGDHRLRNYCSIITVH 364
           +++G+ G+G+++ V +    P  T   R++ +D+        I+GGD  L +    I  H
Sbjct: 43  VLEGNGGVGTIKNVTLGDATPFNTMKTRIDAIDEHAFTYTYTIIGGD-ILLDIIESIENH 101

Query: 365 PEVIDGRPGTMVIESFVVD------VPEGNTRDET 451
            +++    G+ + ++ + +      +PE N +D T
Sbjct: 102 FKIVPTDGGSTITQTTIYNTIGDAVIPEENIKDAT 136
```

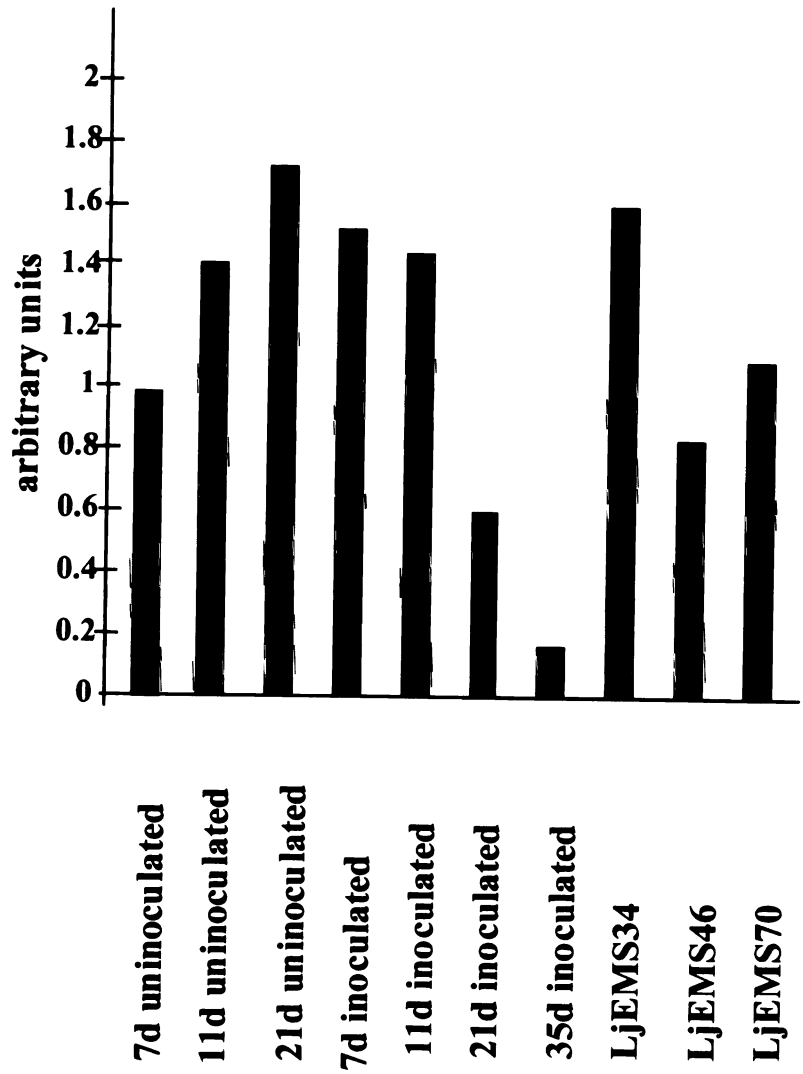
Figure 4.3. The LBP cDNA encodes a protein with a weak similarity to ginseng ribonuclease 2. (**B**) The similarity between LBP (query) and ginseng ribonuclease 2 (subject; Moiseyev et al., 1997) as detected by the BLASTX search algorithm. This match ranks the highest among the similarities found between LBP and proteins of known functions. The putative P-loop motif is indicated by the line drawn above it.

Figure 4.4. Expression of LBP mRNA at different stages of nodule development and in the roots different Nod$^-$ mutant lines of *L. japonicus*. (**A**) Total RNA from wild-type *L. japonicus* roots and nodules harvested at different times after inoculation with rhizobia, control wild-type uninoculated roots, and total RNA from uninoculated roots of the 35day-old LjEMS mutant lines was separated on agarose-formaldehyde gel and transferred to nitrocellulose. The blot was sequentially hybridized with the radioactive probes corresponding to LBP and ubiquitin cDNAs. (**B**) Quantification of the results in panel A. The volume of LBP and ubiquitin signals was quantified using PhosphoImager (Molecular Dynamics). For each lane, the volume of LBP hybridization was divided by the volume of ubiquitin hybridization and expressed in arbitrary units, with the LBP/ubiquitin ratio for the 7d uninoculated roots set as 1.

A


Uninoculated root (wt)
Days after inoculaton (wt)
7
11
21
7
11
21
35
LjEMS 34
LjEMS 46
LjEMS 70
LBP
ubiquitin

B


arbitrary units
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
7d uninoculated
11d uninoculated
21d uninoculated
7d inoculated
11d inoculated
21d inoculated
35d inoculated
LjEMS34
LjEMS46
LjEMS70

4.6. REFERENCES

Adler E, Donella-Deana A, Arigoni F, Pinna LA, Stragier P (1997) Structural relationship between a bacterial developmental protein and eukariotic PP2C protein phosphatase. Mol Microbiol 23: 57-62

Armstrong F, Leung J, Grabov A, Brearly J, Giraudat J, Blatt M (1995) Sensitivity to abscisic acid of guard-cell K^+ channels is suppressed by *abi1-1*, a mutant *Arabidopsis* gene encoding a putative protein phosphatase. Proc Natl Acad Sci USA 92: 9520-9524

Bertauche N. Leung J, Giraudat J (1996) Protein phosphatase activity of abscisic acid insensitive 1 (ABI1) protein from *Arabidopsis thaliana*. Eur J Biochem 241: 193-200

Bögre L, Ligterink W, Meskiene I, Barker PJ, Heberle-Bors R, Huskisson NS, Hirt H (1997) Wounding induces the rapid and transient activation of a specific MAP kinase pathway. Plant Cell 9: 75-83

Braun DM, Stone JM, Walker JC (1997) Interaction of the maize and *Arabidopsis* kinase interaction domains with a subset of receptor-like protein kinases: implications for transmembrane signaling in plants. Plant J 12: 83-95

Broughton WJ, Dilworth MY (1971) Control of leghemoglobin synthesis in snake beans. Biochem J **125:** 1075-1080

Chin-Sang, I.D., and Spence, A.M. (1996) *Caenorhabditis elegans* sex-determining protein FEM-2 is a protein phosphatase that promotes male development and interacts directly with FEM-3. Genes and Dev 10: 2314-2325

Cohen P (1989) The structure and regulation of protein phosphatases. Annu Rev Biochem 58: 453-508

Cohen P, Cohen PTW (1989) Protein phosphatases come of age. J Biol Chem 264: 21435-21438

Davies S P, Helps NR, Cohen PT, Hardie GT (1995) 5'-AMP inhibits dephosphorylation, as well as promoting phosphorylation, of the AMP-activated protein kinase. Studies using bacterially expressed human protein phosphatase-2C alpha and native bovine protein phosphatase-2AC. FEBS Lett 377: 421-425

Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature 340: 245-246

Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, van Neerven RJJ, Schou C, Lowenstein H, Spangfort MD (1996) X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. Nature Struct Biol 3: 1040-1045

Handberg, K., Stiller, J., Thykjear, T., and Stougaard, J. (1994) Transgenic plants: *Agrobacterium* mediated transformation of the diploid legume *Lotus japonicus*. In: Cell Biology: A Laboratory Handbook, Celis JE, ed, 1, 119-127

Jarvis BDW, Pankhurst CE, Patel JJ (1982) *Rhizobium loti*, a new species of legume root nodule bacteria. Int J Syst Bact **32:** 378-380

Jonak C, Kiegerl S, Ligterink W, Barker PJ, Huskisson NS, Hirt H (1996) Stress signaling in plants: a mitogen-activated protein kinase pathway is activated by cold and drought. Proc Natl Acad Sci USA 93: 11274-11279

Kapranov P, de Bruijn FJ, Szczyglowski K (1997) A novel, highly expressed late nodulin gene *LjNOD16* from *Lotus japonicus*. Plant Physiol 113: 1081-1090

Kapranov P, Jensen TJ, Poulsen C, de Bruijn FJ, Szczyglowski K (1999) A protein phosphatase 2C gene, *LjNPP2C1*, from *Lotus japonicus* induced during root nodule development. Proc Natl Acad Sci USA 96: 1738-1743

Kuromori T, Yamamoto M (1994) Cloning of cDNAs from *Arabidopsis thaliana* that encode putative protein phosphatase 2C and a human Dr-1 like protein by transformation of the fission yeast mutant. Nucl Acids Res 22: 5296-52301

Leung J, Bouvier-Durand M, Morris P-C, Guerrier D, Chefdor F, Giraudat J (1994) *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phopshatase. Science 264: 1448-1452

Leung J, Merlot S, Giraudat J (1997) The *Arabidopsis Abscisic Acid-Insensitive 2* (*ABI2*) and *ABI1* genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. Plant Cell 9: 759-771

Meskiene I, Bogre L, Glaser W, Balog J, Brandstotter M, Zwerger K, Ammerer G, Hirt H (1998) MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. Proc Natl Acad sci USA 95: 1938-1943

Maeda T, Wurgler-Murphy SM, Saito H (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature 369: 242-245

Meyer K, Leube MP, Grill E (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. Science 264: 1452-1455

Moiseyev GP, Fedoreyeva LI, Zhuravlev YN, Yasnetskaya E, Jekel PA, Beintema JJ (1997) Primary structure of two ribonucleases from ginseng calluses. FEBS Lett 407: 207-210

Ota IM, Varshavsky A (1993) A yeast protein similar to bacterial two-component regulators. Science 262: 566-569

Pilgrim D, McGregor A, Jackle P, Johnson T, Hansen D (1995) The *C. elegans* sex-determining gene *fem-2* encodes a putative protein phosphatase. Mol Biol Cell 6: 1159-1171

Rodriguez PL (1998) Protein phosphatase 2C (PP2C) function in higher plants. Plant Mol Biol 38: 919-927

Rodriguez PL, Leube M, Grill E (1998) Molecular cloning in *Arabidopsis thaliana* of a new protein phosphatase 2C (PP2C) with homology to ABI1 and ABI2. Plant Mol Biol 38: 879-883

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop- a common motif in ATP- and GTP-binding proteins. TIBS15: 430-434

Shiozaki K, Russell P (1995) Conteractive roles of protein phosphatase 2C (PP2C) and a MAP kinase kinase homolog in the osmoregulation of fission yeast. EMBO J 14: 492-502

Smith RD, Walker JC (1996) Plant protein phosphatases. Annu. Rev Plant Physiol Plant Mol Biol 47: 101-125

Stiller J, Martirani L, Tuppale S, Chian R-J, Chiurazzi M, Gresshoff PM (1997) High frequency transformation and regeneration of transgenic plants in the model legume *Lotus japonicus*. J Exp Bot 48: 1357-1365

Stone JM, Collinge MA, Smith RD, Horn MA, Walker JC (1994) Interaction of a protein phosphatase with an *Arabidopsis* serine-threonine receptor kinase. Science 266: 793-795

Szczyglowski K, Hamburger D, Kapranov P, de Bruijn FJ (1997) Construction of a *Lotus japonicus* late nodulin EST library and identification of novel nodule-specific genes. Plant Physiol. 114: 1335-1346

Szczyglowski K, Shaw SR, Wopereis J, Copeland S, Hamburger D, Kasiborski B, Dazzo FB, de Bruijn FJ (1998a) Nodule organogenesis and symbiotic mutants of the model legume *Lotus japonicus*. Mol Plant-Microbe Interact 11: 684-697

Wang CS, Huang JC, Hu JH (1999) Characterization of two subclasses of PR-10 transcripts in lily anthers and induction of their genes through separate signal transduction pathways. Plant Mol Biol 40: 807-814

Williams, RW, Wilson JM, Meyerowitz EM (1997) A possible role for kinase-associated protein phosphatase in the *Arabidopsis* CLAVATA1 signaling pathway. Proc Natl Acad Sci USA 94: 10467-10472

CHAPTER 5

The *Lotus japonicus LjNOD70* nodulin gene encodes a protein with similarities to transporters[3]

5.1. ABSTRACT

A novel nodule-specific gene, *LjNOD70*, associated with late stages in *Lotus japonicus* nodule development and/or functioning was characterized. The *LjNOD70* gene is a member of a small family of closely related *L. japonicus* genes. Two major mRNA species corresponding to the *LjNOD70* gene were identified in nodules and shown to be the result of a mechanism resembling alternative splicing. The longer, presumably unspliced, mRNA species was shown to contain a single open reading frame (ORF), encoding a polytopic hydrophobic protein, LjN70, with a predicted molecular mass of 70 kD. The second, presumably spliced, mRNA species was shown to be less abundant in nodules. The absence of the presumptive "intron" was found to divide the reading frame into an upstream- and a downstream ORF encoding the partial N- and C- terminal regions of the LjN70 protein, respectively. The predicted amino-acid sequence of nodulin LjN70 revealed structural features characteristic of transport proteins, and was found to share similarity with the oxalate/formate exchange protein of *Oxalobacter formigenes*.

---

[3] This chapter was published in Szczyglowski K, Kapranov P, Hamburger d, de Bruijn FJ (1998) Plant Mol Biol 37: 651-661.

Therefore, we postulate that the *L. japonicus LjNOD70* gene family encodes nodule-specific transport proteins, which may have evolved as a result of exon-intron shuffling.

5.2. INTRODUCTION

A carefully orchestrated integration of regulatory mechanisms of plant host cells and endophytic bacteria of the genera *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium*, and *Azorhizobium* is responsible for the formation of nitrogen-fixing nodules. These nodules develop as a result of a specific and compatible interaction between plant and microbe and involve the highly coordinated induction of both bacterial- and plant-encoded genes. Lipochito-oligosaccharide signal molecules of bacterial origin, the Nod factors, initiate nodule morphogenesis [3, 36]. This process involves the redirection of differentiated root cortical cells towards formation of the nodule primordium [36; and references therein]. The induction and/or enhancement of expression of plant genes called early nodulin genes has been found to be associated with early events in nodule ontogeny [for recent reviews, see 24, 32]. It has been postulated that early nodulin gene products play a role in preinfection, infection and nodule structure formation [24]. However, the exact function and requirement for nodulation of the majority of the early nodulin genes have not yet been elucidated [4]. Moreover, expression of some of the early nodulin genes has been shown to persist in fully developed nodules [e.g., ENOD40; 22] suggesting their involvement in processes, which also occur during late stages of nodule development and/or in nodule functioning.

The "late" events in nodule organogenesis initiate with the successful colonization of root cortical cells by symbiotic bacteria. This, in turn, leads to the final differentiation

of the infected and uninfected cells of the nodule and the commencement of nitrogen fixation [37]. The major molecular and biochemical alterations occurring in nodules prior to, or concurrent with the initiation of nitrogen fixation, appear to be involved in supporting the physiological requirements for nitrogen fixation and ammonia assimilation [24, 44]. Since the physiological environment within the nodule is clearly different from that in other parts of the plant, several new functions have had to evolve to facilitate symbiotic nitrogen fixation. On the molecular level, these adaptive changes manifest themselves in the expression of a number of plant genes associated with late stages of nodule development and/or nodule functioning. These genes have been classified as late nodulin genes [5, 25]. Members of this group include genes encoding enzymes, or subunits of enzymes, involved in carbon and nitrogen metabolism; proteins associated with peribacteroid membrane biogenesis and functioning; oxygen carrying leghemoglobins; putative plant transcription factors; and a number of proteins of unknown function [7, 8, 12, 13, 17, 39].

A continuous molecular cross-talk between plant cells and their microbial organelle-like structures, the symbiosomes, must occur to ensure an appropriate partitioning of the molecular clues and metabolic components involved in symbiotic nitrogen fixation [27, 41]. In addition, an appropriate translocation of organic substances, derived from photosynthetically active tissues, to nodules is required to provide the essential carbon and energy sources for symbiotic nitrogen fixation [41, 43]. Mechanisms for the metabolism and transport of fixed nitrogen from the nodules to other parts of the plant are also essential [41]. It is apparent that the proteins involved in the transport of

different compounds to and from the nodules, as well as within the nodules, are largely unknown. Recently, it has been suggested that the soybean symbiosome membrane protein nodulin-26 forms a pore responsible for the uptake of ions or small metabolites [23, 45-47]. Moreover, a soybean nodulin cDNA encoding a putative sulphate transporter has been described [18, 31]

We have previously described the identification of a range of novel expressed sequence tags (ESTs) of the model legume *Lotus japonicus* associated with late developmental events during nodule organogenesis in [17, 39]. The deduced amino-acid sequence of one of these nodulins, named Nlj65, revealed a significant similarity to the ATPTR2B peptide transporter from *Arabidopsis* thaliana [35, 38], suggesting its involvement in metabolite translocation [39]. Here we report the molecular characterization of an additional gene from *L. japonicus*, *LjNOD70*, encoding a protein with distinct topological features of transport proteins, that may be involved in nodule-specific metabolite transport.

5.3. MATERIALS AND METHODS

Plant material

Seeds of *Lotus japonicus* B-129-S9 ecotype Gifu and B-581 ecotype Funakura (kindly provided by Dr. Jens Stougaard, Aarhus University, Denmark) were germinated according to the procedure described by Kapranov et al. [17]. One week old *L. japonicus* Gifu seedlings were inoculated with *Rhizobium loti* strain NZP2235 [39], and grown under previously described conditions [17]. For the developmental Northern blot analysis segments of control (uninfected) roots and corresponding portions of infected roots or nodules were harvested 7, 11, 13, 17, and 21 days after sowing and/or inoculation with *Rhizobium*, frozen in liquid nitrogen, and stored at -80°C. *L. japonicus* Gifu flower, leaf, stem, root, and nodule tissues were harvested from 21 day old infected plants. The leaf material used for genomic DNA preparations was collected from mature, 8 weeks old, *L. japonicus* Gifu and Funakura plants.

Nucleic acid isolation, Southern and northern analyses

Genomic DNA was extracted from plant tissues by the method of Rogers and Bendich [26]. Southern blot analysis was carried out following standard protocols [30].

For northern analyses, 10 μg of total RNA was separated on agarose gels under denaturing conditions, transferred to nitrocellulose membranes (Fisher Scientific), and hybridized using previously described conditions [17, 39]. Random primer kits

(Boehringer Mannheim) were used to label probes with [α-^{32}P]dATP, following the manufacturer's instructions.

RT-PCR procedures

Reverse transcription (RT)-PCR was performed as follows: 0.6 μg of DNAseI treated total RNA derived from uninfected control roots (7 and 21 day old), and infected roots and nodules of *L. japonicus* Gifu (7, 11, and 21 days after inoculation) was reverse transcribed in the presence of 30 pM of a reverse PCR primer (p2) 5'-GGGTGAAGTTGAAAGATATTTGCA-3', in RT buffer (100 mM Tris HCl, pH 8.3, 10 mM $MgCl_2$, 10 mM DTT, 50 mM KCl), in the presence of 1 mM dNTPs, 40 units of placental RNAse inhibitor (RNasin, Boehringer Mannheim) and 4 U of AMV reverse transcriptase (Gibco), in a total volume of 20 μl. Negative control reactions were carried out under identical conditions, in the absence of AMV-RTase. For the PCR amplification reactions, 3 μl of the first strand cDNA synthesis products, 6 pmol of the forward (p1) 5'-CTTAGAAAATGATGTTTTGTGCAC-3', and reverse (see above) PCR primers, were incubated in the presence of 100 mM Tris pH 8.3, 75 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 1 unit of Taq polymerase. For semi-quantitative RT-PCR experiments the amplification reactions were carried out for 20 cycles (1 min, 94°C; 1 min, 58°C; 1 min, 72°C,); otherwise the amplification was allowed to proceed for 35 cycles. Products of the semi-quantitative RT-PCR were resolved on 1.2% agarose gels, transferred to Nylon membranes (DuPont) and hybridized with the radiolabelled insert of cDNA 48-23.

Intensities of the hybridizing fragments were compared using a PhosphoImager (Molecular Dynamics)

cDNA library screening and DNA sequence analysis

A *L. japonicus* cDNA library prepared from mature nodule mRNA, kindly provided by Dr. Jens Stougaard (Aarhus University, Denmark), was used. The library was constructed in the UniZAP lambda vector (Stratagene) using an oligo-dT primer. Screening of the library for full copy cDNAs corresponding to the EST *LjN48* product was carried out following standard techniques [30; Stratagene manual, Stratagene].

DNA sequencing was performed using Sequenase 2.0 kits (United States Biochemical) following the manufacturer's instructions. Homology searches were performed using BLAST software [2]. Sequence alignments were carried out using the GCG package (Genetics Computer Group, Madison, WI). Hydropathy analysis and secondary structure predictions of protein sequences were carried out using LASERGENE (DNASTAR, Madison, WI) and PredictProtein software [29], respectively.

Generation of anti-LjN70 antibodies

Antibodies were generated against a polypeptide, LjN70-C, comprised of the amino acids 469-575 of the nodulin LjN70 fused to the histidine-tag. To construct this fusion, the 0.6 kb BsrGI-XhoI fragment of the *LjNOD70* cDNA was isolated and blunt-ended with Klenow DNA Polymerase. The fragment was then cloned blunt-end into the

pET15B vector (Novagen), previously digested with XhoI, filled-in with Klenow DNA Polymerase and dephosphorylated. The LjN70-C polypeptide was initially purified by nickel-affinity chromatography following the standard protocols provided by manufacturer (Novagen). Subsequently, the polypeptide was ran on a preparative 15% SDS-PAGE and stained with Coomassie blue. The region of the gel containing the LjN70-C polypeptide was excised and sent to Cocalico Biologicals (Reamstown, PA) for antibody production in rabbits following the standard protocols used by this company. The anti-LjN70 serum was affinity purified using the LjN70-C polypeptide immobilized on nitrocellulose membrane as described in Harlow and Lane [11].

Preparation of peribacteroid membranes

The following experiments were performed by Jim Guenther in the laboratory of Dr. Dan Roberts, University of Tennessee, Knoxville. Nodules were harvested from approximately nine week old plants and processed as in Weaver and Roberts, 1992 [47]. Briefly, the nodules were crushed in ice cold extraction buffer (25 mM Mops NaOh, pH 7.0, 350 mM mannitol, 10 mM $MgSO_4$, 5 mM dithiothreitol, 1% (w/v) polyvinylpyrrolidone-40, 1 % (w/v) BSA, 17 mM Na-isoacorbate, and 10 mM EDTA) and the resultant extract was fractionated on a stepwise Percoll gradient as described previously [45, 47]. The 60-80% Percoll interface was collected and transferred to a 30 ml Corex tube and symbiosomes were pelleted at low speed (650 x g, 3.5 min). Bacteroids were then released by vortexing for two minutes, and removed by

centrifugation at 7,000 x g for ten minutes at 4° C. Membranes were then pelleted from the supernatant by centrifugation for 1 hour at 110,000 x g in the Ti 50 rotor.

Western Blotting

Lotus japonicus nodule membrane samples were separated on 15% SDS-PAGE gel and transferred onto PVDF membranes overnight using a trans blot cell (BioRad) at 120mA. Blots were blocked with 10% non fat dry milk in PBS for 1.5 hours while shaking at 37°C, then washed three times with PBS-Tween (PBS containing 0.05% Tween 20), ten minutes each time, at room temperature, while shaking. The blots were placed in a 1:250 dilution of the affinity-purified anti-LjN70 antibody in PBS Tween with 1% non fat dry milk for 1 hour while shaking at 37°C. Blots were then washed as above with PBS Tween, and incubated in a 1:2000 dilution of alkaline phosphatase labeled goat anti rabbit secondary antibody (supplied by ICN, Aurora western blotting kit), and incubated under the same conditions as above for the primary antibody incubation. Blots were washed three times as above in PBS Tween, and then three final washes were performed for 15 minutes each, at room temperature, in PBS Tween with .1% Triton X100, 0.05% SDS, and 1% goat serum. The blots were then incubated in the Aurora kit assay buffer and chemiluminescent substrate solution containing opti-membrane™ substrate, according to the manufactures suggestions (Aurora™ Western blot chemiluminescent detection system, ICN). Blots were visualized by exposure to Biomax ML film (Kodak).

5.4. RESULTS

EST *LjN48* represents a late nodulin gene

A range of novel expressed sequence tags (ESTs) associated with late developmental events during nodule organogenesis in the legume *L. japonicus* has been identified, including *LjN48* [17, 39].

In order to determinate the expression pattern and the length of the mRNA species corresponding to the *LjN48* EST, developmental northern blot analysis was performed, using total RNA isolated from *L. japonicus* roots and nodules harvested at different time points after infection with *Rhizobium* (Fig. 5.1). A mRNA species of approximately 2100 nt was detected in *L. japonicus* root segments 11 days after infection, but not in uninfected control roots. The corresponding mRNA was found to accumulate gradually during nodule development reaching its highest level 21-days after infection (Fig. 5.1). Thus, the pattern of gene expression observed closely resembles that of leghemoglobin mRNA and therefore is consistent with our previous classification of EST *LjN48* as a nodulin gene associated with late developmental events in nodule organogenesis [39].

Isolation and characterization of a full copy cDNA corresponding to EST *LjN48*

A *L. japonicus* nodule specific cDNA library was screened using radiolabelled *LjN48* EST DNA as a probe. Of 26 independent cDNA clones identified, the 3' ends of 11 clones were partially sequenced, and shown to have an identical nucleotide structure

(data not shown). The entire DNA sequence of the longest cDNA clone isolated (clone 48-23) was determined. The 1903 bp cDNA sequence obtained was found to contain a 1299 nt open reading frame (frame +3) starting with an ATG codon at nucleotide position 306 (ATG_{306}; relative to the 5' end of the cDNA), preceeded by several in frame stop codons. However, a careful analysis of 5' sequences, upstream of the ATG_{306} codon, revealed the presence of two additional short open reading frames. The relative location of all three ORFs found in the 48-23 cDNA is shown schematically in Figure 2. The first upstream ORF (uORF1; frame +1) was found to be 264 bp in length, starting at the first ATG codon (nucleotide position 46; ATG_{46}) and terminating at a TGA stop codon at position 310 (Fig. 5.2). The ATG_{46} codon was found to be preceded by two in frame stop codons. The second short upstream ORF (uORF2; frame +3), was found to initiate at ATG codon at position 123 and to terminate at a TAG codon at position 171.

The amino-acid sequence deduced from the nucleotide sequence of all three ORFs did not reveal significant similarity to sequences in the databases. However, the unusual architecture of the 48-23 cDNA prompted us to carry out a more detailed molecular analysis of this gene.

The 48-23 cDNA appears to correspond to a product of alternative splicing.

In order to examine the possibility that the observed structure of the 48-23 cDNA sequence was a result of a cloning artifact, several experiments were performed. The DNA sequence of the region containing the overlapping ORFs was determined multiple times

on both strands, minimizing the possibility of sequencing error(s) (see below). Moreover, a 246 bp *Eco*RI-*Hin*dIII 5'-fragment of the 48-23 cDNA, containing the uORF1, was used as DNA probe in northern analyses of RNA isolated from 21 day old uninfected roots and nodules of *L. japonicus*. The probe specifically detected a message of 2100 nt present in nodules but not in uninfected roots, suggesting that this fragment constitutes a part of the mRNA species recognized by the full cDNA clone 48-23 (data not shown). This conclusion was further confirmed and extended by amplifying *L. japonicus* mRNA via reverse transcription (RT)-PCR using a pair of primers (p1; p2) derived from sequences flanking the 5'-region of the cDNA (see Fig. 5.2). If the 48-23 cDNA indeed corresponded to a mRNA species present in *L. japonicus* nodules, the corresponding RT-PCR product would be expected to be 587 bp in length. In order to test this hypothesis, RT-PCR reactions were carried out using total RNA from uninfected and infected roots, as well as nodules harvested at different time-points after infection. Control amplification reactions using the same RNA samples, but omitting the reverse transcriptase, were carried out in parallel. After 35 PCR cycles two DNA products, 587 and 753 nt in length, were detected from RNA of 11 day old infected roots and 21 day old nodules (data not shown).

In order to estimate the relative abundance of the two distinct RT-PCR products in nodules (see below), the amplification procedure was repeated using a reduced number of PCR cycles. After 20 PCR cycles two DNA products, 587 and 753 nt in length, were detected from 21 day-old nodule RNA, whereas a single larger product, 753 bp in length, was detected with RNA sample derived from root segments, harvested 11 days after

infection. Shorter and longer RT-PCR bands were found to hybridize specifically with the radiolabelled 48-23 cDNA probe (Fig. 5.3). RT-PCR products derived from 21 day-old nodules were gel purified, cloned and the DNA sequence of several independently isolated inserts was determined. The nucleotide sequence of the shorter, 587 bp long, RT-PCR product was found to correspond to the 5' region of the 48-23 cDNA clone. The longer RT-PCR product (753 bp) contained the same 587 bp long DNA sequence as the shorter product, but carried a 166 bp insert at the boundary of the short overlap between the uORF1 and dORF. Using the same set of specific primers flanking the 5'-region of the cDNA (see above), only the 753 bp long PCR product could be amplified from *L. japonicus* genomic DNA (data not shown). The nucleotide sequence of the genomic fragment was found to be indistinguishable from that of the 753 bp RT-PCR product, and to contain the same 166 bp insert (data not shown).

The relative abundance of the two distinct RT-PCR products was estimated using PhosphoImager (Molecular Dynamics, Boulder, CO) analysis. A 5:1 ratio of the longer, insert containing product versus the shorter, insert lacking RT-PCR product, was found in RNA samples from 21-old *L. japonicus* nodules. This result suggests that the 753 bp long RT-PCR product corresponds to the major form of *LjN48* RNA species in *L. japonicus* nodules. In spite of the higher abundance of the insert-containing mRNA, we were unable to isolate its corresponding cDNA clone from the *L. japonicus* library. Therefore, the presence of an insert containing mRNA corresponding to the *LjN48* of *L. japonicus* nodules was further investigated by determining the sequence of the 2069 bp

long RT-PCR product from nodule RNA and found to be identical to the cDNA clone 48-23, including the 166 bp insert (see Fig. 5.5A).

DNA sequence analysis of several independently isolated PCR fragments derived from both RT-PCR reactions and genomic DNA, revealed that in addition to PCR products containing the exact 48-23 cDNA sequence, at least one more DNA fragment of identical length, but containing several nucleotide substitutions was present (Fig. 5.4). Since such divergent DNA sequence could be generated using both mRNA and genomic DNA as templates, it is unlikely that it is the result of base pair changes introduced by the PCR. Therefore, we conclude, that the two distinct classes of PCR products observed correspond to mRNAs species encoded by different members of a closely related gene family present in *L. japonicus* genome. This conclusion was further substantiated by carrying out Southern blotting analysis (see below).

The 166 bp insert appears to be a plant gene intron

A detailed sequence analysis of the 166 bp intervening sequence revealed that it displayed several features characteristic of plant introns. The size of the putative intron is relatively small, which is characteristic of most of the higher plant introns [34]. Virtually all naturally occuring plant introns have invariant dinucleotide 5' /GU, and 3'AG/ splice sites [34], and so does the intervening sequence of the RT- and genomic PCR products of the *LjN48* gene. In addition, the extended 5'- (AG/GUAAGU) and 3'-(UGYAG/GU) splice sites consensus motifs in higher plants match the sequences at the 5'- and 3'-ends of the 166 bp insert (Fig. 5.4).

The AU-richness, or U-richness, of sequence constitute additional characteristics of higher plant introns [34]. However, in the case of the 166 bp sequence, the AU content was found to be 54.8%, and not to be significantly different from that of the flanking 5'-, and 3'- "exon" sequences. The U-richness of the insert sequence (36.7%) was found to be only slightly higher then that of the flanking regions (30% and 33%, respectively), and the characteristic 3' splice site-proximal polypyrimidine tract was not found. Nevertheless, all the shorter RT-PCR products sequenced revealed the absence of the 166 nt long insert and to correspond to the shorter form of *LjN48* mRNA. Based on these result, we conclude that the 166 bp long insert is likely to represent an intervening sequence (intron).

The presumptive unspliced 48-23 mRNA encodes a single polypeptide

The presence of the intron-like sequence in the 48-23 cDNA results in fusion of the uORF1 and dORF, to form a single 1725 nucleotide long ORF (Fig. 5.5A). The latter initiates at the ATG_{46} codon of uORF1, proceeds through the intron and the dORF and generates a single 575 amino-acid long putative polypeptide, with a calculated molecular mass of approximately 70 kD (Fig. 5.5A). PredictProteins software analysis of the amino-acid sequence of this putative polypeptide revealed a polytopic hydrophobic protein with 12 putative transmembrane α-helices (Fig. 5.5A&B). In addition, a central hydrophilic region of 63 amino-acid residues was detected, located between the predicted transmembrane domains 7 and 8 (Fig. 5.5B). Homology searches using the BLAST

algorithm [2] revealed similarity of the putative 70 kD *L. japonicus* protein, LjN70, to the protein product deduced from the *Arabidopsis* EST 126K15 (accession T44566), and oxalate/formate exchange protein of *Oxalobacter formigenes* [1]. The latter protein represents the vectorial component of a proton-motive metabolic cycle and has been shown to catalyse the oxalate self-exchange reaction and the electrogenic exchange of oxalate and formate [1]. A pairwise comparision of the protein sequences using the BestFit algorithm of the GCG software (version 8.1) revealed that the LjN70 protein shares 51.5% similarity and 22.2% identity with oxalate/formate exchange protein of *Oxalobacter formigenes,* and 58.9% similarity and 33.7% identity with *Arabidopsis* protein (data not shown). Figure 6 shows a multiple alignment between these three proteins.

Based on these observations, we conclude that the intron containing *L. japonicus* nodule associated mRNA encodes a 70 kD protein with structural features characteristic of transport proteins [14, 21]. The corresponding *L. japonicus* gene was named *LjNOD70* accordingly.

Tissue-specific expression of the *LjNOD70* gene

When total RNA, isolated from different *L. japonicus* tissues, was probed with a radiolabelled 48-23 cDNA fragment, hybridization signals could only be detected only in nodules, but not in the other tissues tested (Fig. 5.7). This result indicates that the *L. japonicus LjNOD70* gene is expressed in a nodule-specific fashion, hence implicating its putative protein product(s) in symbiosis specific processes. In contrast, northern blot

hybridization of RNA from different *A. thaliana* tissues, using EST 126K15 as a probe, revealed that the approximately 2200 nt mRNA species was present at similar levels in all tissues tested (data not shown).

LjNOD70 is a member of a closely related gene family

In order to determine the copy number of the *LjNOD70* gene in *L. japonicus* a Southern blot analysis was performed. Genomic DNA from two *L. japonicus* ecotypes, B-129 Gifu and B-581 Funakura, representing polymorphic partners in the *L. japonicus* genome mapping effort [15], was digested using different restriction enzymes and hybridized under stringent conditions with radiolabelled PCR-48 cDNA [30]. A number of hybridizing restriction fragments could be detected for each restriction digest tested, which is indicative of a family of related *L. japonicus* genes (Fig. 5.8). Interestingly, in the case of the Funakura plants, a lower number of hybridizing bands was detected as compared to the ecotype Gifu, suggesting that fewer members of the *LjNOD70* related gene family may be present in this *L. japonicus* ecotype.

LjN70 is a peribacteroid membrane protein

Since LjN70 appears to be an integral membrane protein with a potential transporter function (see above), it was quite interesting to test whether the LjN70 protein resides in the peribacteroid membrane (PBM). The protein extract from the PBM-containing fraction isolated from *L. japonicus* nodules was subjected to a western blot analysis with the anti-LjN70 antibodies (see Materials and Methods). As a result,

two cross-reacting bands migrating at 64 kD and 27 kD could be detected in this fraction (Fig. 5.9A). The upper 64 kD band most likely corresponds to LjN70, while the identity of the lower band is less clear (see Discussion). In a parallel experiment, the protein extract from the same PBM-containing fraction was probed with antibodies against LIMP2 protein, the *L. japonicus* homolog of a well-characterized PBM protein, nodulin 26 [45-47]. As expected, LIMP2 could also be detected in the same fraction (Fig. 5.9B). These results strongly argue that LjN70 is localized in the PBM compartment of the infected cells of *L. japonicus* nodules.

5.5. DISCUSSION

We have identified and characterized a novel late nodulin gene (*LjNOD70*) from the model legume *L. japonicus*. The pattern of *LjNOD70* expression was found to be similar to that of the *L. japonicus* leghemoglobin gene [39], suggesting that it is likely to be associated with late stages of nodule development or nodule functioning.

The *LjNOD70* gene was found to be a member of a small gene family. Sequencing of a number of PCR products, generated by RT-PCR or amplification of genomic sequences revealed that the members of this gene family are closely related. The RT-PCR experiments also revealed the presence of at least two *LjNOD70* mRNA species in *L. japonicus* nodules. The nucleotide sequence of the first mRNA species was identical to the 5' region of the 48-23 cDNA, and did not contain a 166 bp insert sequence. The second, more abundant mRNA species, contained a 166 bp long insert localized in the 5' region of the mRNA. Since genomic sequences corresponding to the "spliced" mRNA species was not found, we conclude that the shorter transcript is created by removal of an 166 bp intron-like sequence from the longer mRNA species.

The presence of the intron results in the formation of a mRNA species containing a single long open reading frame (ORF), encoding a 70 kD polypeptide (LjN70). The amino acid sequence of the LjN70 protein revealed structural features typical of transport proteins, including the presence of 12 putative transmembrane domains, and a hydrophilic region located between transmembrane domains 7 and 8. In addition, the LjN70 protein shares similarity with the oxalate/formate exchange protein of *Oxalobacter formigenes* [1].

High concentrations of oxalate (about 80% of the organic acid content) have been detected in the organic fraction of broad bean (*Vicia faba* L.) nodule cytosol and a role for oxalate as carbon substrate for N_2 fixation has been postulated [40]. In general, the dicarboxylic organic acids, primarily malate and succinate, have been long known to be the source of plant carbon for the nitrogen fixing bacteroids [reviewed in 16; also see Chapter 1 of this thesis). The transporter activity for these organic acids have been detected in the PBM, but the corresponding gene(s) have not yet been cloned [42]. In this respect, the localization of the LjN70 protein to the PBM compartment suggests that this protein may be involved in the transport of the organic acids into the nitrogen fixing bacteroids.

The functional significance of the postulated alternative splicing of the *LjNOD70* mRNA remains unclear. The presence of both short and long forms of the corresponding mRNA suggests that a specific splicing mechanism may be involved in the regulation of *LjNOD70* gene expression in *L. japonicus* nodules. In fact, alternative splicing has been shown to be involved in the alteration of the function of protein kinases in animal cells [33], and a similar mechanism has recently been suggested for the control of *Arabidopsis* ANP1 protein kinase activity [26]. On the other hand, the LjN70 protein appears to be synthesized from the intron containing template. Since the presence of the intron appears to unite rather then split the open reading frame of the *LjNOD70* gene, it is possible that the coding region of the *LjNOD70* gene has evolved due to an exon-intron shuffling event [9]. The 166 bp intron sequence could have been adopted to function as part of the coding region of the *LjNOD16* gene. The limited level of splicing of the intron sequence observed (20%), would therefore reflect the past, rather than present, function of the 166 bp

intron-like sequence. It is also possible that the low efficiency of splicing of the 166 bp sequence observed may be due to its overall low AT content (54.8%) [10, 34]. However, we cannot rule out the possibility that the spliced form of *LjNOD70* mRNA encodes a functional subdomain of the LjN70 protein, and that the presence of the uORFs is involved in co-translational regulation of expression of the downstream open reading frame (dORF) as observed in other cases [6, 20]. It is interesting, in this respect, that the western blot with the anti-LjN70 antibodies could in fact detect a second protein species of a lower molecular weight. The latter could potentially be a product of translation of the spliced form of *LjNOD70* mRNA initiating at an internal AUG codon.

In any event, the identification of a novel, late nodulin gene sharing similarities with transporter proteins, provides us with another useful tool and molecular marker to elucidate the molecular basis of late nodulation events in this emerging model legume species.

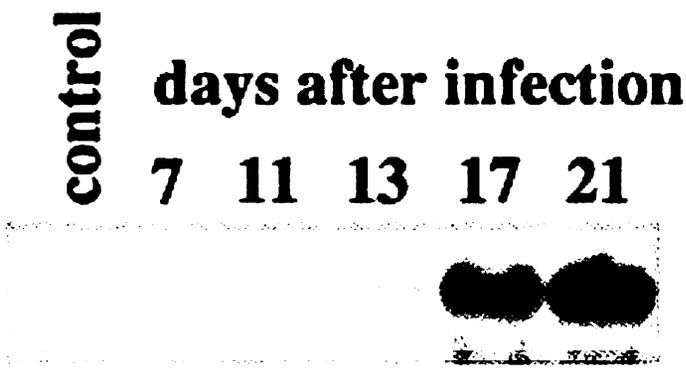


Figure 5.1. RNA gel blot analysis of *L. japonicus* *LjN48* EST. Ten micrograms of total RNA isolated from uninfected roots (control), root segments or nodules harvested 7, 11, 13, 17, and 21 days after infection were analysed, using the radiolabeled *LjN48* sequence as a probe.

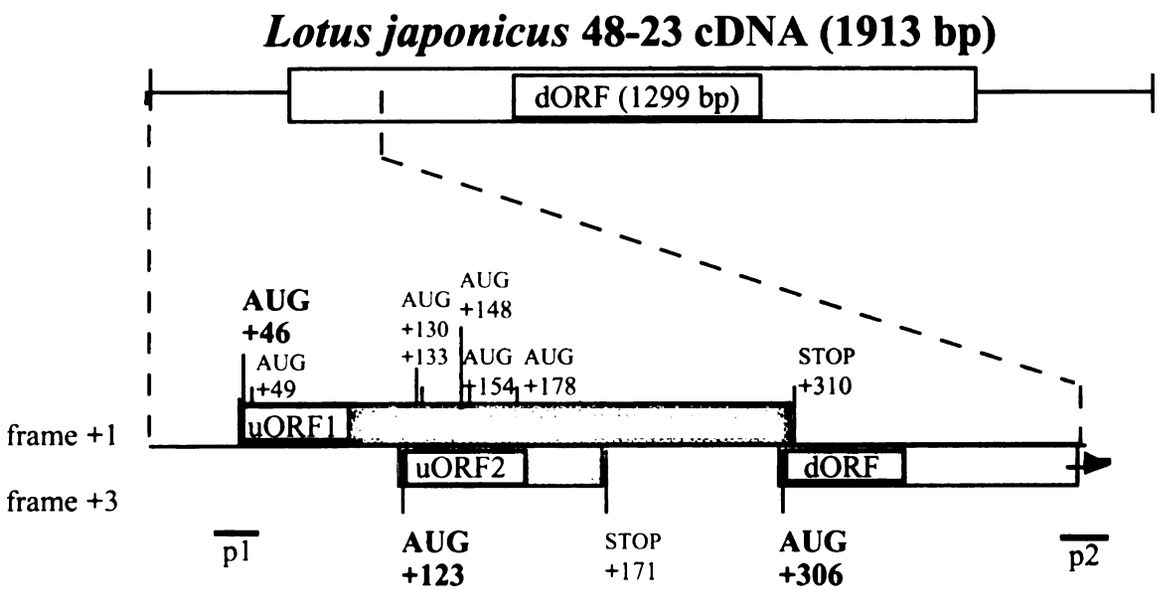


Figure 5.2. A schematic structure of *L. japonicus* cDNA 48-23: uORF1 and uORF2 denote the upstream open reading frames, whereas dORF refers to the long, continuous downstream open reading frame. Putative AUG initiation codons for the three ORFs are

designated in bold. A cluster of AUG codons in the uORF1 is also shown. Positions of PCR-primers, p1 and p2, are indicated.

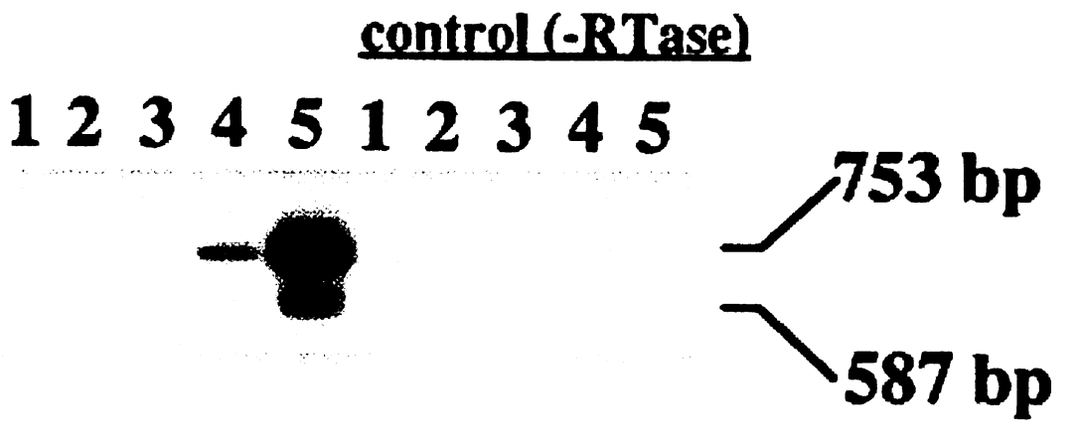


Figure 5.3. Southern blot hybridization of the RT-PCR products. Total RNA (0.6 μg) derived from 7 (lane 1) and 21 (lane 2) days old uninfected roots, root segments or nodules harvested 7 (lane 3), 11 (lane 4) and 21 (lane 5) days after infection was subjected to RT-PCR (see Material and Methods). PCR products of 20 amplification cycles were transfered onto nitrocellulose filter and hybridized with the α-^{32}P-dATP labelled 48-23 cDNA insert.

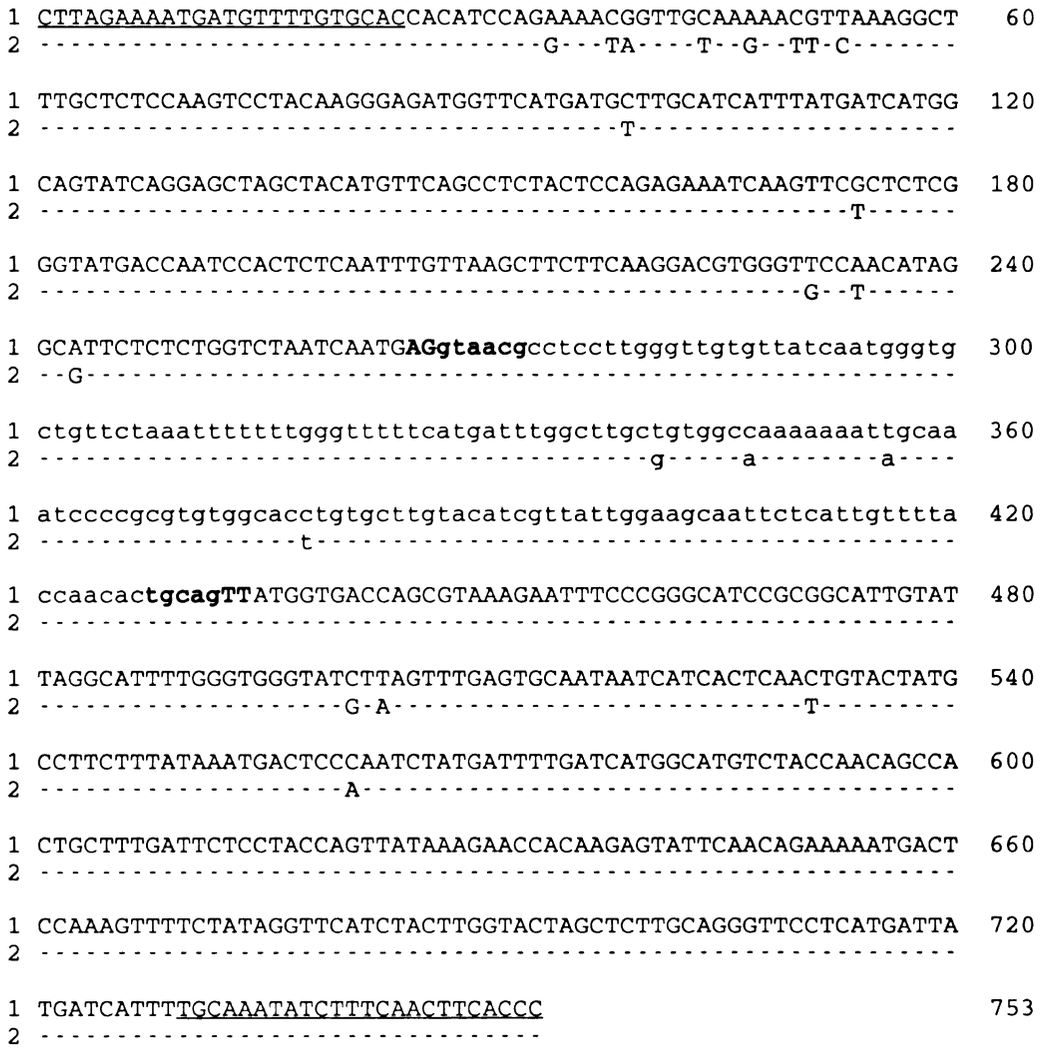
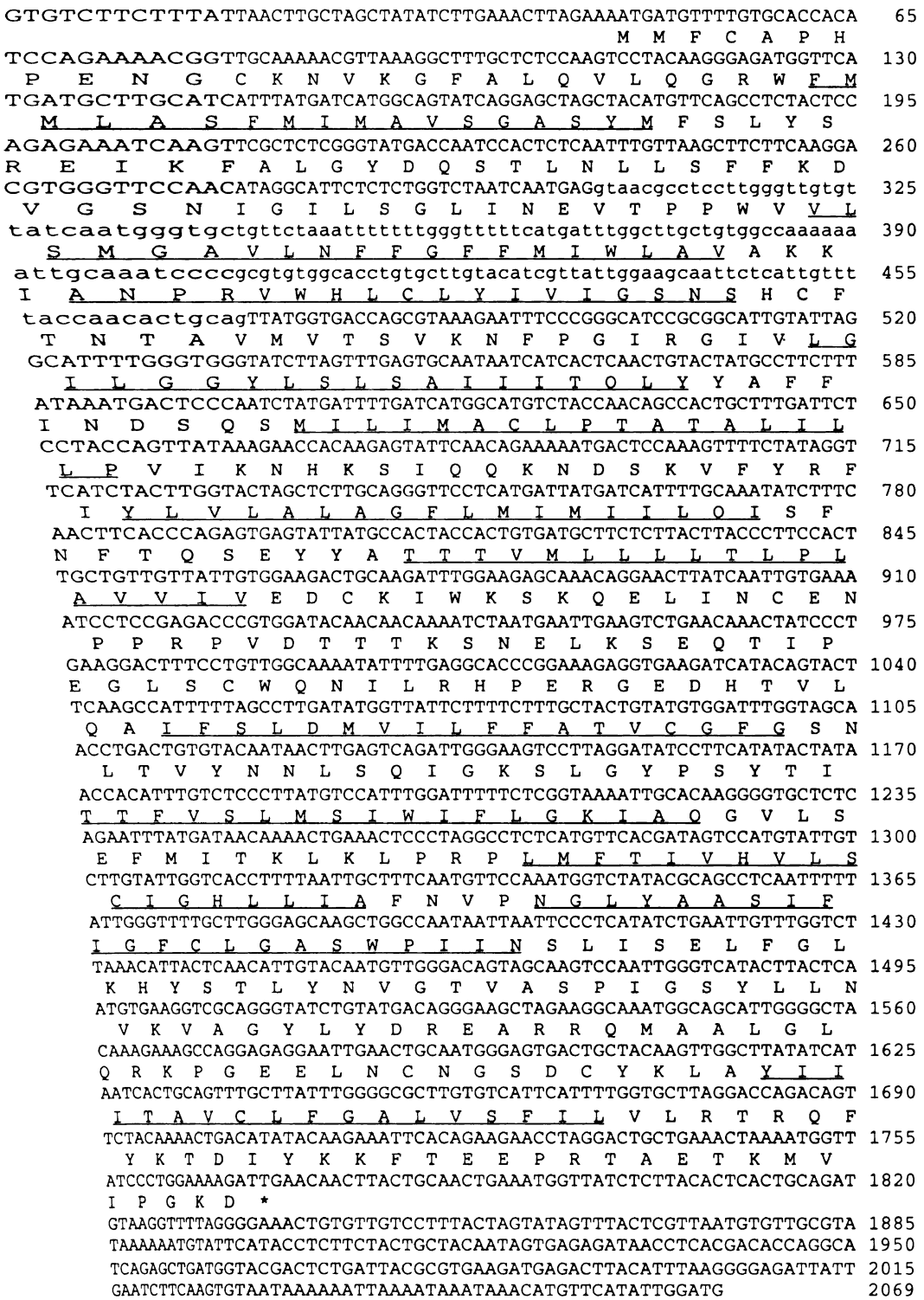


Figure 5.4. Nucleotide sequence of genomic- and RT-PCR derived *LjN48* products. The sequences labelled 1 and 2 correspond to amplified product derived from *L. japonicus* genomic DNA and total RNA from nodules. The presumptive intron sequences are indicated by lower case letters; dashed lines indicate the identical sequences; The DNA sequences of the primers used for the amplification are underlined, whereas the more extended 5'- and 3'- splice sites are shown in bold.

Figure 5.5. (**A**) Nucleotide and deduced amino-acid sequences of the 48-23 cDNA containing the 166 bp insert sequence. The insert sequence is indicated by lower case letters. The 12 transmembrane domains predicted are underlined, and the asterisk denotes the predicted stop codon.



GTGTCTTCTTTATTAACTTGCTAGCTATATCTTGAAACTTAGAAAATGATGTTTTGTGCACCACA 65
M M F C A P H
TCCAGAAAACGGTTGCAAAAACGTTAAAGGCTTTGCTCTCCAAGTCCTACAAGGGAGATGGTTCA 130
P E N G C K N V K G F A L Q V L Q G R W F M
TGATGCTTGCATCATTTATGATCATGGCAGTATCAGGAGCTAGCTACATGTTCAGCCTCTACTCC 195
M L A S F M I M A V S G A S Y M F S L Y S
AGAGAAATCAAGTTCGCTCTCGGGTATGACCAATCCACTCTCAATTTGTTAAGCTTCTTCAAGGA 260
R E I K F A L G Y D Q S T L N L L S F F K D
CGTGGGTTCCAACATAGGCATTCTCTCTGGTCTAATCAATGAGgtaacgcctccttgggttgtgt 325
V G S N I G I L S G L I N E V T P P W V V L
tatcaatgggtgctgttctaaatttttttgggtttttcatgatttggcttgctgtggccaaaaaa 390
S M G A V L N F F G F F M I W L A V A K K
attgcaaatccccgcgtgtggcacctgtgcttgtacatcgttattggaagcaattctcattgttt 455
I A N P R V W H L C L Y I V I G S N S H C F
taccaacactgcagTTATGGTGACCAGCGTAAAGAATTTCCCGGGCATCCGCGGCATTGTATTAG 520
T N T A V M V T S V K N F P G I R G I V L G
GCATTTTGGGTGGGTATCTTAGTTTGAGTGCAATAATCATCACTCAACTGTACTATGCCTTCTTT 585
I L G G Y L S L S A I I I T Q L Y Y A F F
ATAAATGACTCCCAATCTATGATTTTGATCATGGCATGTCTACCAACAGCCACTGCTTTGATTCT 650
I N D S Q S M I L I M A C L P T A T A L I L
CCTACCAGTTATAAAGAACCACAAGAGTATTCAACAGAAAAATGACTCCAAAGTTTTCTATAGGT 715
L P V I K N H K S I Q Q K N D S K V F Y R F
TCATCTACTTGGTACTAGCTCTTGCAGGGTTCCTCATGATTATGATCATTTTGCAAATATCTTTC 780
I Y L V L A L A G F L M I M I I L Q I S F
AACTTCACCCAGAGTGAGTATTATGCCACTACCACTGTGATGCTTCTCTTACTTACCCTTCCACT 845
N F T Q S E Y Y A T T T V M L L L L T L P L
TGCTGTTGTTATTGTGGAAGACTGCAAGATTTGGAAGAGCAAACAGGAACTTATCAATTGTGAAA 910
A V V I V E D C K I W K S K Q E L I N C E N
ATCCTCCGAGACCCGTGGATACAACAACAAAATCTAATGAATTGAAGTCTGAACAAACTATCCCT 975
P P R P V D T T T K S N E L K S E Q T I P
GAAGGACTTTCCTGTTGGCAAAATATTTTGAGGCACCCGGAAAGAGGTGAAGATCATACAGTACT 1040
E G L S C W Q N I L R H P E R G E D H T V L
TCAAGCCATTTTTAGCCTTGATATGGTTATTCTTTTCTTTGCTACTGTATGTGGATTTGGTAGCA 1105
Q A I F S L D M V I L F F A T V C G F G S N
ACCTGACTGTGTACAATAACTTGAGTCAGATTGGGAAGTCCTTAGGATATCCTTCATATACTATA 1170
L T V Y N N L S Q I G K S L G Y P S Y T I
ACCACATTTGTCTCCCTTATGTCCATTTGGATTTTTCTCGGTAAAATTGCACAAGGGGTGCTCTC 1235
T T F V S L M S I W I F L G K I A Q G V L S
AGAATTTATGATAACAAAACTGAAACTCCCTAGGCCTCTCATGTTCACGATAGTCCATGTATTGT 1300
E F M I T K L K L P R P L M F T I V H V L S
CTTGTATTGGTCACCTTTTAATTGCTTTCAATGTTCCAAATGGTCTATACGCAGCCTCAATTTTT 1365
C I G H L L I A F N V P N G L Y A A S I F
ATTGGGTTTTGCTTGGGAGCAAGCTGGCCAATAATTAATTCCCTCATATCTGAATTGTTTGGTCT 1430
I G F C L G A S W P I I N S L I S E L F G L
TAAACATTACTCAACATTGTACAATGTTGGGACAGTAGCAAGTCCAATTGGGTCATACTTACTCA 1495
K H Y S T L Y N V G T V A S P I G S Y L L N
ATGTGAAGGTCGCAGGGTATCTGTATGACAGGGAAGCTAGAAGGCAAATGGCAGCATTGGGGCTA 1560
V K V A G Y L Y D R E A R R Q M A A L G L
CAAAGAAAGCCAGGAGAGGAATTGAACTGCAATGGGAGTGACTGCTACAAGTTGGCTTATATCAT 1625
Q R K P G E E L N C N G S D C Y K L A Y I I
AATCACTGCAGTTTGCTTATTTGGGGCGCTTGTGTCATTCATTTTGGTGCTTAGGACCAGACAGT 1690
I T A V C L F G A L V S F I L V L R T R Q F
TCTACAAAACTGACATATACAAGAAATTCACAGAAGAACCTAGGACTGCTGAAACTAAAATGGTT 1755
Y K T D I Y K K F T E E P R T A E T K M V
ATCCCTGGAAAAGATTGAACAACTTACTGCAACTGAAATGGTTATCTCTTACACTCACTGCAGAT 1820
I P G K D *
GTAAGGTTTTAGGGGAAACTGTGTTGTCCTTTACTAGTATAGTTTACTCGTTAATGTGTTGCGTA 1885
TAAAAAATGTATTCATACCTCTTCTACTGCTACAATAGTGAGAGATAACCTCACGACACCAGGCA 1950
TCAGAGCTGATGGTACGACTCTGATTACGCGTGAAGATGAGACTTACATTTAAGGGGAGATTATT 2015
GAATCTTCAAGTGTAATAAAAAATTAAAATAAATAAACATGTTCATATTGGATG 2069

B

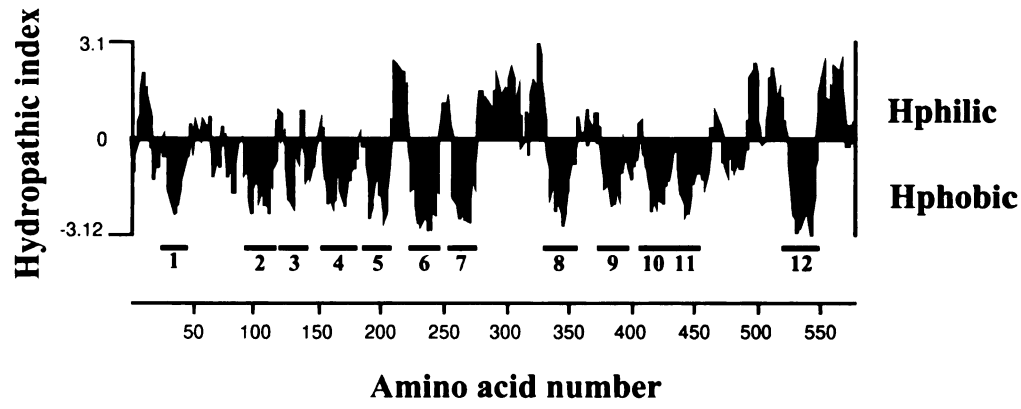


Figure 5.5. (**B**) Hydropathy profile of nodulin LjN70 derived using the method of Kyte and Doolittle [19].

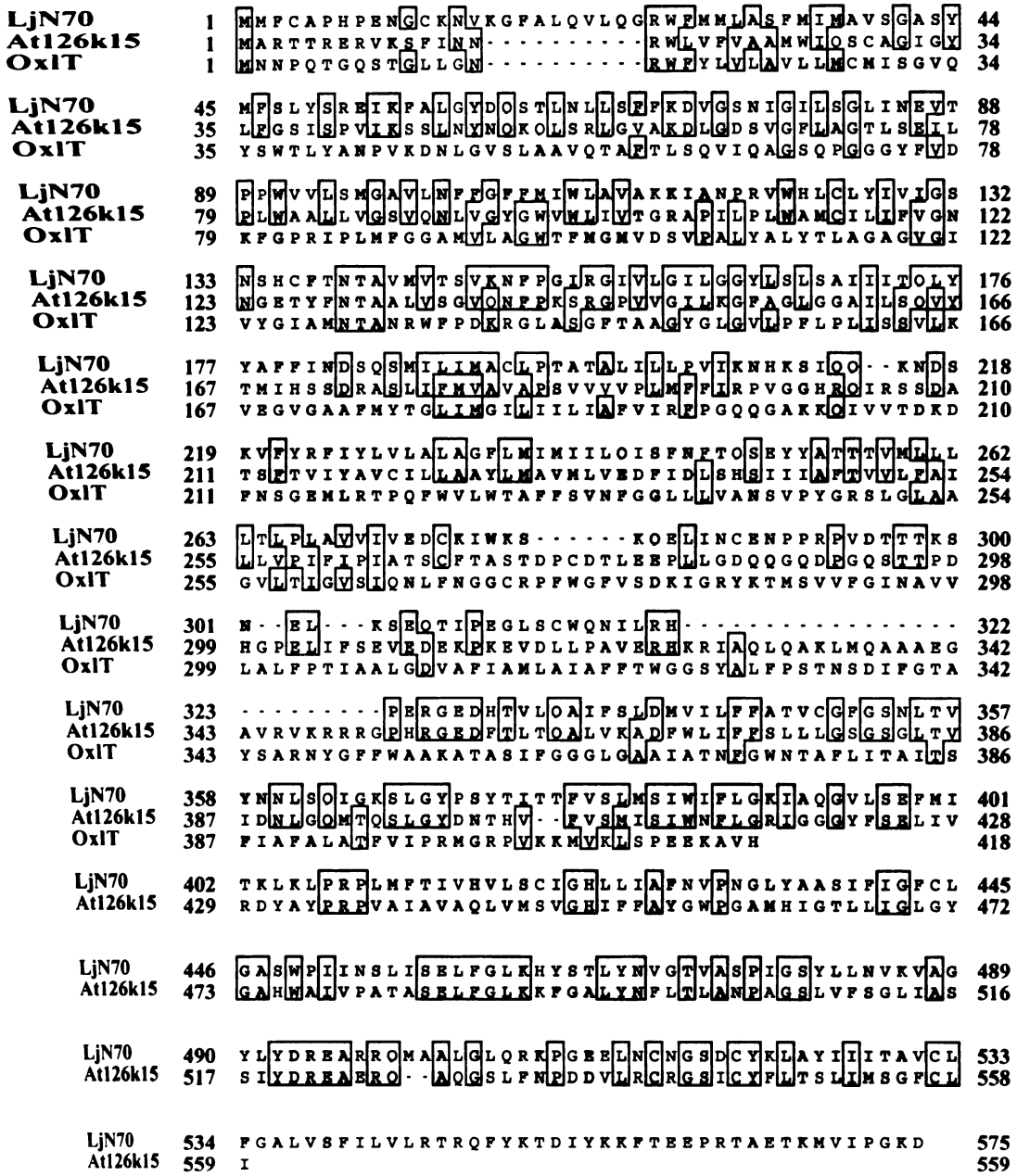


Figure 5.6. Alignment of the LjN70 protein sequence with the deduced partial amino-acid sequence of *Arabidopsis* EST 126K15 (accession. No. T44566), and the oxalate/formate exchange protein of *Oxalobacter formigenes* [1]. Identical residues are shaded and boxed; conservative substitutions are shaded.

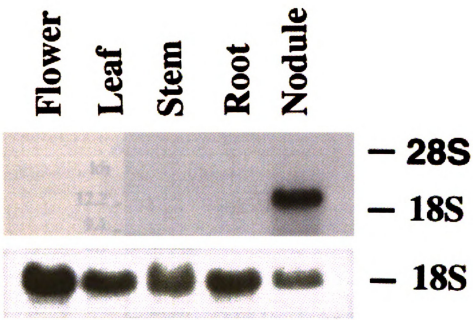


Figure 5.7. Expression of the *LjNOD70* gene in different *L. japonicus* tissues. Ten micrograms of total RNA from *L. japonicus* flowers, leaves, stems, uninfected roots, and nodules was probed with radiolabeled 48-23 cDNA insert (Top Panel). Equal RNA loading was verified by probing the same filter with an 18S DNA probe (Bottom Panel). Positions of the 28S and 18S rRNAs are as indicated.

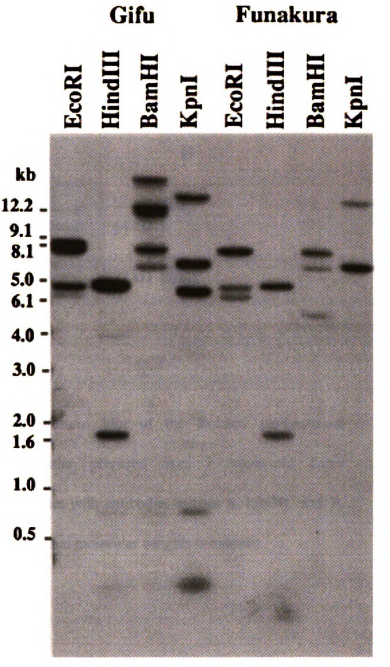


Figure 5.8. Southern hybridization analysis of *L. japonicus* ecotypes Gifu and Funakura using the *LjNOD70* gene as probe. Ten micrograms of *L. japonicus* genomic DNA was

digested with the restriction enzymes indicated, and probed with an [α-^{32}P] dATP labeled *LjN48* EST insert.

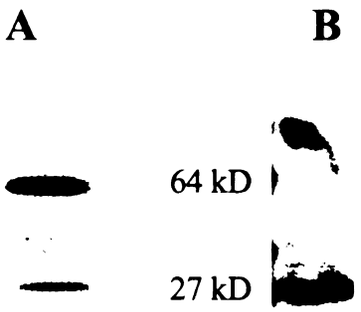


Figure 5.9. Western blot of the isolated peribacteroid membrane fraction prepared from 9 weeks-old *Lotus japonicus* nodules with antibodies against **A.** LjN70 and **B.** LIMP2. Calculated molecular weights are shown.

5.6. ACKNOWLEDGEMENTS

We thank Dr. Dominique Tremousaygue (INRA-CNRS, Tolouse, France), for providing us with a preliminary sequence of the *Arabidopsis* EST 126K15. We also thank Kurt Stepnitz for help with preparing the figures. The research described in this manuscript was supported by grant no. DE-FG02-91ER20021 from Department of Energy and grant no. NSF-09630189 from the National Science Foundation.

5.7. REFERENCES

1. Abe K, Ruan Z-S, Maloney PC: Cloning, sequencing, and expression in *Escherichia coli* of OxlT, the oxalate:formate exchange protein of *Oxalobacter formigenes.* J Biol Chem 27: 6789-6793 (1996).

2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 215: 403-410 (1990).

3. Carlson RW, Price NPJ, Stacey G: The biosynthesis of rhizobial lipo-oligosaccharide nodulation signal molecules. Mol Plant-Microbe Interact 7: 684-695 (1995).

4. Csanadi G, Szecsi J, Kalo P, Kiss P, Endre G, Kondorosi A, Kondorosi E, Kiss GB: *ENOD12*, an early nodulin gene, is not required for nodule formation and efficient nitrogen fixation in alfalfa. Plant Cell 6: 201-213 (1995).

5. de Bruijn FJ, Schell J: Regulation of plant genes specifically induced in developing and mature nitrogen-fixing nodules: *cis*-acting elements and *trans*-acting factors. In Verma DPS, (ed), Control of Plant Gene Expression, pp. 241-258. CRC Press, Boca Raton (1992).

6. Damiani RD, Wessler SR: An upstream open reading frame represses expression of *Lc*, a member of the *R/B* family of maize transcriptional activators. Proc Natl Acad Sci USA 90: 8244-8248 (1993).

7. Delauney AJ, Verma DPS: Cloned nodulin genes for symbiotic nitrogen fixation. Plant Mol Biol Rep 6: 279-285 (1988).

8. Gamas P, Niebel F, de C, Lescure N, and Cullimore J: Use of a substractive hybridization approach to identify new *Medicago truncatula* genes induced during root nodule development. Mol Plant Microbe Interact 9: 233-242 (1996).

9. Gilbert W: Why genes in pieces? Nature 271: 501 (1978).

10. Goodall GJ, Filipowicz W: The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell 58: 473-483 (1989).

11. Harlow E, Lane D: Antibodies. A laboratory manual. Cold Spring Harbor Laboratory (1988).

12. Heard J, Caspi M, Dunn K: Evolutionary diversity of symbiotically induced nodule MADS box genes: characterization of *nmhC5*, a member of a novel subfamily. Mol Plant-Microbe Inter 5: 665-676 (1997).

13. Heard J, Dunn K: Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc Natl Acad Sci USA 92: 5273-5277 (1995).

14 Henderson PJF: The 12-transmembrane helix transporters. Curr Opin Cell Biol 5: 708-721 (1993)

15. Jiang Q, Gresshoff PM: Classical and molecular genetics of the model legume *Lotus japonicus*. Mol Plant-Microbe Interact 1: 59-68 (1997).

16. Kahn ML, McDermott TR, Udvardi MK: Carbon and nitrogen metabolism in *Rhizobia*. In Spaink HP, Kondorosi A, Hooykaas PJJ (eds), The *Rhizobiaceae*. Kluwer Academic Publishers, Dordrecht, pp. 461-485(1998)

17. Kapranov P, de Bruijn FJ, Szczyglowski K: A novel, highly expressed late nodulin gene *LjNOD16* from *Lotus japonicus*. Plant Physiology 113:1081-1090 (1997).

18. Kouchi H, Hata S: Isolation and characterization of novel cDNA representing genes expressed at early stages of soybean nodule development. Mol Gen Genet: 238: 106-119 (1993).

19. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-132 (1982).

20. Lohmer S, Maddaloni M, Motto M, Salamini F, Thompson RD: Translation of the mRNA of the maize transcriptional activator *opaque-2* is inhibited by upstream open reading frames present in the leader sequence. Plant Cell 5: 65-73 (1993).

21. Marger MD, Saier Jr MH: A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. TIBS 18: 13-20 (1993).

22. Matvienko M, Van De Sande K, Yang W-C, Van Kammen A, Bisseling T, Franssen H: Comparison of soybean and pea *ENOD40* cDNA clones representing genes expressed during both early and late stages of nodule development. Plant Mol Biol 26: 487-493 (1994).

23. Miao G-H, Verma DPS: Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterologous plants. Plant Cell 5: 781-794 (1993).

24. Mylona P, Pawlowski K, Bisseling T: Symbiotic nitrogen fixation. Plant Cell 7: 869-885 (1995).

25. Nap J-P, Bisseling T: Developmental biology of a plant-prokaryote symbiosis: the legume root nodule. Science 250: 948-954 (1990).

26. Nishihama R, Banno H, Kawahara E, Irie K, Machida Y: Possible involvement of differential splicing in regulation of the activity of *Arabidopsis* ANP1 that is related to mitogen-activated protein kinase kinases (MAPKKKs). Plant J 12: 39-48 (1997).

27. Quispel A, Van Brussel AAN, Hooymans JJM, Priem WJE, Staal HJM: Intersymbiotic transport processes in nitrogen-fixing root nodules. In Ludden PW, Burris JE (eds), Nitrogen Fixation and CO_2 Metabolism, pp. 193-202. Elsevier Science Publishing Co., Inc (1985).

28. Rogers SO, Bendich AJ: Extraction of DNA from plant tissues. In Gelvin SB, Schilperoort RA (eds), Plant Molecular Biology Manual, pp. A6/1-A6/10. Kluwer Academic Publishers, Dordrecht (1988).

29. Rost B, Casadio R, Fariselli P, Sander C: Prediction of helical transmembrane segments at 95% accuracy. Protein Science 4: 521-533 (1995).

30. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning. A Laboratory Manual, ED2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

31. Sandal NN, Marcker KA: Similarities between a soybean nodulin, *Neurospora crassa* sulphate permease II and a putative human tumor suppressor. Trends Biol Sci 19: 19 (1994).

32. Schultze M, Kondorosi E, Ratet P, Buire M, Kondorosi A: Cell and molecular biology of *Rhizobium*-plant interaction. Int Rev Cytol 156: 1-75 (1994).

33. Shi E, Kan M, Xu J, Wang F, Hou J, McKeehan WL: Control of fibroblast growth factor receptor kinase signal transduction by heterodimerization of combinatorial splice variants. Mol Cell Biol 13: 3907-3918 (1993).

34. Simpson GG, Filipowicz W: Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinary. Plant Mol Biol 32: 1-41 (1996).

35. Song W, Steiner HY, Zhang L, Naider F, Stacey G, Becker J: Cloning of a second *Arabidopsis* peptide transport gene. Plant Physiol: 110: 171-178 (1996).

36. Spaink HP: Regulation of plant morphogenesis by lipo-chitin oligonucleotides. Crit Rev Plant Sci 15: 559-582 (1996).

37. Sprent JI: Which steps are essential for the formation of functional legume nodules? New Phytol 111:129-153 (1989).

38. Steiner HY, Song W, Zhang L, Naider F, Becker JM, Stacey G: An *Arabidopsis* transporter is a member of a novel family of membrane transport proteins. Plant Cell 6: 1289-1299 (1994).

39. Szczyglowski K, Hamburger D, Kapranov P, de Bruijn FJ: Construction of a *Lotus japonicus* late nodulin EST library and identification of novel nodule-specific genes. Plant Physiology 144: 1335-1346 (1997).

40. Trinchant J-C, Guerin V, Rigaud J: Acetylene reduction by symbiosomes and free bacteroids from broad bean (*Vicia faba* L.) nodules. Plant Physiol 105: 555-561 (1994).

41. Udvardi MK, Day DA: Metabolite transport across symbiotic membranes of legume nodules. Annu Rev Plant Physiol Plant Mol Biol 48: 493-523 (1997).

42. Udvardi MK, Price GD, Gresshoff PM, Day DA: A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. FEBS Lett 231: 36-40 (1988)

43. Vance CP, Heichel GH: Carbon in N_2 fixation: limitation or exqusitive adaptation. Annu Rev Plant Physiol Plant Mol Biol 42: 373-392 (1991).

44. Verma DPS: Signals in root nodule organogenesis and endocytosis of *Rhizobium*. Plant Cell 4: 373-382 (1992).

45. Weaver CD, Crombie B, Stacey G, Roberts DM (1991). Calcium-dependent phosphorylation of symbiosome membrane proteins from nitrogen fixing root nodules: Evidence for the phosphorylation of nodulin-26. Plant Physiol. 95: 222-227.

46. Weaver CD, Shomer NH, Louis CF, Roberts DM: Nodulin 26, a nodule-specific symbiosome membrane protein from soybean, is an ion channel. J Biol Chem 269: 17858-17862 (1994).

47. Weaver CD, Roberts DM (1992) Determination of the site of phosphorylation of nodulin 26 by the calcium-dependent protein kinase from soybean nodules. Biochemistry 31: 8954-8959.

CHAPTER 6

Novel, highly expressed late nodulin gene (*LjNOD16*) from *Lotus japonicus*[4]

6.1. ABSTRACT

A *Lotus japonicus* cDNA was isolated and shown to correspond to a highly abundant, late nodule-specific RNA species that encodes a polypeptide with a predicted MW of 15.6 kD. The protein and its corresponding gene were designated Nlj16 and *LjNOD16*, respectively. The *LjNOD16* was found to be expressed only in the infected cells of *L. japonicus* nodules. Related DNA sequences could be identified in the genomes of both *Glycine max* and *Medicago sativa*. In the latter case, a homologous mRNA species was detected in the nodules. Unlike *LjNOD16*, its alfalfa homologs appear to represent low-abundant mRNA species. However, the proteins corresponding to the *LjNOD16* and its alfalfa homolog could be detected at similar levels in nodules, but not in roots of both legume species. The predicted amino acid sequence analysis of nodulin Nlj16 revealed the presence of a long α-helical region and a positively charged C-terminus. The former domain has a very high propensity to form a coiled-coil type structure, indicating that nodulin Nlj16 may interact with as-yet-unidentified protein target(s) in the nodule-infected cells. Homology searches revealed no significant similarities to any known

---

[4] This chapter was published in Kapranov P, de Bruijn FJ, Szczyglowski K (1997) Plant Physiology 113: 1081-1090.

sequences in the databases, with the exception of two related, anonymous *Arabidopsis* ESTs.

6.2. INTRODUCTION

Symbiotic nitrogen fixation is a unique example of a complex and subtly regulated biological process that takes place in a specialized plant organ - the nodule. Upon infection with specific strains of symbiotic bacteria belonging to the genera *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium*, root cortical cells of legume plants undergo a dedifferentiation process that eventually leads to the formation of a nodule meristem. A highly organized and controlled series of events thereafter culminates in the formation of a fully functional, nitrogen-fixing nodule. Nodule ontogeny appears to be predominantly controlled by a plant morphogenetic program (for a recent review see Gresshoff, 1993). Multiple signals, derived from both the host plant and the symbiotic bacteria, specify the induction and coordination of this organogenic process which involves the activation of specific genes in both symbiotic partners (for a review see Schultze et al., 1994). A collection of genes encoding plant-nodule-specific proteins, collectively referred to as nodulin genes (van Kammen, 1984), has been isolated from various species of legumes (for a review see Mylona et al., 1995). These genes have been traditionally classified as early or late nodulin genes, reflecting the developmental time point of their expression (Nap and Bisseling, 1990). Induction of the early nodulin genes has been correlated with early morphogenetic processes, such as preinfection, infection, and cortical cell division (Franssen et al., 1992; Nap and Bisseling, 1990; Cook et al., 1995). However, the

functions and requirement for nodulation (Csanadi et al., 1994) of none of the early nodulin genes have been elucidated.

Significant progress has been made in unraveling the cascade of both bacterial and plant signals, which initiate and direct early morphogenic events. The discovery of a set of flavonoid inducers of plant origin and *Rhizobium*-derived lipochito-oligosaccharide signal molecules, known as Nod factors, has led to exciting progress in this area (Horvath et al., 1987; Lerouge et al., 1990; Spaink et al., 1987; Spaink, 1992). Although research on the early stages of nodulation has progressed rapidly, the developmental cues and molecular events responsible for the final steps of nodule formation and functioning remain largely unknown (de Bruijn and Schell, 1992). These late stages include important events such as central nodule tissue formation, bacterial release from infection threads and plant cell colonization, production of peribacteroid membranes (PBM), bacteroid differentiation, and eventually, commencement of nitrogen fixation (Sprent, 1989). The latter process is accompanied by major molecular and biochemical alterations, which create and support the physiological requirements for nitrogen fixation and ammonia assimilation (Nap and Bisseling, 1990).

In spite of the highly complex nature of the late developmental events, only a limited number of late nodulin genes have been identified and characterized. Typical members of this group include genes encoding enzymes involved in specific biochemical pathways, e.g., the γ subunit of Gln synthetase involved in ammonia assimilation (Lara et al., 1983; Gebhardt et al., 1986; Bennett et al., 1989; Boron et al., 1989; Boron and

Legocki, 1993), uricase II (Legocki and Verma, 1979; Bergmann et al., 1983), xanthine dehydrogenase (Triplett, 1985), sucrose synthase (Thummler and Verma, 1987), peribacteroid membrane proteins (Fortin et al., 1985; 1987; Verma, 1992; Miao and Verma, 1993), leghemoglobins (Brisson et al., 1982), and a number of proteins the functions of which remain to be identified (Delauney and Verma, 1988). Recently, a symbiotically induced MADS-box containing gene (*nmh*7) has also been identified in alfalfa root nodules (Heard and Dunn, 1995). It has been suggested that the protein encoded by this gene could be involved in cellular activities specific to the differentiation of the infected cells (Heard and Dunn, 1995).

Although a number of late nodulin genes have been isolated, it appears likely that additional genes remain to be identified. Further cloning and detailed characterization of genes coding for novel late nodulins is crucial for understanding the molecular and biochemical details of late nodule morphogenic events and nodule functioning. Therefore, we have initiated a systematic search for novel late nodule-specific transcripts in the model legume *Lotus japonicus* (Handberg and Stougaard, 1992; Jiang and Gresshoff, 1993), using the RNA differential display technology described by Liang and Pardee, (1992; and also Szczyglowski et al., 1997). We report here the isolation and molecular characterization of *LjNOD*16, a novel late nodulin gene from *L. japonicus*.

6.3. MATERIALS AND METHODS

Plant Material

Lotus japonicus GIFU B-129-S9 seeds were kindly provided by Dr. Jens Stougaard, Aarhus University , Denmark. Seeds were surface sterilized and germinated as described by Handberg and Stougaard (1992). One-week-old *L. japonicus* seedlings were inoculated with *Rhizobium loti* strain NZP2235 (Jarvis et al., 1982) and transferred to pots (30 plants per pot) containing a 6:1 mixture of vermiculite and sand. All plants were grown in cabinets with a controlled environment: an 18h/6h day/night cycle, a light intensity of 246 μE sec^{-} m^{-2}, a $22^0C/18^0C$ day/ night temperature regime, and a 70% humidity level. Both inoculated and uninoculated control plants were watered using B&D nutrient solution (Broughton and Dilworth, 1971) containing 0.5 mM KNO_3. This low concentration of combined nitrogen supported growth of the uninfected control plant but did not affect nodule formation on the inoculated *L. japonicus* plants (data not shown). For the initial stages (3, 7, and 11 days) root segments were harvested, whereas only nodules were collected for the 21-day time-point.

The transgenic *L. japonicus* plants were inoculated with *Rhizobia loti* strain 2235 (Jarvis et al., 1982) before moving from tissue culture to soil and grown under the same conditions. Transgenic plants were watered with B&D nutrient solution (Broughton and Dilworth, 1971) supplemented with 1 mM KNO_3. The nodulation and general phenotype of transgenic plants were evaluated starting from 2-3 weeks after inoculation. Fully

mature nodules from each transgenic line were harvested for western blot to assess the levels of Nlj16 protein. Subsequently the transgenic plants were moved to a rich soil and grown for seed production.

The *Arabidopsis thaliana* Landsberg plants were kindly provided by Jacqueline Chernys from Dr. Hans Kende's laboratory, Plant Research Laboratory, Michigan State University. *Medicago sativa* Cardinal seeds were surface sterilized and germinated on 1% agar in water. Three-day-old seedlings were infected with *Rhizobium meliloti* strain 1021 and transferred to soil [3 vermiculite (grade2): 3 vermiculite (grade 3): 1 sand]. *M. sativa* plants were grown under the same controlled environment as *L. japonicus* plants and watered with B&D solution or distilled water. *M. sativa* plants were grown for 21 days before harvesting. The plant tissues were immediately frozen in liquid nitrogen and stored at -80°C until use.

Nucleic Acid Isolation, Southern and Northern Analyses

Plant genomic DNA was isolated following the procedure described by Rogers and Bendich (1988). Total RNA isolation was performed according to the method of Verwoerd et al. (1989), except that the extraction buffer was as described in Hall et al. (1978). The poly $(A)^+$ fraction of total RNA was isolated using oligo(dT) cellulose spin column kits (5 Prime→3 Prime, Inc., Boulder, CO) following the manufacturer's instructions.

For Southern blot analysis, 10 μg of genomic DNA was completely digested using the appropriate restriction enzymes, separated on 0.8% agarose gel, and transferred to nylon filters according to standard procedures (Sambrook et al., 1989). Membranes were prehybridized and hybridized under high stringency conditions (2xSSC) or low stringency conditions (4xSSC), 0.5% SDS, 5X Denhard's solution, 100 μg/ml denatured sheared salmon sperm DNA at 65°C. Washes were carried out under low or high stringency, as described by Sambrook et al. (1989).

For Northern analysis, 10 μg of total RNA was separated on 1.2% agarose-formaldehyde gels in MOPS buffer, as described by Sambrook et al. (1989) and transferred to Nitro Plus supported nitrocellulose membranes (Fisher Scientific, Pittsburgh, PA). Prehybridization and hybridization reactions were performed in 0.5M phosphate buffer, pH 7.2, 7% SDS and 1% BSA at 65°C, according to the procedure described by Church and Gilbert (1984). Filters were washed twice for 15 min in 2X SSC, 0.1% SDS, and once for 15 min at 0.3X SSC, 0.1% SDS at 65°C. Probes were labeled with [α-^{32}P]dATP using the random prime kit from Boehringer-Mannheim (Indianapolis, IN), following the manufacturer's instructions.

RNA Differential Display

Differential Display was carried out using RNAmap™ kits from the GenHunter Corporation (Brookline, MA), essentially following the manufacturer's instructions. cDNA synthesis was performed using 0.5 μg of total RNA isolated from 21-day-old *L.*

japonicus roots, or nodules harvested 21 days after rhizobial infection. Differentially expressed bands were purified from the polyacrylamide gel and reamplified using the procedure described in the manual of the RNAmap kit. The DNA fragments were purified on a 1% agarose gel, blunt ended using the Klenow fragment of DNA polymerase I, and cloned into the *Sma* I-digested vector pK18 (Pridmore, 1987).

cDNA Library Screening

The cDNA library from mature nodules of *L. japonicus* was kindly provided by Dr. Jens Stougaard, Aarhus, Denmark. The library was constructed with oligo-dT primer in the lambda UniZAP vector from Stratagene (La Jolla, CA). Screening for full-copy cDNAs corresponding to the PCR-5 product was performed following standard procedures (Sambrook et al., 1989; Stratagene manual).

DNA Sequencing and Computer Analyses

DNA sequencing was performed using the Sequenase 2.0 kit from the United States Biochemical, Inc. (Cleveland, OH), according to the manufacturer's instructions. Computer analysis of DNA sequences was carried out using the SeqEd software (Applied Biosystems, Foster City, CA). Analysis of predicted protein sequences was performed using the GCG (Genetics Computer Group, Madison, WI, USA), PHDsec (Rost and Sander, 1993, 1994) and COILS version 2.2 (Lupas, 1996) programs. Homology searches were performed using BLAST software (Altschul et al., 1990).

In Situ Hybridization

Twenty-one-day-old nodules were fixed, dehydrated, and embedded into paraffin according to procedures described by van de Wiel et al. (1990). Nodule sections of 7 μm were hybridized with digoxigenin-UTP labeled antisense and sense RNA probes, using the conditions reported by Engler et al. (1994). The full-copy cDNA encoding nodulin Nlj16 and a cDNA representing a *L. japonicus* leghemoglobin gene were used to prepare RNA probes, using the DIG-RNA labeling kit (Boehringer Mannheim, Indianapolis, IN) following the manufacturer's instruction. Washing and detection conditions were as described by Engler et al. (1994). Dry slides were mounted using polymount (Polysciences Inc., Warrington, PA) and examined by dark- and bright-field microscopy using an Axiophot microscope (Zeiss, Oberkochen, Germany).

Production of Recombinant Nlj16 Protein in *Escherichia coli*

The 470-bp *Bgl* II-*Hinc* II fragment of the full length *LjNOD16* cDNA, containing the entire coding region of the nodulin *LjNOD16* gene was blunt ended using Klenow enzyme and cloned into the dephosphorylated pET 15B expression vector (Novagen, Madison, WI), which had been digested with *Xho* I and blunt ended with Klenow enzyme. The construct with the insert cloned in the sense orientation was identified by digestion with *Xho* I and *Bam* HI. In frame fusion of the coding region of *LjNOD16* to the N-terminal histidine tag of pET 15B was confirmed by DNA sequencing. The construct was transformed into *E. coli* strain BL21 (DE3) (Novagen, Madison, WI), for expression

of the recombinant protein. The His-tag-Nlj16 fusion protein was purified under denaturing conditions using His-Bind® Resin (Novagen, Madison, WI), following the manufacturer's instructions. The histidine tag was cleaved off by digestion with thrombin (Novagen, Madison, WI) and the Nlj16 recombinant protein was further purified by passing it through the His-Bind® Resin column after digestion. The flow through and wash fractions were collected. Nlj16 was refolded by dializing against phosphate-buffered saline (PBS; Sambrook et al., 1989) overnight. The refolded protein was mixed with TiterMax® adjuvant (CytRx® Corporation, Norcross, GA) and used for immunization of rabbits following the instructions of the TiterMax® manual.

Western Blot Analysis

The western blot described in figure 9A&B was performed as follows. The total protein extracts from *L. japonicus* and *M. sativa* roots and nodules were obtained by grinding tissues in liquid nitrogen and boiling for 5 min in extraction buffer (100mM Tris, pH 6.8, 5% SDS, 0.5% β-mercaptoethanol). For Western blot analysis, 50 μg of total protein was separated by SDS-PAGE (Laemmli, 1970), using 15% acrylamide gels, and electroblotted overnight onto Protran™ nitrocellulose membranes (Schleicher and Schuell, Keene, NH) in Towbin buffer (Towbin et al., 1979), supplemented with 0.05% SDS. Blocking, binding and washes were performed in PBS supplemented with 0.3% Tween 20 (Sigma, St. Louis, MO). Pre-immune and immune sera from a rabbit immunized with Nlj16 recombinant protein were used in a 1: 2000 dilution. Antibody detection was

performed using goat anti-rabbit antibodies conjugated to alkaline phosphatase (Kirkegaard & Perry Laboratories, Gaithersburg, MD).

The western blots in Figures 10 and 11 were performed with the affinity purified antibody. The anti-Nlj16 serum was affinity-purified over a column containing recombinant Nlj16 protein conjugated to the Affi-Gel 10 matrix (Bio-Rad Laboratories) following the manufacturer's instructions. The affinity-purified antibody was used in1:500 dilution.

Generation of *L. japonicus* transgenic plants

The *LjNOD16* cDNA used for transgenic work contains the entire 423 bp Nlj16 coding region with 21 bp 5'UTR and 404 bp 3'UTR. The cDNA was isolated as EcoRI-XhoI fragment and both sites were filled-in with Klenow DNA polymerase. The *LjNOD16* cDNA region was then cloned blunt-end into the pLP14 binary vector (Szczyglowski et al., 1994) digested with KpnI-SacI, treated with T4 DNA Polymerase to blunt both sites and dephosphorylated. The pLP14 binary vector contains the 2 kb promoter sequence of *Sesbania rostrata* leghemoglobin gene *glb3* fused to the GUS reporter gene (Szczyglowski et al., 1994). The *LjNOD16* cDNA was placed under the control of the *glb3* promoter, in place of the GUS gene, in both sense (construct **pCR5Sense**) and anti-sense (construct **pCR5Anti**) orientations.

Both constructs were conjugally transferred into *Agrobacterium tumefaciens* strain LBA4404. The transgenic *L. japonicus* plants were generated using the modification of

the original hypocotyl-transformation method (Handberg et al., 1994) essentially as described by Stiller et al., 1997.

6.4. RESULTS

Identification and cloning of the *L. japonicus* PCR-*5* cDNA

The recent development of RNA arbitrarily primed fingerprinting techniques, also known as RNA differential display, has resulted in an experimental tool with exciting potential for the detection of differential gene expression during complex biological processes (Liang and Pardee, 1992; Welsh et al., 1992). We have employed this technique to construct a library of *L. japonicus* late nodulin ESTs (Szczyglowski, Hamburger, Kapranov and de Bruijn, manuscript in preparation). Our experiments involved a simple comparison of RNA profiles from two samples: mature 21-day uninfected roots and fully developed nodules harvested at 21 days after infection. Thirteen different combinations of PCR primers were initially used. Five bands appeared to be differentially expressed on display gels. These were excised, reamplified by PCR, and analyzed using northern blotting and DNA sequencing. Three out of five differentially displayed cDNA products corresponded to nodule-specific mRNAs, as determined by northern blotting (data not shown). One of the differentially expressed cDNAs, designated PCR-5, was found to be

607 bp long and showed no significant homology to DNA sequences in the databases (data not shown). We chose this cDNA for further detailed molecular characterization.

Northern blot analysis was used to correlate specific phases in nodule development with the expression pattern of the gene corresponding to the PCR-5 cDNA. This experiment revealed that the PCR-5 product corresponded to a highly abundant late nodulin mRNA species. The mRNA corresponding to PCR-5 was found to be detectable between seven and eleven days after infection and to accumulate gradually to a high level in fully developed 21-day-old nodules (see Fig. 6.1). The observed kinetics of mRNA accumulation resembled very closely the pattern of *L. japonicus* leghemoglobin gene expression (data not shown), justifying the categorization of this gene as a late nodulin gene.

PCR-5 cDNA corresponds to the gene encoding *L. japonicus* nodulin Nlj16

PCR-5 cDNA was used as a probe to screen a *L. japonicus* nodule-specific cDNA library (kindly provided by Dr. Jens Stougaard, University of Aarhus, Denmark). Approximately forty hybridizing cDNA clones were purified. DNA sequence analysis of the six longest cDNA clones (length range 790 bp to 1010 bp) showed that they all carried identical DNA sequences, but had different length of 3'-nontranslated regions. Six different sites of polyA addition were identified in the cDNAs. The sequence of the longest cDNA clone is shown in Figure 6.2A (accession number U64964). The different polyadenylation sites, derived from the sequences of the corresponding cDNA clones, are as indicated. Further analysis of the DNA sequence of the longest cDNA clone revealed

that it contained a 41-bp 5'-untranslated region, followed by an open reading frame 423 nucleotide-long, and a 546-bp 3'-untranslated region. The stop codon (TAA) was found to be located 6 bp upstream of the putative AUG initiation triplet, indicating that the cDNA contains the entire coding region (Fig. 6.2A). The deduced protein sequence was found to correspond to a polypeptide of 141 amino acid residues and a molecular weight of approximately 15.6 kD. Therefore, we propose that our cDNA represents a late *L. japonicus* nodule-specific gene, designated *LjNOD16*, encoding a 15.6-kD nodule-specific protein (nodulin Nlj16).

Nodulin Nlj16 contains a long α-helical domain and is homologous to two anonymous ESTs from *Arabidopsis thaliana*

A hydropathy profile of nodulin Nlj16 was derived using the algorithm of Kyte and Doolittle (1982) and LASERGENE software (DNASTAR, Madison, WI). This analysis indicated that the protein was mostly hydrophilic, except for a hydrophobic stretch of 16 amino acids, with tetrad repeats of proline residues at the N-terminal end (Fig. 6.2B; see also Fig. 6.2A). Secondary structure analysis using the PHDsec program predicted with a high probability (7-9 on a scale from 0 to 10) the presence of two adjacent α-helical regions, divided by a loop or turn (Fig. 6.2A). The relatively long α-helical domain, spanning 80 out of 141 amino-acid residues of the Nlj16 protein, would be expected to form a coiled-coil type protein structure, as predicted by the COILS version 2.2 program (Lupas, 1996). A homology search using the BLAST algorithm revealed no

significant similarities to any gene with known function, except for a limited homology to two anonymous *Arabidopsis* ESTs: EST168K8 (accession R64923) and 110G16 (accession T42081). The corresponding *A. thaliana* cDNA clones were obtained from Tom Newman's laboratory at the MSU-DOE Plant Research Laboratory and from the Arabidopsis Biological Resource Center (Ohio State University), respectively, and their entire DNA sequence was determined. The complete DNA sequences for ESTs 168K8 and 110G16 have been deposited in GenBank under accession numbers U64965 and U64966, respectively. The deduced amino acid sequences of EST 168K8 (751 bp; A.t. EST1; Fig. 6.3) and EST 110G16 (644 bp; A.t. EST2; Fig. 6.3), which represent 200 and 150 amino acids, respectively, from the C-terminal ends of the proteins, showed a high level of homology to each other throughout the whole amino acid sequence (Fig. 6.3). In addition, a significant similarity between the *A. thaliana* ESTs and the amino acid sequence of the Nlj16 protein, especially in its α-helical domain, was found (Fig. 6.3). Interestingly, the region of the 168K8 EST with the highest similarity to the α-helical domain of Nlj16 also has a high probability to form two α-helices and to participate in generating coiled-coil structures, as predicted by the PHDsec and COILS programs. A computer search of the PROSITE and BLOCKS databases did not reveal any distinct conserved amino acid motifs in nodulin Nlj16 or in the protein sequences derived from the *Arabidopsis* ESTs.

Tissue-specific expression of the *LjNOD16* gene

To gain further insight into the tissue specificity of *LjNOD16* gene expression, total RNA isolated from *L. japonicus* flower, leaf, stem, root, and fully developed nodule tissues was analyzed using northern blot hybridization (Fig. 6.4). The *LjNOD16* gene was found to be expressed exclusively in the nodule tissues as revealed after a brief, 7.5-h exposure of the blot (data not shown). However, a prolonged exposure of the blot, over a period of 7 d, revealed the presence of a very low level of hybridizing mRNA species in uninfected roots and flowers. In the latter case, an mRNA species of approximately 2500 nucleotides could be detected, as opposed to the 900- to 1200-nucleotide-long transcripts consistently found in *L. japonicus* nodules. Therefore, we conclude that the *LjNOD16* gene represents a nodulin gene, and that a related gene(s) may exist in the *L. japonicus* genome. Northern blot analysis was also performed using total RNA isolated from various organs of *A. thaliana* plants and A.t. EST1 or A.T. EST2 cDNAs were used as probes. An example of this analysis using A.t. EST1 is shown in Figure 6.5. Both ESTs hybridized to an approximately 3000 nucleotide-long mRNA in all tissues tested.

Organization of *LjNOD16* gene in genomes of *L. japonicus* and other legume plants

To determine the number of *LjNOD16* genes in the *L. japonicus* genome, Southern blot analysis was performed. Genomic DNA isolated from *L. japonicus* plants was digested to completion using four different restriction enzymes and hybridized with an

[α-^{32}P]dATP-labeled probe, representing the PCR-5 cDNA sequence. Under high-stringency conditions (see Materials and Methods) a single hybridizing band was observed for three out of four different restriction digests (note that a *Hind* III site is present in the PCR-5 cDNA sequence), indicating that the *LjNOD16* gene is represented by a single or low copy gene number in the *L. japonicus* genome (Fig. 6.6A). To determine whether gene(s) homologous to *L. japonicus LjNOD16* were present in the genomes of other legumes, a Southern hybridization was performed using the full-length cDNA for *LjNOD16* as a probe. Several hybridizing bands were observed in genomic DNA isolated from *Lotus corniculatus*, *Glycine max*, and *Medicago sativa* plants. Three additional hybridizing fragments were detected in the *L. japonicus* genome under moderate stringency conditions using the full-length cDNA as a probe (Fig. 6.6B). The latter results indicate the presence of *LjNOD16*-related genes in the *L. japonicus* genome, providing additional support to the conclusion drawn above with regard to expression of a related message of high molecular weight in flowers.

Localization of *LjNOD16* transcripts in the *L. japonicus* nodule

Cellular localization of the *LjNOD16* mRNA in nodules was analyzed using non-radioactive *in situ* hybridization. Semi-thin sections (7 μm) of *L. japonicus* 21-d-old nodules were hybridized with digoxigenin-UTP-labeled RNA probes transcribed from the *LjNOD16* and *L. japonicus* leghemoglobin cDNAs. Both sense and antisense transcripts were used as probes. Upon examination of several consecutive sections derived from

different *L. japonicus* nodules, we observed that expression of *LjNOD16* was restricted to the infected cells of the nodule (Fig. 6.7 B&C), which closely resembled the expression pattern of the *L. japonicus* leghemoglobin mRNA (Fig. 6.7D). No expression of *LjNOD16* was detected in uninfected interstitial cells, nor in the peripheral tissues, such as the cortex and endodermis. In addition, no hybridization signals were observed when digoxigenin-UTP-labeled sense RNA transcripts corresponding to *LjNOD16* was used as a probe (Fig. 6.7A).

The *Medicago sativa LjNOD16* homologue is expressed specifically in nodules

Since the *L. japonicus LjNOD16* cDNA probe hybridized to DNA sequences in the *M. sativa* genome, we investigated the tissue-specific expression of the corresponding gene(s) using northern blot analysis. Ten micrograms of the total RNA isolated from different *M. sativa* tissues was hybridized with the radiolabeled *L. japonicus LjNOD16* cDNA insert. Unexpectedly, no detectable signal could be found in any of the *M. sativa* tissues analyzed when total RNA was used (data not shown). We hypothesized, therefore, that the corresponding gene(s) might be expressed at a low level in the indeterminate nodules of *M. sativa* plants. We tested this assumption using 4 μg of poly(A)$^+$ RNA derived from leaves, stems, roots, and nodules of the same 21-d-old *M. sativa* plant (Fig. 6.8). Two hybridizing bands corresponding to mRNA species of

approximately 800 nt and 1800 nt in length, respectively, could be detected specifically in nodules, but not in any other *M. sativa* tissues analyzed.

The *L. japonicus* Nlj16 protein and its *M. sativa* homologue can be detected in nodules but not in the other tissues of *L. japonicus*

Polyclonal antibodies were raised against the recombinant *L. japonicus* Nlj16 protein, and used to carry out a western blot analysis using total protein extracts from roots and nodules of *L. japonicus* and *M. sativa* plants (Fig. 6.9). The recombinant Nlj16 protein was loaded on the SDS-PAGE gel as a positive control. Five to six protein bands in the size range of 16 to 19 kD could be consistently detected in the extract derived from *L. japonicus* nodules, but not from *L. japonicus* roots (Fig. 6.9A). In comparison, a 10-kD protein was specifically detected in *M. sativa* nodule extract (Fig. 6.9A). Intense protein bands were also observed in *M. sativa* root extracts. Since they could also be detected using the pre-immune serum, it is highly unlikely that they reflect specific reactions with the anti-LjN16 antibody (Fig. 6.9B). The pre-immune serum failed to detect the control recombinant Nlj16 protein. In addition, neither the *L. japonicus* 16 to 19-kD protein bands nor the 10-kD protein from *M. sativa* nodule extracts were recognized by pre-immune serum (Fig. 6.9B). The same nodule proteins were, however, specifically detected with the antibodies purified from the anti-Nlj16 serum by affinity chromatography over a column with a recombinant Nlj16 protein (Fig. 6.10). On the other hand, the immuno-depleted fraction of the serum can no longer efficiently recognize the 16-19 kD protein

bands in *L. japonicus* nodules (data not shown). Based on these results, we conclude that the proteins detected specifically in the nodule extracts of both plants analyzed are likely to correspond to nodulin Nlj16 and its *M. sativa* homolog, respectively.

To further analyze the distribution of Nlj16 in different tissues of *L. japonicus*, the western blot with protein extracts from flowers, leaves and stems, uninfected roots and nodules of *L. japonicus* was probed with the affinity-purified anti-Nlj16 antibody (Fig. 6.10). As expected, the Nlj16 protein could be detected only in the nodules (Fig. 6.10). The 40 kD nodule-specific band represents an unknown *Rhizobial* protein, also present in free living *Rhizobia loti* and *E.coli* cultures (data not shown). A doublet of bands around 60 kD could be detected in all the tissues of *L. japonicus* with a higher level in flowers (Fig. 6.10, the doublet is very faint in the nodules on this figure). The 60 kD protein doublet most likely represents the product(s) of the larger, ~2500 nt transcripts, cross-hybridizing with the *LjNOD16* probe (Fig. 6.4, see the next chapter of this thesis)

Analysis of the transgenic *L. japonicus* plants carrying constructs pCR5Anti and pCR5Sense

In total, 32 independent *L. japonicus* transgenic lines, carrying pCR5 sense construct, and 44 lines carrying pCR5antisense construct, were generated and analyzed. The nodules from 4 independent transgenic *L. japonicus* lines (pCR5Anti-3, 11, 18 and 25) were to found to contain significantly diminished levels of Nlj16 protein compared with the wild-type nodules (Fig. 6.11). The nodules from the transgenic line pCR5Anti-1contain the typical amount of Nlj16 protein present in the nodules from the majority of

L. japonicus transgenic plants, lower than in the wild-type *L. japonicus* nodules 21dai (Fig. 6.11). The nodules from the line pCR5Anti-25 did not contain detectable amounts of Nlj16 protein as judged by the western blot analysis (Fig. 6.11). However, regardless of the level of Nlj16 nodulin, the nodules from all transgenic pCR5Anti lines had a normal appearance as judged by normal size, shape and a pink color indicative of the presence of late nodulin leghemoglobin. Examination of the sections from pCR5Anti-18 and pCR5Anti-25 nodules using light microscopy did not reveal any significant structural differences from the wild-type nodules (data not shown). Also, the plants from both pCR5Anti-18 and pCR5Anti-25 transgenic lines did not show chlorosis, a typical sign of nitrogen starvation, thus further indicating that the nodules on these plants fix normal amount of nitrogen. Therefore, *L. japonicus* nodules with significantly diminished levels of Nlj16 protein appear to develop and function normally. None of the tested pCR5sense transgenic *L. japonicus* plants had nodules with elevated levels of Nlj16 protein comparing to the wild-type nodules (data not shown).

6.5. DISCUSSION

We have isolated and characterized a cDNA clone from the model legume *L. japonicus* representing a novel late nodulin gene, designated *LjNOD16*. One intriguing feature of the *LjNOD16* is its very high level of nodule-specific transcriptional activity, which is similar to the transcriptional activity of the leghemoglobin genes in *L. japonicus* nodules. This property inspired us to perform a more detailed molecular analysis of the *LjNOD16* gene.

The expression pattern of the *LjNOD16* gene, based on the developmental northern blot analysis, showed distinct characteristics of late-nodulin genes. The corresponding mRNA was first detectable in the *L. japonicus* roots around 11 d after *Rhizobium* infection and accumulated to a very high level in fully developed nodules. This expression pattern may suggest that the product of the *LjNOD16* gene is involved in relatively late stages of nodule ontogeny and/or nodule functioning. This assumption was further supported by the localization of transcripts corresponding to the *LjNOD16* gene in infected cells of *L. japonicus* nodules. Infected-cell-specific expression has also been demonstrated for leghemoglobin (*lb*) genes (Szczyglowski et al., 1994). Since *in situ* localization experiments were performed using fully developed *L. japonicus* nodules, it is not clear whether the induction of *LjNOD16* gene expression precedes bacterial colonization of the plant cells, or occurs concomitantly with the release of symbiotic bacteria from the infection threads and commencement of nitrogen fixation. The latter expression pattern is characteristic of the majority of late-nodulin genes identified thus far (Nap and Bisseling, 1990; Govers et al., 1987). Taking into account the time point after *Rhizobium* infection when the *LjNOD16* gene becomes activated, we propose that it is a late nodulin gene, appearing to be coordinately induced just prior to, or concomitant with, the commencement of nitrogenase activity.

Northern blot analysis using different *L. japonicus* tissues revealed the presence of very low levels of mRNAs hybridizing to *LjNOD16* in uninfected roots and flowers, in addition to nodules. A faint band corresponding to the size of the *LjNOD16* mRNA was detected in the uninfected roots after a prolonged exposure of the blot. A

different mRNA species of approximately 2500 nt was detected in flowers, suggesting that there may be more than one gene related to *LjNOD16* in the *L. japonicus* genome. This assumption was supported by the data obtained from our Southern blot analyses. Several hybridizing bands of different intensities were observed under moderately stringent conditions. However, when high stringency hybridization conditions were used, and the 3'-portion of the *LjNOD16* cDNA (PCR5) was employed as a probe, single hybridizing bands were detected, indicating that the *LjNOD16* gene is likely to be represented by a single or low-copy gene number in the *L. japonicus* genome. A full-copy *LjNOD16* cDNA hybridized with several genomic fragments derived from different legume species. This result suggests that the sequences related to the *L. japonicus* *LjNOD16* gene are present in the genomes of different legume plants, and may play a role in both determinate and indeterminate nodules. This assumption was extended by the result of the northern analysis using mRNAs derived from different *M. sativa* tissues. The *LjNOD16* cDNA probe hybridized specifically with two distinct nodule-specific mRNA species, indicating that homologous transcripts are present in the indeterminate nodules of *M. sativa* plants.

Interestingly, the level of alfalfa mRNAs appeared to be significantly lower than the level of *LjNOD16* mRNA in *L. japonicus* nodules. In the former case, 4 μg of poly(A)$^+$ RNA and a long exposure time of 7 to 10 d produced detectable hybridization signals, clearly contrasting with the situation observed in the *L. japonicus* nodules. The apparent low abundance of alfalfa *NOD16* mRNAs is not likely to be due to

the fact that a heterologous *L. japonicus LjNOD16* probe was used, since the same DNA insert gave a clearly detectable signal when hybridized with *M. sativa* genomic DNA (Fig. 6.6B). Although this remains to be proven, it is possible that a different mode of gene regulation may account for the observed difference in the abundance of homologous mRNAs in nodules of *L. japonicus* and *M. sativa* plants. It is noteworthy here that in the case of the *L. japonicus* nodule extract, an array of closely migrating proteins was consistently detected using anti-Nlj16 antibody. Although the basis for this phenomenon is not clear, it may reflect the presence of different forms of Nlj16 protein in the *L. japonicus* nodules, perhaps due to post-translational modifications.

A detailed analysis of the amino acid sequence of the *L. japonicus* protein, deduced from the full-length cDNA sequence, revealed the presence of several interesting domains. Nlj16 was predicted to be a soluble protein, as suggested by Kyte-Doolittle hydropathy analysis. However, Nlj16 contains a hydrophobic domain, consisting of a regularly spaced prolines, the function of which is not clear. Secondary structure predictions strongly indicated the presence of an extended α-helical domain, in which two α-helices are separated by a putative turn. Homology searches using the predicted amino-acid sequence of Nlj16 revealed a significant similarity to two closely related anonymous *Arabidopsis* ESTs mostly in the predicted α-helical domain. Interestingly, both *L. japonicus* and *Arabidopsis* α-helical sequences showed a very high propensity to form a coiled-coil-type structure, which may indicate that the nodulin Nlj16, as well as the *Arabidopsis* proteins, may interact with as yet unknown proteins in the plant cells.

With regard to the size of corresponding mRNAs, the *Arabidopsis* ESTs seem to represent a counterpart of the flower-specific *L. japonicus* mRNA. The mRNA corresponding to the nodule-specific gene is significantly smaller than either *Arabidopsis* or *L. japonicus* flower mRNA species. It is tempting, therefore, to speculate that the nodulin Nlj16 may represent a truncated version of the otherwise constitutively expressed proteins (as exemplified by the ubiquitous expression of the *Arabidopsis* ESTs), primarily harboring the conserved domain, which was adopted during evolution for specific symbiotic function(s). However, the exact role of nodulin Nlj16 in the infected cells of *L. japonicus* nodules remains to be determined.

The *L. japonicus* nodules containing significantly reduced levels of nodulin Nlj16 appear to function normally. However, the residual levels of Nlj16 protein may be sufficient to fulfill its normal function(s) in *L. japonicus* nodules. Therefore to unequivocally address the importance of Nlj16 for nodule development and functioning it is necessary to generate a loss-of-function mutation in the corresponding gene. The over-expression of Nlj16 protein in *L. japonicus* nodules was attempted as a complementary approach to investigate the function of this nodulin. However, an increase in the Nlj16 levels was not achieved. Furthemore, an attempt to over-express Nlj16 in *Arabidopsis* was not successful either. Transgenic *Arabidopsis* plants carrying the *LjNOD16* cDNA under the control of the 35S CAMV promoter failed to produce any detectable quantities of Nlj16 protein despite the fact that *LjNOD16* mRNA was produced at high levels in these plants (data not shown). It is therefore possible that the level of Nlj16 protein is

controlled by post-transcriptional mechanisms, such as regulation of mRNA translation or protein stability.

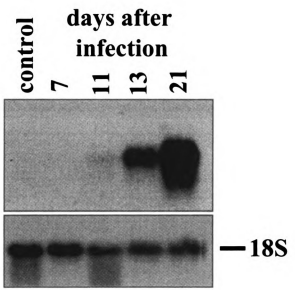


Figure 6.1. Northern blot analysis of PCR5 expression. 10 μg of total RNA isolated from 21-d-old uninfected roots (control), and root segments and nodules harvested 7, 11, 13 and 21 d after infection, were analyzed using the radiolabeled PCR-5 product as a probe (top panel). The blot was re-probed with 18S ribosomal DNA, as an RNA loading control (bottom panel).

Figure 6.2. Nucleotide and deduced amino acid sequences of the *LjNOD16* cDNA. (**A**) The positions of the multiple polyadenylation sites are indicated by the bold underlined nucleotides. Predicted α-helical domains of nodulin Nlj16 are underlined, and the four prolines of the N-terminal hydrophobic stretch are indicated in bold. The asterisk denotes the predicted stop codon.

A

```
CTCGTCGAATTCGGCACGAGGTTGTATATAGATAAAATTATATGAAGATCTTGCAGCTTGTAGGT   65
                                                M  K  I  L  Q  L  V  G
CCTTCTGAGCATATAGAGTTTGTCCCTGCTGCTAAACTTTCCAAGAACGTGGACGTAATCCCTGT  130
P  S  E  H  I  E  F  V  P  A  A  K  L  S  K  N  V  D  V  I  P  V
GGCTATCCCCGTGGGTGTCCCGGTGGCTGTTCCAGCGGCTGACAAAAATGCATCGAAGAAAGTTG  195
 A  I  P  V  G  V  P  V  A  V  P  A  A  D  K  N  A  S  K  K  V  G
GTCAGAATGACACAACGTCCAAAGAGTTTACAACTGTGATGAAACGCATGGCTGAGTTGGAAGAG  260
  Q  N  D  T  T  S  K  E  F  T  T  V  M  K  R  M  A  E  L  E  E
AAAATGACCACCATGAATCATCAGCCTGCTACCATGCCGCCAGAGAAGGAGGAAATGCTGAATGC  325
K  M  T  T  M  N  H  Q  P  A  T  M  P  P  E  K  E  E  M  L  N  A
TACTATAAGTCGAGCGGATGTCTTAGAGAAACAACTTATGGACACCAAGAAGGCTTTGGAGGATT  390
 T  I  S  R  A  D  V  L  I  K  Q  L  M  D  T  K  K  A  L  E  D  S
CGCTTGCTAAGCAAGAGGTGCTTTCAGCTTATGTTGAGAAAAAGAAACAGAAGAAGAAGACGTTT  455
  L  A  K  Q  E  V  L  S  A  Y  V  E  K  K  K  Q  K  K  K  T  F
TTCTGCTGTTAAGTGCGAAATAGTGGATGCCAAACAAGAGGTCCATATATTAAAGGAGAGTTGAC  520
F  C  C  *
TTTTACTTTAAGCTTTTTCAGGAGACTCCCAAAGTGCCCTCTTACCATAAGTGTGCAATATGCGT  585
ACTAGTAAATCTACATGACGCCAAGAAAAGGATATGCTCAATGTGTGAATTGTATATGCTAATCT  650
GAATTTTTGTCTCTACTAGGTTAGGACTACTTGCTCTCTATGTATATAGGGCTATGCACAATATG  715
CAAAGCCAGCCCTGTGATGTCCTTACAGGAACAATACATCTAGTGCAAAATTAGAGAGTATGTTT  780
GATTTTTATATACATGGAAATCTACCTCCAACGAATGAGCTCTTTCCTTTTGCTTATGTGACAAG  845
AAAAAAGCATTACTAGTTGTGAGAAGTACAATTGGCCATATTTGTGTGTGAACTGAGCTGTAAAT  910
GAAGCAAAGTTTTATTTAAATATAGAAAATGTGACTCTTGTAAACATTTGTTCAATTACCACAAT  975
CAGGGGCTGACATGAACTCATGAAGGCTTTTGGCT                                1010
```

B

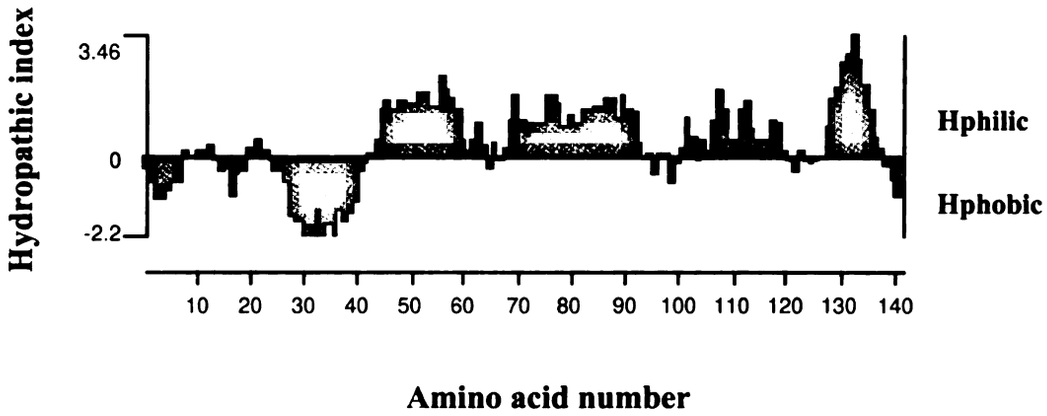


Figure 6.2. Nucleotide and deduced amino acid sequences of the *LjNOD16* cDNA. (**B**) Hydropathy profile of nodulin Nlj16 determined by the method of Kyte and Doolittle (1982) with a window of 7 residues.

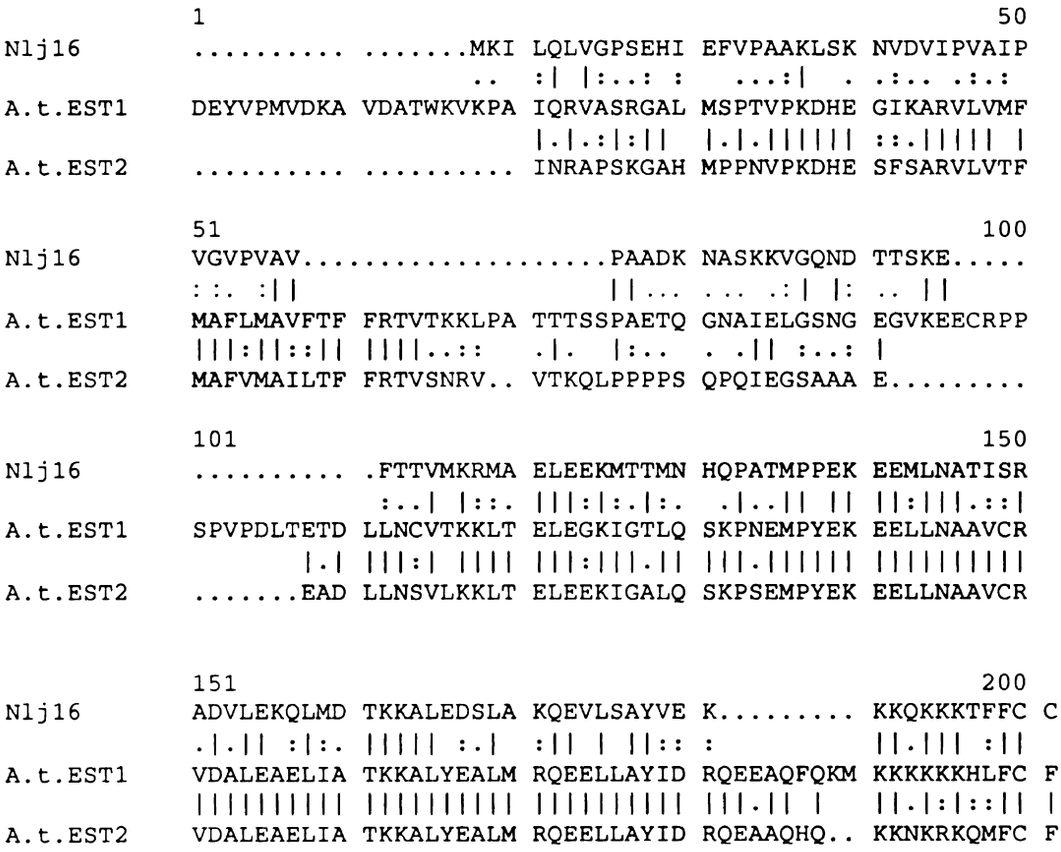


Figure 6.3. Alignment of Nlj16 with the deduced aminoacid sequences of *Arabidopsis* ESTs 168K8 (A.t. EST1) and 110G16 (A.t. EST2). The alignment was performed using the GCG package. Vertical bars indicate identical amino acids, semicolons represent conservative substitutions and dots indicate semiconservative substitutions.

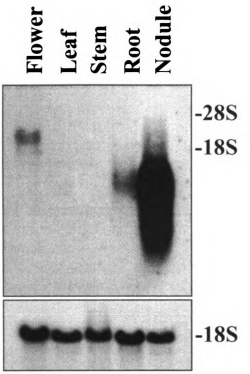


Figure 6.4. Expression of *LjNOD16* in different tissues of *L. japonicus*. Ten micrograms of total RNA from *L. japonicus* flowers, leaves, stems, uninfected roots and nodules were probed with radiolabeled *LjNOD16* cDNA. The blot was exposed for 7 days to visualize the less intensely hybridizing mRNA species from flowers and uninfected roots, respectively (top panel). Equal RNA loading was verified by probing the same filter with the 18S DNA probe (bottom panel). Positions of 28S and 18S rRNAs are as indicated.

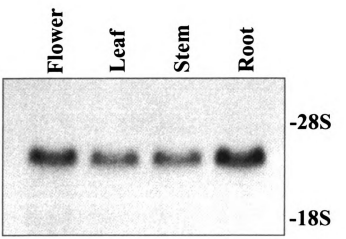


Figure 6.5. Expression pattern of *Arabidopsis* EST 168K8. Ten micrograms of total RNA from flowers, leaves, stems and roots of *Arabidopsis thaliana* were separated by agarose-formaldehyde gel, transferred to nitrocellulose and hybridized to 168K8 DNA labelled with [α-^{32}P]dATP. Positions of 28S and 18S rRNAs are as indicated.

Figure 6.6. Organization of the *LjNOD16* gene in the genomes of *L. japonicus* and other legume species. (A) Ten micrograms of *L. japonicus* genomic DNA were completely digested with the restriction enzymes indicated, separated on 0.8% agarose gel and probed with [α-^{32}P]dATP labelled PCR-5 DNA. Hybridization and washes were performed under high stringency conditions (see Materials and Methods). (B) Southern blot hybridization with *Eco*RI digested DNA (10μg) from *L. japonicus* cv. Gifu, *L. corniculatus* cv. Rodeo, *Medicago sativa* cv. Cardinal and *Glycine max* cv. Dimon. A full length *LjNOD16* cDNA was used as a probe. Hybridization was performed under low stringency conditions (see Materials and Methods). The ~4.0 kb *Eco*RI fragment corresponds to the *LjNOD16* gene.

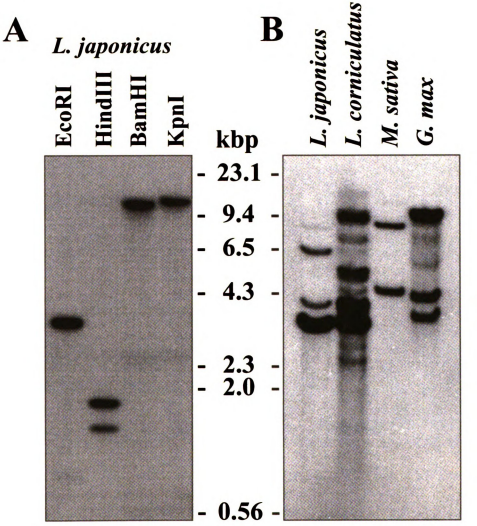
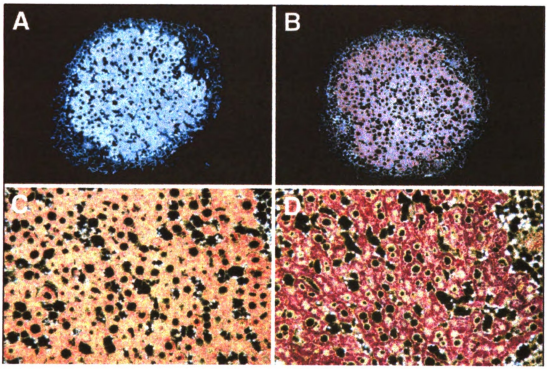
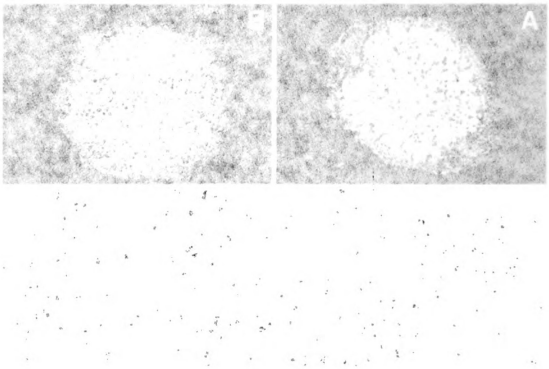

A
L. japonicus
EcoRI
HindIII
BamHI
KpnI
kbp
23.1
9.4
6.5
4.3
2.3
2.0
0.56
B
L. japonicus
L. corniculatus
M. sativa
G. max

Figure 6.7. *In situ* localization of *LjNOD16* and leghemoglobin transcripts in sections of 21-d-old *L. japonicus* nodules. Dark field micrographs are shown in which the hybridization signal appears as pink color. (**A**) Section hybridized with sense *LjNOD16* RNA probe. (**B**) Section hybridized with the anti-sense *LjNOD16* RNA probe. (**C**) Detailed view of a region containing both infected and uninfected cells from the section shown in panel B. (**D**) Detailed view of a region containing both infected and uninfected cells from a section hybridized with anti-sense probe corresponding to the *L. japonicus* leghemoglobin gene. This image is presented in color.


A
B
C
D


A

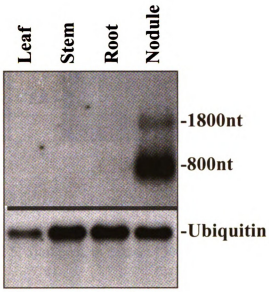


Figure 6.8. Expression of putative homologue(s) of *LjNOD16* in different tissues of *M. sativa*. Four micrograms of poly $(A)^+$ fraction of total RNA isolated from leaves, stems, roots, and nodules of 21-d-old nodulated alfalfa plants were separated on 1.2% agarose-formaldehyde gel, blotted on nitrocellulose, and probed with radiolabelled *LjNOD16* cDNA fragment. Two hybridizing bands with approximate length of 800 nt and 1800 nt, detected under low stringency conditions (see Materials and Methods), are shown. The blot was stripped and rehybridized with an ubiquitin cDNA from *Sesbania rostrata* as a loading control.

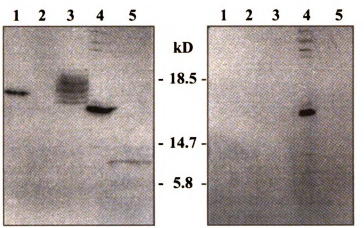


Figure 6.9. Detection of Nlj16 protein and its putative alfalfa homologue in nodules of *L. japonicus* and *M. sativa*. Lane 1, recombinant Nlj16 protein (12.5 ng); lanes 2 and 4, total protein (50 μg) extracted from roots of *L. japonicus* and *M. sativa* plants, respectively; lanes 3 and 5, total protein extract (50 μg) from nodules of *L. japonicus* and *M. sativa* plants, respectively. (A) Western blot analysis with anti-Nlj16 serum. (B). Western blot analysis with pre-immune serum.

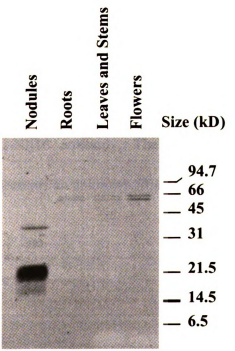


Figure 6.10. Expression of Nlj16 in different tissues of *L. japoniucs*. Equal amount of total protein extracted from nodules, uninfected roots, leaves and stems and flowers of *L. japonicus* were probed with the affinity-purified anti-Nlj16 antibody.

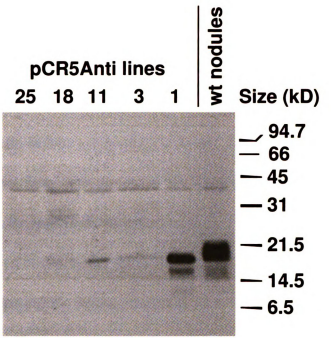


Figure 6.11. The levels of Nlj16 protein in the nodules from the different pCR5Anti transgenic lines of *L. japonicus*. Equal amounts of total nodule protein from the indicated lines and wild-type (wt) nodules were probed with the affinity-purified anti-Nlj16 antibody. The cross-hybridizing protein species migrating at ~40 kD represents an unknown *Rhizobial* protein and can also serve as a loading control.

6.6. ACKNOWLEDGMENTS:

During the course of this work, Krzysztof Szczyglowski visited the laboratory of Dr. Ton Bisseling. We thank Drs. Katharina Pawlowski and Ton Bisseling for their help and many useful suggestions with regard to the *in situ* hybridization procedure. We wish to thank Dr. Marcelle Holsters from the University of Gent, Belgium for kindly providing the ubiquitin clone from *Sesbania rostrata*. We thank Drs. Lee McIntosh, John Wilson, and Leslie Kuhn for helpful discussions. We thank Scott Shaw for the preparation of the plant material used in this work. We also thank Kurt Stepnitz and Marlene Cameron for their help in preparing the figures, Karen Bird for expert editorial assistance, and Susan Fujimoto, David Silver, Dirk Hamburger, and Rujin Chen from our laboratory for their advice and helpful comments on the manuscript.

6.7. REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

Bennett MJ, Lighftoot DA, Cullimore JV (1989) cDNA sequence and differential expression of the gene encoding the glutamine synthetase γ-polypeptide of *Phaseolus vulgaris* L. Plant Mol Biol **12:** 553-565

Bergmann H, Preddie E, Verma DPS (1983) Nodulin-35: a subunit of specific uricase (uricaseII) induced and localized in the uninfected cells of soybean nodules. EMBO J **2:** 2333-2339

Boron L, Szczyglowski K, Konieczny A, Legocki AB (1989) Glutamine synthetase in *Lupinus luteus*. Identification and preliminary characterization of nodule-specific cDNA clone. Acta Biochim Polon **36:** 295-301

Boron L , Legocki AB (1993) Cloning and characterization of a nodule-enhanced glutamine synthetase-encoding gene from *Lupinus luteus*. Gene **136:** 95-102

Brisson NA, Pombo-Gentile A, Verma DPS (1982) Organization and expression of leghemoglobin genes. Can J Bot **60:** 272-278

Broughton WJ, Dilworth MY (1971) Control of leghemoglobin synthesis in snake beans. Biochem J **125:** 1075-1080

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA **81**: 1991-1995

Cook D, Dreyer D, Bonnet D, Howell M, Nony N, VandenBosch K (1995) Transient induction of a peroxidase gene in *Medicago truncatula* precedes infection by *Rhizobium meliloti*. Plant Cell **7:** 43-55

Csanadi G, Szecsi J, Kalo P, Kiss P, Endre G, Kondorosi A, Kondorosi E, Kiss GB (1994) ENOD12, an early nodulin gene, is not required for nodule formation and efficient nitrogen fixation in alfalfa. Plant Cell **6:** 201-213

de Bruijn FJ, Schell J (1992) Regulation of plant genes specifically induced in developing and mature nitrogen-fixing nodules: *cis*-acting elements and *trans*-acting factors. *In* DPS Verma, ed, Control of Plant Gene Expression, CRC Press, Boca Raton, pp 241-258

Delauney AJ, Verma DPS (1988) Cloned nodulin genes for symbiotic nitrogen fixation. Plant Mol Biol Rep **6:** 279-285

Engler JA, Van Montagu M, Engler G (1994) Hybridization in situ of whole mount messenger RNA in plants. Plant Mol Biol Rep **12**: 321-331

Fortin MG, Zalechowska M, Verma DPS (1985) Specific targeting of membrane nodulins to the bacteroid enclosing compartment in soybean nodules. EMBO J **4:** 3041-3046

Fortin MG, Morison NA, Verma DPS (1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. Nucleic Acids Res **15:** 813-824

Franssen HJ, Vijn I, Yang WC, Bisseling T (1992) Developmental aspects of the *Rhizobium*-legume symbiosis. Plant Mol Biol **19:** 89-107

Gebhardt C, Oliver JE, Forde BG, Saarelainen R, Miflin BJ (1986) Primary structure and differential expression of glutamine synthetase genes in nodules, roots and leaves of *Phaseolus vulgaris*. EMBO J **5:** 1429-1435

Govers F, Nap JP, van Kammen A, Bisseling T (1987) Nodulins in the developing root nodule. Plant Physiol Biochem **25:** 309-322

Gresshoff PM (1993) Molecular genetic analysis of nodulation genes in soybean. Plant Breeding Reviews **11:** 275-318

Hall TC, Ma Y, Buchbinder BU, Pyne JW, Sun SM, Bliss FA (1978) Messenger RNA for G1 protein of French bean seeds: cell-free translation and product characterization. Proc Natl Acad Sci USA **75:** 3196-3200

Handberg K, Stougaard J (1992) *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. Plant J **2:** 487-496

Handberg, K., Stiller, J., Thykjear, T., and Stougaard, J. (1994) Transgenic plants: *Agrobacterium* mediated transformation of the diploid legume *Lotus japonicus*. In: Cell Biology: A Laboratory Handbook, Celis JE, ed, 1, 119-127

Heard J, Dunn K (1995) Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc Natl Acad Sci USA **92:** 5273-5277

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

Jarvis BDW, Pankhurst CE, Patel JJ (1982) *Rhizobium loti*, a new species of legume root nodule bacteria. Int J Syst Bact **32:** 378-380

Jiang Q, Gresshoff PM (1993) *Lotus japonicus*: a model plant for structure-function analysis in nodulation and nitrogen fixation. Curr Topics Plant Mol Biol **2:** 97-110

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol **157**: 105-132

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685

Lara M, Cullimore JV, Lea PJ, Johnston AWB, Lamb JW (1983) Appearance of a novel form of plant glutamine synthetase during nodule development in *Phaseolus vulgaris* L. Planta **157:** 254-258

Legocki RP, Verma DPS (1979) A nodule-specific plant protein (Nodulin-35) from soybean. Science **205:** 190-193

Lerouge P, Roche P, Faucher C, Maillet F, Truchet G, Prome JC, Dénarié, J (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulfated and acylated glucosamine oligosaccharide signal. Nature **344:** 781-784

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science **257:** 967-971

Lupas A (1996) Prediction and analysis of coiled-coil structures. Methods of Enzymology **266:** 513-525

Miao G-H, Verma DPS (1993) Soybean nodulin-26 gene encoding a channel protein is expressed only in the infected cells of nodules and is regulated differently in roots of homologous and heterologous plants. Plant Cell **5:** 781-794

Mylona P, Pawlowski K, Bisseling T (1995) Symbiotic nitrogen fixation. Plant Cell **7:** 869-885

Nap J-P, Bisseling T (1990) Developmental biology of a plant-prokaryote symbiosis: the legume root nodule. Science **250:** 948-954

Pridmore RD (1987) New and versatile cloning vectors with kanamycin resistance marker. Gene **56:** 309-312

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. *In*: SB Gelvin, RA Schilperoort, eds, Plant Molecular Biology Manual. Kluwer Academic Publishers, Dordrecht, **A6** 1-10

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol **232:** 584-599

Rost B, Sander C (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins **19:** 55-77

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A Laboratory Manual, ED2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schultze M, Kondorosi E, Ratet P, Buire M, Kondorosi A (1994) Cell and molecular biology of *Rhizobium*-plant interaction. Int Rev Cytol **156:** 1-75

Spaink HP, Wijffelman CA, Pees E, Okker RJH, Lugtenberg BJJ (1987) *Rhizobium* nodulation gene *nodD* as a determinant of host specificity. Nature **328:** 337-340

Spaink HP (1992) Rhizobial lipooligosaccharides: answers and questions. Plant Mol Biol **20:** 977-986

Sprent JI (1989) Which steps are essential for the formation of functional legume nodules? New Phytol **111:**129-153

Stiller J, Martirani L, Tuppale S, Chian R-J, Chiurazzi M, Gresshoff PM (1997) High frequency transformation and regeneration of transgenic plants in the model legume *Lotus japonicus*. J Exp Bot 48: 1357-1365

Szczyglowski K, Szabados L, Fujimoto S, Silver D, de Bruijn FJ (1994) Site-specific mutagenesis of the nodule-infected cell expression (NICE) element and the AT-rich element ATRE-BS2* of the *Sesbania rostrata* leghemoglobin *glb3* promoter. Plant Cell **6:** 317-332

Thummler F, Verma DPS (1987) Nodulin-100 of soybean is the subunit of sucrose synthetase regulated by the availability of free heme in nodules. J Biol Chem **262:** 14730-14736

Towbin H, Staehelin T, Gordon T (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets. Proc Natl Acad Sci USA **76**: 4350-4354

Triplett EW (1985) Intercellular nodule localization and nodule specificity of xanthine dehydrogenase in soybean. Plant Physiol. **77:** 1004-1005

Van de Wiel C, Scheres B, Franssen H, Van Lierop MJ, Van Lammeren A, Van Kammen A, Bisseling T (1990) The early nodulin transcript ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. EMBO J **9:** 1-7

van Kammen A (1984) Suggested nomenclature for plant genes involved in nodulation and symbiosis. Plant Mol Biol Rep **2:** 43-45

Verma DPS (1992) Signals in root nodule organogenesis and endocytosis of *Rhizobium*. Plant Cell **4:** 373-382

Verwoerd TC, Dekker BMM, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res **17:** 2362

Welsh J, Changa K, Dalal SS, Cheng R, Ralph D, McClelland M (1992) Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Res **20:** 4965-4970

CHAPTER 7

Lotus japonicus phosphatidylinositol transfer-like proteins with a plasma membrane targeting domain expressed separately in root nodules

7.1. ABSTRACT

Phosphatidylinositol transfer proteins (PITPs) are the essential components of vital cellular functions in eukaryotes, such as phosphoinositol-mediated signaling and vesicle-traficking. Here the characterization of a gene family from *Lotus japonicus,* encoding a novel class of plant PITP-like proteins, called LjPLPs, is described. The members of this gene family were identified based on their nucleotide sequence homology to a cDNA, *LjNOD16*, encoding *L. japonicus* late nodulin, Nlj16. LjPLPs contain an N-terminal domain related to the yeast sec14p PITP that is also capable of complementing a mutation in the *sec14* gene. In addition, LjPLPs contain a C-terminal domain represented by either Nlj16, or a highly related amino-acid sequence, involved in targeting LjPLPs to the plasma membrane. Furthermore, nodule-specific *LjNOD16* mRNA was shown to be the result of an unusual transcriptional event, mediated by a nodule-specific, bi-directional promoter present in an intron of a member of the *LjPLP* gene familly. Finally, we suggest that nodulin Nlj16 may exert a dominant negative effect in nodules, and discuss its possible role in a mechanism that inactivates the function of LjPLPs.

7.2. INTRODUCTION

Phosphatydilinositol transfer proteins (PITPs) represent an interesting group of eukaryotic proteins that play important roles in diverse cellular processes involving phospholipid metabolism (Cockcroft, 1998; Kearns et al., 1998a). PIPTs were originally discovered based on their unique biochemical properties, such as binding and catalyzing the exchange of phosphatidylinositol (PI) or phosphatidylcholine (PC) moieties between membrane bilayers *in vitro* (Cleves et al., 1991a). The first clue about their physiological functions emerged from genetic studies in *Saccharomyces cerevisiae* (Aitken et al., 1990; Bankaitis et al., 1989; Bankaitis et al., 1990), showing that a mutation (*sec14-1^{ts})* in the gene encoding the major PITP transfer protein resulted in the impairment of secretory function of the Golgi apparatus and seriously affected yeast viability. The *sec14* locus was postulated to have a regulatory role in this process due to its ability to maintain a critical level diacylglycerol (DAG) pool in the Golgi, which, in turn, has been assumed to be required for the formation of secretory vesicles in Golgi membranes (McGee et al., 1994; Kearns et al., 1997). Experimental evidence suggests that yeast sec14p PITP acts as a sensor of phospholipid concentrations in the cell and indirectly helps to preserve a necessary level of DAG in the Golgi, by modulating the rates of biosynthesis or turnover of PC and PI. The PC-bound form of sec14p inhibits the rate-limiting enzyme of the PC biosynthetic pathway, cholinephosphate cytidyltransferase, thus preventing consumption of Golgi-derived DAG (Skinner et al., 1995). On the other hand, the PI-bound form of

Sec14p is believed to enhance the rate of PI turnover, thus increasing DAG levels in Golgi membranes (Kearns et al., 1998a).

Similarly to yeast sec14p, reconstitution studies using permeabilized cells have implicated PITPs in trans-Golgi trafficking events in mammals as well (Ohashi et al., 1995). In addition, PITPs have been postulated to play an important role in cellular mechanisms that compartmentalize and coordinate the synthesis of phosphatidylinositol(4,5)biphosphate (PIP_2). The latter compound (PIP_2) is required for regulated exocytosis in mammalian cells, and has also been shown to constitute a substrate for phospholipase C (PLC) and PI 3-kinase activities. In agreement with the postulated function in PIP_2 biosynthesis, PITPs have been identified as "priming " factors for exocytosis in mammalian cells, as well as crucial components of PLC and PI 3-kinase mediated signaling pathways (Hay and Martin, 1993; Hay et al., 1995; Thomas et al., 1993; Kauffmann-Zeh et al., 1995; Cunningham et al., 1996). One possible function of PITP in these processes could be the transport of PI lipids from the location of their synthesis in the endoplasmic reticulum (ER), to the plasma membrane, where PIP_2 synthesis and hydrolysis take place (Cockcroft, 1998; Kearns et al., 1998a). However, strong evidence also exists that the role of PITPs may be more complex than that of a simple shuttle protein (Cockcroft, 1998; Kearns et al., 1998a). In fact, it has been suggested that PITPs may transfer phospholipids to specific regions within the cellular membranes or to specific protein complexes, mediating differentially regulated signal transduction events (Cockcroft, 1998; Kearns et al., 1998a; see also Discussion).

Plant proteins similar to fungal PITPs have also been identified and characterized. These include the Ssh1p and Ssh2p proteins from soybean and AtSEC14 from *Arabidopsis* (Kearns et al, 1998b; Jouannic et al., 1998). Plant PITPs share approximately 25% amino acid identity and 36-50% similarity with the *S. cerevisiaea* sec14p protein, but do not have any sequence similarity to mammalian PITPs (Kearns et al, 1998b; Jouannic et al., 1998). Ssh2p and AtSEC14 are capable of transfering PI, but not PC, *in vitro*, and thus appear to have different biochemical properties in comparison to the corresponding yeast and metazoan PITPs (Kearns et al, 1998b; Jouannic et al., 1998). Moreover, Ssh1p displayed no detectable PI or PC transfer activity under the conditions tested. Interestingly, Ssh1p was found to undergo phosphorylation in response to various environmental stress conditions, including hyperosmotic stress, suggesting its physiological role in plant osmoprotection (Kearns et al., 1998b).

As a part of an extensive search for expressed sequence tags (ESTs) correlated with late stages of symbiotic root nodule development in *L. japonicus,* we previously identified and characterized a novel nodule-specific cDNA, *LjNOD16*, encoding nodulin Nlj16 (Kapranov et al., 1996; Szczyglowski et al., 1997). Here we report that *LjNOD16* transcripts originate in nodule tissues as a result of unusual transcription events governed by an intron-localized promoter sequence in the *L. japonicus LjPLP*-IV gene. Furthermore, we show that *LjPLP*-IV gene is a member of a gene family encoding a novel class of PITP-like proteins, sharing on average 40% identity and 60% similarity with the *S. cerevisiae* sec14p protein and capable of complementing the temperature-sensitive phenotype of a yeast *sec14-1*^{ts} mutant strain. This new family of PITP-like

proteins can be distinguished from previously described PITPs by the presence of a C-terminal extension comprised of either Nlj16 derived or other, highly related amino-acid sequences. We also present evidence for a role of the Nlj16 C-terminal extension in targeting LjPLPs to the cell plasma membrane. Finally, we suggest that nodulin Nlj16 may exert a dominant negative effect in nodules, and discuss its possible role in a mechanism that inactivates the function of PLPs.

7.3. RESULTS

***LjNOD16* shares a region of similarity with a novel class of PITP-like genes.**

We previously described the isolation and preliminary characterization of a *Lotus japonicus* cDNA, *LjNOD16,* corresponding to a highly abundant mRNA species present in nitrogen fixing root nodules (Kapranov et al., 1997). *LjNOD16* mRNA was localized to the infected, bacteroid containing, cells of *L. japonicus* nodules, and was found to encode a 15.5 kD protein, termed nodulin 16 (Nlj16; Kapranov et al., 1997). Analysis of the predicted amino-acid sequence of nodulin Nlj16 revealed several interesting features, including the presence of two α-helical regions with a high propensity to form coiled-coil structures, and a positively charged C-terminus (Kapranov et al., 1997). The latter domain, together with two cystein residues present at the penultimate and ultimate positions of the Nlj16 protein, showed a striking resemblance to known plasma membrane targeting motifs of the $p21^{K\text{-}ras(B)}$ protein (Hancock et al. 1991). Interestingly, higher molecular weight mRNA species hybridizing to the *LjNOD16* probe were detected in *L. japonicus* flowers (Kapranov et al., 1997). In addition, the deduced amino-acid

sequence of Nlj16 protein was found to share a significant level of similarity with C-terminal regions of much larger proteins represented by two anonymous *Arabidopsis* ESTs (Kapranov et al., 1997). Based on these observations we postulated that nodulin Nlj16, in addition to its specific function in nodules, may constitute a functional module of a much larger protein(s) present in legume and nonlegume plants (Kapranov et al., 1997). Southern analysis, using *LjNOD16* as a probe, confirmed that related genes were present in the *L. japonicus* genome (Kapranov et al., 1997), which we set out to isolate and characterize.

Initially, we focused on the isolation of *LjNOD16*-related cDNA species from *L. japonicus* nodules, and characterized one class of the cDNA clones isolated in detail. The nucleotide sequence of the longest cDNA clone of this class (pCR5h-24) was found to be 2453 bp long. A conceptual translation of this cDNA, as well as a homology search using the BLAST algorithm, revealed that the predicted protein was composed of two distinct amino-acid domains (Fig. 7.1). The C-terminal domain of the protein showed significant similarity to the entire putative coiled-coil domain of nodulin Nlj16 (87% similar and 75% identical on the amino-acid level). The N-terminal domain shared significant level of similarity with a number of PITPs, including the sec14p PITP from *S.cerevisiae* (39% identity and 59% similarity; Bankaitis et al., 1989). Based on this observation we designated this cDNA and its corresponding gene as the *L. japonicus* PITP-like protein gene (*LjPLP*-I; Fig. 7.1).

An unusual feature of the *LjPLP*-I cDNA was the presence of an in-frame TAG stop codon within the PITP-like domain at the codon 222 (TAG_{222}) relative to the first ATG codon. In order to rule out the possibility that the TAG_{222} stop codon was the product of a cloning artifact, corresponding genomic DNA fragment, encompassing the TAG_{222} codon, was PCR amplified from *L. japonicus* genomic DNA via PCR, and the nucleotide sequence of four independent PCR fragments was determined. Two of these fragments were shown to be identical to the corresponding region of the *LjPLP*-I cDNA

and to contain TAG stop codons at equivalent nucleotide positions. The two other genomic PCR products were found to be highly similar but not identical to the *LjPLP*-I cDNA and to contain a CAG codon in place of the TAG_{222} stop codon present in *LjPLP*-I, suggesting the presence of another *LjPLP* gene in the *L. japonicus* genome (*LjPLP*-II; Fig. 7.1).

In order to further characterize the *LjPLP*-II gene and its potential protein product, a 5.5 kb genomic DNA fragment containing the entire *LjPLP*-II coding region was cloned and its DNA sequence determined (see Materials and Methods). The corresponding region was found to be 98% identical to the *LjPLP*-I cDNA at the nucleotide level, and to contain an un-interrupted open reading frame of 550 amino acids, comprising both PITP- and Nlj16-like domains (Fig. 7.1). Taken together, these results show that *LjPLP*-I and -II genes constitute a pair of genes with nearly identical DNA sequence in the *L. japonicus* genome. However, while the *LjPLP*-II appears to encode a PITP-like protein, the *LjPLP*-I gene most likely does not encode a functional protein product due to the presence of the TAG_{222} stop codon within its PITP-like domain(Fig. 7.1).

Southern blot hybridization experiment using a DNA probe corresponding to the *LjPLP*-I PITP domain was used to estimate the complexity of the *LjPLP* gene family in *L. japonicus*. Low-stringency hybridization revealed that the *LjPLP*-I and *LjPLP*-II genes are members of a gene family (data not shown)

A novel class of PITP-like proteins is present in legumes and non-legume plants

After the initial report describing the isolation of *LjNOD16* cDNA (Kapranov et al., 1997), database search using the *LjNOD16* sequence as a query identified a number of plant genes encoding PITP-like proteins containing C-terminal domains highly similar with the putative coiled-coil domain of Nlj16 protein (data not shown). All of these genes, with the exception of a maize EST (accession # AJ006545), were derived from the

Arabidopsis genome sequencing project (e.g., accession # Z99708.1, AC007212.6, AL023094.2, AC006841_29, T08565). In fact, the two Arabidopsis ESTs (168K8 and 110G16), previously shown by us to share similarity to the Nlj16 protein (Kapranov et al., 1997, also see above), were later found to represent PITP-like proteins with the accession # AC006841_29 and T08565, respectively (data not shown). Thus, the Nlj16-type coiled-coil domain, found originally as a major structural component of nodulin Nlj16, turned out to constitute an integral part of what appears to be a novel class of PITP-like proteins. The members of this class are distinguished by the presence of the C-terminal domain, not found in the previously characterized PITP and PITP-like proteins from plants and other eukaryotes (Fig. 7.1; Cockcroft, 1998; Kearns et al., 1997; Kearns et al., 1998; Jouannic et al., 1998). We therefore, refer to this new class of proteins as plant PITP-like proteins (PLPs).

Antisense *LjPLP* transcripts.

To identify additional members of the *LjPLP* gene family, a *L. japonicus* nodule-specific cDNA library was screened with a DNA probe corresponding to the PITP-like domain of the *LjPLP*-I cDNA (see Materials and Methods). Two additional classes of cDNAs were identified (*LjPLP*-III and *LjPLP*-IV; Fig 7.1).

The longest cDNA corresponding to the *LjPLP*-III gene was found to be 2256 bp in length and to contain a 625 amino acids long open reading frame starting with an ATG codon at nucleotide position 77. The predicted structure of the deduced protein product was found to be identical to the overall two-domain composition of the PLPs describe above (Fig. 7.1). Therefore, the *LjPLP*-III cDNA represents another member of the *LjPLP* gene family in *L. japonicus*.

The second class of cDNAs identified were unusual and appeared to correspond to endogenous anti-sense transcripts derived from yet another gene (*LjPLP*-IV; Fig 7.1). A search for open reading frames (ORFs) in *LjPLP*-IV cDNAs revealed the presence of

multiple short ORFs on the non-coding strands, interrupted by intron sequences (see below). Conceptual translation of these ORFs revealed amino acid sequences showing a high similarity to the LjPLPs, and also to PITPs from other plant species (data not shown). The regions of similarity were confined to the PITP-like domains of these proteins (Fig. 7.1). The *LjPLP*-IV cDNAs varied in size (0.5-1.5 kb), in the positions of their 5' and 3' ends, and contained poly(A) sequence (Fig. 7.1).

Differential expression of the *LjPLP* genes

Having identified several different classes of *LjPLP* genes, *LjPLP*-I, -II, -III and -IV, it was of interest to analyze their expression patterns in different *L. japonicus* tissues. The *LjPLP*-I mRNA, containing the TAG_{222} stop codon, could be conveniently distinguished from the *LjPLP*-II mRNA, containing a CAG codon at the equivalent position (CAG_{222}), by digestion with *Eco*57I (recognition site for *Eco*57I is CTTCAG). RT-PCR amplification was carried out using total RNA samples derived from different *L. japonicus* tissues, and a pair of primers (see Materials and Methods) designed to specifically anneal to the *LjPLP*-I or *LjPLP*-II mRNAs, but not to the *LjPLP*-III or -IV mRNAs. These primers were predicted to amplify the 400bp region containing either the TAG_{222} or CAG_{222} codons, respectively. *Eco*57I was expected to cut the RT-PCR product derived from the *LjPLP*-II mRNA in the middle, generating a doublet of approximately equal size bands of 200 bp. The aliquotes of the PCR reactions were separated on an agarose gel side by side with the samples digested with the *Eco*57I endonuclease, transferred to nitrocellulose filter and hybridized with a radiolabelled *LjPLP*-I cDNA probe under stringent conditions (Fig. 7.2). RT-PCR products of the expected length (400bp) that hybridized with the probe were present in undigested control samples derived from *L. japonicus* flowers, roots and nodules, but not from the shoot tissues. Interestingly, only the flower-derived RT-PCR products were digestable with *Eco*57I, generating two hybridizing bands of 400bp and 200bp. These results show that while the

LjPLP-I mRNA is present in different *L. japonicus* tissues, the expression of *LjPLP*-II gene appears to be limited to flowers.

The transcripts hybridizing under stringent conditions to the RNA probe derived from the *LjPLP*-III cDNA could be detected at a similar level in *L. japonicus* flowers, uninfected roots, and nodules, but were absent in the shoot tissues as determined by northern blot analysis (data not shown).

The expression pattern of *LjPLP*-IV gene was investigated using radiolabelled strand–specific RNA probes, complementary to either antisense transcripts or the hypothetical sense transcripts of the *LjPLP*-IV gene. *LjPLP*-IV antisense transcripts were found to be present predominantly in *L. japonicus* nodules (Fig. 7.3). The broad size distribution of hybridizing antisense transcripts is in agreement with the heterogeneity in length observed for the corresponding cDNAs (see above). Interestingly, the apparent sense transcripts of the same gene could be detected in the *L. japonicus* flowers and were found to be represented by a range of closely migrating hybridizing bands (Fig. 7.3). The reason for the size heterogeneity of the *LjPLP*-IV sense transcripts is presently unknown.

The *LjPLP*-IV antisense transcripts and the *LjNOD16* mRNA are derived from the same gene.

We speculated that the antisense *LjPLP*-IV transcripts and the *LjNOD16* mRNA might represent the products of a divergent transcription from a single *L. japonicus* gene (*LjPLP*-IV). This notion was based on the following observations: 1. Both the *LjPLP*-IV antisense and *LjNOD16* transcripts are present almost exclusively in the same *L. japonicus* tissue, the mature nitrogen-fixing nodule. 2. The predicted protein products of the short ORFs present on the minus strands of the antisense transcripts and the Nlj16 protein represent the two domains always found together in the PLP proteins.

Two complementary approaches were undertaken in order to test this hypothesis. First, a λ phage clone, containing the *LjNOD16* gene, was isolated from a *L. japonicus*

genomic DNA library. DNA sequence analysis revealed the presence of a region encoding a PITP-like domain, with nucleotide sequence of the non-coding strand identical to the *LjPLP*-IV anti-sense transcripts, immediately upstream of the *LjNOD16* gene (Fig. 7.4). This observation established a close physical link between the DNA regions encoding the PITP-like and Nlj16 domains in the *L. japonicus* genome. An RT-PCR approach was used to confirm that these two domains are present in a single transcriptional unit. For this purpose, an upstream RT-PCR primer was designed, based on the nucleotide sequence of the genomic region, 77 bp upstream from the presumed ATG initiation codon of the PIPT-like domain, while the downstream primer was derived from nucleotide sequence of the 3'-UTR of the *LjNOD16* mRNA (Fig. 7.4; see also Material and Methods). RT-PCR product of 1903 bp was amplified from total RNA of *L. japonicus* nodule and flower tissues, and was found to contain an ORF of 482 amino acids. A stop codon (TAA) was found to be located 36 bp upstream from a putative ATG initiating triplet suggesting that the RT-PCR product contains the entire coding region of the LjPLP-IV protein. The predicted amino-acid sequence of this protein was found to encompass the PIPT-like domain, sharing ~40% identity and 60% similarity with yeast sec14p PITP, and a C-terminal Nlj16 domain, identical to the amino acids 16-141 of nodulin Nlj16 (Figures 7.4 and 7.5; see also next section). The first exon of the *LjNOD16* mRNA, containing the 5' UTR and the amino acids 1-15 of Nlj16 protein, was found to be derived from the intron 10 of the *LjPLP*-IV gene and was absent from the mRNA encoding the full LjPLP-IV protein (Fig. 7.4; see also next section).

The exon-intron organization of *LjPLP*-IV gene

The exon-intron structure of the coding region of *LjPLP*-IV gene was determined based on the comparison between corresponding nucleotide sequences of the genomic

clone and the product of the RT-PCR amplification. The *LjPLP*-IV gene contains at least 14 exons and 13 introns (Fig. 7.4). The largest intron of the gene (intron number 10), subdivides the predicted protein structure into an N-terminal PIPT-like and a C-terminal Nlj16 domain. Since the transcription of the antisense *LjPLP*-IV RNAs and the *LjNOD16* mRNA initiates within the sequence of the intron 10 of the *LjPLP*-IV gene, we speculated that a bi-directional nodule-specific promoter must be present in this intron. The putative promoter region of the intron was defined as a 581 bp fragment located between the 5' ends of the longest antisense *LjPLP*-IV and *LjNOD16* transcripts (Fig. 7.4). Analysis of the nucleotide sequence of this putative promoter region revealed the presence of several potential regulatory elements (Fig. 7.6). Two TATA-box-like sequences were found approximately 40 bp upstream of the 5' termini of the longest antisense and *LjNOD16* cDNAs. Furthermore, a number of DNA sequence motifs, showing high similarity to nodulin gene consensus sequence, 5'-TTGTCTCTT-3', were present within this putative promoter sequence (Fig. 7.6; Szczyglowski et al., 1994). The latter motifs, and especially the CTCTT core sequences, have been shown to be indispensable for nodule infected-cell-specific expression of late nodulin genes, such as the leghemoglobin genes (Ramlov et al., 1993; Szczyglowski et al., 1994). Similar to TATA box-like sequences, the nodulin-box motifs are located on both strands of the promoter sequence, coinciding with the presumed orientations of the bi-directional gene transcription (Fig. 7.6).

A functional promoter is present within intron 10 of the *L. japonicus LjPLP*-IV gene.

The presence of an active nodule-specific promoter sequence within intron 10 of the *LjPLP*-IV gene was confirmed using a transgenic plant approach. A 581 bp DNA

fragment, encompassing the predicted promoter region of the intron, was fused, in both orientations, to the coding region of a *uidA* reporter gene, encoding β-glucuronidase (GUS). Thus, the p-For construct contained the GUS coding region fused to the *LjNOD16* side of the intron-derived fragment, while in the p-Rev construct the position of this putative promoter region was reversed.

GUS staining of hand-cut nodule sections revealed that the intron fragment directed GUS expression only to the central, infected zone, of the nodules (Fig. 7.7). Other plant tissues, including *L. corniculatus* roots, leaves and flowers, showed no cytological staining for GUS activity (data not shown). The intron sequence was found to be capable of activating the reporter gene expression in an orientation-independent manner (see Fig. 7.7). However, in contrast with the p-For construct reporter gene construct, the p-Rev construct showed a strong histochemical staining also in the nodule vascular bundles (Fig. 7.7). The promoterless *uidA* construct, used as a negative control showed no detectable staining in the central zone of the nodules (Fig. 7.7). However, a relatively weak staining in nodule vascular bundles could be detected in some of these control transgenic lines (data not shown).

The Nlj16 domain contains a functional plasma membrane targeting motif.

A number of proteins are known to be attached to cellular membranes via lipid modification of their C-terminal cysteine residues (reviewed in Zhang and Casey, 1996; Rodriguez-Concepcion et al., 1999). In addition, a polybasic region, present at the C-

terminus of K-Ras(B) protein, has been shown to be required for the localization of this prenylated protein to the inner surface of the plasma membrane (Hancock et al., 1991). Since both of these motifs, namely a polybasic domain of six lysine residues and two terminal cysteine residues, were found to be present at the C-terminal end of the Nlj16 protein we hypothesized that they may also constitute a plasma membrane targeting signal (Figures 7.6 and 7.9). In order to test this hypothesis, cell extracts derived from *L. japonicus* nodules were fractionated by differential centrifugation. The majority of Nlj16 protein was found to be associated with membrane-enriched fractions obtained after differential centrifugation at 10,000g and 100,000g (data not shown). Nlj16 protein could not be detected in the supernatant fraction collected after centrifugation at 100,000g, consistent with a predicted localization of Nlj16 to a cellular membrane compartment. In a parallel experiment, we failed to detect Nlj16 protein in symbiosome- or symbiosome membrane-enriched fractions derived from *L. japonicus* nodule extracts, suggesting that this protein may be anchored to a different membrane compartment in the infected cells of *L. japonicus* nodules (data not shown).

In order to investigate the subcellular localization of Nlj16 and LjPLP-IV proteins in more details, the mGFP5-Nlj16 and mGFP5-LjPLP-IV fusion proteins were transiently expressed in onion epidermal cells. Control cells, expressing mGFP5 alone, displayed a broad distribution of mGFP5-fluorescence in the cell cytoplasm, nucleus, and transvacuolar strands (Fig. 7.8A). In contrast, the C-terminal Nlj16 extension was found to be able to target the mGFP5-Nlj16 chimeric protein to the periphery of the cells only, indicating localization of the Nlj16 protein to the plasma membrane (Fig. 7.8B).

Furthermore, a small subset of transformed onion cells expressing the mGFP5-Nlj16 fusion protein was found to undergo plasmolysis. In these cells the mGFP5 fluorescence was always detected at the periphery of the protoplasts, and was not associated with cell walls or vacuolar membranes (Fig. 7.8D & E). The mGFP5-LjPLP-IV fusion protein was also found to be localized to the plasma membrane in the manner identical to the mGFP5-Nlj16 fusion (data not shown).

The contributions of the polybasic domain and the two C-terminal cystein residues of the Nlj16 protein to targeting the chimeric mGFP5-Nlj16 protein to the plasma membrane were also evaluated. A chimeric protein deprived of the two C-terminal cysteines (mGFP5-Nlj16ΔCC) was localized non-specifically to all accessible intracellular compartments in a manner similar to the mGFP5 alone (Fig. 7.8C). Although required for the proper localization of the chimeric protein, the two cysteine residues were not sufficient for targeting when fused to the C-terminal end of the mGFP5 (data not shown). Instead the localizion of the mGFP5 to the plasma membrane required both, the presence of the polybasic region and the two cysteine residues (mGFP5+ amino acids KKKQKKKTFFCC; data not shown).

As mentioned above, the Nlj16 domain of LjPLP-IV is highly related to the C-terminal domains of the other PLP proteins from L. japonicus and other plant species (Fig. 7.9). The most notable differences include the variations in the nature of the amino acids present at the very C-termini of these proteins, e.g. the CC residues of LjPLP-IV are substituted to the CW residues in LjPLP-III or residues WA in the LjPLP-II (Fig. 7.9). It

is therefore possible that the Nlj16-like domains of the PLP proteins may have different requirements for the plasma mebrane localization. Thus, we evaluated the contribution of the Nlj16-like domain of the LjPLP-III protein (Nlj16-III) to the plasma membrane localization of this protein in the transient expression experiments. As expected, the mGFP5-Nlj16-III fusion was localized to the plasma mebrane in a manner identical to the mGFP5-Nlj16 fusion (Fig. 7.8B, data not shown). Also, the deletion of the last two residues (ΔCW) from the C-terminus of the Nlj16-III domain caused de-localization of the mGFP5-Nlj16-IIIΔCW fusion from plasma membrane to the interior of the onion cells, in a manner identical to the mGFP5-Nlj16ΔCC fusion (Fig. 7.8C, data not shown). However, the C-terminal basic domain of the Nlj16-III region (amino acids RQAEAKLRKKRFCW) was not sufficient to localize the heterologous protein mGFP5 to the plasma membrane (data not shown), in contrast to the analogous region derived from Nlj16 domain (see above).

Functional characterization of the LjPLP-IV protein

Complementation of the growth defects caused by the mutant yeast *sec14* alleles has been a powerful tool to isolate and functionally characterize novel heterologous PITPs (Kearns et al., 1998; Skinner et al., 1993; Tanaka and Hosaka, 1994). Expression of the LjPLP-IV protein from the yeast PGK-promoter failed to rescue a temperature-sensitive phenotype of the CTY1079 yeast strain, carrying *sec14-1*^{ts} allele (Phillips et al., 1999), when grown under non-permissive condition (37°C; construct pGK-IV; Fig. 7.10). Based on our targeting experiments, we postulated that the observed

lack of functional complementation could be due to sequestration of intracellular pool of LjPLP-IV protein to the yeast cell plasma membrane, thereby precluding this protein from substituting for the functions of predominantly cytosolic yeast sec14p protein (Bankaitis et al., 1989). To test this hypothesis, a construct (pGK-IVΔCC) was engineered to express a truncated LjPLP-IVΔCC protein lacking the two C-terminal cysteine residues required for the plasma membrane localization in the plant cells (see above). Expression of this mutant protein could in fact rescue the growth of the CTY1079 yeast strain at the non-permissive temperature (Fig. 7.10). A truncated form of a related protein, LjPLP-III, lacking the entire Nlj16-like domain (LjPLP-IIIΔ) could also complement the temperature sensitive growth of strain CTY1079 (construct pGK-IIIΔ; Fig. 7.10). This result suggests that the Nlj16-like domain is not required for functioning of the LjPLP proteins in yeast. In addition to the *sec14-1*^{ts} allele, the CTY1079 strain also contains a deletion for the phospholipase D (PLD) gene (Phillips et al., 1999), therefore the PLD activity is not required for the observed rescue of the temperature-sensitive phenotype by the LjPLPs.

7.4. DISCUSSION

The biological functions of fungal and animal phosphatidylinositol transfer proteins (PITP) and PIPT-like proteins have been a subject of intensive investigations for more then a decade (reviewed in Cockcroft, 1998; Kearns et al., 1998). On the other hand, plant PITPs, are only beginning to be characterized (Kearns et al., 1998; Jouannic et al., 1998). We report here on identification and characterization of a novel class of plant PITP-like proteins, (PLPs). We show that these proteins are present in both legume and non-legume plants, and describe a set of intriguing observations that suggest that PLP proteins may be connected to the development and/or functioning of *L. japonicus* nodules.

Generally, little is known about the role of plant phospholipid metabolism in the legume-rhizobium symbiosis. Hong and Verma, (1994) have been the only authors to touch on this topic when they describe the expression patterns and biochemical properties of root- and nodule-specific isoforms of soybean PI 3-kinase. PI 3-kinase carries out the phosphorylation of PIP_2, PI, and PI4P, giving rise to 3-phosphorylated inositol lipids that serve as secondry cellular messengers. Interestingly, the expression of the PI 3-kinase gene was shown to be associated with the proliferation of peribacteroid membranes in infected cells of the nodules, suggesting that lipid metabolism may play a role in this process (Hong and Verma, 1994).

We previously identified a highly abundant mRNA species from *L. japonicus* nodules expressed during the developmental transition between nodule ontogeny and the commencement of nitrogen fixation, and its corresponding cDNA (*LjNOD16*; Kapranov

et al., 1997; Szczyglowski et al., 1997). We show here that the *LjNOD16* sequence, encoding late nodulin Nlj16, constitutes a C-terminal portion of a much larger transcriptional unit, namely the *L. japonicus LjPLP*-IV gene. Furthermore, we demonstrate that *LjPLP*-IV is a member of a small gene family in *L. japonicus*, encoding a novel class of phosphatidylinositol transfer-like proteins (LjPLPs).

With exception of the *LjPLP*-I gene, which contains the TAG stop codon within its presumed coding region, the other members of this family isolated thus far (LjPLP II-IV) are predicted to encode PLP proteins that share a two-domain protein structures. The N-terminal portions of these proteins show a significant level of similarity with previously described PITPs from yeast, *Arabidopsis,* and soybean, whereas their C-terminal regions are composed of either NLj16, or highly related amino acid sequences.

The LjPLP-III or IV proteins, when deprived of their plasma membrane targeting signals (see below), are able to complement the temperature sensitive phenotype of yeast *sec14-1*^{ts} mutant. These proteins also contain the conserved amino-acid residues (K85, E227, and K259), known to be necessary for PI binding by the yeast sec14p protein (Sha et al., 1998; Phillips et al., 1999). Furthermore, they are unlike other previously characterized plant PITPs, since they do not require a functional yeast PLD gene for their complementation activity (Phillips et al., 1999). Therefore, it appears that LjPLPs function as phosphatidylinositol transfer protein, in spite of the fact that evidence for their enzymatic function, binding and transport properties, yet remains to be established. We have as yet been unable to isolate sufficient amount of these protein for this purpose, due to their strong toxicity, which prevents their production in *E. coli.*

The presence of Nlj16 or Nlj16-like C-terminal domains of LjPLPs appears to constitute a hallmark of this new class of plant PITP-like proteins. With the exception of the *retinal degeneration-B* (*rdgB*) group of PITPs, which in *Drosophila* have been shown to represent integral membrane proteins, and to contain carboxy-terminal region encompassing six potential transmembrane domains (Vihtelic at el., 1991, 1993), all other previously characterized PITPs are cytosolic, and lack C-terminal extensions (Cockcroft, 1998). We demonstrate here that Nlj16 and Nlj16-III domains of LjPLP-IV and LjPLP-III proteins, respectively, contain functional targeting signals, sufficient for the localization of chimeric GFP-Nlj16 or GFP-Nlj16-III proteins to the plasma membrane of onion epidermal cells. These results are consistent with the predicted localization of Nlj16 to cellular membrane compartments in nodules, and suggest a role of these C-terminal amino-acid sequences in localization of LjPLPs to the plasma membrane.

It is noteworthy that the Nlj16-III domain contains the most divergent amino-acid sequence when compared with the equivalent regions of the other LjPLP proteins. Both the C-terminal polybasic region and two terminal cystein residues (amino-acids 471-482) of the Nlj16 domain are required and sufficient for specific targeting of GFP to the plasma membrane, while the equivalent region of the Nlj16-III domain (amino-acids 612-625) is necessary, but not sufficient. These results indicate that in addition to the polybasic region and the CW motif, the Nlj16-III domain may contain other, as yet not identified signals that are required for plasma membrane targeting activity.

The *L. japonicus LjPLP*-IV gene appears to be the most unusual member of this novel family of phospatidylinositol-transfer-like proteins since it encodes at least three different mRNA species, expressed in *L. japonicus* nodules and/or flowers. A full-length 1.9 kb sense *LjPLP*-IV transcript, encoding the LjPLP-IV protein, accumulates predominantly in *L. japonicus* flowers. The same transcript is also present in nodules, albeit at a significantly diminished level. In contrast two distinct classes of mRNA species derived from the same *LjPLP*-IV gene, anti-sense *LjPLP*-IV transcripts and *LjNOD16*, accumulate to relatively high levels in *L. japonicus* nodules. These transcripts originate as a result of the tissue-specific activity of an internal bi-directional promoter, localized within the largest intron (intron No. 10) of the *LjPLP*-IV gene. By using a transgenic plant approach we were able to show that this intron-derived promoter sequence is sufficient to direct a high level of GUS reporter gene activity specifically to the central, infected cell containing, zone of *L. corniculatus* nodules. The latter observation is entirely consistent with our previous *in-situ* hybridization results, showing a localization of *LjNOD16* mRNA to the infected cell of *L. japonicus* nodules (Kapranov et al., 1996). The nodule-specific nature of this promoter sequence is further indicated by the presence of the nodulin gene promoter consensus-like sequence, CTCTT, which has been shown to be essential for infected-cell specific expression of leghemoglobin (*lb*) genes in nodules (Ramlov et al., 1993; Szczyglowski et al., 1994). Interestingly, the CTCTT motifs, as well as putative TATA box sequences, are present on both strands of the intron-localized promoter region of the *LjPLP*-IV gene. We postulate, therefore, that

these putative regulatory elements determine the bi-directional and tissue-specific nature of the intron-born promoter.

The biological function of LjPLP-proteins remains to be determined. The data presented here strongly suggest that LjPLP proteins are plasma membrane-localized. It is likely that these proteins bind and present phospholipids to plasma membrane-associated proteins or protein complexes, containing enzymes of downstream phospholipid metabolism pathways and/or components of phospholipid-mediated signaling. In this context, it is tempting to speculate that if the recruitment of LjPLPs to a specific location on plasma membrane indeed occurs, it could be mediated, at least in part, by a mechanism involving the α-helical regions of Nlj16 or Nlj16-like domains, since they are predicted to have a high propensity to form coil-coiled structures, and thus are likely to be involved in protein-protein interactions (Kapranov et al., 1997).

Interestingly, in the infected cells of nodules, the Nlj16 domain of *LjPLP*-IV gene is expressed from the intron-borne bi-directional promoter as an independent entity, namely nodulin Nlj16. Since nodulin Nlj16 is targeted to plasma membrane, we speculate that it may occupy (a) specific location(s) therein, normally destined to actively interact with LjPLP proteins, thus preventing the latter from exerting their biological function(s) in this particular subcellular compartment.

The biological role of LjPLPs in nodules remains elusive. However, it has been suggested that soybean PITP (Ssh1p) functions as a component of a stress response pathway, protecting adult plants from osmotic insults (Kearns et al., 1998). Hyperosmotic

conditions have been shown to induce phosphorylation of Ssh1p, and its concomitant mobilization from the plasma membrane to a cytosolic location. In its new cellular location, the cytoplasm, the phosphorylated form of Ssh1, called Ssh1p*, has been proposed to assume an active role in the plant osmoprotective response (Kearns et al., 1998). In this context, it is interesting to consider a possible role of LjPLPs in mechanisms governing osmoregulation in *L. japonicus* nodule tissues. The infected cells of the nodules contain high concentration of sugars, amino-acids, and organic acids that results in four- to five-fold higher osmoticum in these cells (Verma et al., 1978). Consequently, the active employment of osmoregulation may be necessary for optimum functioning of this tissue (Delauney and Verma, 1996), and the LjPLP proteins can potentially be involved in this process. However, the presence of anti-sense *LjPLP*-IV transcripts in nodules argues against this hypothesis. Rather then being actively involved in osmoprotection, the expression of *LjPLPs* genes in nodules seems to be vigorously inactivated. Whether such an inactivation would facilitate, in any way, the osmoprotective response in nodules remains to be established.

There may be yet another explanation for the detection of sense and anti-sense *LjPLP*-IV transcripts in *L. japonicus* nodules. Assuming that both of those transcripts are present in infected cells of the nodules, this could lead to the formation of double-stranded RNAs species (dsRNA). It is tempting to speculate that such dsRNA molecules, in addition to their presumed inhibitory role in *LjPLP* gene expression in nodule, could serve as signaling molecules involved in long distance coordination of plant developmental processes related to symbiotic nitrogen fixation (e.g. signaling between

nodules and flowers), in a manner similar to a phenomenon of posttranscriptional gene silencing (PTGS; Waterhouse et al., 1998). A more detailed comparative analysis of *LjPLP*-IV gene expression in nodules and flowers, and at the different developmental stages, will be required to test this hypothesis.

7.5. MATERIALS AND METHODS

Plant material and growth conditions.

L. japonicus ecotype B-129-S9 Gifu plants were germinated and grown as described previously (Kapranov et al., 1997; Szczyglowski et al., 1997). Nodules, leaves, and stems of *L. japonicus* plants inoculated with *Mezorhizobium loti* strain NZP2235 were harvested 35 days after inoculation (dai), while control uninoculated roots were collected from axenically grown *L. japonicus* plants of the same age. *L. japonicus* flowers were obtained from 2 to 3 month-old plants.

Transgenic *Lotus corniculatus* plants were inoculated with *M. loti* strain 2037 (Jarvis et al., 1982) and subsequently grown in a 6:1 mixture of vermiculate and sand under controlled environmental conditions (18-/6-h day/night cycle, 250 $\mu E\ s^{-1}\ m^{-2}$, 22/18°C day/night temperature). B&D solution (Broughton and Dilworth, 1971), supplemented with 1 mM KNO_3, was used to water these plants. Fully mature nodules, leaves and root segments were harvested from transgenic plants 42-45 dai, and used directly for histochemical analyses.

Screening of *L. japonicus* genomic DNA- and nodule-specific cDNA libraries.

A *L. japonicus* genomic DNA library, and a cDNA library from mature nodules of the same plant species were kindly provided by Dr. Jens Stougaard (Aarhus University, Denmark). The genomic library was constructed in the FIX II λ vector (Stratagene), while the cDNA library was constructed with oligo(dT) primers in the λ-

UniZAP vector (Strategene). Filters carrying the libraries were pre-hybridized and hybridized in a buffer containing 0.5 M sodium phosphate pH 7.2, 7% SDS, and 1% BSA, at 65°C. The filters were washed either at low-stringency (last wash in 2 X SSC, 0.1% SDS at 65°C for 15 minutes), or high-stringency (last wash in 0.1 X SSC, 0.1% SDS at 65°C for 15 minutes) conditions, as specified.

To isolate *LjNOD16*-related genes, *L. japonicus* genomic and cDNA libraries were initially screened under low-stringency conditions, using a 530 bp *Eco*RI-*Hind*II fragment of the *LjNOD16* cDNA, representing the entire coding region of Nlj16, as a probe. Hybridizing phage plaques were also hybridized with a gene-specific probe corresponding to the 3'UTR of *LjNOD16* mRNA (the 370 bp *Hind*II-*Xho*I fragment of *LjNOD16* cDNA) under high-stringency conditions. Plaques hybridizing to the probe derived from the coding region, but not to the gene-specific probe, were assumed to represent *L. japonicus* genes related to, but not identical, to *LjNOD16*. In addition, to isolate *LjPLP*-III and -IV cDNAs, the *L. japonicus* cDNA library was screened with the PIPT-like domain-containing fragment of the *LjPLP*-I cDNA (base pairs 44-1282), under low-stringency conditions.

Nucleic Acid Isolation and Northern Analyses

Genomic DNA, and total RNA from different *L. japonicus* tissues were isolated as described by Kapranov et al., (1997) and Szczyglowski et al., (1997). Northern blot analyses were performed essentially as described (Kapranov et al., 1997; Szczyglowski et al., 1997). For hybridization with strand-specific RNA probes, the filters were pre-

hybridized in 100 mM potassium phosphate buffer pH 6.8, 5X SSC, 1X Denhardt's, 0.1% SDS, 100 μg/ml denatured salmon sperm DNA, at 50°C, for 4 hours. Hybridization was carried out in 70 mM potassium phosphate buffer pH 6.8, 3.6 X SSC, 0.7 X Denhardt's, 7.0 % dextran sulphate, 71 μg/ml denatured salmon sperm DNA, and 50% deionized formamide, at 65°C. The filters were washed for 15 minutes in 2X SSC, 0.1% SDS, 15 minutes in 1X SSC, 0.1% SDS, and 15 minutes in 0.1X SSC, 0.1% SDS, at 65°C.

Radiolabelled RNA probes were prepared as followed. Template DNA (0.5-1 μg), was linearized and incubated in a buffer containing 40 mM Tris pH 7.5, 8 mM $MgCl_2$, 2 mM spermidine, 25 mM NaCl, 10 mM DTT, 40 Units Placental RNAse Inhibitor (BMB), 0.5 mM ATP, 0.5 mM GTP, 0.5 mM CTP, 15 μM UTP, 50 μCi [α-^{32}P]UTP, and 20 units of T3 or T7 RNA polymerase (BMB), in a total volume of 20 μL. The labeling reactions were performed for 1 hr, at 37°C. The DNA template was removed from the reaction mix by adding 10 units of RNAse-free DNAse I (BMB), and incubation at 37°C, for 15 min. Radiolabelled RNA probes were purified on Bio-Spin 6 Chromatography Columns, following the manufacturer's instructions (Bio-Rad).

PCR amplifications and DNA sequencing.

The 400 bp genomic PCR fragments encompassing the regions of the *LjPLP*-I or -II genes containing TAG or CAG codons, respectively, were PCR-amplified from the *L. japonicus* genomic DNA using a pair of common primers, DB588 and DB544 (see Table I). PCR reactions were performed for 30 cycles with the following cycle

profile: 1 min. DNA denaturation step at 94°C, 1 min annealing step at 55 °C, and 1 min. extension step at 72 °C.

A 5.5 kb genomic DNA fragment corresponding to the entire coding region of the *LjPLP* class II gene was PCR-amplified from the corresponding phage lysate using forward (DB562) and reverse (DB561) primers (see Table I). These primers were designed based on the nucleotide sequences of the 5' and 3' UTRs of *LjPLP*-I cDNA, respectively. The PCR was carried out for 30 cycles using *Pfu* DNA polymerase (Promega) under the following cycling conditions: 1 min. denaturation step at 94°C, 1 min. annealing step at 55°C, and 7 min. extension step at 72°C. The PCR product was cloned and sequenced.

RT-PCR procedure.

Total RNA (5 μg) from different *L. japonicus* tissues was denatured for 10 minutes at 65°C, and reverse transcribed for 1 hr at 42°C, in a reaction mix containing 50 mM Tris-HCL pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM of each dNTPs, 40 units of placental RNAse inhibitor (BMB), 100 ng of DB544 primer, and 200 units of Superscript II reverse transcriptase (Gibco), in a total volume of 20 μL. A subsequent PCR amplification step was performed using 4 μL of the original cDNA reaction mix, and 40ng of forward (DB588) and reverse (DB544) primers, in a total volume of 25 μL. The primers (DB588, DB544; see Table I) were designed to amplify a 400bp long DNA fragments encompassing a portion of the PIPT-like domain of both the *LjPLP*-I and -II

genes. The PCR reactions were performed essentially as described above, except that 20 PCR amplification cycles were used. Aliquots of each PCR reaction were digested with *Eco*57I endonuclease, and separated on 1% agarose gel along with the same amount of undigested DNA. The resulting DNA fragments were transferred to a nylon membrane and hybridized to [α^{32}P]-radiolabelled *LjPLP*-I cDNA probe under stringent condition described above. The relative intensities of hybridizing bands were evaluated by using PhosphoImager (Molecular Dynamics).

The RT-PCR approach was also used to amplify a full copy of the *LjPLP*-IV mRNA. The forward N16-4 primer was designed based on the nucleotide sequence of a genomic region position 77bp upstream from the putative ATG codon of the LjPLP-IV protein. The reverse DB641 primer was designed to be complementary to the nucleotide sequence of the 3'UTR of *LjNOD16* cDNA (see Table 1). RT-PCR was carried out using total RNA derived from *L. japonicus* flowers and nodules, essentially as described above, except that two consecutive rounds of PCR amplification were performed. The 1.9 kb PCR product was cloned into the eyast YePlac195PGK expression vector (see below), and its nucleotide sequence was determined.

Chimeric gene constructs and generation of transgenic *L. corniculatus* plants.

Chimeric gene constructs were prepared using standard molecular techniques and the exact details of their construction are available from the authors upon request. Briefly, the 581bp long DNA fragment derived from intron 10 of *LjPLP*-IV gene

was PCR amplified and cloned in both orientations into the unique *Bam*HI restriction site of pBI101 (Clontech) derived binary vector. This resulted in the construction of the p-For and p-Rev binary vectors, carrying the intron sequence in forward (p-For), or reverse (p-Rev) orientation, with respect to the direction of the GUS coding region.

The binary vectors described above were independently transferred into *Agrobacterium rhizogenes* A4 (Tempe and Casse-Delbart, 1989), by using the freeze-thaw method of Hofgen and Willmitzer, (1988). Transgenic *Lotus corniculatus* cv. Rodeo plants were generated as previously described (Szabados et al., 1990, Szczyglowski et al., 1994). GUS activity in the nodule hand sections and other *L. corniculatus* tissues was analysed histochemically (Jefferson et al., 1987; Szczyglowski et al., 1994), using condition described by Malamy and Benfey (1997). Stained tissues were examined using a Wild Heerburgg M420 stereoscope. The images of stained nodule sections were generated using Kodak DC120 digital camera and processed using Adobe Photoshop 5.02 software.

Subcellular localization of mGFP5-Nlj16 fusion in onion epidermal cells

In order to generate an mGFP5-Nlj16 chimeric protein, the *Bgl*II-*Hind*II fragment (470 bp), encompassing almost the entire coding region (except for the first two aminoacids) of *LjNOD16* cDNA, was isolated and blunt ended using Klenow fragment of Polymerase I. The fragment was then fused, in-frame, to the C-terminal end of the coding region of mGFP5 (Siemering et al., 1996) in vector pAVA393. The latter is based on the vector pAVA319 (von Arnim et al., 1998), in which the *gfp* cDNA was replaced with the

cDNA encoding mGFP5, and kindly provided by Dr. von Arnim, the University of Tennessee, Knoxville. The mGFP5-Nlj16ΔCC fusion was constructed in a similar manner as mGFP5-Nlj16, except that a 464 bp *Bgl*II-*Bbs*I fragment of *LjNOD16* cDNA, deprived of the two C-terminal codons encoding cystein residues, was used.

To generate mGFP5 protein with either two C-terminal cysteins residues (mGFP5+CC) or KKKQKKKTFFCC sequence (mGFP5+KCC), the corresponding nucleotide sequences of the *LjNOD16* cDNA were PCR-amplified using two sets of specific primer pairs, NOD16-1/NOD16-5 and NOD16-2/NOD16-5, respectively (see Table I). The resulting PCR products were digested with *Bam*HI-*Xba*I and cloned into vector pAVA393. The Nlj16-like domain of the LjPLP-III protein (amino acids 513-612) and its derivatives were used to construct the fusions mGFP5-Nlj16(III), mGFP5-Nlj16(III)ΔCW and mGFP5-basic domain of Nlj16(III), in a manner similar to the constructs described above.

Plasmid DNA from each constructs was delivered into the onion epidermal cells using 1.6 μM gold particles and the BIOLISTIC PDS-1000 gene transformation system (Dupont), essentially as described by Varagona et al., (1992). The bombarded onion epidermal explants were incubated for 18-24 hrs on light, on a solid MS media (Gibco) containing 30 g/L sucrose and 180 mg/L KH_2PO_4. Fluorescence of mGFP5 protein fusions was analyzed using a Zeiss Axiophot epifluorescent microscope under a filter with following parameters: emission at 470±20 nm, beam splitter at 510 nm and excitation at 540±20 nm. Images were obtained using a Kodak DC120 digital camera and

Adobe Photoshop 5.02 software. Each transfection experiment was repeated at least two times.

Yeast complementation experiments

The yeast strain CTY1079 (*MAT***a** *ura3-52, lys2-801, Δhis3-200, sec14-1*^{ts} Δ*spo14::HIS3*; Phillips et al., 1999) was used for the complementation experiments. The YePlac195PGK yeast expression vector was created by cloning a 2 kb *Hind*III fragment containing the yeast PGK promoter cassette from pHVX2 (Volschenk et al., 1997) into the *Hind*III site of the YePlac195 plasmid (Gietz and Sugino, 1988). A 1.9 kb RT-PCR product containing the entire LjPLP-IV coding region (see above) was cloned into YePlac195PGK under the control of the PGK promoter to generate construct pGK-IV. To remove the two cysteine residues from the C-terminus of LjPLP-IV protein, the 1.5 kb region of the *LjPLP*-IV cDNA was amplified using primers SFX and SRX. The PCR product was inserted into the YePlac195PGK vector, downstream of the PGK promoter to generate construct pGK-IVΔCC. To express a truncated LjPLP-IIIΔ protein, construct pGK-IIIΔ was generated as follows. A 1.6 kb DNA fragment was PCR amplified from the *LjPLP*-III cDNA with the primer pair DB637-DB638 and inserted into vector YePlac195PGK downstream of the PGK promoter.

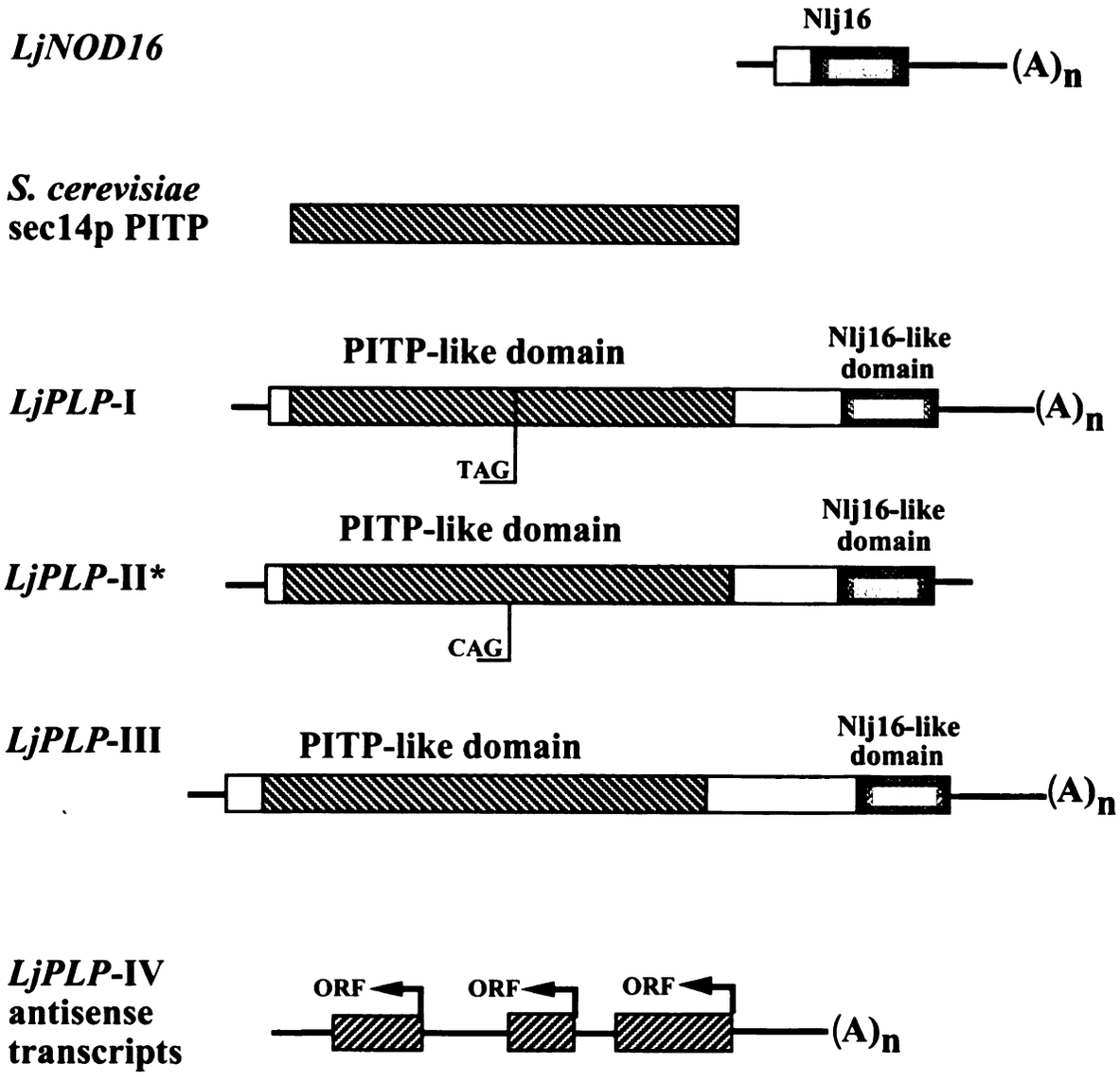
Plasmids YePlac195PGK, pGK-IV, pGK-IVΔCC or pGK-IIIΔ were introduced into the yeast strains following the procedure of Elble, (1992). CTY1079 transformants were initially selected at 28°C on defined yeast media lacking uracil. Complementation of

the temperature-sensitive (ts) phenotype of the *sec14-1*^{ts} allele in the strain CTY1079 was evaluated after 4 days of growth on selective media at 37°C. Five independent CTY1079 transformants containing each plasmid were evaluated.

Table 7.1. The nucleotide sequences of the primers used during different amplification procedures.

DB568	GAACTTCAACAAACATGCCAG
DB544	CAAGCAATTTGCTTTGATAC
DB561	GGGAAGTAGCATTTGGAAAGC
DB562	CATATTAAAATTCAGCAGAAGC
DB588	GTGTACATTGAGAATATAGGC
DB640	GACGACCCGTGTACATAGAGC
DB641	CTTGTCACATAAGCAAAAGG
N16-5'-4	GGGAGTGCTTTTGTTCTCTGC
NOD16-1	GAAGAAGACGGGATCCTGCTGTTAAGTG
NOD16-2	TCAGCTTATGGATCCAAAAAGAAAC
NOD16-5	CACACTTATGTCTAGAGGGCACTTTG
SRX	ATTTCGCACTCGAGATTAGAAAAACGTC
SFX	CTTTTGTTCTCGAGCTCGCATGAT
DB637	CATGATTAGATGTTATTGCAGG
DB638	CAACGCTGCACAAATTCTAGG

Figure 7.1. Schematic diagrams of *LjPLP* cDNAs. Boxes represent the coding regions, while lines correspond to the 5' and 3' UTRs of the cDNAs. The stippled boxes represent either PITP or PITP-like domains, the dotted boxes denotes the Nlj16-like domains, and the white boxes correspond to the regions with no apparent sequence similarity between different cDNAs. The positions of the stop TAG_{222} codon in *LjPLP-I*, and the corresponding CAG_{222} codon in *LjPLP*-II genes, are shown. The asterisk indicates the fact that the *LjPLP*-II cDNA sequence was deduced from the nucleotide sequence of the corresponding *L. japonicus* genomic region. The short open reading frames (ORFs), present in the *LjPLP*-IV antisense transcripts, are represented by the stippled boxes, and the intron sequences are represented by lines. Arrows indicate the direction of ORFs.


LjNOD16
Nlj16
(A)n
S. cerevisiae
sec14p PITP
PITP-like domain
Nlj16-like
domain
LjPLP-I
TAG
(A)n
PITP-like domain
Nlj16-like
domain
LjPLP-II*
CAG
LjPLP-III
PITP-like domain
Nlj16-like
domain
(A)n
LjPLP-IV
antisense
transcripts
ORF
ORF
ORF
(A)n

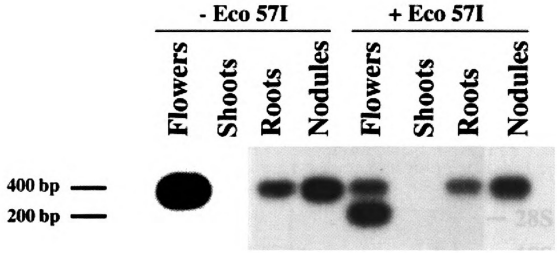


Figure 7.2. Tissue-specific expression of the *LjPLP*-I and *LjPLP*-II genes. The 400 bp fragments, encompassing either UAG_{222} stop codon-, or CAG_{222} triplet-containing regions of *LjPLP*-I or *LjPLP*-II mRNAs, respectively, were amplified using RT-PCR approach. An aliquot of each PCR reaction was digested with Eco 57I (+Eco 57I), separated on agarose gel, along with the undigested control sample (-Eco 57I), and hybridized to the radiolabelled *LjPLP*-I cDNA probe. The 400 and 200 bp fragments of the digested samples correspond to *LjPLP*-I and *LjPLP*-II transcript, respectively.

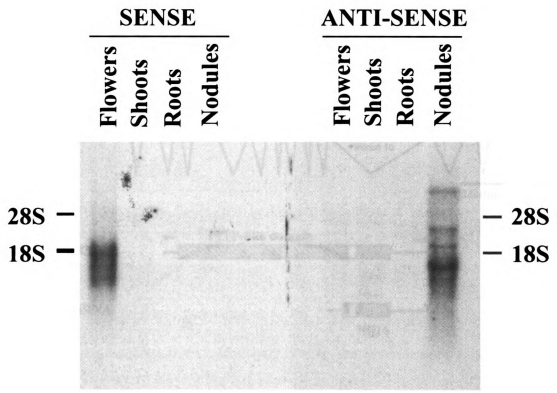


Figure 7.3. Tissue-specific expression of the *LjPLP*-IV transcripts. Ten micrograms of total RNA derived from *L. japonicus* flowers, leaves, uninoculated roots, and fully mature nodules, was separated on the formaldehyde-agarose gel and hybridized with radiolabeled, strand-specific, RNA probes, complementary to the presumed sense or antisense transcripts of the *LjPLP*-IV gene. The positions of 28S and 18S ribosomal RNAs are indicated.

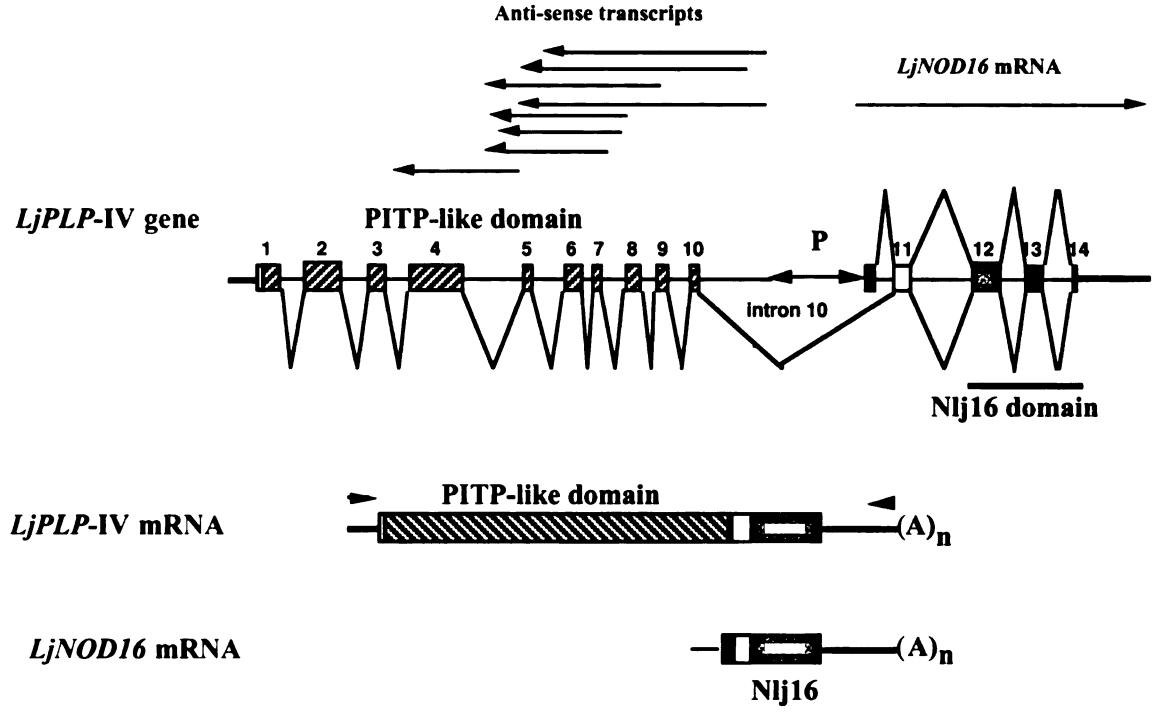
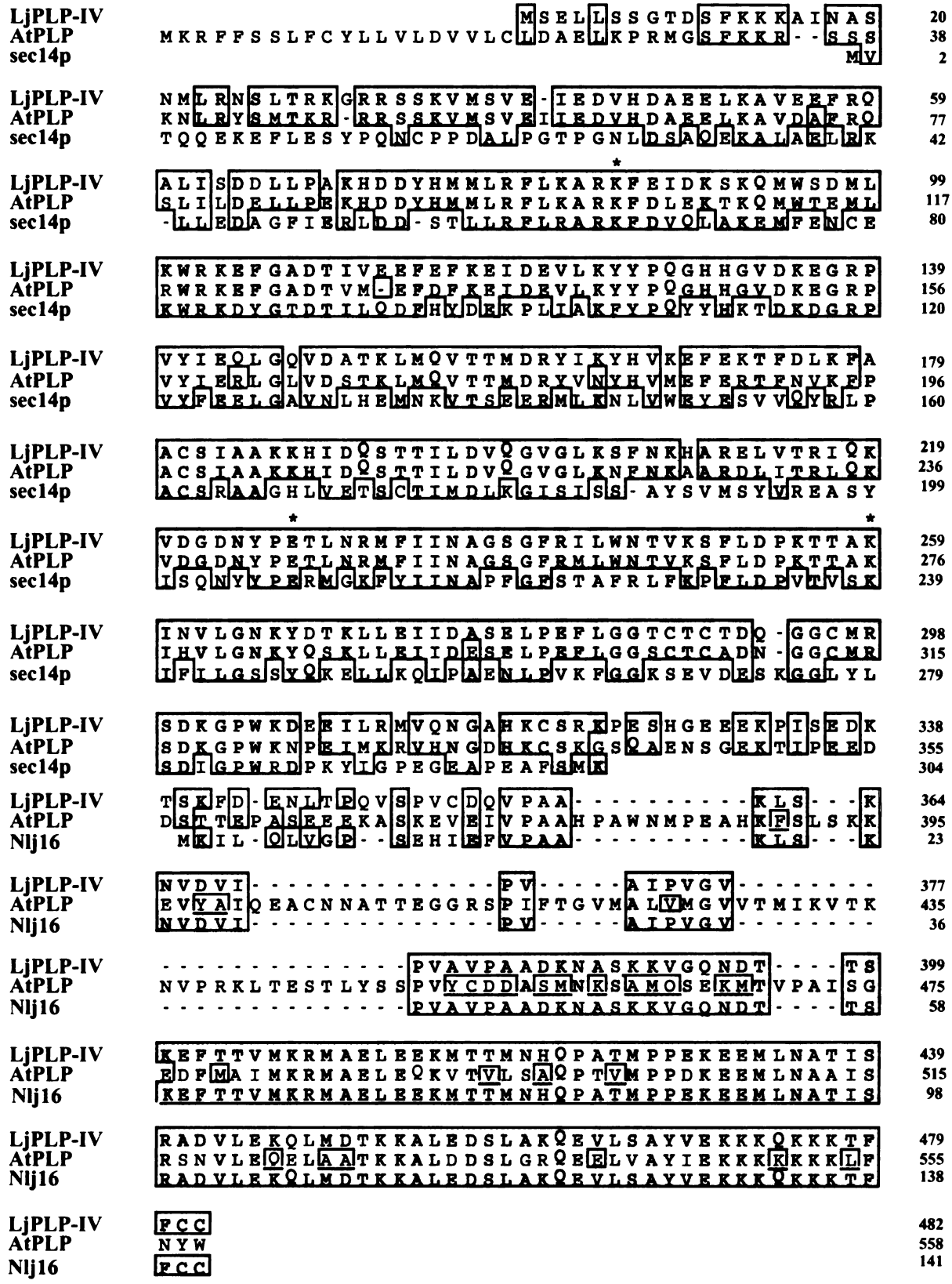


Figure 7.4. Schematic diagrams of *LjPLP*-IV genomic region, and *LjPLP*-IV sense- and *LjNOD16* transcripts. Exons are numbered and represented by boxes, introns are symbolized by the thin lines, whereas thick lines indicate the 5' and 3' UTRs. Differently shaded boxes correspond to either PIPT-like- or Nlj16 domains, as describe in the legend to Figure 7.1. The solid box represents the first exon of the *LjNOD16* transcript. The position of different cDNA clones, corresponding to either *LjPLP*-IV antisense- or *LjNOD16* transcripts, is indicated by arrows above the diagram of the *LjPLP*-IV gene. The localization of the bi-directional promoter (P) within the intron 10 is indicated. The positions of primers used to amplify the *LjPLP*-IV mRNA are indicated by arrowheads.

Fig. 7.5. Amino acids sequence alignment between *L. japonicus* LjPLP-IV, nodulin Nlj16, *Arabidopsis* AtPLP (Acc. No. Z99708.1), and yeast sec14p proteins. The alignment was generated using the PileUp function of the Genetics Computer Group package (GCG, Madison, Wisconsin) and the SeqVu software version 1.0.1. Similar residues are boxed, while the identical aminoacids are boxed and shaded. Dashes represent gaps in the sequences, and asterisks indicate the conserved amino-acid residues required for PI binding activity of yeast sec14p protein (Sha et al., 1998).


LjPLP-IV
AtPLP
sec14p
Nlj16

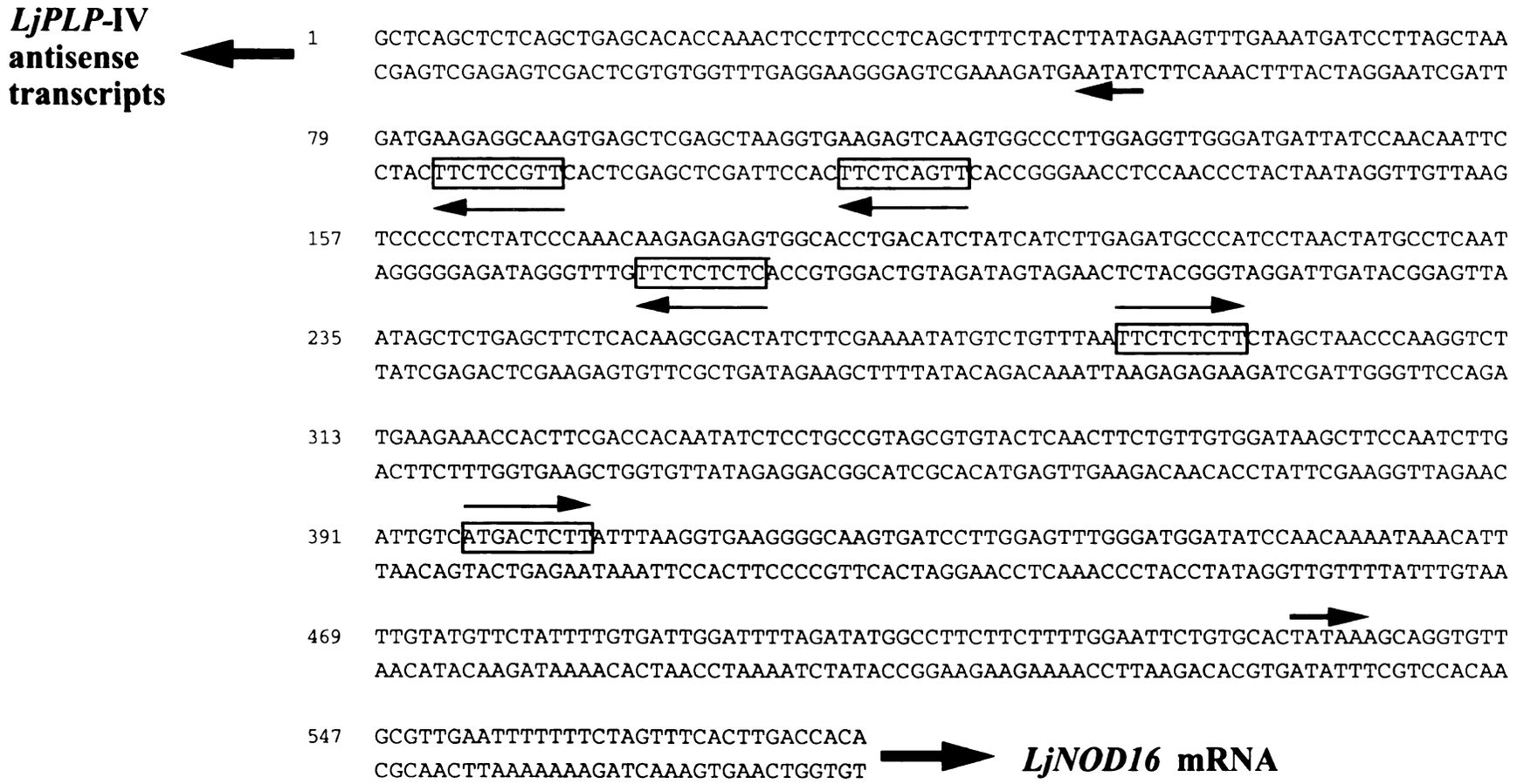
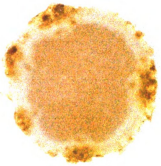
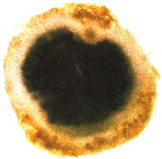


Fig. 7.6. The nucleotide sequence of the intron-localized bi-directional promoter of the *LjPLP*-IV gene. The directions of transcription of the *LjPLP*-IV antisense and *LjNOD16* transcripts are indicated by arrows. The position and presumed orientations of TATA-box-like sequences and nodulin-gene- like motifs are indicated.

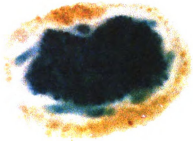
Fig. 7.7. The internal bi-directional promoter of the *LjPLP*-IV gene directs the expression of the GUS reporter gene to the central zone of *L. corniculatus* nodules. Histochemical localization of β-glucuronidase activity in nodule hand-sections derived from transgenic *L. corniculatus* plants harboring the promoter-GUS fusions. This image is presented in color.



control

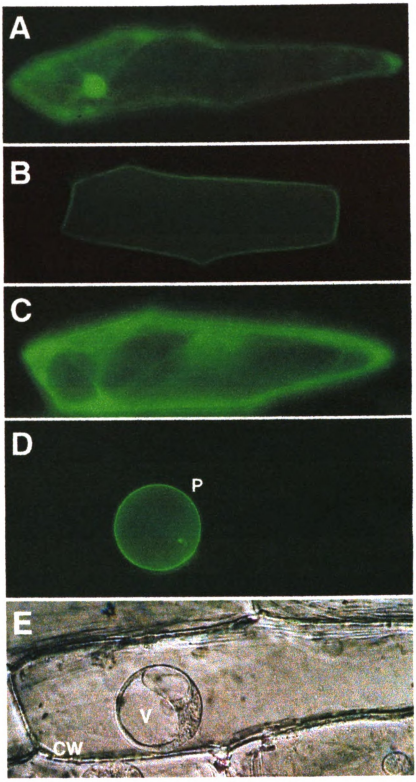


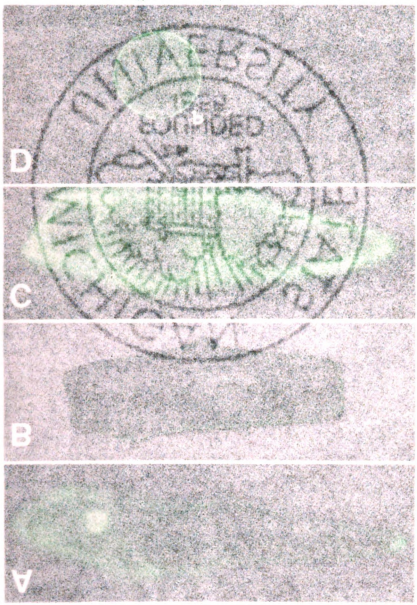
p-For



p-Rev

Fig. 7.8. Subcellular localization of the mGFP5-Nlj16 fusions in the onion epidermal cells. Fluorescent images of onion cells expressing mGFP5 alone (**A**), mGFP5-Nlj16 fusion (**B** & **D**) and mGFP5-Nlj16ΔCC fusion (**C**). (**E**) Bright field image of the cell shown in the panel D. The onion cell shown in the panels D & E has undergone plasmolysis and the GFP fluorescence can be seen on the periphery of the protoplast. This image is presented in color.


A
B
C
D
P
E
V
CW


D
C
B
A

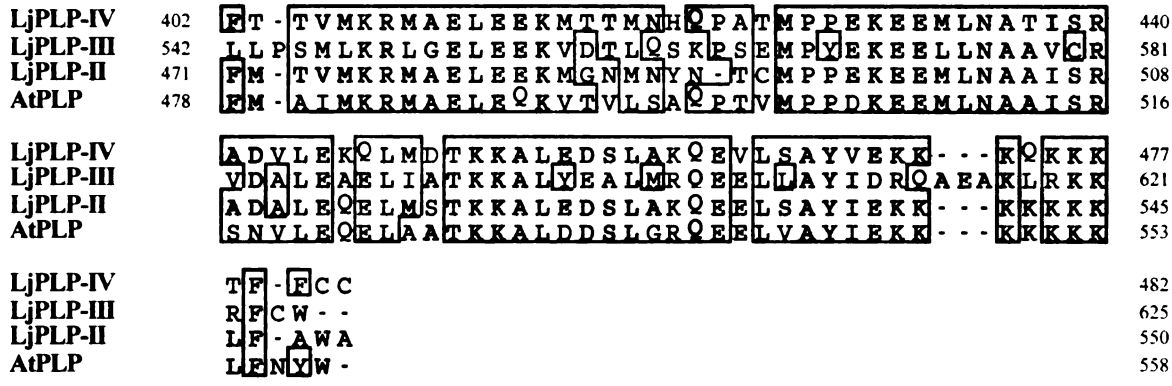
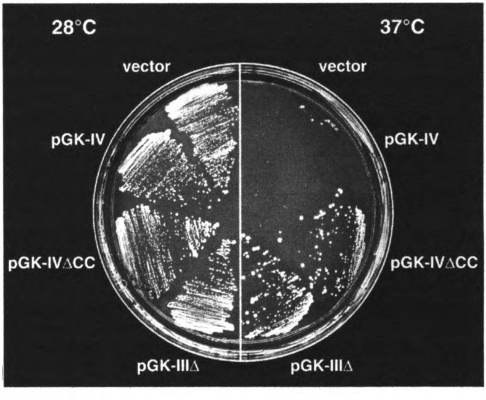
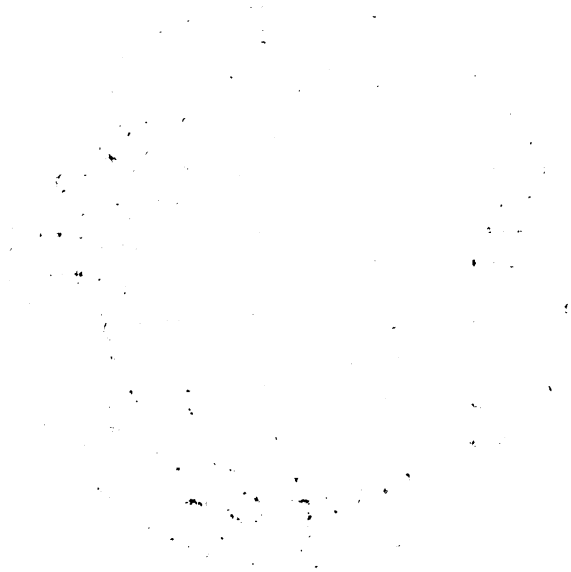


Fig. 7.9. Amino acid sequence alignment of the C-terminal Nlj16-like domains of LjPLP II-IV and Arabidopsis AtPLP (accession # Z99708.1) proteins. The alignment was generated using the PileUp function of the Genetics Computer Group package (GCG, Madison, Wisconsin) and the SeqVu software version 1.0.1. Similar residues are boxed, while the identical aminoacids are boxed and shaded. Dashes represent gaps in the sequences.

Fig. 7.10. The LjPLP proteins complement the temperature-sensitive phenotype of yeast *sec14* mutant. The growth phenotype of the yeast strain CTY1079 (*sec14-1^{ts} Δspo14*) harboring the indicated constructs is shown (see Materials and Methods). Individual transformants were streaked on a selective (-Ura) media and grown at permissive 28°C and non-permissive 37°C temperatures.


28°C
37°C
vector
vector
pGK-IV
pGK-IV
pGK-IVΔCC
pGK-IVΔCC
pGK-IIIΔ
pGK-IIIΔ



7.6. REFERENCES:

Aitken,J.F., van Heusden,G.P.H., Temkin,M. and Dowhan,W. (1990) The gene encoding the phosphatidylinositol transfer protein is essential for cell growth. *J. Biol. Chem.*, **265,** 4711-4717.

Bankaitis,V.A., Aitken,J.F., Cleves,A.E. and Dowhan,W. (1990) An essential role for a phospholipid transfer protein in yeast Golgi function. *Nature*, **347,** 561-562.

Bankaitis,V.A., Malehorn,D.E., Emr,S.D. and Greene,R. (1989) The *Saccharomyces cerevisiae SEC14* gene encodes a cytosolic factor that is required for transport of secretory proteins from the yeast Golgi complex. *J. Cell. Biol.*, **108,** 1271-1281.

Broughton,W.J. and Dilworth,M.Y. (1971) Control of leghemoglobin synthesis in snake beans. *Biochem. J.*, **125,** 1075-1080.

Cockcroft,S. (1998) Phosphatidylinositol transfer proteins: a requirement in signal transduction and vesicle traffic. *BioEssays*, **20,** 423-432.

Cunningham,E., Tan,S.W., Swigart,P., Hsuan,J., Bankaitis,V. and Cockcroft,S. (1996) The yeast and mammalian isoforms of phosphatidylinositol transfer protein can all restore phospholipase C-mediated inositol lipid signaling in cytosol-depleted RBL-2H3 and HL60 cells. *Proc. Natl. Acad. Sci. USA*, **93,** 6589-6593.

Delauney,A.J. and Verma,D.P.S. (1996) Improvement of soybean for nitrogen fixation: molecular genetics of nodulation. In Verma,D.P.S. and Shoemaker,R.C. (eds.), *Soybean Genetics, Molecular Biology and Biotechnology*. CAN International, Wallingford, UK, pp. 219-248.

Dickeson,S.K., Lim,C.N., Schulyer,G.T., Dalton,T.P., Helmkamp,G.M.,Jr. and Yarbrough,L.R. (1989) Isolation and sequence of cDNA clones encoding rat phosphatidylinositol transfer protein. *J. Biol. Chem.*, **264,** 16557-16564.

Elble,R. (1992) A simple and efficient procedure for transformation of yeasts. *Biotechniques*, **13,** 18-20.

Gietz,R.D. and Sugino,A. (1988) New yeast-*Escherichia coli* shuttle vectors constructed with *in vitro* mutagenized yeast genes lacking six-base pair restriction sites. *Gene*, **74,** 527-534.

Hara,S., Swigart,P., Jones,D. and Cockcroft,S. (1997) The first 5 amino acids of the carboxy terminus of phosphatidylinositol transfer protein α (PITPα) play a critical role in inositol lipid signaling: transfer activity of PITP is essential but not sufficient for restoration of phospholipase C signaling. *J. Biol. Chem.*, **272,** 14909-14913.

Hay,J.C., Fisette,P.L., Jenkins,G.H., Fukami,K., Takenawa,T., Anderson,R.E. and Martin,T.F.J. (1995) ATP-dependent inositide phosphorylation required for Ca2+-activated secretion. *Nature*, **372,** 173-177.

Hay,J.C. and Martin,T.F.J. (1993) Phosphatidylinositol transfer protein required for ATP-dependent priming of Ca^{2+}-activated secretion. *Nature*, **366,** 572-575.

Hong,Z. and Verma,D.P.S. (1994) A phosphatidylinositol 3-kinase is induced during soybean nodule organogenesis and is associated with membrane proliferation. *Proc. Natl. Acad. Sci. USA*, **91,** 9617-9621.

Jouannic,N., Lepetit,M., Vergnolle,C., Cantrel,C., Gardies,A.-M., Kader,J.-C. and Arondel,V. (1998) Isolation of a cDNA from *Arabidopsis thaliana* that complements the *sec14* mutant of yeast. *Eur. J. Biochem.*, **258,** 402-410.

Kapranov,P., de Bruijn,F.J. and Szczyglowski,K. (1997) Novel, highly expressed late nodulin gene (*LjNOD16*) from *Lotus japonicus*. *Plant Physiol.*, **113,** 1081-1090.

Kauffmann-Zeh,A., Thomas,G.M.H., Ball,A., Prosser,S., Cunningham,E., Cockcroft,S. and Hsuan,J.J. (1995) Requirement for phosphatidylinositol transfer protein in epidermal growth factor signaling. *Science*, **268,** 1188-1190.

Kearns,B.G., Alb,J.G.,Jr., and Bankaitis,V.A. (1998a) Phosphatidylinositol transfer proteins: the long and winding road to physiological function. *Trends Cell Biol.*, **8,** 276-282.

Kearns,M.A., Monks,D.E., Fang,M., Rivas,M.P., Courtney,P.D., Chen,J., Prestwich,G.D., Theibert,A.B., Dewey,R.E. and Bankaitis,V.A. (1998b) Novel developmentally regulated phosphoinositide binding proteins from soybean whose expression bypasses the requirements for an essential phosphatidylinositol transfer protein in yeast. *EMBO J.*, **17,** 4004-4017.

Kearns,M.A., McGee,T.P., Mayinger,P., Gedvilaite,A., Phillips,S.E., Kagiwada,S. and Bankaitis,V.A. (1997) An essential role for diacylglycerol in protein transport from the yeast Golgi complex. *Nature*, **387,** 101-105.

Lopez,M.C., Nicaud,J.-M., Skiner,H.B., Vergnolle,C., Kader,J.-C., Bankaitis,V.A. and Gaillardin,C. (1994) A phoshatidylinositol/phosphatidylcholine transfer protein is required for differentiation of the dimorphic yeast *Yarrowia lipolytica* from the yeast to the mycelial form. *J. Cell. Biol.*, **125,** 113-127.

Malamy,J.E. and Benfey,P.N. (1997) Organization and cell differentiation in lateral roots of *Arabidopsis thaliana. Development*, **124,** 33-44.

McGee, T.P., Skinner,H.B., Whitters,E.A., Henry,S.A. and Bankaitis,V.A. (1994) A phosphatidylinositol transfer protein controls the phosphatidylcholine content of yeast Golgi membranes. *J. Cell. Biol.*, **124,** 273-287.

Ohashi,M., de Vries,J.K., Frank,R., Snoek,G., Bankaitis,V., Wirtz,K. and Huttner,W.B. (1995) A role for phosphatidylinositol transfer protein in secretory vesicle formation. *Nature*, **377,** 544-547.

Phillips,S.E., Sha,B., Topalof,L., Xie,Z., Alb,J.G., Klenchin,V.A., Swigart,P., Cockcroft,S., Martin,T.F.J., Luo,M. and Bankaitis,V.A. (1999) Yeast sec14p deficient in phosphatidylinositol transfer activity is functional *in vivo*. *Mol. Cell*, **4,** 187-197.

Ramlov,K.B., Laursen,N.B., Stougaard,J. and Marcker,K.A. (1993) Site-directed mutagenesis of the organ-specific element in the soybean leghemoglobin *lbc3* gene promoter. *Plant J.*, **4,** 577-580.

Rodriguez-Concepcion,M., Yalovsky,S., Zik,M., Fromm,H. and Gruissem,W. (1999) The prenylation status of a novel plant calmodulin directs plasma membrane or nuclear localization of the protein. *EMBO J.*, **18,** 1996-2007.

Sha,B., Phillips,S.E., Bankaitis,V.A. and Luo,M. (1998) Crystal structure of the *Saccharomyces cerevisiae* phosphatidylinositol-transfer protein. *Nature*, **391,** 506-510.

Siemering,K.R., Golbik,R., Sever,R. and Haseloff,J. (1996) Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr. Biol.*, **6,** 1653-1663.

Skinner,H.B., McGee,T.P., McMaster,C.R., Fry,M.R., Bell,R.M. and Bankaitis,V.A. (1995) Phosphatidylinositol transfer protein stimulates yeast Golgi function by inhibiting choline-phosphate cytidyltransferase activity. *Proc. Natl. Acad. Sci. USA*, **92,** 112-116.

Szczyglowski,K., Hamburger,D., Kapranov,P. and de Bruijn,F.J. (1997) Construction of a *Lotus japonicus* late nodulin expressed sequence tag library and identification of novel nodule-specific genes. *Plant Physiol.*, **114,** 1335-1346.

Szczyglowski,K., Szabados,L., Fujimoto,S.Y., Silver,D. and de Bruijn,F.J. (1994) Site-specific mutagenesis of the nodule-infected cell expression (NICE) element and the AT-rich element ATRE-BS2* of the *Sesbania rostrata* leghemoglobin *glb3* promoter. *Plant Cell*, **6,** 317-332.

Thomas,G.M.H., Cunningham,E., Fensome,A., Ball,A., Totty,N.F., Troung,O., Hsuan,J.J. and Cockcroft, S. (1993) An essential role for phosphatidylinositol transfer protein in phospholipase C-mediated inositol lipid signaling. *Cell*, **74,** 919-928.

Vanhee-Brossollet,C. and Vaquero,C. (1998) Do natural antisense transcripts make sense in eukaryotes? *Gene*, **211,** 1-9.

Varagona,M.J., Schmidt,R.J. and Raikhel,N.V. (1992) Nuclear localization signal(s) required for nuclear targeting of the maize regulatory protein Opaque-2. *Plant Cell*, **4,** 1213-1227.

Verma,D., Kazain, V., Zogbie,V. and Bal., (1978) Isolation and characterization of the membrane envelope enclosing the bacteroids in soybean root nodules. *J. Cell Biol.*, **78,** 919-939.

Vihtelic,T.S., Goebl,M., Milligan,S., O'Tousa,S.E. and Hyde,D.R. (1993) Localization of *Drosophila* retinal degeneration B, a membrane-associated phosphatidylinositol transfer protein. *J. Cell Biol.*, **122,** 1013-1022.

Volschenk,H., Viljoen,M., Grobler,J., Petzold,B., Bauer,F., Subden,R.E., Young,R.A., Lonvaud,A., Denayrolles,M. and van Vuuren,H.J.J. (1997) Engineering pathways for malate degradation in Saccharomyces serevisiae. *Nature Biotech.*, **15,** 253-257.

von Arnim,A.G., Deng,X.-W. and Stacey,M.G. (1998) Cloning vectors for the expression of green fluorescent protein fusion proteins in transgenic plants. *Gene,* **221,** 35-43.

Waterhouse,P.M., Graham,M. and Wang,M.-B. (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl. Acad. Sci. USA*, **95,** 13959-13964.

FUTURE PERSPECTIVES

The research described in this thesis has led to identification and characterization of a number of novel late nodulin genes from a model legume, *Lotus japonicus*. Three nodulin genes, *LjNPP2C1*, *LjNOD70* and *LjNOD16*, were chosen for detailed characterization, leading to the formulation of separate projects that yielded their own hypotheses and conclusions. These projects also generated a number of questions to be addressed in future, the main one being the biological role of these genes in nodules and other tissues of *L. japonicus*. This question was addressed to various extents for each of these genes using various molecular and biochemical approaches.

In my opinion, the disruptions of, or loss-of-function mutations in, the genes encoding LjNPP2C1 and its potential interacting partner (LBP) will be absolutely required to understand the biological functions of these proteins and assess the relevance of their interaction. A second route to address the function of LjNPP2C1 is to conduct a more extensive yeast two-hybrid screen for putative binding proteins, followed by genetic experiments to test *in vivo* the significance of any observed interactions.

Nodulin LjN70 could potentially be an organic acid transporter in the peribacteroid membrane. It is therefore crucial to functionally characterize the biochemical properties of this protein, to establish whether it has transporter-like activity and to identify its substrates. Such experiments are in progress in collaboration with the laboratory of Dr. Dan Roberts, The University of Tennessee, Knoxville. In addition, it will also be extremely interesting to know if the products of the *LjNOD70* gene and related members of this gene family are required for normal functioning of *L. japonicus*

nodules. The anti sense experiments with transgenic *L. japonicus* plants are in progress to down regulate the expression of LjNOD70 and related genes. It is also important to establish whether the spliced form of the *LjNOD70* mRNA has a specific function in nodules, and, whether it encodes a protein product.

The function of the *LjPLP* gene family and *LjNOD16* late nodulin is quite intriguing. The apparent nodule-flower dichotomy in the gene expression patterns of the different members of this family is very interesting and clearly warrants further investigation. Furthermore, the detailed biochemical characterization of the lipid binding and transfer properties of the LjPLP proteins is necessary to gain an understanding of their function. These experiments are underway in collaboration with the laboratory of Dr. Vytas Bankaitis, University of Alabama, Birmingham. In addition, several important questions remain. What are the functions of Nlj16 protein and the *LjPLP*-IV antisense transcripts? Are they a part of a mechanism designed to inactivate the *LjPLP*-IV and other *LjPLP* genes in nodules? If yes, then why is this gene family inactivated in nodules? What are the functions of the members of *LjPLP* gene family in other plant tissues?