Putative Dicarboxylate and Amino Acid Transporters in Soybean (*Glycine max* L.) - A Molecular Characterisation

This thesis is presented for the degree of Doctor of Philosophy

of The University of Western Australia

THE UNIVERSITY OF WESTERN AUSTRALIA

Helle Martha Christophersen

Candidatus Agronomiae (M. Sc)

Biochemistry and Molecular Biology

School of Biomedical, Biomolecular and Chemical Sciences

April, 2006

Declaration

The work presented in this thesis is my own work except where stated below. This work was carried out in the School of Biomedical, Biomolecular and Chemical Sciences at the University of Western Australia. The material presented in this thesis has not been presented for any other degree.

The 5' and 3' RACE PCRs and the subsequent cloning and sequencing reactions that resulted in a full length sequence of *Gm*AAP5, was performed by a third year student, David Zeelenberg, under my supervision.

Helle Martha Christophersen, 21st April, 2006.

Acknowledgment

Here at the very end of my PhD I would like to thank several people, who have helped me through and been there when times were tough.

Without the financial support from the Danish Research Agency, it would never have been possible for me to come to Perth to undertake my PhD candidature. I am very grateful for being granted such a possibility. I was also awarded a Jill Bradshaw Bursary from the Australian Federation of University Women in Western Australia (AFUW-WA), a UWA completion scholarship and a UWA Scholarship for International Research Fees. For these I am very grateful.

A huge thanks to my principal supervisor, Dr. Martha Ludwig, for taking me under her wing, when my initial principal supervisor Prof. David Day left for Sydney. Martha has strongly supported my scientific research both intellectually and financially, and is an excellent supervisor and a wonderful person. Her encouragement and understanding have been invaluable.

I would like to thank David for his encouragement for me to come to Perth in the first instance, and also for his support in the form of an ad hoc scholarship from UWA, when I was without a Danish living allowance.

I would also like to thank my co-supervisors during my PhD. Currently, Dr. Patrick Finnegan, and previously Dr. Amy Curran, Dr. Liliane Gerhardt, Dr. Penelope Smith and Dr. Christel Norman. It has been a very interesting and challenging experience. I am particularly grateful to Pat for his comments on my thesis.

Numerous people have provided help, advice, encouragement and friendship during my PhD and it is impossible to mention all of them. However, I would like in particular to thank: Vanessa Muuns, Joanne Castelli, Andrew Wiszniewski, David Zeelenberg, Dr.

Ruth Holtsapffel, Dr. Rowena Thomson, Dr. Sophie Monreau, Dr. Brent Kaiser and Dr. Suzanne Long.

Thanks to my parents, Hanne and Stoffer Christophersen, for supporting me in achieving my goals. A huge thanks to my mum for visiting us in Perth so many times, and for all the washing and cleaning and baby sitting, making life with two PhD's and two boys much easier for us.

The love, encouragement and support from my lovely husband, Claus Thagaard Christophersen, have been outstanding. Words just cannot express how grateful I am for that.

Helle Martha Christophersen, April 2006.

List of abbreviations

3’	3 prime
5’	5 prime
Al	aluminium
Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartate
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
bp	base pair(s)
CCCP	corbonylcyanide *m*-chlorophenylhydrazone
cDNA	complementary DNA
cfu	colony forming units
CTAB	cetyltrimethylammonium bromide
C-terminal/terminus	carboxyl terminal/terminus
DIDS	4,4’-diisothiocyano stilbene-2,2’disulphonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EDTA	ethylene diamine tetra-acetic acid
EMM	Edinburg minimal medium
EST	expressed sequence tag
G418	geneticin
GFP	green fluorescence protein

Gln	glutamine
Glu	glutamate
GPI	glycosylphosphatidylinisotol
hrs	hours
IPTG	isopropylthio-ß-D-galatoside
kDa	kiloDalton(s)
LB medium	Luria-Bertani medium
Leu	leucine
min	minute(s)
mRNA	messenger RNA
N	nitrogen
N- terminal/terminus	amino terminal/terminus
Na	sodium
NCBI	National Center for Biotechnology Information
nm	nanometers
nodulin	nodule enhanced transcript
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
Pi	inorganic phosphate
PTR	peptide transporter
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNase	ribonuclease
RT	reverse transcriptase

s	second(s)
SDS	sodium dodecyl sulphate
SM	Symbiosome membrane
TAE	Tris-acetate/EDTA electrophoresis buffer
TE	Tris EDTA
Thr	threonine
TIGR	The Institute for Genomic Research
Tris	tris[hydroxymethyl]aminomethene
U	units
UTR	untranslated region
V	volt(s)
w/v	weight/volume
WT	wild type
X-gal	5-bromo-4-chloro-3-indol-ß-D-galactoside
YES medium	yeast extract supplemented medium

Abstract

Some plants, such as legumes, are able to use atmospheric nitrogen as a nitrogen source due to the nitrogen-fixing bacteria residing in specialised root structures called nodules. The exchange of carbon and nitrogen between the host plant (legume) and the nitrogen-fixing micro-symbiont is vital for biological nitrogen fixation. In particular, transport of C4-dicarboxylates, mainly malate, from the plant to the micro-symbiont, and the reverse transport of fixed nitrogen in the form of ammonium are essential for symbiotic nitrogen fixation. In the legume nodule, the symbiosome membrane (SM) surrounds the bacteroid and all exchanges of metabolites and nutrients that occur between the plant and the micro-symbiont must cross this membrane. Recently it has been established that cycling of amino acids across the SM is also critical for optimal symbiotic nitrogen fixation. Therefore to fully understand this agriculturally significant phenomenon, the mechanisms facilitating these exchanges need to be investigated.

The major aim of this study was to increase the understanding of nutrient exchange within the nodule at the molecular level by isolating and characterising genes encoding transporters responsible for malate and amino acids transport in soybean (*Glycine max*, L.), with particular interest in genes significantly or highly expressed in nodules. A combination of molecular and biochemical techniques was used to achieve this.

A malate transport mutant of the fission yeast *Schizosaccharomyces pombe* was created to obtain a molecular tool for screening legume cDNA libraries by functional complementation and characterisation of putative malate transporters.

Four soybean cDNAs encoding proteins named *Gm*DT1, *Gm*DT2, *Gm*DT3 and *Gm*DT4, with high similarity to the Arabidopsis tonoplast dicarboxylate transporter,

*At*tDT, were identified, isolated and characterised. These represent the first legume genes discovered that belong to the SLC13 gene family, which is a solute carrier family. The expression patterns of the *GmDT* genes were investigated in different soybean tissues and the function of the proteins was examined in two different heterologous systems. Southern blot analysis showed that a small gene family of up to five members encodes these proteins in soybean.

A full-length cDNA, designated *GmAAP5*, was isolated that encodes a novel, putative amino acid transporter. Molecular characterisation of this cDNA and that of *GmAAP1* (GenBank Accession no: AY029352), a previously identified putative amino acid transporter gene, was done. Expression analyses showed relatively high expression of *GmAAP5* in soybean nodules compared to that in leaf and root tissues, while *GmAAP1* showed uniformly high expression in root, leaf and nodule tissues.

Phylogenetic analysis of the deduced amino acid sequences of known functional AAPs from dicotyledonous plants revealed that *Gm*AAP1 is most closely related to AAP2 from *V. faba*, while *Gm*AAP5 is more closely related to AAPs from non-leguminous plants than from leguminous plants. Based on the functional characterisation of the AAPs with which *Gm*AAP1 and *Gm*AAP5 cluster, it is likely that both transporters are neutral and acidic amino acid transporters within the AAP subfamily.

Table of contents

Chapter One: General introduction

Introduction

This chapter reviews the significance and important aspects of biological N_2 fixation in general, and symbiotic N_2 fixation in particular. The focus will be on the legume/rhizobia interaction. The main areas in respect to symbiotic N_2 fixation that will be discussed are: The symbiosome membrane (SM), carbon and nitrogen metabolism of root nodules and transport of carbon and nitrogen compounds across the SM. Transport of C4-dicarboxylates (in particular malate) in the legume root nodule was the main focus of this study. Because of the attempt to identify putative homologs of the SM dicarboxylate transporter, a section on plant malate transporters is included. Translocation of amino acids and other organic nitrogen compounds within the nodule and the plant is also crucial to N_2 fixation, and a section on transporters of organic nitrogen in legumes is included in this chapter. Since the identification of putative transporters was attempted by functional complementation in heterologous systems, the transport of malate in yeast and bacteria, and identification of plant transporter proteins using these heterologous systems are described.

Significance of biological nitrogen fixation

Nitrogen (N) is an essential component in macromolecules such as DNA, RNA and proteins, and is, therefore, an essential element for life. Although plants live and grow in an atmosphere that is 79 per cent atmospheric nitrogen (N_2), their growth and development are limited more by the availability of N than of any other essential element (Date, 1973). Since higher plants are unable to assimilate N_2 directly, they rely on accessible N in the soil (Lee *et al.*, 1983). For most crop species N is frequently a limiting nutrient and consequently N fertilisation is required for optimal growth. Plants can utilize a range of N forms, but by far the most common form of N acquired from the soil are the inorganic forms; nitrate and ammonium (Marschner, 1995). However under

nitrogen-poor conditions, certain legume and actinorrhizal species are capable of forming symbiotic associations with N_2-fixing microorganisms, which convert the N_2-fixing into a plant-accessible N source (Marschner, 1995; Schubert, 1986).

Biological N_2 fixation provides about 65% of the biosphere's available N (Newton, 2000), with nodulated legumes in symbiosis with rhizobia among the leading N_2-fixing systems in agriculture (Marschner, 1995). Not only in agricultural systems do legumes have great significance. In natural systems the inflow of N by legumes is of great value for their habitats, like tropical forests (Dixon & Wheeler, 1986).

The importance of legumes in agriculture is not only due to their ability to fix N_2. In addition legume plants and seed are high in protein, which makes legumes nutritionally important for humans, and forage legumes supply a high protein diet for livestocks (Dixon & Wheeler, 1986; Hardarson & Atkins, 2003). Worldwide, legumes provide at least 33% of humankind's need for N, but in the tropics and subtropics plants provide up to 80 % of the dietary requirement (Vance *et al.*, 2000).

Effective management of N in the environment is one of the driving forces in agricultural sustainability (Vance, 2002). For example when N_2-fixing species are introduced to the cropping system, a reduction in requirement of N fertilisers and an increase of soil quality is achieved (Vance, 2002). In addition, a reduction in crop-production cost is achieved and, since biologically fixed N is bound in organic matter, a decrease in the risk of contamination of water resources is also achieved (Hardarson & Atkins, 2003; Vance, 2002). Furthermore, nodulated legumes have the potential to provide the entire N required for their growth, and the energy to fuel biological N_2 fixation is derived mainly from renewable plant resources. In contrast, chemical N_2 reduction requires depletion of non-renewable fossil fuel supplies. Therefore the agricultural, economic, humanitarian and ecological significance of biological N_2 fixation cannot be disputed (Vance, 2002).

The catalysing complex, nitrogenase

Biological N_2 fixation involves the conversion of N_2 to NH_3 and requires energy in the form of ATP and reducing equivalents (Marschner, 1995). This dinitrogen reduction, catalysed by the enzyme complex nitrogenase, takes place in the bacteria, and follows the reaction:

$$N_2 + 8H^+ + 8e^- + nMgATP \rightarrow 2NH_3 + H_2 + nMgADP + nP_i,$$

where $n \geq 16$ (Dixon & Wheeler, 1986; Halbleib & Ludden, 2000; Marschner, 1995). Nitrogenase is found only in certain prokaryotes, including members of the family Rhizobiaceae that form symbioses with legumes (Hardy & Havelka, 1975). Nitrogenase is extremely sensitive to oxygen, which inactivates it irreversibly, but at the same time efficient N_2 fixation has a high-energy demand that can only be met by ATP generated by aerobic respiration (Marschner, 1995). However, N_2-fixing organisms have evolved several strategies to overcome this problem. In legume root nodules an oxygen diffusion barrier in the cortex provides mechanical protection and the presence of leghemoglobin provides enzymatic protection (Becana & Rodriguez-Barrueco, 1989).

The plant-microbe endosymbiosis

Symbiosis is the close and often obligatory association of two different species living together, and in its broadest definition includes both parasitic and mutually beneficial partnerships. Two types of symbiosis exist that is ecto- and endosymbioses, where one organism lives on or in the other organism, respectively. Endosymbioses can be further characterised depending on whether the microsymbiont (smaller partner) is intra- or

extracellular with respect to the host (Whitehead & Day, 1997). A feature of many intracellular symbioses, for example the mutual relationship between rhizobia/legume symbiosis, is that the symbiont is surrounded by a membrane of host origin formed by endocytosis of the host cell membrane (Whitehead & Day, 1997).

Two types of nodule-forming symbioses exist between N_2-fixing soil bacteria and higher plants. First, the symbiosis between legumes and soil bacteria predominantly from the genera *Rhizobium* and *Bradyrhizobium* (following referred to as rhizobia). Second, the symbiosis between actinomycetes of the genus *Frankia* and a diverse group of angiosperms, collectively called actinorhizal plants (Dixon & Wheeler, 1986; Schubert, 1986). In both cases a membrane separating the two controls the exchange of nutrients between the host and the microsymbiont (Pawlowski & Bisseling, 1996). Therefore transport proteins are needed in this membrane for the symbiosis to be maintained (Pawlowski & Bisseling, 1996).

The nodule

Infection and nodulation

The legume nodule is a specialized plant organ, which the rhizobia inhabit as nitrogen fixing endo-symbionts. The early stages of infection include bacterial gene expression, division and growth as well as differentiation in the host. All of these events are mediated by signal exchanges between the eukaryotic host and the prokaryotic symbiont (Long, 1996).

Roots are known to exude organic compounds that attract the soil bacteria. Pea roots, for example, exude an unusual amino acid, homoserine, which is a preferred carbon and N source for the pea root symbiont, *Rhizobium leguminosarum* (Hopkins, 1995). Phenolic compounds in root exudates, such as flavonoids, flavones, flavonones and isoflavones, act as transcriptional activators of nod genes in the rhizobia (Peters *et al.*,

1986). Flavonoids and isoflavonoids are known to impact biosynthesis and transport of auxin in plants and may not only signal the rhizobia but also lead to root cell division (Vance, 2002). Some of the nod genes encode bacterial signal molecules (oligosaccharides) called Nod factors, which stimulate the expression of corresponding "nodulation" genes in the plant (Fisher & Long, 1992). Purified Nod factors have been demonstrated to be the sole bacterial compound needed to initiate morphogenic changes in the host plant (Journet *et al.*, 1994).

The infection is initiated when the bacteria come into contact with legume root hairs and penetrate the roots. Apparently in response to signals from the rhizobia, the roots of the plant host are stimulated to enhance root hair production and the development of shorter, thicker roots that curl around the bacteria (Rolfe & Gresshoff, 1988). One of the plant responses to bacterial Nod factors is tissue de-differentiation and division of root cortex cells, resulting in the formation of a nodule primordium (Fisher & Long, 1992; Whitehead & Day, 1997). Rhizobia invade the host root by digesting the root-hair cell wall and forming an infection thread, which itself represents an involution of the plant cell wall. The rhizobia continue to divide as the infection thread elongates towards the root cortex (Callaham & Torrey, 1981; Dixon & Wheeler, 1986). The infection thread branches to penetrate numerous cortical cells as visible nodules develop on the root (Dixon & Wheeler, 1986). Once the bacteria have stopped dividing they undergo differentiation into their symbiotic form, the bacteroids (Bergersen, 1958). Rhizobia are taken up into the cells of the nodule by endocytosis of the infection thread membrane (Verma *et al.*, 1978). Endocytotic vesicles called symbiosomes are formed (Roth *et al.*, 1988). These structures are the fundamental N_2-fixing units of the symbiosis (Udvardi & Day, 1997).

In the mature nodule the infected cells are enlarged due to the presence of thousands of symbiosomes (Goodchild & Bergersen, 1966). In the nodules of temperate legumes, SM

division, resulting in single-bacteroid symbiosomes, always accompanies bacterial division (Robertson & Lyttleton, 1984). However in tropical legumes, the symbiosomes contain numerous bacteriods because after a certain point in development only the bacteria continue to divide (Goodchild & Bergersen, 1966) or the small symbiosomes fuse (Fedorova *et al.*, 1999).

Determinate and indeterminate nodules

Determinate and indeterminate nodules are two major types of nodules present on roots of herbaceous legumes and on leguminous shrubs (Dixon & Wheeler, 1986; Hansen, 1994). What type of nodule develops depends on the plant and in general only one type of nodule is formed on each plant species (Dixon & Wheeler, 1986; Hansen, 1994). Indeterminate nodules have a persistent meristem that generates cylindrical nodules that may occasionally bifurcate to form a coralloid structure. Consequently, all stages of nodule development are represented in each indeterminant nodule because an age gradient exists from the distal meristem to the proximal point of attachment to the legume root (Dixon & Wheeler, 1986; Hansen, 1994). This type of nodule is developed on pea (*Pisum Sativum*) roots. In determinate nodules, cell division is only transient and the final shape of the nodule is more or less spherical, and these are found, for example, on soybean (*Glycime max*) (Dixon & Wheeler, 1986; Hansen, 1994).

The symbiosome membrane (SM)

The biochemical properties of the SM from the initial endocytosis to the mature symbiosome have been studied for decades (reviewed by (Whitehead & Day, 1997). However, most of the research have been concentrated on the structure of the mature SM, which is defined as the membrane that surrounds differentiated, N_2-fixing

bacteroids during the period when the symbiosis is fully functional (Whitehead & Day, 1997).

During biogenesis of the SM a number of transport activities that aid in establishment and maintenance are acquired. The mechanism by which proteins are targeted to the SM during symbiosome formation, is unknown (Whitehead & Day, 1997). Catalano *et al*., (2004) proposed a model, based on their own and previously reported data, that suggest four pathways for symbiosome formation: 1) Translation of proteins from free cytosolic ribosomes and targeting via N-terminal signal peptides; 2) proteins translated on endoplasmatic reticulum (ER)-bound ribosomes, modified by glycosylation or an GPI achor addition in the ER, transported to the Golgi and secreted to the SM by membrane-bound vesicles; 3) transition of proteins directly from the ER derived membrane and 4) symbiosome membrane proteins derived from the bacteroid (Catalano *et al.*, 2004).

There is no doubt that the SM has a major role in the symbiosis. This membrane of plant origin effectively segregates the bacteroids from the plant cytoplasm, and it selectively regulates all transport of metabolites between the symbiotic partners (Price *et al.*, 1987). Therefore, rhizobia are totally dependent upon their plant hosts for nutrients when living within the nodule.

Any signalling or metabolite exchange in the symbiosis must proceed via the interface between the symbionts, which consists of the symbiosome membrane, the symbiosome space and the bacteroid outer membrane (Roth *et al.*, 1988). Compound exchange between the symbionts may be mediated either by diffusion down a concentration gradient or by active transport via transport proteins present in the membrane of the interface (Udvardi & Day, 1997).

During biogenesis of the SM a number of transport activities that aid in establishment and maintenance are acquired (Vincill *et al.*, 2005). These essentially include dicarboylate transport activity (Herrada *et al.*, 1989; Ouyang *et al.*, 1990; Ouyang &

Day, 1992; Price *et al.*, 1987; Rosendahl *et al.*, 1992; Udvardi *et al.*, 1988a) and transport pathways for the efflux of fixed N in the form of NH_3 (Niemietz & Tyerman, 2000) or NH_4^+ (Day *et al.*, 2001b; Tyerman *et al.*, 1995; Roberts & Tyerman, 2002). These transport mechanisms will be discussed later. Other transport activities that have been reported on the mature SM include; a proton pump that generates both a membrane potential (positive inside) and a pH gradient (Blumwald *et al.*, 1985; Fedorova *et al.*, 1999; Udvardi & Day, 1989; Udvardi *et al.*, 1991) and has the potential to profoundly affect the exchange of other ions across the SM (Tyerman *et al.*, 1995; Udvardi *et al.*, 1991); an aquaglyceroporin activity (Dean *et al.*, 1999; Rivers *et al.*, 1997), proposed to be involved in osmoregulation (Guenther *et al.*, 2003) and transporters for iron and zinc uptake (Kaiser *et al.*, 2003; Moreau *et al.*, 2002).

Carbon and nitrogen metabolism of root nodules

The nodule and the subtending root system represent one of the major sinks for carbon, receiving at least 15-30 % of the net photosynthate of the plant (Schubert, 1986). The photosynthate transported via the phloem is used to support: the growth and maintenance of the nodule tissue; the nitrogenase activity in the bacteroid and assimilation of the reduced N in the host cytosol; and the synthesis of N-containing organic compounds for export from the nodule (Schubert, 1986).

As mentioned, within the specialised environment of the nodule, bacteroids reduce N_2 to ammonia (NH_3), which is supplied to the host (see later). In return the plant provides the bacterioids primarily with photo assimilates (see later) and other important nutrients, which support growth and function of the nodule (Schubert, 1986). Consequently a close link between the carbon and N metabolism in host-plant cells of legume nodules with the N_2-fixing activity in the rhizobia is needed for an optimal output.

The primary reduced carbon source translocated to the legume nodules is sucrose. Nodule metabolism subsequently generates significant amounts of glucose and fructose and several organic acids including malate and succinate (Streeter, 1991; Stumpf & Burris, 1979). In the symbiosome the carbon source is used to fuel bacteriod metabolism (Udvardi & Day, 1997). It is generally accepted that C4-dicarboxylates, particular malate and succinate, are the main carbon-containing substrates supplied to the bacteroid for the support of N_2 fixation in most legumes (see Udvardi & Day, 1997).

Nitrogen fixed by the bacteroid needs to leave the symbiosome to become accessible for assimilation in the plant. As mentioned, in the bacteroid N_2 is reduced to NH_3, which is driven by the concentration gradient to rapidly diffuse into the symbiosome space (Bergersen & Turner, 1967). The acidic pH in the symbiosome space, created by proton pumps both on the SM and the bacteroid membrane, favours protonation to NH_4^+ (Blumwald *et al.*, 1985; Brewin, 1991), which is then transported across the SM to the cytoplasm. Recently, new evidence has been found that amino acids are secreted from the bacteroids and play a crucial role in the pea symbiosis (Lodwig *et al.*, 2003) (see later). However, ammonium presumably still accounts for one of the major N source provided to the cell *in vivo* (Day *et al.*, 2001b). The ammonium originated from the nitrogenase activity should mostly be available to the plant as the enzymes of ammonium assimilation are repressed in bacteroids (Brown & Dilworth, 1975), while induced in the host-plant tissue (see Schubert, 1986). In addition the *R. elit* gene, *amtB*, encoding an ammonium transporter, is down regulated at an early stage of bacteroid differentiation (Tate *et al.*, 1998). This down regulation of *amtB* is vital for the symbiosis, since forced expression of *amtB* in the wild-type strain of *R. elit* leads to a strongly delayed or arrested bacteroid differentiation process (Tate *et al.*, 1998). There is a strong consensus that ammonium entering the plant cytosol across the SM

(regardless of mechanism) is rapidly assimilated into glutamine as a result of the very high activity of cytosolic glutamine synthetase (GS) (Streeter, 1991). Subsequently the glutamine is then converted into glutamate, a reaction catalysed by glutamate synthase, GOGAT (Streeter, 1991). The assimilated N needs to be translocated to the rest of the plant via xylem. The form in which it is transported out of the nodule depends on the legume species. In legumes with indeterminate nodules glutamine is mainly metabolised to asparagine in the nodule, and mainly asparagine gets exported (Marschner, 1995; Snapp & Vance, 1986). In contrast, in nodules of determinate nodules a more complex biosynthetic pathway of glutamine into purine derivates known as ureides (mainly allantoin and allantoic acid) get exported (Marschner, 1995).

Recently it have been shown that transport of amino acids either in or out of the pea bacteroid is critical for an optimal symbiosis at least in pea (Lodwig *et al.*, 2003). This very strongly supports previous reports in pea that either showed: secretion of aspartate and alanine from isolated bacteroids incubated with glutamate and aspartate (Rosendahl *et al.*, 1992); aspartate transport activity in pea SM in-side out vesicles (Rosendahl *et al.*, 2001; Rudbeck *et al.*, 1999); or mixed secretion of ammonium and alanine by bacteroids incubated with N_2 (Allaway *et al.*, 2000). However these studies do not rule out ammonium as the major N_2 fixation product in pea. Indeed it was shown that when alanine synthesis was disturbed in bacteroids of *R. leguminosarum* or *Mesorhizobium loti* there was no change in N_2 fixation in nodules of pea (indeterminate nodules) or *Lotus corniculatus* (determinate nodules) (Allaway *et al.*, 2000; Kumar *et al.*, 2005; Prell & Poole, 2006). In addition, a study with isolated soybean bacteroids concluded that NH_3 *ex planta* is the principal soluble product of N_2 fixation and it was argued that ammonium should be considered as the main product of N_2 fixation *in vivo* (Li *et al.*, 2002).

Transport of carbon and nitrogen containing compounds across the SM

The nutrient exchange between the symbiotic partners requires membrane transporters for the carbon compounds and trace elements that flow from the plant to the microsymbiont, along with transporters for the fixed N that flows in the opposite direction (Day *et al.*, 2001a; Pawlowski & Bisseling, 1996; Udvardi & Day, 1997) and possibly also in the same direction (Lodwig *et al.*, 2003).

Physiological and biochemical studies of solute transport, where the solute in vivo move from the plant to the microsymbiont, are facilitated by the fact that intact symbiosomes can be isolated from legume nodules (Day *et al.*, 1989; Rosendahl *et al.*, 1992) and used to follow the in vitro uptake of radioactively labelled compounds. Transport of dicarboxylates, zinc, and iron from soybean nodules have been characterized in this way (Kaiser *et al.*, 2003; Moreau *et al.*, 2002; Udvardi *et al.*, 1988a). Transport of solutes going out of bacteroids and symbiosomes is more difficult to determine as it requires isolation of high purity SM and the reconstitution of inside-out vesicles, which can then be used in transport studies or alternatively patch clamping techniques. The transport of ammonium and aspartate has been characterized from pea using inside out SM vesicles (Mouritzen & Rosendahl, 1997; Rudbeck *et al.*, 1999), whereas the transport of ammonium from soybean symbiosomes was characterized using the patch clamping technique (Tyerman *et al.*, 1995).

The SM of soybean nodules possesses a dicarboxylate transporter capable of mediating a rapid flux of dicarboxylate anions, such as malate and succinate, to the bacteroids. The preferred anion transported by this carrier is monovalent dicarboxylate anions (Udvardi *et al.*, 1988a). Malate and succinate competively inhibit the transport of the other and therefore are likely to be transported by the same carrier (Ouyang *et al.*, 1990; Udvardi *et al.*, 1988a). Other inhibitor studies showed that malate transport across the SM in

soybean nodules was inhibited by oxaloacetate, fumarate, and oxoglutarate (Ouyang & Day, 1992).

The malate transporters on the SM and the bacteroid membrane differ in their sensitivity to known mitochondrial organic acid transport inhibitors that inhibit malate uptake into symbiosomes but have no effect on malate uptake by bacteroids (Udvardi *et al.*, 1988a). In addition there are differences in the metabolites that are transported across the SM and the bacteroid membrane, which leaves little doubt that separate transport mechanisms exist on the SM (Ouyang *et al.*, 1990; Udvardi *et al.*, 1988a).

A major 26 kDa protein (nodulin 26), which is only found on the SM has been identified as a membrane-spanning channel protein (Miao *et al.*, 1992; Miao & Verma, 1993). Ouyang *et al*., (1991) showed that phosphorylation of nodulin 26 is correlated with malate transport, and studies with this protein incorporated into artificial lipid bilayers indicated that the protein is an anion channel (Weaver *et al.*, 1994). However, it is now evident that nodulin 26 encodes an aquaporin likely to play a role in osmoregulation (Dean *et al.*, 1999; Rivers *et al.*, 1997). It is therefore not likely that nodulin 26 plays a role in dicarboxylate uptake into symbiosome. Thus the task to identify the SM dicarboxylate transporter still remains

Physiological and biochemical studies have identified the presence of a voltage-gated monovalent cation channel, which is capable of transporting ammonium across the SM, in soybean (Tyerman *et al.*, 1995) and pea SM (Mouritzen & Rosendahl, 1997). Even though the two transport mechanisms show similar kinetics, differences in the regulation of the transport mechanisms were reported. Increased transport of ammonium by the putative pea channel to low extra-vesicular pH was not reported for the soybean putative channel (Mouritzen & Rosendahl, 1997). In addition, the inhibition of the soybean channel by Ca^{2+} was only observed in pea inside-out SM vesicles under very

high and physiologically irrelevant concentrations. The implication that ammonium transport should be regulated by acidification of the symbiosome space is very possible, due to the presence of the ATPase on the SM and the respiratory transport chain on the bacteroid membrane, both of which contribute to the acidification (Mouritzen & Rosendahl, 1997). In addition, the rapid assimilation of ammonium in the plant cytosol would ensure the presence of a electrochemical gradient that would allow ammonium transport through the channel (Mouritzen & Rosendahl, 1997).

Although amino acid transport between the plant and the bacteroid in the legume nodule has been shown to be important, and despite extensive studies, the forms in which fixed N is provided to the plant, following N_2 fixation in the bacteroid, remains unclear. While a large body of evidence suggests that ammonium *in vivo* is supplied to the plant (Day *et al.*, 2001b) some results as described above indicate that, at least in pea, this is supplemented by an exchange of amino acids between the symbiotic partners. Biochemical evidence for an aspartate mechanism on the SM in pea nodules has been reported (Rudbeck *et al.*, 1999). Aspartate transport activity was measured in inside-out SM vesicles, indicating the presence of an aspartate specific transporter responsible for transport of aspartate from the bacteroid to the plant (Rosendahl *et al.*, 2001; Rudbeck *et al.*, 1999). However, in spite of the biochemical and physiological evidence for exchange of amino acids across the SM, the genes encoding amino acid transporters on the SM still remain to be identified.

Amino acids may pass the bacteroid membrane in either direction by the general amino acid permease (Walshaw *et al.*, 1997). The general amino acid transporters, *aap* and *bra*, on the bacteroid membrane involved in this secretion have since been cloned (Hosie *et al.*, 2002; Poole *et al.*, 1985; Walshaw & Poole, 1996) and mutants in either do not affect the symbiosis (Lodwig *et al.*, 2003). Interestingly, the *aap*/*bra* double

mutants of *R. leguminosarum* are able to nodulate pea roots and can fix nitrogen, but the N is not transferred to the plant efficiently and the plant becomes N starved (Lodwig *et al.*, 2003). These results strongly indicate that amino acid transport is involved in effective symbiosis. Whether this transport is into or out of the bacteroid or both remains to be determined, as these transporters have been shown to transport amino acids in both directions (Hosie *et al.*, 2002; Walshaw & Poole, 1996). In any event, the above results indicate that there is a more complex cycling of amino acids between the symbiotic partners than previously thought and suggest that complementary amino acid transporters may be present on the SM. Therefore the research into amino acid transporter genes expressed in legume nodules is important.

Malate transporters in plants

In plants, many of the proteins involved in transport of dicarboxylates have not yet been identified at the molecular level. Until recently (see later) the only genes identified were encoding plant dicarboxylate transporters in mitochondrial and chloroplast. (Picault *et al.*, 2002; Taniguchi *et al.*, 2002; Taniguchi *et al.*, 2004; Weber *et al.*, 1995). Tonoplast dicarboxylate transporters have been described physiologically and biochemically for some plant cells (Martinoia & Rentsch, 1994; Bettey & Smith, 1993; Blackford *et al.*, 1990; Pantoja & Smith, 2002), but their characteristics are somewhat different to those of the SM dicarboxylate transporter. On the other hand, the plasmalemma and the SM share common biochemical properties (Verma & Hong, 1996), and it may be that SM and tonoplast transporters share common features as well. In fact one of the ATPase enzymes (the Mg^{2+}/H^{+} ATPase) present on the SM, is also present in the Golgi and the tonoplast membranes (Verma & Hong, 1996). In addition, the symbiosome space

contains protease, acid trehalase, α-mannosidase II and protease inhibitors, which are all typically found in vacuoles (Verma & Hong, 1996).

Recently, three dicarboxylate transporters have been identified at the molecular level; Emmerlich *et al*., (2003) were the first to describe the *Arabidopsis thaliana* tonoplast dicarboxylate transporter (*At*tDT), while Sasaki *et al*., (2004) identified a wheat gene (*ALMT1*), which encodes a transporter that facilitates organic anion efflux from the cell, and Jeong *et al*., (2004) identified a gene (*AgDCAT1*), encoding a nodule-specific dicarboxylate transporter from alder. The last is believed to be located on the invaginated part of the plasmalemma, where the N_2 fixation takes place in alder nodules (Jeong *et al.*, 2004).

One common feature of these three transporters is that they are believed to facilitate transport of one or more dicarboxylates out of the cytosol. The orientation of malate transport into the symbiosome in the legume nodule is also directed out of the cytosol. This might be a feature of particular interest for the identification of the malate transporter located on the SM of legume symbiosomes. Moreover, none of these three transporters exhibit much amino acid sequence homology to the dicarboxylate transport proteins identified in plant mitochondria or chloroplasts. In addition they all belong to different transporter families.

The *A. thaliana* tonoplast dicarboxylate transporter (Emmerlich *et al.*, 2003) was the first plant tonoplast dicarboxylate transporter to be identified and characterised. *At*tDT is a member of the solute carrier 13 (SLC13) gene family (see Chapter 3). Very recently it has been proposed that *At*tDT is a transporter different from the vacuolar malate channel previously reported (Hurth *et al.*, 2005) and that *At*tDT is critical for pH homeostasis (Hurth *et al.*, 2005). The fact that *At*tDT facilitates transport of malate and that this transport is inhibited by the uncoupling agent carbonyl cyanide m-

chlorophenylhydrazone (Emmerlich *et al.*, 2003) makes it a possible homolog to the SM malate transporter. However, the expression of *AttDT* seems to be highest in photosynthetic source tissues like young leaves, with root showing less expression (Emmerlich *et al.*, 2003).

ALMT1 was isolated by subtractive hybridisation of cDNA from near-isogenic wheat lines that differ in aluminium tolerance and the full-length cDNA obtained by RACE-PCR. *ALMT1* is constitutively expressed in root apices of both lines but in greater amounts in the aluminium-tolerant line (Sasaki *et al.*, 2004). The ALMT1 amino-acid sequence shows highest similarity (69%) to a rice putative protein of unknown function and no homologs have so far been found in animal and microbial genomes (Sasaki *et al.*, 2004). ALMT1 and its putative rice homolog do not belong to an identified protein family (Sasaki *et al.*, 2004). Heterologous expression in Xenopus oocytes, rice and cultured tobacco cells and transgenic barley confirmed *ALMT1* encodes an Al-activated transporter that facilitates the efflux of malate but not citrate (Delhaize *et al.*, 2004; Sasaki *et al.*, 2004). Additionally, ALMT1 increased the tolerance of tobacco cells to Al treatments (Sasaki *et al.*, 2004). Recently the evidence for root plasma membrane localisation of ALMT1 was reported (Yamaguchi *et al.*, 2005). The activation pattern reported for ALMT1 is not similar to that of the malate transporter on the SM and ALMT1 is therefore not considered a possible homolog of the SM transporter.

AgDCAT, encodes a protein that belong to the Peptide Transporter (PTR) family. It was isolated from an alder root nodule cDNA library by differential screening of cDNA from nodules versus uninoculated roots (Jeong *et al.*, 2004). Southern blot analysis of genomic DNA indicated that a small gene family encodes the *AgDCAT1* gene (Jeong *et al.*, 2004). *AgDCAT1* mRNA was detected solely in nodules, but not in roots, flowers, or developing fruits (Jeong *et al.*, 2004). The protein was localised to the interphace

between the plant cell and the bacteria. The expression of *AgDCAT1* in *E. coli* showed that AgDCAT1 mediates transport of dicarboxylates, like malate, succinate, fumarate and oxaloacetate (Jeong *et al.*, 2004). The transport of dicarboxylates by AgDCAT1 is energised by a membrane potential, suggesting that the anionic forms of dicarboxylates are transported (Jeong *et al.*, 2004). This energisation is similar to the soybean SM dicarboxylate transporter (Ouyang *et al.*, 1990). A similar affinity for malate, fumarate and succinate was reported for AgDCAT1 (Jeong *et al.*, 2004) but malate and oxaloacetate are preferred over succinate and fumarate by the soybean SM dicarboxylate transporter (Ouyang *et al.*, 1990). In addition the soybean dicarboxylate transporter has a much lower affinity for dicarboxylates than AgDCAT1. It is important to note that studies on AgDCAT1 were done with *AgDCAT1* expressed in *E. coli*, and the studies of the soybean SM dicarboxylate transporter was done on symbiosomes. This could possibly account for the much lower affinity towards dicarboxylates observed for AgDCAT1 compared to the soybean SM dicarboxylate transporter (Jeong *et al.*, 2004).

Since the function of AgDCAT1 *in vivo* probably involves transfer of dicarboxylates from the host plant to the N_2-fixing microsymbiont, AgDCAT1 is a very likely homolog of the legume SM dicarboxylate transporter.

Malate transport in heterologous systems

Yeast

Yeast species recognised for their ability to metabolise extracellular L-malic acid are classed as either K(-) or K(+), depending on their ability to utilise malate as sole carbon source. Both *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* are classified as a K(-) species, since they utilize TCA cycle intermediates only in the presence of

glucose or another carbon source reviewed by (Volschenk *et al.*, 2003). *S. cerevisiae* is not a very efficient consumer of malate, even in the presence of another carbon source (Camarasa et al, 2001), whereas *Sch. pompe* has been found to be the most potent yeast for malate degradation (Mayer & Temperli, 1963). In general, the ability to metabolise L-malic acid is dependent on an effective uptake system and the presence of a malate-converting enzyme, i.e. fumarase, malolactic enzyme, malate dehydrogenase or malic enzyme (Volschenk et al, 2003).

Studies have shown that malate transport in *S. cerevisiae* is not carrier mediated (Salmon, 1987) and only one mitochondria dicarboxylate transporter has been purified and characterised from *S. cerevisiae* reviewed by (Volschenk *et al.*, 2003). In contrast, evidence for carrier-mediated transport of dicarboxylates in *Sch. pombe* has been reported (Osothsilp & Subden, 1986a) and the gene *mae1* encoding a permease for malate and other C4-dicarboxylates has been isolated (Grobler *et al.*, 1995). DNA sequence analysis revealed that the *mae1* open reading frame is 1314 bp, which translates into a putative protein of 438 amino acids. The hydropathy profile of the predicted protein revealed a protein with hydrophilic N- and C- termini and 10 membrane-spanning helices or associated domains, typical of membrane-transport proteins from both prokaryotes and eukaryotes (Grobler *et al.*, 1995). Grobler *et al*. (1995) showed that the *mae1* gene of *Sch. pombe* encodes a general permease for L-malate, succinate and malonate and confirmed that it is constitutively expressed and not subject to catabolite repression.

Bacteria

The C4-dicarboxylate transporters from bacteria can be classified by protein sequence similarities into five coherent groups, the dicarboxylate transport (DctA), dicarboxylate uptake (Dcu)AB, DcuC, CitT and the tripartite ATP-independent periplasmic

transporter (TRAP) families (Janausch *et al.*, 2002). Dicarboxylates like succinate, fumarate and malate, and dicarboxylic amino acids like aspartate are accepted by most of these transporters (Janausch *et al.*, 2002).

The focus on the following will be on the C4-dicarboxylic transport (Dct) system, since this system is of particular importance for the symbiosis between legumes and rhizobium (Engelke *et al.*, 1987; Finan *et al.*, 1983; Jording *et al.*, 1992; Rastogi *et al.*, 1992). The DctA family comprises a large number of carriers from aerobic or facultative anaerobic bacteria and these carriers catalyse uptake of C4-dicarboxylates during aerobic growth (Janausch et al 2002). The Dct system consists of three genes, *dctA*, *dctB* and *dctD*. A putative transport protein encoded by *dctA*, while dctB and dctD activate transcription of *dctA* in response to the presence of dicarboylates (Engelke *et al.*, 1987; Jording *et al.*, 1992; Watson, 1990; Yarosh *et al.*, 1989). DctA is a typical membrane transport protein that comprises 12 membrane spanning regions, with both the N-terminus and the C-terminus located on the cytoplasm side (Jording & Puhler, 1993). Strains mutated in *dctA* have been shown to form ineffective nodules and the isolated bacteroids lack the ability to take up C4-dicarboxylates. (Engelke *et al.*, 1987; Finan *et al.*, 1983; Janausch *et al.*, 2002; Jording *et al.*, 1994).

Transport of organic nitrogen in legumes

As in any other plant translocation of organic N in legumes is vital for plant growth. Amino acids are the building blocks of protein synthesis, they are precursors of all other nitrogen-containing compound, including nucleic acids, growth regulators, photosynthetic pigments, alkaloids, and a wide rage of other essential compounds (Bush, 1999), like the ureides, allantoin and allantoic acid, the major long-distance transport forms of organic N in determinate nodules (Marschner, 1995). Consequently transporters of amino acids and other organic nitrogen-containing compounds, present

in the nodule are vital for the function of the symbiosis. Nitrogen fixed in the bacteroid and N assimilated into amino acid either in the bacteroid or in the plant cytosol needs to be translocated from the nodule to the leaves, where it can participate in the N metabolism of the whole plant. Within the nodule tissue, it would therefore be expected to find transporters of amino acid or ureides at least in the bacteroid membrane, the SM, the plasma membrane and in membranes of the vascular tissue. As described above genes encoding the general amino acid transporters *aap* and *bra* on the bacteroid membrane of *R. leguminosarum* have been identified. However, no genes encoding amino acid transporters localised in nodule cells have yet been reported. This is in spite of physiological evidence for a unidirectional not very selective release of amino acids (mainly asparagine, 62% of total amino acids) by infected cells and accumulation of amino acids into uninfected cells of pea nodules (Peiter *et al.*, 2004).

So far only one gene (*PvUPS1*) encoding an allatoin transporter from French bean nodules have been functionally characterised (Pelissier *et al.*, 2004). *PvUPS1* showed strongest expression in nodulated roots, consistent with the site of synthesis and translocation of allantoin, strongly indicating a function in these tissues (Pelissier *et al.*, 2004). In addition PvUPS1 mRNA was localised to the nodule and vascular endodermis and nodule phloem, which supports the proposed role of PvUPS1 in allantoin transport (Pelissier *et al.*, 2004).

Genes encoding amino acid transporters from legumes other than soybean have been identified and some of them have been functionally characterised (Bick *et al.*, 1998; Marvier *et al.*, 1998; Miranda *et al.*, 2003; Miranda *et al.*, 2001; Montamat *et al.*, 1999; Rolletschek *et al.*, 2005; Tegeder *et al.*, 2000). Surprisingly, expression of none of these genes was tested in nodules. Since none of these amino acid or putative amino acid transporters have been localised to the nodule, they will not be further discussed.

One putative amino acid transporter, *GmAAT-1* (GenBank, Accession no: AAK33098), has been identified during a genomic survey of soybean. No functional analysis or expression studies have been reported for this putative soybean amino acid transporter.

Identification of plant transport proteins by expression in heterologous systems

Demonstrating the function of a protein encoded by a putative transporter gene may be efficiently done by expression of the gene in a heterologous host cells. Heterologous expression is crucial for the analysis of proteins like transport proteins that have been shown to be difficult to manipulate biochemically (Frommer & Ninnemann, 1995). Heterologous expression of plant genes is of special value for the functional analysis of genes for which no plant mutants are available and no screening scheme or phenotype is predictable (Frommer & Ninnemann, 1995).

Heterologous gene expression in yeast

The experimental history of the budding yeast Saccharomyces cerevisiae is long and therefore the biology of *S. cereviciae* is well understood (Forsburg 1999). Naturally, *S. cerevisiae* has become a common molecular tool. Somewhat later the fission yeast *Sch. pombe* has been developed as a model system (Forsburg 1999). The use of two distinct yeasts can be very informative; if both yeasts solve the same problem the same way, the conclusion will be that this solution is general and possibly also occurring in other eukaryotes (Forsburg, 1999)

The level of gene expression, folding and targeting of the protein, and the proper orientation in the preferred compartment is vital to the success of heterologous

expression of transport proteins (Dreyer *et al.*, 1999; Frommer & Ninnemann, 1995). Although heterologous membrane proteins expressed in *S. cerevisiae* fail to reach their normal cellular location and instead accumulate in stacked internal membranes or the endoplasmatic reticulum (de Kerchove d'Exaerde *et al.*, 1995; Villalba *et al.*, 1992; Wright *et al.*, 1988), the use of heterologous complementation of yeast transport mutants has enabled isolation and functional characterisation of for example many genes encoding amino acid transporters (Frommer *et al.*, 1994a; Hsu *et al.*, 1993; Marvier *et al.*, 1998; Montamat *et al.*, 1999). This provides evidence that some hydrophobic membrane proteins appear to be suited for expression in yeast. However it is likely that experiments where functional expression of plant proteins in yeast has failed remain unpublished.

Heterologous gene expression in *E. coli*

The advantage of heterologous expression in *E. coli* is the availability of well-established molecular tools, defined mutants, high growth rates and high yield of over-produced proteins (Frommer & Ninnemann, 1995). Many plant genes have been identified by complementation of *E. coli* mutants (Frommer & Ninnemann, 1995). Bacteria, however, lack organelles and the mechanisms responsible for the type of RNA and protein processing found in eukaryotes, so that eukaryotic ploypeptides expressed in *E. coli* often denature and aggregate (Marston, 1986). In addition, many membrane proteins, even those from prokaryotes, are toxic when overexpressed in bacteria (Schertler, 1992). Despite these shortcomings, *E. coli* stain CBT315 was recently used to characterize the alder gene, *AgDCAT1*, that encodes a nodule-specific dicarboxylate transporter.

General aim of the thesis

The exchange of carbon and nitrogen metabolites within the nodule tissue, and in particular across the SM, is of major importance for the success of nitrogen fixation.

The major aim of this study was to increase the understanding of this exchange at the molecular level by isolating and characterising genes encoding transporters responsible for malate and amino acids exchange in soybean (*Glycine max*, L.), with particular interest in genes significantly or highly expressed in nodules. A combination of molecular and biochemical techniques was used to achieve this.

Chapter Two: Creation of an *Schizosaccharomyces pombe* malate transporter mutant for functional complementation with soybean cDNAs

Introduction

From studies with both plants and bacteroids, there is overwhelming evidence that C4 dicarboxylic acids such as malate and succinate are the main product of sucrose degradation supplied to the bacteroid in support of the bacteroid nitrogenase activity in legume nodules (Day & Copeland, 1991; Udvardi & Day, 1997). Biochemical studies provide evidence for a malate transport mechanism on the SM of soybean (Ouyang *et al.*, 1990; Price *et al.*, 1987; Udvardi *et al.*, 1988a), siratro (*Macroptilium atropurpureum*) (Ouyang & Day, 1992), pea (*Pisum sativum*) (Rosendahl *et al.*, 1992) and french bean (*Phaeseolus coccineus*) (Herrada *et al.*, 1989), however, the genes encoding the transporters remain unidentified.

Yeasts mutated in uptake mechanisms have proved to be invaluable molecular tools in the identification and characterisation of genes encoding vital transport proteins (Frommer *et al.*, 1995). One-step gene disruption in yeast is a very useful and relatively simple technique that requires a single transformation. The method can be applied to (a) determine whether a cloned DNA fragment contains a specific gene; (b) determine whether a cloned gene is essential; and (c) alter or completely delete a specific region (Rothstein, 1983). The method has been used to create mutants in both budding, i.e., *Saccharomyces cerevisiae* and fission, i.e., *Schizosaccharomyces pombe,* yeast, (Duenas *et al.*, 1999; Seron *et al.*, 1999; Toya *et al.*, 2001). The method takes advantage of the finding that during yeast transformation free DNA stimulates recombination by interacting directly with homologous sequences in the genome (Orr-Weaver *et al.*, 1981). Wach *et al*., (1994) further developed this technique, by construction of the pFA6a-kanMX6.

Malate degradation is glucose (or other glucolytically related compounds e.g. fructose or glycerol) dependent both in *Sch. pombe* and *S. cerevisiae* (Osothsilp & Subden, 1986b; Rodriquez & Thorton, 1990). In *S. cerevisiae* the uptake of malic acid has been shown to occur by simple diffusion and not to be carrier mediated (Salmon, 1987). In contrast studies in *Sch. pombe* have shown that malic acid uptake is carrier mediated (Osothsilp & Subden, 1986a). Sousa *et al*. (1992) concluded that the dicarboxylic carrier in *Sch. pombe* is a proton-dicarboxylate symporter and that the mono-anionic form of malic acid was the preferential substrate. The *Sch. pombe mae1* gene was cloned and identified on the basis of its ability to restore the ability of a yeast mutant to take up malate (Grobler *et al.*, 1995). It was shown to encode a transporter responsible for the transport of malate, succinate and malonate in *Sch. Pombe* (Grobler *et al.*, 1995). On the basis of these findings, *Sch. pombe* would be the preferred yeast to use when a malate transporter mutant was to be created, and subsequently used in functional complementation studies to isolate soybean malate transporter cDNAs.

Typically when cDNA libraries are screened for functional complementation in yeast, transformants are identified by their ability to grow on media containing a nutrient source that the mutant is unable live on. Thus only cells that have gained a gene encoding a transporter of that particular nutrient can grow. Since glucose is a carbon source as well as malate and since the presence of glucose is essential to malate metabolism in *Sch. pombe*, a typical screening assay is not suitable here. A malate glucose indicator medium (MGI) was developed based on the findings that malate degradation by *Sch. pombe* resulted in the increase in growth medium pH when a suitable amount of glucose was added to the medium (Osothsilp & Subden, 1986b). The presence of the pH indicator bromcresol green in the MGI medium enabled wild-type malate-utilising colonies to be distinguished from non-malate utilising colonies by a

simple colour change, where malate-utilising colonies would show blue colour and non-malate utilising colonies would have a cream colour (Osothsilp & Subden, 1986b).

The aims of this part of the work were to: i) create a *Sch. pombe* malate transporter deficient mutant; ii) synthesise a soybean (*Glycine max*) nodule/seedling cDNA library; 3) identify the gene(s) encoding putative soybean malate transporters by screening the cDNA library through functional complementation of the *Sch. pombe* malate transporter deficient mutant; and iv) functionally complement the *Sch. pombe* malate transporter mutant with the *LjNOD70* gene from *Lotus japonicus*, encoding the LjN70 protein postulated to represent a symbiosis-associated transport protein involved in translocation of carbon substrates in *L. japonicus* nodules (Szczyglowski *et al.*, 1998).

Materials and methods

General reagents

All reagents used were molecular biology grade, analytical grade or equivalent from commercial suppliers including Ajax Finechem (Seven Hills, Australia), Amersham Biosciences (Sydney, Australia), Astral Scientific (Caringbah, Australia), BDH (Crown Scientific, Perth, Australia), Clontech (Mountain View, Canada), Difco (Kansas City, USA), Merck (Melbourne, Australia), Molecular Probes (Eugene, USA), Roche (Sydney, Australia), Sigma-Aldrich (Sydney, Australia), Spectrum Chemical (New Burnswick, USA). Zymolase from *Arthrobacter luteus* was obtained from Scikagaku Corporation, Japan.

Radio-labelled malate and related products

^{14}C malic acid was supplied by Amersham Biosciences. Whatman glass microfibre filters GF/F 25 mm were obtained from Crown Scientific, Poly Q mini vials and caps

from Fisher Biotech (Perth, Australia) and Emulsifier-Safe LSC-scintillation cocktail from Packard Bioscience (Melbourne, Australia).

Primers

Oligonucleotides were synthesised by Sigma-Proligo (Lismore, Australia) or Invitrogen (Mount Waverley, Australia).

Yeast and bacterial strains

Sch. pombe strain *Sp*.011 (ade6-704, leu1-32, ura4-D18, h-) was kindly provided by Dr. Mary Albury (University of Sussex, Brighton, UK). *Escherichia coli* Epicurian Coli XL10-Gold and *E. coli* XL10 GOLD ultra-competent cells supplied by Stratagene (La Jolla, Canada).

Yeast media

Yeast extract plus supplements (YES) and Edinburgh minimal medium (EMM) were made according to Moreno *et al*., (1991). EMM-low medium is EMM medium containing only half the amount of glucose in the EMM medium. Malate-glucose indicator medium (MGI) was modified from Osothsilp & Subden (1986b) and composed of 1% (w/v) L-malic acid, 1% glucose (w/v), and 0.005 % (w/v) bromocresol green in EMM. Adjustment of EMM to pH 3.5 was accomplished by omitting Na_2HPO_4, adding 30 mM NaCl, and using H_3PO_4 for pH adjustment as described by (Saleki *et al.*, 1997). The pH in MGI media increased 0.5 pH units when agar was added to solidify the media.

Plant material and growth conditions

Soybean (*Glycine max* (L.) cv. Stevens) seeds were inoculated with *Bradyrhizobium japonicum* strain USDA 110 in a peat-based form and grown in a natural illuminated greenhouse in winter in Perth, Australia. Soybean seeds were soaked with inoculated

peat in tap water (Ratio 2:1:4). Seeds were sown at a 2 cm depth in 8 litre pots (19 seeds per pot) containing washed river sand. For the initial two weeks after planting, watering was solely with tap water and only when the sand in the top of the pot was dry. After two weeks, plants were watered with modified nitrogen-free Herridge solution (Herrigde, 1977) three days of the week and with tap water every day.

Plasmids

pFA6a-kanMX6 was used to construct the *mae1* gene disruption cassette. The kanMX module encodes the *E. coli* transposon Tn903 *kan*^{r} open reading frame and transcriptional and translational control sequences of the *TEF* gene from the filamentous fungus *Ashbya gossypii* (Wach *et al.*, 1994). This module permits efficient selection of transformants resistant to geneticin, hereafter G418, (Astral Scientific) and was used for gene disruption.

The *Sch. pombe* expression vectors pREP1, pREP41 and pREP81 (Basi *et al.*, 1993) have a *LEU2* auxotrophic marker gene and also contain the *nmt1* promoter/terminator sequences native to *Sch. pombe*. The *nmt1* promoter is tightly regulated by thiamine, and truncations of the TATA box result in different expression levels, with pREP1 having the highest expression level and pREP81 the lowest (Basi *et al.*, 1993). These vectors were kindly provided by Dr. Van Vuuren, University of British Columbia, Canada. Cloning vectors used were pCR2.1 (Invitrogen) and pGEMT (Promega, Annandale, Australia).

PCR

The Expand High Fidelity System (Roche) was used for all reactions where a proof reading polymerase was needed. PCR volumes were 20 - 50 µl and contained 1.5 mM $MgCl_2$, 200 µM of dNTP, 300 µM of each primer, 2.6 U of Expand High Fidelity PCR system enzyme mix and 0.05 – 1 µg of template.

In PCRs where a polymerase with proof reading ability was not necessary, *Taq* polymerase (Promega) was used. PCR volumes of 20-50 µl contained 1.5 mM $MgCl_2$, 200 µM of dNTPs, 300 µM of each primer, 0.5 U of *Taq* polymerase and 0.05 – 0.2 µg of template.

PCR parameters used were: After 2 min at 94 ºC, 30 cycles of 94ºC for 15 s, 55 or 57 ºC for 30 s, and 72 ºC for 2 min were carried out. A final extension step at 72 ºC for 6 min was included.

Colony PCR was used to verify presence of gene or DNA fragment in the plasmid of the transformed cells. Single bacterial or yeast colonies were selected and resuspended in 10 µl of sterile water. The PCR contained 12.5 µl 2X PCR Master Mix (Promega), 3.75 µl of cell suspension, and forward and reverse primers added to a final concentration of 300 nM in a total volume of 25 µl. The PCR program was: After 2 min at 94 ºC, 30 cycles of 94 ºC for 15 s, 45-57 ºC for 30 s, and 72 ºC for 2 min were carried out. A final extension step at 72 ºC for 6 min was included.

Purification of PCR products and gel electrophoresis

An aliquot of or the total PCR was resolved on a gel of 1-2% (w/v) agarose in TAE buffer (40 mM tris-acetate, 1mM EDTA pH 8 (Sambrook *et al.*, 1989) containing 1 µg ml^{-1} ethidium bromide. PCR products were purified directly using the QIAquick PCR purification kit (Qiagen) or from excised gel pieces using the QIAquick gel extraction kit (Qiagen) according to manufacturers instructions. DNA concentration was estimated by spectrophotometric measurements at 260 nm.

Ligation reactions and bacterial transformation

Ligation reactions contained 25-50 ng of vector. The molar ratio of insert DNA to vector was 5:1 (unless stated otherwise). One – 3 U T4 DNA ligase (Roche or Promega) was used in 10 µl reactions according to manufactures instructions. To transform cells,

1-5 µl of ligation reaction was gently added to 100 µl of chemically competent cells prepared using the method of (Inoue *et al.*, 1990). The cells were held on ice for 30 min, then heated for 1 min at 37 ºC and then placed back on ice for a minimum of 2 min. Cells were then spread onto LB medium (Sambrook *et al.*, 1989) containing 100 µg ml^{-1} ampicillin together with 40 µl of 50 mM isopropyl-2-D-thiogalactopyranoside (IPTG) and 25 µg $µl^{-1}$ 5-bromo-chloro-3-indoyl-2-D-galactopyranoside (X-gal) solution. Cells were grown at 37 ºC for at least 16 hours.

Plasmid isolation

Single colonies were selected and used to inoculate 2 ml of LB medium (Sambrook *et al.*, 1989) containing 100 µg ml^{-1} ampicillin and grown at 37 ºC overnight with shaking at 200-250 rpm. Plasmids were isolated using the Qiaprep Spin Mini Kit (Qiagen) according to the manufacturer's instructions.

DNA sequencing

An ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit version 1 or 3.1 with AmpliTaq DNA polymerase, FS (PE Biosystems Group, Applied Biosystem, Foster City, USA) was used for cycle sequencing. Each 5 µl reaction contained 2 µl reaction mix, 300 nM primer and 100-150 ng plasmid or 5-20 ng of PCR product DNA. Cycle sequencing consisted of an initial step for 1 min at 96 ºC followed by 30 cycles of 96 ºC for 10 sec, 50 ºC for 5 sec, 60 ºC for 4 min (Perkin Elmer, Gene Amp PCR System 2400). After the extension reactions, DNA was precipitated in 0.1 volume of 3 M sodium acetate, pH 5.2 and 2.5 volumes of ethanol. DNA was pelleted by centrifugation at 12000x*g*, washed with 70% (v/v) ethanol and dried. Separation of extension products was done at the West Australian Genome Resource Centre at the Royal Perth Hospital using an ABI Prism 3730 48 capillary sequencer (PE Biosystems Group, Applied Biosystem, Foster City, USA).

Construction of the Sch. pombe mae1 *disruption cassette*

A PCR method for producing a *mae1* disruption cassette (Figure 2.1, A) was developed after (Wach, 1996) and (Krawchuk & Wahls, 1999). Four primers were designed that bind either to the 5' region or to the 3' region of the *mae1* gene (Table 2.1); Mae1-WF and Mae1-XR, positioned 298-321 and 109-129 bases upstream from the *mae1* start codon, respectively, and Mae1-YF and Mae1-ZR, positioned 31-50 and 220-242 bases downstream from the *mae1* stop codon, respectively. These primers were used in PCRs with *Sch. pombe* genomic DNA as template. A polymerase with proofreading ability was used to minimise PCR introduced mutations. Primers Mae1-XR and Mae1-YF were designed as hybrid primers that also contain regions of homology to the pFA6a-kanMX6 module.

The fragments synthesized in the PCR, therefore, encoded a 213 or 212 base pair *mae1* region and a 26 or 27 base pair pFA6a-kanMX6 region. The pFA6a-kanMX6 module was used as template in a PCR containing these hybrid fragments and the outer most primers, Mae1-WF and Mae1-ZR. This resulted in a *kan*-containing disruption cassette flanked on each side by long tracts of the *mae1* sequence. In order to verify the construct, the final PCR product was inserted into pCR 2.1 and sequenced.

Table 2.1: Primers used in construction of disruption cassette.

Underlined sequence marks the part of the oligonucleotide that is complementary to the pFA6-MCS in plasmid pFA6-kanMX6 (Wach, 1996).

Primer name	Nucleotide sequence
Mae1-WF	AGCTTATTTGTTGCTGCACTAGA
Mae1-XR	GGGGATCCGTCGACCTGCAGCGTACG AAAAGACGGAGGAAGGAAAGA
Mae1-YF	GTTTAAACGAGCTCGAATTCATCGAT AATCTGTGTGCCCTGCTCTT
Mae1-ZR	GATTCATATGCAAGACGCATACA

Construction of mae1 *deletion mutant*

Sch. pombe cells were transformed with the kan-mae cassette using a protocol modified from (Bähler, 1998). Wildtype cells were grown at 32ºC in YES medium to a cell density of 10^5 cells ml^{-1} and 20 ml of cell suspension were used for the transformation. Cells were harvested by centrifugation at 3220x*g* for 5 min and washed twice with a double volume of sterile water, once with one volume of sterile LiAc/TE (10 X LiAc: 1 M lithium acetate, adjusted to pH 7.5 with dilute acetic acid; 10 X TE 0.1 M Tris-HCL, 0.01 M EDTA, pH 7.5). The cell pellet was then resuspended in LiAc/TE to a cell density of 2 X 10^9 cells ml^{-1}. Twenty µg of sheared herring testes DNA (Sigma-Aldrich) and 7.5 µg of kan-mae fragment were gently mixed with the cells. After a 10 min incubation at ambient temperature (20-25 °C), 600 µl of pre-warmed (30 ºC) 40% (w/v) PEG4000 in LiAc/TE were added and the cell suspension was gently mixed and incubated at 30 ºC for 60 min. DMSO (100 µl) was added and the cells were then heat shocked at 42 ºC for 5 min. The transformants were washed once with 1 ml of YES broth, resuspended in 1 ml of YES broth and incubated for 3 hrs at 30 ºC with shaking at 200 rpm before spreading on plates of solid YES medium containing 200 µg ml^{-1} G418. After incubation for 3 days at 30 ºC, the larger colonies (stable transformants) were spread onto fresh YES plates containing 200 µg ml^{-1} G418 and grown at 30 ºC for 3 days and then stored at 4 ºC.

Isolation of Sch. pombe *genomic DNA*

The Wizard Genomic DNA Purification Kit (Promega) was used for isolation of genomic DNA from *Sch. pombe* wild-type cells and mutant cells according to the manufacturer's instructions, except for the following modifications. *S. pombe* cells were grown in YES broth (including 200 µg ml^{-1} of G418 for transformants) to an OD_{595} of

2-3. Cells from 2 ml of this culture were used as starting material for the isolation procedure.

Isolation and subcloning of mae1 *and* LjNOD70

Mae1 was amplified by PCR from *Sch. pombe* genomic DNA using the primer pair Mae1-NdeI/Mae1-BamHI (Appendix 1) and using the primer pair LjNOD70-NdeI/LjNOD70-BamHI (Appendix 1) *LjNOD70* was amplified by PCR from a *Lotus japonicus* nodule cDNA library prepared from mature nodule mRNA, kindly provided by Dr. Jens Stougaard (Aarhus University, Denmark). These primers were designed to incorporate restriction sites for directed sub-cloning. The library was constructed in the UniZap Lambda vector (Stratagene) using an oligo-dT primer. Purified *mae1* and *LjNOD70* PCR fragments were ligated into pCR2.1 and further sub-cloned into the shuttle vectors pREP1, pREP41 and pREP81 using the introduced restriction sites *Nde*I and *Bam*HI.

Yeast Growth curves

Cultures grown for at least 16 hours were used to inoculate fresh medium to an OD_{595} of 0.03. The cultures were grown with shaking at 150-200 rpm at 32 ºC. Aliquots were taken at time intervals and OD measured at 595 nm. In order to ensure a linear relationship between OD measured a cell density, dilutions of the aliquots were measured so that all OD_{595} were in the range of 0.1 to 0.3.

Growth of Sch. pombe *wild-type cells and transformants on malate-glucose indicator media*

Single colonies were streaked onto plates containing solid MGI medium. The plates were incubated at 30 ºC and colour development monitored by periodic viewing.

Uptake studies

Yeast cells grown to log phase in YES broth at 32 ºC with shaking at 200 rpm were collected by centrifugation for 5 min at 3220x*g* and washed twice in wash buffer (0.1 M Na-phosphate buffer, pH 5, containing 3% (w/v) glucose). Cells were then resuspended to an OD_{595} of approximately 25. Cells were incubated at 30 ºC for 10 min before initiation of uptake by addition of an equal volume uptake buffer (4 mM malate; 3% (w/v) glucose; 0.1 M Na-phosphate, pH 5, and trace amounts of [L-^{14}C] malic acid (specific activity 0.2 µCi/µmol malate). At intervals, 100-200 µl aliquots were taken and filtered through GF/F glass micro-fibre filters. Cells were then washed with 3 x 5 ml of ice-cold wash buffer. Subsequently filters were transferred to a Poly Q mini vial containing 4 ml of scintillation fluid and the vials were mixed thoroughly. Radioactivity was measured by counting beta radiation (LS 6500 Multi-purpose Scintillation Counter, Beckman Instruments, Fullerton, Canada). Uptake was calculated as nmol malate taken up per mg dry weight of yeast cells. To determine the dry weight, the cells were filtered onto 0.22 µM filters (Millipore, Bedford, USA). Filters dried in an oven at 70 ºC for at least 16 hours. The dry weight was calculated as the difference between the weight of the filter and cells and the filter only.

Construction of pREP41b

The expression vector pREP41 was modified to give pREP41a (Appendix 2), which was further modified to give pREP41b (Appendix 2). Primers were designed (pREP-MCS-F and pREP-MCS-R) to amplify parts of the poly-linker from expression vector pYES3 (Smith *et al.*, 1995) and to add a stop codon, directly after and in frame with the start codon in pREP41 (Appendix 2). The amplified fragment was PCR purified as described above and ligated into the pREP41 vector by means of the restriction enzymes sites *Nde*I and *Sma*I. As the stop codon after the *Nde*I site in pREP41a was not

desirable, pREP41a was digested with *Nde*I and *Xho*I, and re-ligated with the synthesised oligonucleotides (MCS-pREP41b-A and MCS-pREP41b-B). Prior to ligation the oligonucleotides were resuspended in sterile water to a molar concentration of 1 mM. Equal amounts were mixed and the suspension heated to 95 ºC for 2 min, and then cooled down to RT. In the ligation reaction the molar ratio was 1/10 of vector to insert.

RNA Extraction and mRNA purification

Nodules from 7-week-old soybean plants, and 13-day-old seedlings, were harvested, quickly frozen in liquid nitrogen and stored at –80 ºC. For RNA extraction, whole nodules and seedlings were ground to a fine powder in liquid nitrogen and equal amounts by weight of each were mixed. Total RNA from the mixed tissue was extracted using the RNeasy Plant Mini Kit (Qiagen) and poly A+ mRNA was purified from the total RNA using the Oligotex mRNA Mini Kit (Qiagen) according to manufacturers instructions. A sample of purified mRNA was visualised on an ethidium bromide-stained 1% (w/v) agarose gel in TAE. The mRNA was estimated by spectrophotometric readings at 260 and 280 nm.

Construction of a soybean nodule/seedling cDNA library

A soybean nodule/seedling cDNA library was constructed using the mRNA from mixed tissue and the cDNA inserted in the sense orientation downstream of the *Sch. pombe nmt1* thiamine repressible promoter in pREP41b. The protocol from the Lambda ZAP-CMV XR Library Construction Kit (Stratagene) was used to synthesize the cDNA, with the following changes: the *Eco*RI adapter was omitted and a *Bam*HI adapter used instead; consequently the cDNA was digested with *Bam*HI and *Xho*I. Ligation of digested cDNA into pREP41b, transformation of *E. coli* XL10 GOLD ultra-competent cells (Stratagene) and amplification of the primary library was done following the

instructions given in the pCMV-ScriptTM XR cDNA Library Construction Kit manual (Stratagene). The amount of cDNA was quantitated using the picoGreen dsDNA Quantification Kit (Molecular Probes). Sample fluorescence was detected using a FluorImager SI (Vistra Fluorescence) and analysed using IQMac v1.2 software (Visual Molecular Dynamics).

Yeast transformation

A simplified lithium acetate method described by Elble (1992) was used to transform *Sch. pombe* wild-type cells and *mae1* mutant cells when a high number of transformants was not necessary. When the *Sch. pombe mae1* mutant cells were transformed with the soybean nodule/seedling cDNA library, a high-frequency transformation protocol for *Sch. pombe* using lithium acetate modified from (Okazaki *et al.*, 1990) was used. The cells were grown with shaking at 200 rpm at 32 ºC in EMM-low broth with Leu (+Leu) to a cell density of 10^7 cells ml^{-1}. Cells were harvested by centrifugation at 3220x*g* for 5 min, washed twice with sterile water, resuspended in 1 M lithium acetate (adjusted to pH 4.9 with acetic acid) to a cell density of 10^9 cells ml^{-1} and incubated at 32 ºC for 1-2 hrs. The cDNA Library (1 µg in 15 µl TE pH 7.5) was added to 100 µl of resuspended cells and then 290 µl of pre-warmed (30 ºC) 50% (w/v) PEG 4000 was added. The suspension was mixed gently, incubated for 60 min at 30 ºC and then heat shocked at 43 ºC for 15 min. After cooling to ambient temperature (20-25 ºC), the cells were pelleted (2370x*g* for 2 min), resuspended in 10 ml 1/2 strength YES broth and incubated with shaking (200 rpm) at 32ºC for 60 min. Cells were subsequently pelleted as above, washed in EMM (-Leu) broth and resuspended in 1 ml EMM (-Leu) broth before spreading them on MGI (-Leu) agar plates. The plates were incubated for 5 - 10 days at 30 ºC and subsequently stored at 4 ºC.

Results

Synthesis of a mae1 *disruption cassette*

The *mae1* disruption cassette was synthesized using PCR in order to generate a *Sch. pombe* malate transporter deficient mutant. Products with regions homologous to each side of the *Sch. pombe mae1* gene as well as the *kan*^{r} gene were initially amplified. The expected size PCR products, 213 bp for the 5' end of *mae1* and 212 bp for the 3' end of *mae1*, were obtained (Figure 2.1 A and B). A subsequent PCR, using these PCR products and the pFA6a-KanMX6 module as template, was then done to "glue" the *mae1* regions to the *kan*^{r} gene. A PCR product of the expected size (1.9 kb) was obtained (Figure 2.1 A and C). The PCR product was following cloned and transformed into *E. coli*.

Colony PCRs of 24 colonies with the Mae1-WF and Mae1-ZR primers (Figure 2.2) were used to screen for transformants containing the *mae1* disruption cassette, two colonies were found that produced an intense PCR product of the expected size, which was 1.9 kb (Figure 2.2, lanes 15 and 18). A faint PCR product of expected size was also evident in the remaining lanes, since these fragments also appeared in the negative control. This indicates contamination of the PCRs. However, DNA sequencing of isolated plasmids from colony number 15 and 18 verified that they contained the hybrid *mae1* disruption cassette (Appendix 3). Plasmids extracted from colony number 15 were subsequently used as template in a set of PCRs with the primer set Mae1-WF/Mae1-ZR to amplify the *mae1* disruption cassette.

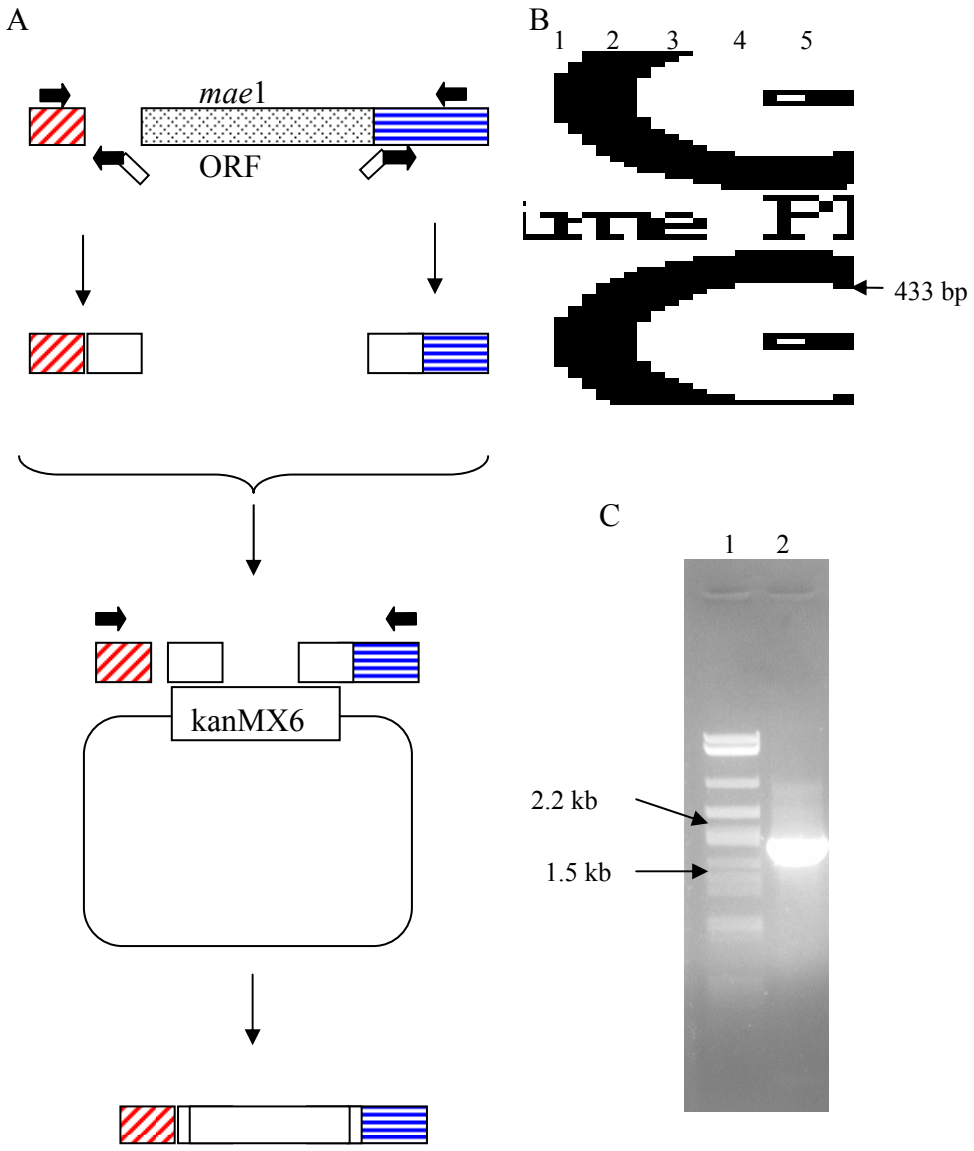


Figure 2.1: Generation of the *mae1* disruption cassette.

A) Schematic diagram showing the PCR synthesis of the *mae1* disruption cassette. PCRs were done to amplify 213 and 212 bp long DNA amplicons encoding the 5'- (red) and 3'- (blue) untranslated regions of *mae1* from *Sch. pombe* genomic DNA. The fragments also encoded 26 and 27 bp extensions, respectively homologous to the kanMX6 marker gene (white). Each of these fragments were then used in a subsequent PCR, where one strand of each served as primer in a PCR using the kanMX6 fragment as a template, in addition the primers Mae1-WF and Mae1-ZR were present in this reaction (black arrows). B) PCR products encoding the *mae1* 5'- and 3'-ends. Lanes 1 and 3: the 213 bp 5'-end product, using 0.5 and 1 µg of template DNA, respectively. Lanes 2 and 4: the 212 bp 3'-end product, using 0.5 and 1 µg of template DNA, respectively. Lane 5: DNA marker: λ genomic DNA digested with *Ava*II. One-fifth of each PCR was separated on 2% (w/v) agarose gel. C) PCR to amplify the *mae1* disruption cassette. Lane

1: DNA marker: λ genomic DNA digested with *Ava*II. Lane 2: The expected 1.9 kb PCR product encoding the hybrid kan-mae fragment. One-fifth of the PCR was separated on 1% (w/v) agarose gels.

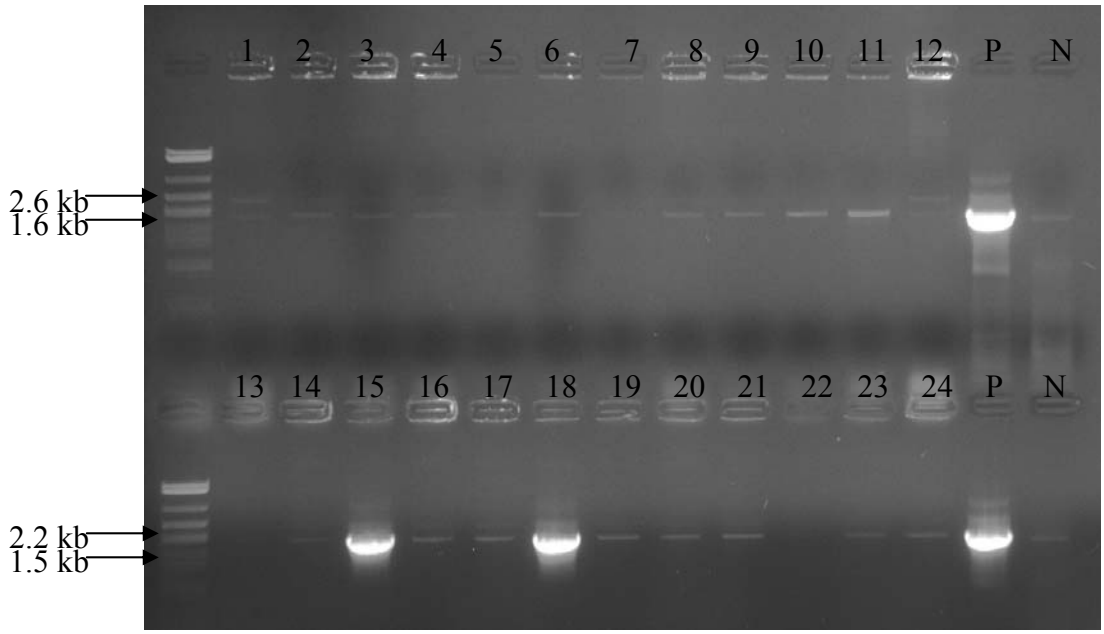


Figure 2.2: Colony PCR of bacterial transformants to verify presence of *mae1* disruption cassette

The first lane on the left in each row is DNA marker: λ genomic DNA digested with *Ava*II. The sizes of selected bands are indicated on the left. Number 1-24: colony number, P: *mae1* disruption cassette positive control, N: no template control. The total 25 µl of each PCR was separated on a 1 % (w/v) agarose gel. The primer set Mae1-FW/Mae1-ZR was used to amplify the *mae1* disruption cassette from bacteria transformed with the ligation reaction of pGEMT and PCR product of *mae1* disruption cassette. Two of the 24 colonies screened show similar sized fragments as the positive control. Sequencing verified that these clones contained the *mae1* disruption cassette.

Selection of Sch. pombe *transformants and identification of* mae1 *mutants*

Four days after transformation of *Sch. pombe* with the *mae1* disruption cassette, hundreds of transformants were present on YES plates containing 200 µg ml^{-1} G418. Some colonies grew 2 to 4 times larger than others (data not shown). The smaller colonies were probably non-stable transformants (Bähler, 1998) and were omitted from further analysis. Thirteen large colonies were selected and re-streaked onto fresh

YES/G418 plates. Within 3 days, all transformants grew, whereas wild-type cells grown as control did not.

Colony PCR with the primers Mae1iIII and Mae1iVI (Appendix 1) and the yeast transformants that grew on YES containing G418, produced a PCR product of the expected size (464 bp) in eight samples (Figure 2.3: Colony PCR of yeast transformants. This result indicated that eight colonies contained an intact *mae1* gene that had not been disrupted by the *mae1* disruption cassette. However since the cells grew on medium containing G418, the *mae1* disruption cassette must have been inserted elsewhere in the genome. These eight colonies were therefore excluded from further study.

A PCR strategy was designed (Table 2.2 and Figure 2.4) to verify that the remaining 5 colonies now designated *mae1* mutant T1, T3, T8, T10 and T11, contained cells with a disrupted *mae1* gene. Four different PCRs were set up. Genomic DNA extracted from the different transformants was used as template in each PCR.

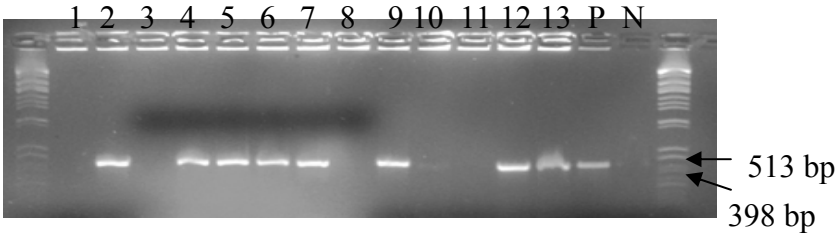


Figure 2.3: Colony PCR of yeast transformants.

The numbers, 1-13, on top of the lane indicate colony number; P: positive control (wild-type colony) and N: no template control. DNA marker: λ genomic DNA digested with *Ava*II, on each side. The size of two DNA fragments of the marker is indicated on the left. The total 25 µl of each PCR was separated on a 1% (w/v) agarose gel. The primers Mae1iIII and Mae1iVI produced a 464 bp PCR product in transformants containing part of or an entire *mae*1 gene. These transformants were excluded for further investigation.

The primer set mae1-usF/kanmaeR3 generated a PCR product of expected size (1.2 kb) in all five transformants No positive control was available for this PCR and no amplification was found in the negative control (Figure 2.4, primer set 1). This indicates that all five transformants have a disrupted *mae1* gene. The primers kanmaeF/kanmaeR2 amplified a PCR product of 933 bp as expected for transformants. The *mae1* disruption cassette served as positive control and the band seen in the negative control was due to overflow from the positive control sample (Figure 2.4, primer set 2). This set of PCRs serves as a positive control for the PCR. The primer set mae1iV/mae1exR was designed to amplify a part of the *mae1* gene that should have been knocked out and therefore no PCR product should be detected in this PCR for transformants with a disrupted *mae1*. In fact no PCR product was generated in any of the transformants. Using genomic DNA from the wild-type cells as a positive control, the expected size fragment (748 bp) was generated. The negative control showed no PCR products (Figure 2.4, primer set 3). The fourth PCR was a multiplex PCR with six primers present in the same reaction (Table 2.2). From this PCR two PCR products of 230 bp and 1.3 kb were generated as expected from all five transformants (Figure 2.4, primer set 4). The 230 bp product was very intense while the 1.3 kb band was faint. A reasonable explanation for this is that the same forward primer, kanmaeF, was used for each PCR product and since smaller products are easier to amplify, this would be the most abundant product. No products were amplified in the negative control. In the positive control where the *mae1* disruption cassette was used as template, the 230 bp expected size product was amplified. However, where genomic DNA from wild-type cells was used only the smaller 464 bp expected size product was amplified. Here again the same forward primer was used for amplification of both PCR products and then only the smaller product was amplified. In conclusion, this experiment verified that all five

transformants had been disrupted in their *mae1* gene and further analyses of the transformants could proceed.

Table 2.2: PCR strategy to verify insertion of the *mae1* disruption cassette into the *Sch. pombe mae1* locus.

For a schematic explanation of this table see Appendix 4.

Primers names* (PCR set number):	Expected size fragment:	Interpretation:	Positive control (positive control indication on gel):
mae1-usF/kanmaeR3 (1)	1230 bp	Desired transformant	-
kanmaeF/kanmaeR2 (2)	933 bp	Any transformant	*mae1*-disruption cassette (2P)
mae1iV/mae1exR (3)	748 bp	WT	genomic DNA WT (3P)
kanmaeF/kanmaeR3 (4)	230 bp	Any transformant	Mae1-disruption cassette (4P2)
mae1iIII/mae1iVI (4)	464 bp	WT/false transformants	genomic DNA WT (4P1)
kanmaeF/mae1exR (4)	1308 bp	Desired transformant	-
mae1iIII/mae1exR (4)	1035 bp	WT/false transformants	genomic DNA WT (4P1)

* For sequence information on the primers used see Appendix 1.

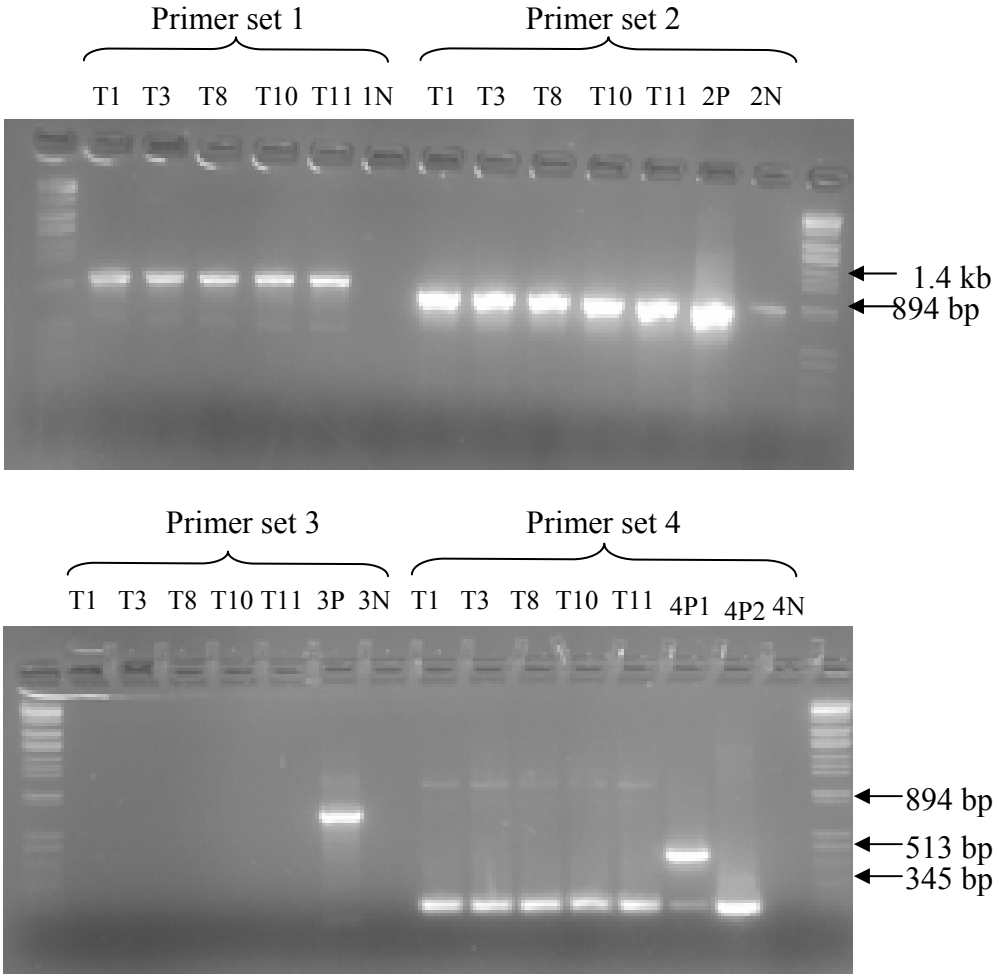


Figure 2.4: Verification of Sch. pombe *mae1* mutants using PCR.

PCR's using genomic DNA extracted from transformant T1, T3, T8, T10, T11 and WT. Four different PCR's were done to verify that the mae1 disrution cassette was introduced at the *mae1* locus, thereby deleting the *mae1* ORF. DNA marker: λ genomic DNA digested with *Ava*II, shown on each side. The sizes of particular DNA fragments of the marker are indicated on the left. N: no template control, P: positive control (see Table 2.2). The total 25µl PCR was separated on 1% (w/v) agarose.

Functional characterisation of the mae1 *mutant cells*

Growth curves showed that the *mae1* mutant cells had a longer lag phase than wild-type *Sch. pombe* cells (Figure 2.5) In addition the mean generation time was longer for the *mae1* mutant cells than for wild-type cells, 3.2 hrs (*mae1* mutant T1) and 2.7 hrs, respectively. When grown on solid MGI medium, the *mae1* mutants were cream to

green in colour (same colour as medium), whereas the wild-type colonies appeared blue (Figure 2.6). These results indicated that no malate degradation is accomplished by the *mae1* mutant cells and growth-medium pH remains unchanged, which suggest that the *mae1* mutants no longer were able to take up malate from the medium. Cream colonies were also obtained when *S. cerevisiae* strain R757 was streaked on MGI plates. *S. cerevisiae* has been reported to be inefficient in malate degradation (Osothsilp & Subden, 1986b; Volschenk *et al.*, 2003). In addition *S. cerevisiae* has been reported to lack an active transport system for malate (Salmon, 1987). Absorbance at 620 nm of the bromocresol was followed in the MGI medium where either wild-type cells or *mae1* mutants grew (data not shown). An increase in absorbance over time was measured from medium of wild-type cells, but no change in absorbance was measured in the medium of *mae1* mutants. This experiment verifies the plate assay.

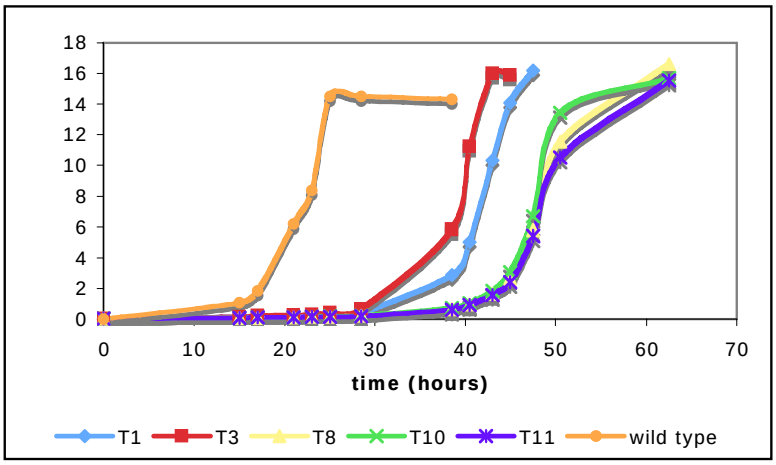


Figure 2.5: Growth curves of *mae1* mutant and wild type cells.

The different *mae1* mutants (T1, T3, T8, T10 and T11) and wild-type cells were grown in YES medium. Growth curves were performed in duplicate. The data points shown are the mean of two samples. Which growth curve corresponds to what yeast is indicated in the figure.

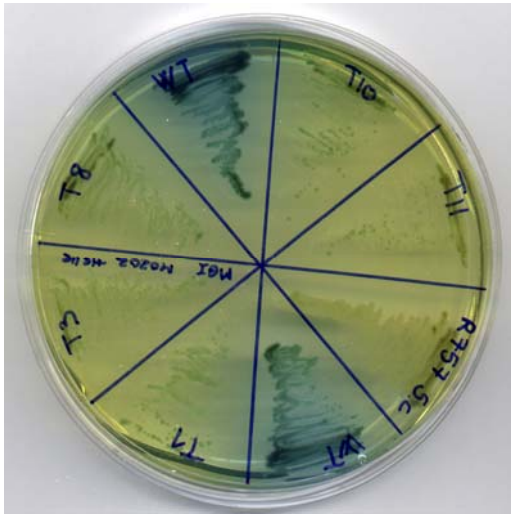


Figure 2.6: Growth of *mae1* disruption mutant and wild-type cells on MGI plates.

The *mae1* mutants T1, T3, T8, T10 and T11 after 4 days of growth on MGI medium. Wild-type strain *Sp*011 (WT). *Saccharomyces cerevisiae* strain R757 (R757 Sc). Blue colour indicates malate uptake and degradation, followed by medium pH increase. Cream- green colour indicates medium pH is unchanged, due to lack of malate degradation.

To confirm that the *mae1* mutants indeed were not taking up malate, ^{14}C-malate uptake studies were done. Initial uptake studies showed that all the *mae1* mutants lacked the ability to accumulate malate and *mae1* mutant T1, hereafter called *mae1* mutant, and was chosen for further analysis. Wild-type cells accumulated malate over time and this uptake was found to continue over 60 min, although the rate of uptake was higher in the first 10 min than the rest of the time (Figure 2.7). In contrast over the same period the *mae1* mutant did not show a similar uptake pattern (Figure 2.7). Even after 60 min incubation, T1 cells showed very little uptake of ^{14}C-malate, in fact 16 times less than the wild-type cells.

Transformation of the *mae1* mutant with a plasmid containing the *Sch. pombe mae1* ORF restored the cells' ability to take up malate (Figure 2.8). This result indicated that only the lack of a functional *mae*1 gene was responsible for the mutant phenotype.

The late nodulin gene, *ljNOD70* (Szczyglowski *et al.*, 1998), from the model legume *L. japonicus*, which encodes a protein with amino acid sequence similar to the oxalate/formate exchange protein from *Oxalobacter formigenes*, was tested for its ability to complement the *mae1* deletion (Figure 2.8). However, no restoration of malate uptake was observed as determined both by growth on MGI medium, where no colour change was observed (data not shown), and by ^{14}C-malate uptake.

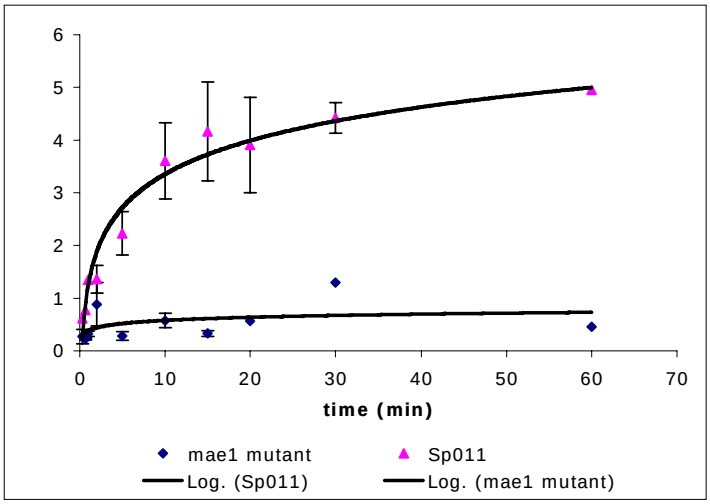


Figure 2.7: Accumulation of malate wild-type Sch. pombe cells and the *mae1* mutant cells.

Malate accumulation into wild-type cells (*Sp*011, ▲) and *mae*1 mutant cells (mae1 mutant, ◆) incubated with 2 mM ^{14}C-malate (specific activity 0.2 µCi/µmol). Values are expressed as nmol malate per mg yeast cell dry weight (dw). Where error bars are present data are means± SE of five separate experiments otherwise data represent a single experiment.

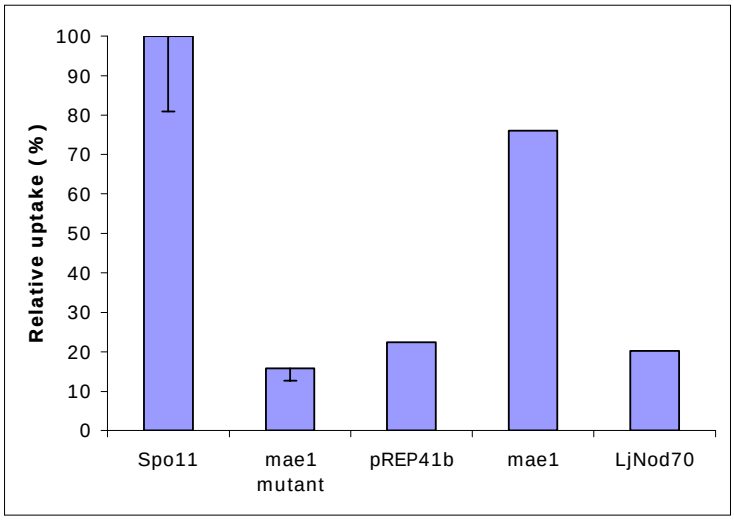


Figure 2.8: Uptake of malate in *mae1* mutant cells transformed with empty vector, *mae1* and *LjNOD70*.

^{14}C-malate uptake is given in percent uptake relative to that of wild-type cells. Wild-type cells (Spo11) and *mae1* mutant transformed with: *mae1*::pREP41b (mae1), *LjNOD*70::pREP41b (LjNOD70) or pREP41b (pREP41b) and then incubated with 2mM malate (specific activity 0.2 µCi/µmol) for 10 min. Where error bars are present data are means± SE from five separate experiments otherwise data represent a single experiment.

Screening the soybean nodule/seedling cDNA library for inserts complementing the Sch. pombe mae1 *mutation*

The soybean nodule/seedling primary cDNA library consisted of 2 x 10^5 cfu in *E. coli*. After amplification the library contained 3 x 10^{13} cfu in *E. coli*. Transformation of the *mae1* deletion mutant with the amplified cDNA library resulted in a total of 5 x 10^5 cfu being screened. However, none of the 5 colonies that appeared blue on the initial MGI screening plate remained blue when re-streaked onto fresh MGI plates. This indicated that no soybean cDNA fragments had complemented the *mae1* mutation. PCR analysis of 50 randomly chosen *E. coli* colonies, using pREP41b specific primers to amplify the soybean cDNA inserts, showed 60% of the clones contained inserts smaller than 600 bp

(data not shown). This result indicates that full-length cDNA clones were likely to be under-represented in the library.

Discussion

The first aim of this chapter was to create a yeast malate transporter mutant. This aim was achieved with the fission yeast *Sch. pombe*. In *Sch. pombe* homologous integration efficiencies have been shown to vary greatly, depending on the sizes of the flanking homologous regions. With a 40 bp flanking region, homologous integration efficiencies of 1-3 % have been obtained and efficiencies increased to greater than 50 % with 500 bp of flanking homology (Krawchuk & Wahls, 1999). A PCR-based strategy that generated more than 200 bp flanking homology to the *mae1* locus on each side of the *kan*^{r} gene was chosen for this work. This approach worked well. Initial verification of the yeast transformants was done by colony PCR. However, because some of the primers did not perform well in colony PCR, I chose to isolate genomic DNA from *mae1* mutants T1, T3, T8, T10 and T11 and do a new series of PCR. It was then confirmed that 38 % of selected transformants did have a disrupted *mae1* gene.

The inability of the *mae1* mutants to transport malate was shown both by growth on MGI media, where no colour change was observed for the mutants, and by ^{14}C- malate uptake studies, where the wild-type cells were found to take up malate at a higher rate than the *mae1* mutant cells. Even after 60 min incubation with the radiolabelled precursor, the *mae1* mutant still contained a very low level of malate. These uptake studies were performed at 30 ºC and the non-specific binding of ^{14}C-malate to the yeast cells was not accounted for. The non-specific binding of the precursor to the cells is important for the kinetic characterization of a transport mechanism; however, in this case where no uptake or very low uptake was observed, it was not necessary to measure non-specific binding.

Further evidence that the inability of the mutant to transport malate was due to disruption of the *mae1* gene was the restoration of malate transport to nearly wild-type levels when the mutant was transformed with a plasmid containing the *mae1*. The mutant regained 75% of wild-type ability to take up malate. The *nmt* promoter controls expression of *mae1* from pREP41b. This promoter might not allow expression of *mae1* at the same level as in the wild-type cells, and could account for the difference in uptake by the wild-type cells and the *mae1* mutant cells transformed with *mae1*::pREP41b.

The second and main aim of this chapter was to synthesise a soybean cDNA library and screen this library by functional complementation of the *mae1* deletion mutant. The soybean nodule/seedling primary cDNA library that was constructed contained 2 x 10^5 cfu. Screening of the soybean nodule/seedling cDNA library failed to identify any cDNAs encoding putative malate transporters. The quality of the cDNA likely played a role to this outcome, because analysis of the cDNA found that 60% of the clones contained cDNAs ≤ 600 bp in length. This result indicated that full-length cDNAs were strongly under-represented in the library and that the likelihood of isolating a full-length clone would be small. However, we do not know what size cDNA could be expected.

Several other factors inherent to the system probably also contributed to making the identification of putative malate transporters difficult. First, the inability of *Sch. pombe* to metabolise malate as a sole carbon source (Osothsilp & Subden, 1986b; Rodriquez & Thorton, 1990; Sousa *et al.*, 1992) complicates screening of the cDNA library (Osothsilp & Subden, 1986b). Typically when cDNA libraries are screened for transport activities, transformants are identified by growth on medium containing a nutrient source on which the mutant cannot live; thus only cells that have gained a gene encoding a transporter enabling them to take up the nutrient can grow. When screening on MGI medium, all yeast cells containing a plasmid grow, because malate is not

provided as the sole carbon source. The screen then relies on a pH change accompanying malate uptake for a transformant to be identified. Second, these experiments relied on a heterologous system where a plant gene(s) needs to be transcribed, translated and functionally integrated into yeast cells. Plant transporter proteins that complement yeast uptake-deficient mutants must be localised on the plasma membrane. Thus the yeast targeting machinery must either recognize the protein correctly or the proteins must reach their destination via unspecific mechanisms (Frommer & Ninnemann, 1995). Although many transporters have been identified this way (Frommer *et al.*, 1994a; Frommer *et al.*, 1993; Hsu *et al.*, 1993; Marvier *et al.*, 1998; Ninnemann *et al.*, 1994) reports have shown that when plant genes are expressed in yeast, the encoded protein can end up in stacked membranes or the endoplasmatic reticulum and not reach the plasma membrane (de Kerchove d'Exaerde *et al.*, 1995; Palmgren & Christensen, 1993; Palmgren & Christensen, 1994; Villalba *et al.*, 1992; Wright *et al.*, 1988). There is also evidence that only a fraction of the heterologous expressed protein, which reaches the yeast plasma membrane, is functional (Rentsch *et al.*, 1998). Third, the orientation of the protein in the plasma membrane is crucial. In theory, thermodynamically all carriers should function in a reversible manner; however, *in vivo* the proton co-transporters, in particular, function as import systems and whether these identified transporters in an uncoupled mode also function in metabolite export, remains to be tested (Rentsch *et al.*, 1998). It has been shown that the dicarboxylate carrier on the SM of soybean nodules has a high affinity for the monovalent malate anion, and that the energisation of the membrane controls the transport rate (Ouyang *et al.*, 1990; Udvardi *et al.*, 1988a). However it has not been shown that malate is transported in symport with protons across the SM. If the malate transporter on the SM *in vivo* only functions as an import mechanism, and if this transporter in the heterologous yeast system is orientated, so that malate can only be transported out of the

cell, this transporter would not be identified in the particular screening system that I used.

Malate transport systems are present through out the plant and found on; the inner mitochondria membrane, the inner envelope of chloroplasts and the tonoplast, and possibly also on the membrane of glyoxysomes and peroxisomes, and the plasma membrane (Martinoia & Rentsch, 1994). However, if transporters are located on the plasma membrane, at least in roots, these may only function to export malate (Martinoia & Rentsch, 1994). That is, plant cell membranes may not have uptake malate transporters. However to maximize the chances of getting complementation a nodule/seedling cDNA library and not only a nodule cDNA library was constructed.

The third aim of this part of the project was to test if *LjNOD70*, when transformed into the *mae1* deletion mutant, would result in restoration of the mutant's ability to take up malate. The *LjNOD70* nodule-specific gene encodes a poly-topic hydrophobic protein, LjN70, which was postulated to represent a symbiosis-associated transport protein involved in translocation of carbon substrates in *L. japonicus* nodules (Szczyglowski *et al.*, 1998). However, no complementation of the malate uptake deficiency of the *mae1* mutant was found following transformation with the *LjNOD70* ORF. Recently it was reported that *LjNOD70* and *GmN70*, its soybean ortholog, encode inorganic anion transporters of the symbiosome membranes, with enhanced preference for nitrate (Vincill *et al.*, 2005).

Another approach towards identification of genes encoding malate transporters in the soybean genome would be to search for soybean Expressed sequence tags (EST) that show sequence similarity to malate transporters identified in other plant species. This is pursued in a subsequent chapter.

Chapter Three: Characterisation of putative dicarboxylate transporters from soybean

Introduction

Malate occupies a central role in plant metabolism. It is an intermediate in the tricarboxylic acid and glyoxylate cycles. In C_4 and Crassulacean acid metabolism plants, CO_2 is stored as malate (Martinoia & Rentsch, 1994). Extracellular chelation by organic acids, such as malate and citrate, is important in mechanisms of aluminium tolerance (Larsen *et al.*, 1998; Ryan *et al.*, 1995). Malate is also involved in the regulation of stomatal aperture by maintaining osmotic pressure and charge balance (Emmerlich *et al.*, 2003). In addition malate plays a key role in the legume-*rhizobium* symbiosis, where it is believed to be the major carbon source supplied to the bacteroid in return for reduced nitrogen (Udvardi & Day, 1997).

Biochemical studies have provided evidence for a malate transport mechanism on the SM in different legumes species (Herrada *et al.*, 1989; Ouyang *et al.*, 1990; Ouyang & Day, 1992; Price *et al.*, 1987; Rosendahl *et al.*, 1992; Udvardi *et al.*, 1988a) however, the genes encoding these transporters remain unidentified.

Until recently genes encoding plant dicarboxylate transporters had only been identified from mitochondria and chloroplasts (Catoni *et al.*, 2003; Taniguchi *et al.*, 2002; Taniguchi *et al.*, 2004; Weber *et al.*, 1995). Tonoplast malate transporters have been described at physiological and biochemical levels for some plant cells (Cerana *et al.*, 1995; Martinoia *et al.*, 1991; Pantoja & Smith, 2002) and their characteristics are somewhat different to those of the SM dicarboxylate transporter. However, the SM shares common biochemical properties to both the plasmalemma and the tonoplast of the host and appears to be a mosaic membrane (Verma & Hong, 1996). Thus transporters located on those membranes might share common features. Furthermore,

transport of solutes either into the symbiosomes or the vacuole is similarly directed, namely it represents export from the cytosol.

In addition to these mitochondrial and chloroplast dicarboxylate transporters, three dicarboxylate transporters have recently been characterised at the molecular level. Emmerlich *et al.* (2003) were the first to describe the *Arabidopsis thaliana* tonoplast dicarboxylate transporter, *At*tDT, while Sasaki *et al.* (2004) identified a wheat gene, *ALMT1*, which encodes a transporter that facilitates organic anion efflux from roots, and Jeong *et al*. (2004) identified a gene encoding a nodule-specific dicarboxylate transporter from alder (*AgDCAT1*). One common feature of these three transporters is that they are believed to facilitate transport of one or more dicarboxylates out of the cytosol. The direction of the transport out of the cytosol is of particular interest for the identification of the malate transporter located on the SM of legume symbiosomes, since malate is transported out of the cytosol across the SM and into the symbiosomes. Moreover, none of the above transporters exhibit considerable sequence similarity at nucleic acid or protein levels to the dicarboxylate transporters identified in plant mitochondria and chloroplasts.

Vacuoles isolated from leaf discs of *Arabidopsis AttDT* knock-out mutants were shown to have a decreased level of malate uptake compared to vacuoles isolated from leaf discs of wild-type plants. *AttDT* was also the first plant gene belonging to the solute carrier 13 (SLC13) gene family to be identified. The vertebrate SLC13 genes encode proteins with 8-13 transmembrane domains and have a wide tissue distribution (Markovich & Murer, 2004). In addition they are DIDS-insensitive, sodium-coupled symporters, with substrate preference for divalent anions such as sulphate and/or Krebs cycle intermediates, including mono-, di- and tri-carboxylates (Markovich & Murer, 2004). Recently more members of the SLC13 family have been identified in insects, nematode worms, plants and bacteria (Pajor, 2006). The cellular distribution and

transport mechanism of these transporters is rather different from that of the vertebrate transporters (Pajor, 2006), meaning that they do not necessarily co-transport sodium and they have been found in other membranes than the plasma membrane.

Even though expression of *AttDT* seems to be highest in source tissues like young leaves, with roots showing less expression (Emmerlich *et al.*, 2003), the fact that the *Arabidopsis* tonoplast dicarboxylate transporter facilitates transport of malate, and that this transport is inhibited by carbonylcyanide *m*-chlorophenylhydrazone (CCCP) (Emmerlich *et al.*, 2003) makes it a possible homologue for the SM malate transporter.

The aims of this chapter were to i) identify soybean expressed sequence tags (ESTs) similar to the *A. thaliana* tonoplast dicarboxylate transporter, ii) isolate full-length cDNA clones of identified ESTs, and iii) functionally characterise the genes by complementation studies in yeast and bacteria malate transporter mutants.

Materials and methods

General reagents

All reagents used were molecular biology grade, analytical grade or equivalent from commercial suppliers listed in Chapter two.

Radionucleotides

Described in Chapter two.

Yeast and bacterial strains

The *Schizosaccharomyces pombe* wild-type strain used was *Sp*.011 (ade6-704, leu1-32, ura4-D18, h^-), kindly provided by Dr. Mary Albury (University of Sussex, Brighton, UK). The *Sch. pombe mae1* mutant (*ade6-704*, *leu-32*, *ura-D18*, h^-, *mae1*Δ*::KanMX6*)

was generated as described in Chapter 2. *E. coli* Epicurian Coli XL10-Gold cells were obtained from Stratagene (La Jolla, USA), while *E. coli* K-12 wild-type and *E. coli* strain CBT315 (F^-, *sdh-2*, *rpsL129*(strR), *dctA5*, *thi-1*) were obtained from the *E. coli* Genetic Stock Centre (http://cgsc.biology.yale.edu).

cDNA libraries

A nodule cDNA library constructed in the pYES3 expression vector was kindly provided by Dr. Brent Kaiser (University of Adelaide, Australia).

Yeast Media

The *Sch. pombe* wild-type strain *Sp.*011 and the *Sch. pombe mae1* mutant were grown in either YES, EMM or MGI medium (see Chapter 2). Cells were agitated during growth in an orbital shaker at 200 rpm at 32 ºC.

Plant material and growth conditions

Plants material and growth conditions were as described in Chapter two.

Plasmids

Expression vector pREP41b was constructed as described in Chapter 2. Expression vector p*Trc*99A was obtained from Pharmacia Biotech (USA). Cloning vector pGEMT was obtained from Promega.

DNA sequencing

DNA sequencing was done as described in Chapter 2.

Construction of cDNA libraries

Poly-A+ RNA was isolated, as described in Chapter 2, from nodules of 4-7 week-old, and leaves of 3 week-old soybean plants. One µg of poly-A+ RNA was used as template for the first strand cDNA synthesis. 5' and 3' RACE cDNA libraries were

made using the BD SMART RACE cDNA amplification kit (BD Biosciences Clontech) according to the manufacturer's instructions.

Rapid amplification of cDNA ends

The Expand High Fidelity System (Roche) was used for all RACE PCRs. Each 50 µl reaction contained 1.5 –2 mM $MgCl_2$, 200 µM dNTPs, 300 nM gene specific primer, 1X Universal Primer A Mix (UPM) or 200 nM Nested Universal Primer (NUP), 0.875 U Expand High Fidelity PCR system enzyme mix and 2.5 µl 5'- or 3'-RACE-Ready cDNA. For sequence information on primers see Table 3.1. The PCR parameters were as follows: 2 min at 94 ºC; 5 cycles of 94 ºC for 30 s, 72 ºC for 2 min; 5 cycles of 94 ºC for 30 s, 70 ºC for 2 min, 72 ºC for 2 min; and 25-30 cycles of 94 ºC for 30 s, 68 ºC for 2 min, 72 ºC for 2 min. A final extension step at 72 ºC for 6 min was included. For nested PCR the program used was: 2 min at 94 ºC; 30-35 cycles of 94 ºC for 30 s, 65-68 ºC for 2 min, 72 ºC for 2 min and a final extension step at 72 ºC for 6 min. For *GmDT2* gene it was necessary to do one more nested PCR. The PCR parameters were: 2 min at 94 ºC; 5 cycles of 94 ºC for 30 s, 56 ºC for 2 min, 72 ºC for 2 min; and 30 cycles of 94 ºC for 30 s, 50 ºC for 2 min, 72 ºC for 2 min. A final extension step at 72 ºC for 6 min was included

Standard and colony PCR

Standard and colony PCRs were done as described in Chapter two.

Gel electrophoresis and purification of PCR products

PCR products were separated by electrophoresis and purified as described in Chapter two.

Table 3.1: Primers used in RACE PCR

Primer name	Nucleotide sequence	Specific to
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGG TATCAACGCAGAGT	BD smart II A oligonucleotide
NUP	AAGCAGTGGTATCAACGCAGAGT	BD smart II A oligonucleotide
T7	TAATACGACTCACTATAGGG	Part of UPM primer
PYESFw*	AACCCCGGATCGGACTACTA	pYES3
PYESRv*	CTTTTCGGTTAGAGCGGATG	pYES3
NaDcR1(r)	TCAGCAACCCAAACACAGAC	*GmDT1*
NaDcF1(r)	CGCGACAATATTCTTCGTGA	*GmDT1*
RacemixR1	ATCTTCCAAGAAATCCAAGGCTCTTGAC	*GmDT4*
RacemixR2	CATCAGCTATGGCAAAACCAGCTCCC	*GmDT4*
RaceseedR1	AAGAATATCGTCGCGCACAGCCCCAG	*GmDT3*
RaceseedF1	CGGTTAATCCAGCGCTGTTACTGCTG	*GmDT3*
RaceSeedF2	GGTCACGCAGGCCGTGCCAC	*GmDT3*
RaceSeedF3	TGCACGAGCAGTCGGAGGCGG	*GmDT3*
RaceNodR1	TTCCATCACCAACAAGACCATGGAAGC	*GmDT2*
RaceNodR2	CAAGGTAACTTCTTGCGTTCATTCCAATC	*GmDT2*
RaceNodF1	CAAGAAACAAGAGGGGGAGAAGCTAATGG	*GmDT2*
RaceNodF2	TTATTTTGCTTCTTGGAGCAGGGTTTGCT	*GmDT2*
NaDcF4_R	ACATTCCCGGATGGGGAGCTT	*GmDT2*

* Primers designed by Dr. Brent Kaiser (University of Adelaide).

Generation of chemically competent cells

E. coli Epicurian *Coli* XL10-Gold cells were made competent for DNA uptake and transformation using the method of Inoue *et al.* (1990). *E. coli* K-12 wild-type and *E. coli* strain CBT315 were made competent by means of $CaCl_2$, using a protocol modified from the method of (Dagert & Ehrlich, 1979): A bacterial culture was grown to log phase in 50 ml LB broth and chilled on ice for 10 min. Cells were collected at 3200x*g*

for 10 min at 4 ºC and resuspended in ice-cold $CaCl_2$ solution (60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES, pH 7.0), collected as above, and resuspended in 4 ml of ice-cold $CaCl_2$ solution. Aliquots of competent cells were snap frozen in liquid nitrogen and stored at -80 ºC.

Ligation, bacterial transformation

Ligation and transformation was done as described in Chapter two. However when *E. coli* strain CBT315 or the wild-type strain K12 were transformed with *GmDT1*::p*Trc*99A, *GmDT2*::p*Trc*99A, or empty p*Trc*99A vector, the transformed cell were plated on M9 medium (Sambrook *et al.*, 1989) containing either 2% glucose or 10 mM malate as the sole carbon source, 25 µg ml^{-1} thiamine and100 µg ml^{-1}Ampicillin.

Plasmid isolation

Plasmid isolation was done as described in Chapter 2.

Cloning and subcloning

The full-length cDNAs for *GmDT1*, *GmDT2* and *GmDT3* were amplified using primers (Table 3.2) containing restriction sites in frame with either the ATG or the stop codon of the respective genes. The PCR products were purified and ligated into pGEMT. The full-length sequence of each *GmDT* was verified by sequencing before the cDNA was sub-cloned into the *Sch. pombe* pREP41b expression vector and the *E. coli* p*Trc*99A expression vector by means of the introduced restriction sites (*Bam*HI/*Xho*I for pREP41b and *Bam*HI/*Xba*I for p*Trc*99A). Once in the expression vectors, the sequences of the *GmDT* inserts were verified.

Yeast transformation

A simplified transformation method using lithium acetate described by Elble (1992) was used to transform *Sch. pombe* wild-type cells and the *mae1* deletion mutant cells.

Transformed yeast cells were plated on EMM (-Leu) or MGI (-Leu) media and incubated for 5-10 days at 30 ºC. The cells were then stored at 4 ºC.

Table 3.2: Primers designed for cloning.

The sequence underlined represents the restriction enzyme introduced for sub-cloning. (F) and (R) in the primer name designates if the primer is a forward or reverse primer respectively.

Primer name	Nucleotide sequence	Specific to
GmDT1-BamHI (F)	CTAGAGGATCCATGGTCGGAGAACACAGTAC	*GmDT1*
GmDT1-XbaI (R)	CTAGATCTAGATTATTGAGTATCATTGTTTGTTCC	*GmDT1*
8A-XhoI (R)	CTAGACTCGAGTTATTGAGTATCATTGTTTGTTCC	*GmDT1*
GmDT2-BamHI (F)	CTAGAGGATCCATGGTTCTTCTTCGGTTTTCC	*GmDT2*
GmDT2-XbaI (R)	CTAGATCTAGATCAATTTCTCAACCAAGG	*GmDT2*
GmDT2-XhoI (R)	CTAGACTCGAGTCAATTTCTCAACCAAGG	*GmDT2*
GmDT3-BamHI (F)	CTAGAGGATCCATGGTCGGAGAACACG	*GmDT3*
GmDT3-XbaI (R)	CTAGACTCAGACTTATTGAGTACCAGTGCCC	*GmDT3*
GmDT3-XhoI (R)	CTAGACTCGAGCTTATTGAGTACCAGTGCCC	*GmDT3*

Reverse transcription (RT)-PCR

Total RNA was extracted from nodules, leaves, and roots of 3 week-old soybean plants as described in Chapter two.

Yeast cells grown to log phase were used for RNA extractions with the Qiagen RNeasy kit. Log phase cells were pelleted by centrifugation at 3220x*g* for 5 min and the cell wall (approximately 10^8 cells) digested for 30 min at 30 ºC with 150 U zymolase from *Arthrobacter luteus* (Scikagaku Corporation, Japan) in 1M sorbitol, 0.1 M EDTA pH 8, 0.1% (v/v) ß-mercaptoethanol. After digestion, the Qiagen RLT buffer was added to the cells together with 250 µl acid-washed glass beads (BDH) and the suspension was vortexed. The manufacturers standard protocol was then followed. If additional DNase

treatment was necessary, DNaseI (Ambion, Austin, USA) was used as per the manufacturer's instructions.

RT-PCR was performed with: 1 µg of DNase-treated total RNA. The RNA was reverse transcribed for 1h at 42 ºC in Expand reverse transcriptase buffer (Roche) containing 1 mM dNTPs (Promega), 10 µM oligo $(dT)_{15}$ (Sigma), 10 mM DTT, 20 U RNase Inhibitor (Roche) and 20 U Expand reverse transcriptase (Roche), in a total of 40 µl for yeast samples and 20 µl for plant samples. Negative control reactions were carried out under identical conditions, in the absence of either reverse transcriptase or RNA in separate reactions. For PCR amplification, 2 µl of the first-strand cDNA synthesis product and 150 nM of the forward and reverse PCR primers (Table 3) were incubated in the presence of 50 mM KCl, 10 mM Tris-HCl, pH 9, 0.1% (v/v) Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.625 U *Taq* polymerase (Promega). For the yeast samples, PCR cycling was as follows: 2 min at 94 ºC; 5 cycles at 94 ºC for 15 s, 53 ºC for 30 s, 72 ºC for 1 min; 30 cycles at 94 ºC for 15 s, 58 ºC for 30 s, 72 ºC for 1 min; and a final step at 72 °C for 6 min. For the plant samples, the cycling was: 2 min at 94 ºC; 30 cycles 94 ºC for 15 s, 55 ºC for 30 s, 72 ºC for 1 min; and a final step at 72 ºC for 6 min. Products were separated on 1% (w/v) agarose gels in TAE buffer containing 1 µg ml^{-1} ethidium bromide.

Table 3.3: Primers used in RT-PCR.

Primer name	Nucleotide sequence	Specific to
NaDCR1(r)	TCAGCAACCCAAACACAGAC	*GmDT1*
NaDCF1(r)	CGCGACAATATTCTTCGTGA	*GmDT1*
GmDT2R2	TAGCTCCTAGAGTTGGCATGA	*GmDT2*
GmDT2F1	GGTGCTTCGTCTGTTTCCTC	*GmDT2*
NaDCR1(s)	AGACCACCGTCAGAATCACC	*GmDT3*
NaDCF1(s)	CCACAAGCCAGAAAATGCTT	*GmDT3*
KanmaeF1	GATTGTATGGGAAGCCCGATGCG	*kan*
KanmaeR2	GATGGCGGCGTTAGTATCGA	*kan*

Southern blot analysis

Probe synthesis

A digoxigenin (DIG)-labelled DNA probe spanning the entire sequence of *GmDT2* was PCR amplified from plasmid containing the *GmDT2* cDNA using the PCR DIG Probe Synthesis kit (Roche) as per the manufacturer's instructions. The labelling efficiency of the PCR-generated probe was estimated by gel electrophoresis. In order to do so a PCR reaction that produced an unlabelled version of the probe was performed under similar conditions and the PCR products were separated on a 1% (w/v) agarose gel in TAE buffer containing 1μg ml^{-1} ethidium bromide. The probe was considered efficiently labelled if the following conditions were met: 1) DIG-labelled probe migrated slower in the gel than the unlabelled probe, 2) unlabelled probe ran to predicted size, 3) intensity of the DIG-labelled probe was equal to or slightly less than the intensity of the unlabelled DNA.

Probe specificity and dot blot hybridisation

To test for probe specificity 10 ng, 1 ng, 0.1 ng, 10 pg, 1 pg and 0.1 pg of unlabelled *GmDT1, GmDT2* and *GmDT3* PCR products were dot blotted onto a positively charged

nylon membrane (Roche). A 2 µl aliqout of the PCR product was heated to 99 ºC for 2 min and then placed on ice before 1 volume chilled 20x SSC (1x SSC: 0.15M NaCl, 15 mM tri-sodium citrate, pH 7) was added. The DNA solution was applied to the membrane in 2 µl aliqouts and air-dried. The DNA was then denatured for 5 min on filter paper saturated with denaturing solution (0.5 M NaOH, 1.5 M NaCl), before neutralisation for 1 min on filter paper saturated with neutralisation solution (0.5 M Tris-HCl, 1.5 M NaCl, pH 7.5). Then the DNA was cross-linked to the membrane by UV light (2 x 120 Joules with a UVItec cross-linker, Cambrigde, UK). Hybridisation of the DIG-labelled *GmDT2* probe to the target DNA on the membrane was done overnight at 42 ºC, followed by two washes, each for 5 min, at room temperature in 2x SSC/0.1% SDS and then two washes, each for 15 min, at 45 ºC in 0.5x SSC/0.1 % (w/v) SDS.

Southern hybridisation

Genomic DNA for Southern blot analysis was isolated from leaves of 2 week-old soybean plants according to a CTAB protocol (Ausubel *et al.*, 1994). Integrity of the genomic DNA was tested on 1 % agarose gel in TAE buffer containing 1 µg g ml^{-1} ethidium bromide.

Genomic DNA was digested with *Eco*RI, *Eco*RV, *Spe*I and *Xba*I, which do not have recognition sites within the ORF of *GmDT1*, *GmDT2*, *GmDT3* or the partial ORF of *GmDT4*. Digestions were done overnight at 37 ºC. Each digestion was done in a total volume of 10 µl containing 10 µg of genomic DNA, 10 U of restriction enzyme and 1 volume of 10X digestion buffer (supplied by manufacturer with each enzyme). The next day an additional 10 U of restriction enzymes, 1 volume of 10X digestion buffer and water to a total volume of 20 µl was added and digestion continued overnight at 37 ºC. The digested DNA was subsequently precipitated as described in Chapter 2 (DNA

sequencing) and resuspended in sterile water. Approximately 40 µg of fully digested genomic DNA was separated on a 0.8% (w/v) agarose gel in TAE buffer and transferred onto a positively charged nylon membrane by a capillary transfer method as described in the DIG Application Manual for filter hybridisation (Roche). The DNA was cross-linked to the hybridisation membrane as described for the dot blot above. Hybridisation of the DIG-labelled *GmDT2* probe to the target DNA on the membrane occurred overnight at 42 ºC, followed by 2 washes, each for 5 min, at room temperature in 2x SSC/0.1% SDS and then 2 washes, each for 15 min, at 45 ºC in 0.5x SSC/0.1 % (w/v) SDS for high stringency washes or 2 washes, each of 5 min, at room temperature in 0.2x SSC/0.1% SDS for low stringency washes. Labelled DNA fragments were visualised using an anti-DIG antiserum conjugated to alkaline phosphatase and the CDP-*Star* chemiluminescent substrate (Roche) as per the manufacturer's instructions.

Uptake studies

Uptake studies were performed as described in Chapter two, except that yeast were incubated with ^{14}C-malate for 5 min only.

Amino-acid sequence analysis

Predictions of transmembrane helices in deduced amino acid sequences were done using the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/). The method for prediction of version 1 is described in Sonnhammer *et al*., (1998). During an evaluation of methods for predicting transmembrane helices by (Moller *et al.*, 2001), TMHMM was found to be the best performing algorithm

Signal sequence predictions and targeting prediction were done using the SignalP 3.0 and the TargetP 1.1 algorithms, respectively (http://www.cbs.dtu.dk/services/) (Bendtsen *et al.*, 2004; Emanuelsson *et al.*, 2000), and pair-wise alignment of amino acid sequences the software program BestFit (http://www.angis.org.au).

Results

Identification of soybean ESTs homologous to AttDT

A tblastn search compares a protein query sequence against a nucleotide sequence database translated in all reading frames. Eight soybean expressed sequence tags (ESTs), homologous to *Att*DT (GenBank accession no: CAC19845), were identified by a tblastn search of the Soybean TIGR gene index (http://www.tigr.org). Four ESTs (Gene Bank accession numbers AI442425, AI973637, BU764236, and contig number TC183135) had a significant similarity with a probability ≤ 1.3e-42 and were chosen for further investigations. Data from the Soybean TIGR gene index concerning these four ESTs is presented in Table 3.4.

Table 3.4: ESTs with similarity to *Att*DT.

Designated name (this study)	TC #	EST clones which comprise consensus	Source of ESTs	Age of tissue used in library
GmDT1	TC211158	AI442425	Entire root from seedling	8 days
		CF805923	*Phytophora sojae*-infected hypocotyl	48 hr. post infection
GmDT2	TC233074	AI973637	Root nodules	2.5 month old plants
		AI442898	Entire root from seedling	8 days
GmDT3	-	BU764236	Seed coats	
GmDT4	TC210329	BE397935	Whole seedling without cotyledons	2 weeks
		BQ630204	Hypocotyl from etiolated seedling	9-10 days
		AW348139	Entire root from seedling	8 days
		BM085683	Leaf and shoot tip from salt stressed seedlings	2 weeks

Primer sets (Table 3.5) specific for each EST were designed to amplify a fragment of each EST from nodule cDNA and for subsequent use in RACE PCR. The four ESTs were designated *GmDT1*, *GmDT2*, *GmDT3* and *GmDT4*, respectively to the Gene Bank

accession numbers and the contig number given above. Each EST showed more than 57% identity to the *At*tDT amino acid sequence spanning 136-186 amino-acid residues. PCR products obtained from nodule cDNA using the gene specific primers (Table 3.5) were of the expected sizes, 575 bp, 191 bp, 458 bp and 484 bp respectively for *GmDT1*, *GmDT2*, *GmDT3* and *GmDT4*. The PCR product of *GmDT3* was however very faint (Figure 3.1). The sequences of the PCR products were determined directly and found to be identical to the sequences of *GmDT1*, *GmDT2* and *GmDT4* (data not shown). The amplicon generated with *GmDT3*-specific primers was not sequenced at this stage due to the very low yield.

Table 3.5: Primers used to test expression of *GmDT*s in nodules tissue.

Primer name	Sequence	Designed towards
NaDcF1(r)	CGCGACAATATTCTTCGTGA	*GmDT1*
NaDcF1(n)	ATGCCACTGCCACTGCCACA	*GmDT2*
NaDcF1(s)	CCACAAGCCAGAAAATGCTT	*GmDT3*
NaDcF1(m)	TTGCTTTCTGGCTTCCAACT	*GmDT4*
NaDcR1(r)	TCAGCAACCCAAACACAGAC	*GmDT1*
NaDcR1(n)	CGCCAATTTTGAGCATGTCTAGG	*GmDT2*
NaDcR1(s)	AGACCACCGTCAGAATCACC	*GmDT3*
NaDcR2(m)	GCAAGAAAGAAAGAGATGGAGA	*GmDT4*

In the tblastn search, the partial *GmDT1* and *GmDT3* translated sequences overlapped each other and aligned with *At*tDT towards its the amino terminal part while the partial *GmDT2* and *GmDT4* translated sequences overlapped each other and aligned with *At*tDT near the very end of the carboxy-terminal. In order to verify that *GmDT1* or *GmDT3* was not part of *GmDT2* or *GmDT4* and in fact were from distinct ESTs, PCRs with the forward gene-specific primers of *GmDT1* and *GmDT3* were combined with

reverse gene-specific primers of *GmDT2* and *GmDT4*. These PCRs did not generate amplicons, suggesting that the four *GmDT*s are in fact distinct (data not shown).

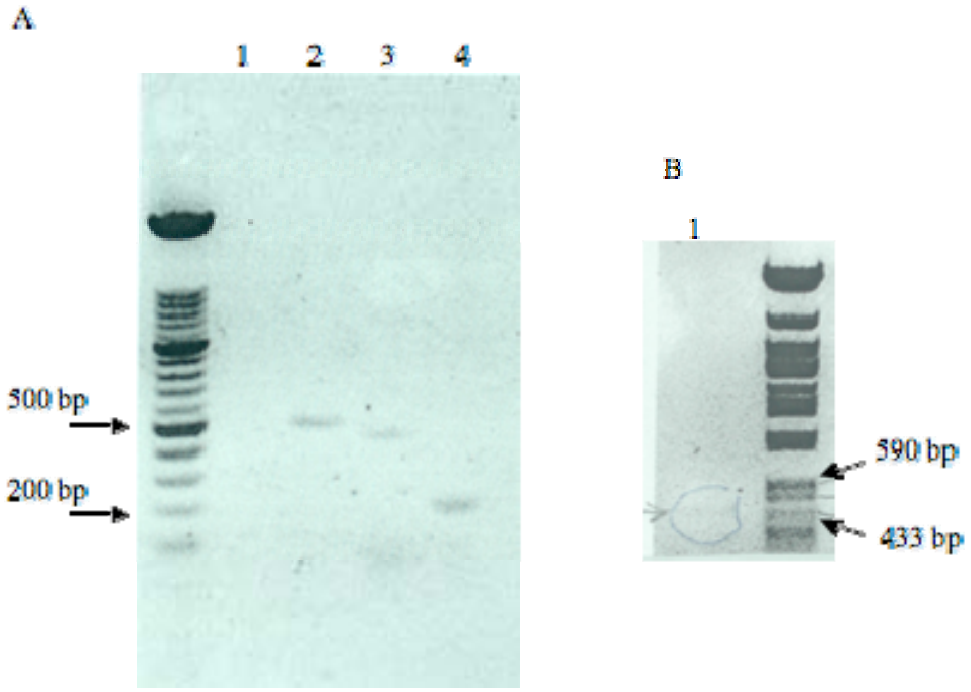


Figure 3.1: Amplification of *GmDT* ESTs from nodule cDNA.

PCR products amplified from cDNA generated from 1 µg of mRNA isolated from mature soybean nodules. PCR products were separated on a 1% agarose gel in TAE buffer containing 1 µg ml^{-1} ethidium bromide. A) PCR products produced using gene specific primers for *GmDT1* (lane 2), *GmDT4* (lane3) and *GmDT2* (lane 4). No product was visual for the PCR with gene specific primers of *GmDT3* (lane 1). One tenth of the PCR was separated on the gel. A 100 bp ladder ran on the left side, fragment size 500 bp and 200 bp is indicated to the left. B) When 40% of the PCR product generated with *GmDT3* gene specific primers was separated on a similar gel, a faint amplicon was just visible. Marker on the right side is λDNA digested with *Ava*II, the size of two fragments is indicated to the right.

Expression profiles of GmDT1 to GmDT4 using qualitative RT-PCR

Expression profiles of the four *GmDT*s were studied in different soybean tissues by RT-PCR (Figure 3.2). The expected size products were 472 bp, 194 bp, 458 bp and 433 bp for *GmDT1*, *GmDT2*, *GmDT3* and *GmDT4*, respectively. *GmDT* mRNA was detected at

some level in all the tissues examined; *GmDT1* showed highest expression in root tissue followed by nodule and then leaf tissue. *GmDT2* and *GmDT4* showed highest expression in nodules and similar low expression in leaf and root tissue, while *GmDT3* showed low expression in all tissues examined. However it should be emphasised that RT-PCR is not really quantitative and the expression levels shown in Figure 3.2 are relative.

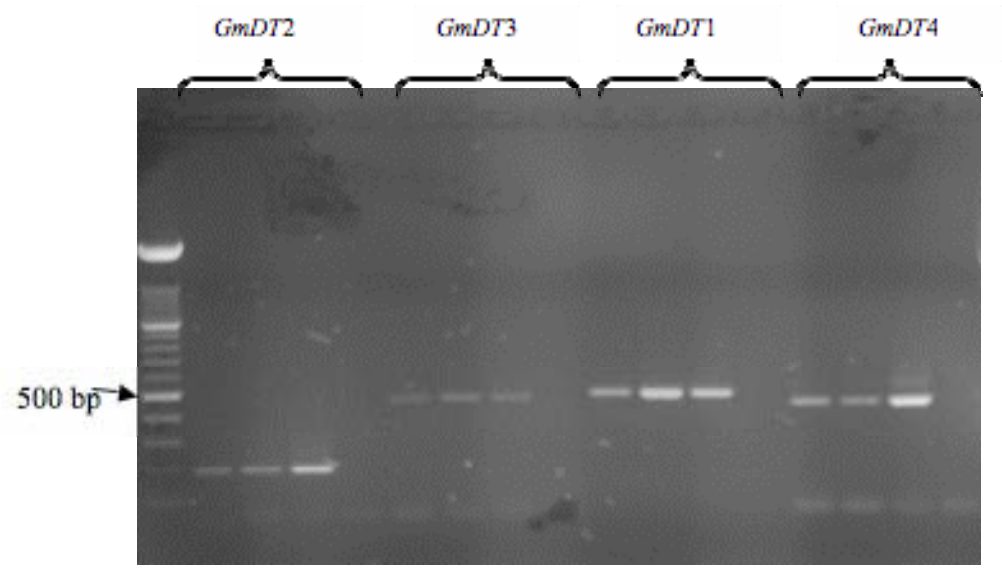


Figure 3.2: Relative expression profiles of *GmDT1, GmDT2, GmDT3* and *GmDT4*.

RT-PCR products obtained from using specific primers for *GmDT1* to *4* and leaf, root and nodule cDNAs. The PCR products were separated on a 2% agarose gel in TAE buffer containing 1 µg ml^{-1} ethidium bromide. Template in the different lanes are: lanes 1, 5, 9 and 13 leaf cDNA; lanes 2, 6, 10 and 14 root cDNA; lanes 3, 7, 11 and 15 nodule cDNA; lanes 4, 8, 12 and 16 no template negative control. A 100 bp DNA marker was separated on the left; the thick 500 bp band is indicated on the left.

Isolation of full-length cDNAs encoding GmDTs

Contiguous sequences of *GmDT1*, *GmDT2*, *GmDT3* and *GmDT4* were constructed from consensus sequences of fragments amplified from three soybean libraries; a nodule cDNA library constructed in the pYES3 expression vector, a leaf BD SMART RACE library and a nodule BD SMART RACE library.

Initially 5' and 3' RACE products were amplified from the nodule cDNA library constructed in the pYES3 vector using vector specific primers and gene specific primers (Table 3.1). Each RACE reaction resulted in several PCR products (data not shown). Each PCR product was extracted from the gel and used as the template in a subsequent PCR with gene specific primers. The RACE PCR fragments for which a PCR fragment of expected size was obtained were then cloned into pGEMT and the different RACE PCR fragments sequenced. The full-length cDNA sequence of *GmDT1*, the 3' end of *GmDT4* and an extension of the 5' sequence from *GmDT2 and GmDT4* were obtained. After unsuccessful attempts to obtain full-length cDNAs of *GmDT2*, *GmDT3* and *GmDT4* from the nodule cDNA library constructed in the pYES3 expression vector, 5' and 3' BD SMART RACE cDNA libraries were synthesised to provide full-length *GmDT* cDNAs. The advantage of this kit compared to other cDNA synthesis kits it that the BD Powerscript reverse transcriptase switches template from the mRNA molecule to the BD SMART oligo, generating a complete cDNA copy of the original RNA with the additional BD SMART sequence at the end. Gene specific primers specifically for use in SMART RACE PCRs were designed (Table 3.1). RACE PCRs were set up using the 5' and 3' RACE-ready cDNA constructed from leaf and nodule mRNA using the universal primers UMP and NUP (for nested PCR) together with the gene specific primers. No RACE PCR products were obtained in the initial PCR with the gene specific primers, so a nested approach was taken. This resulted in amplification of the full-length *GmDT2* and *GmDT3* cDNAs from the nodule and leaf cDNA library, respectively. The full-length sequence of *GmDT4* was not obtained. A PCR product of expected size was obtained from the 5' nested RACE PCR for *GmDT4*, but this product turned out to be non-specific, and time constraints did not allow further investigation.

Nucleotide Sequence analysis and analysis of predicted proteins of the GmDTs

GmDT1

Analysis of the *GmDT1* nucleotide sequence (Appendix 5) identified an ORF of 536 amino acids encoding a protein with a calculated molecular mass of 58,373 kDa. Only a very short (10 bp) 5' untranslated region (UTR) was found and it was not possible to identify stop codons upstream and in-frame with the proposed initiation codon. Analysis of the 3' UTR identified one possible poly(A) addition signal, AATGAA, (Wu *et al.*, 1995) positioned 88-93 nucleotides downstream from the stop codon. The *Gm*DT1 protein is predicted to contain 13 membrane-spanning domains, with the N- and the C-termini predicted to be extracellular and intracellular, respectively (Figure 3.4). No intracellular targeting signal sequence was identified upon analysis of the *Gm*DT1 amino-acid sequence with the SignalP 3.0 program (Bendtsen *et al.*, 2004).

GmDT2

Upon translation of the *GmDT2* nucleotide sequence (Appendix 6) an ORF of 252 amino acids was identified. Stop codons upstream and in-frame with the initiation codon were identified, thereby confirming that the entire *GmDT2* ORF was present. The stop codons upstream and in-frame with the initiation codon were also identified in cDNA from root and leave tissue (data not shown). A poly-A tail was found approximately 125 nucleotides downstream from the stop codon. However, no poly(A) addition signals were identified. The GmDT2 protein is predicted to contain 6 transmembrane spanning regions (Figure 3.4), with the N- and C-termini intracellular, and has a calculated molecular mass of 27,214 kDa. Using the TargetP program (Emanuelsson *et al.*, 2000), *Gm*DT2 was predicted to be targeted to the secretory pathway, i.e. the sequence contains a signal peptide, with an output score of 0.927, and a reliability class (RC) of 2, indicating a strong prediction (Emanuelsson *et al.*, 2000). Signal peptide prediction using the SignalP 3.0 program (Bendtsen *et al.*, 2004)

identified a signal peptide of 17 amino acids in the *Gm*DT2 protein, and the most likely cleavage site was predicted between Gly_{17} and Cys_{18}.

GmDT3

Analysis of the *GmDT3* nucleotide sequence (Appendix 7) identified an ORF of 536 amino acids with a calculated molecular mass of 57.149 kDa. No upstream stop codons in-frame with the proposed initiation codon was found. Two Poly(A) addition signals, AATGAA, (Wu *et al.*, 1995) were identified in the 3' UTR, 88-93 and 256-261 nucleotides downstream from the stop codon and a Poly A tail was identified 20 nucleotides down stream from the latter poly(A) addition signal. The putative *Gm*DT3 protein is predicted to contain 12 transmembrane regions and both termini predicted to be intracellular (Figure 3.4). Upon analysis of the *Gm*DT3 amino-acid sequence no signal sequence was found using the SignalP 3.0 program (Bendtsen *et al.*, 2004).

GmDT4

A partial cDNA sequence (Appendix 8) of 989 bp was obtained for *GmDT4*. The first 661 bp represent a partial ORF of 220 amino acids, encoding the C-terminal part of the ORF. Two poly(A) addition signals, AATGAA, (Wu *et al.*, 1995) were identified in the 3' UTR, 88-93 and 256-261 nucleotides downstream from the stop codon. The poly A-tail is positioned 48 nucleotides downstream from the last poly(A) addition signal.

Similarity between the full-length GmDT1, GmDT2 and GmDT3 amino-acid sequences with AttDT

An alignment of the predicted amino-acid sequences of *Gm*DT1, *Gm*DT2 and *Gm*DT3 with that of *At*tDT showed that 73-75% similarity exists between the soybean and Arabidopsis proteins, and the similarity between the soybean proteins was 88-95%. (Table 3.6). Alignment of the partial *Gm*DT4 predicted protein sequence over the region it shares with *At*tDT showed 70% similarity to the *At*tDT. On the basis of the high

sequence similarity between the *Gm*DTs individually and *Att*DT (Figure 3.4) it was concluded that the *Gm*DTs encode putative dicarboxylate transporters from soybean.

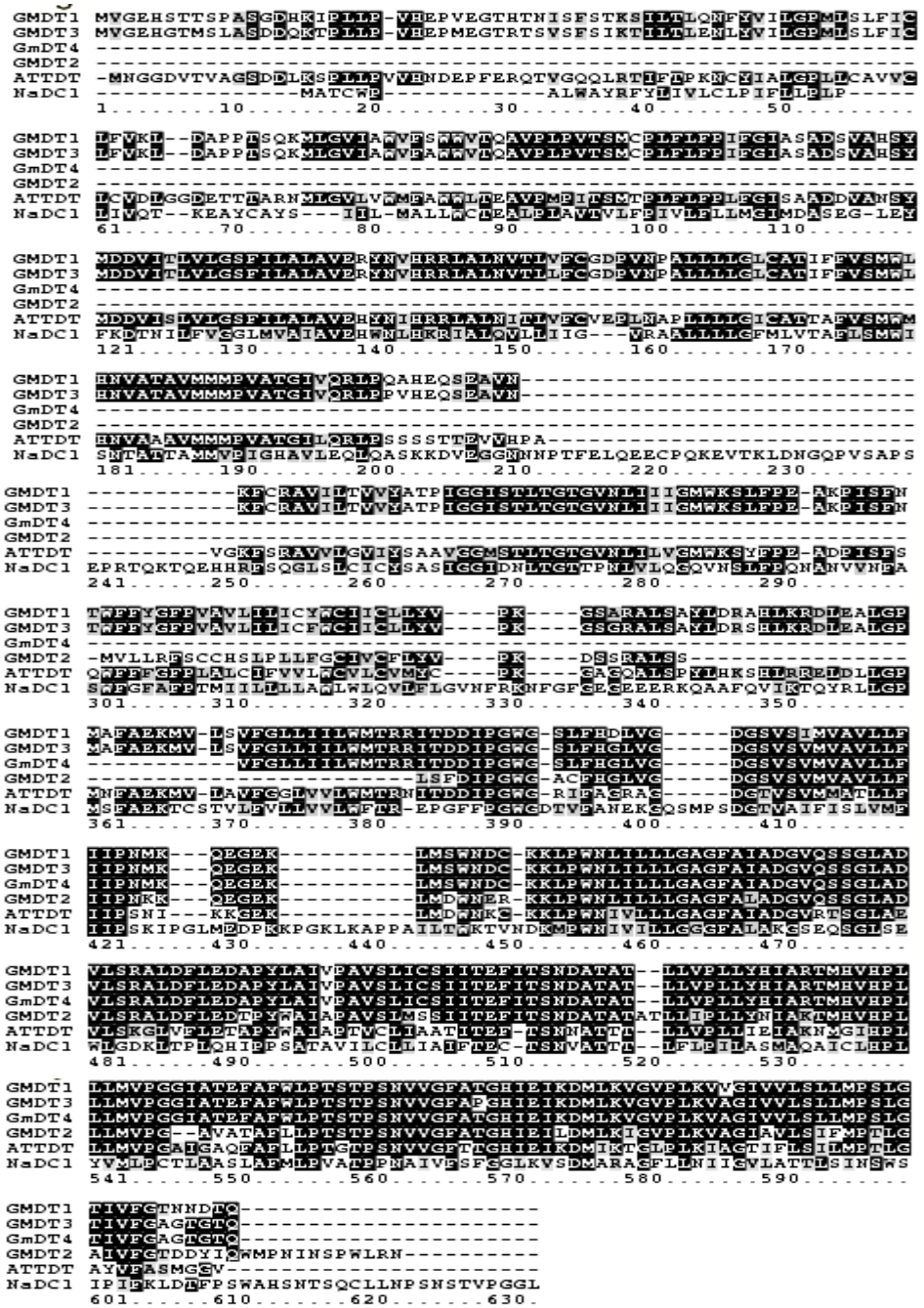


Figure 3.3: Alignment of the predicted amino acid sequences of *Gm*DT1, *Gm*DT2, *Gm*DT3, *Gm*DT4 and *Att*DT and the human NaDC-1.

Identical residues are boxed in black and similar residues are in grey. Hyphens indicate gaps introduced to facilitate alignment. Note that the translated *GmDT4* is partial.

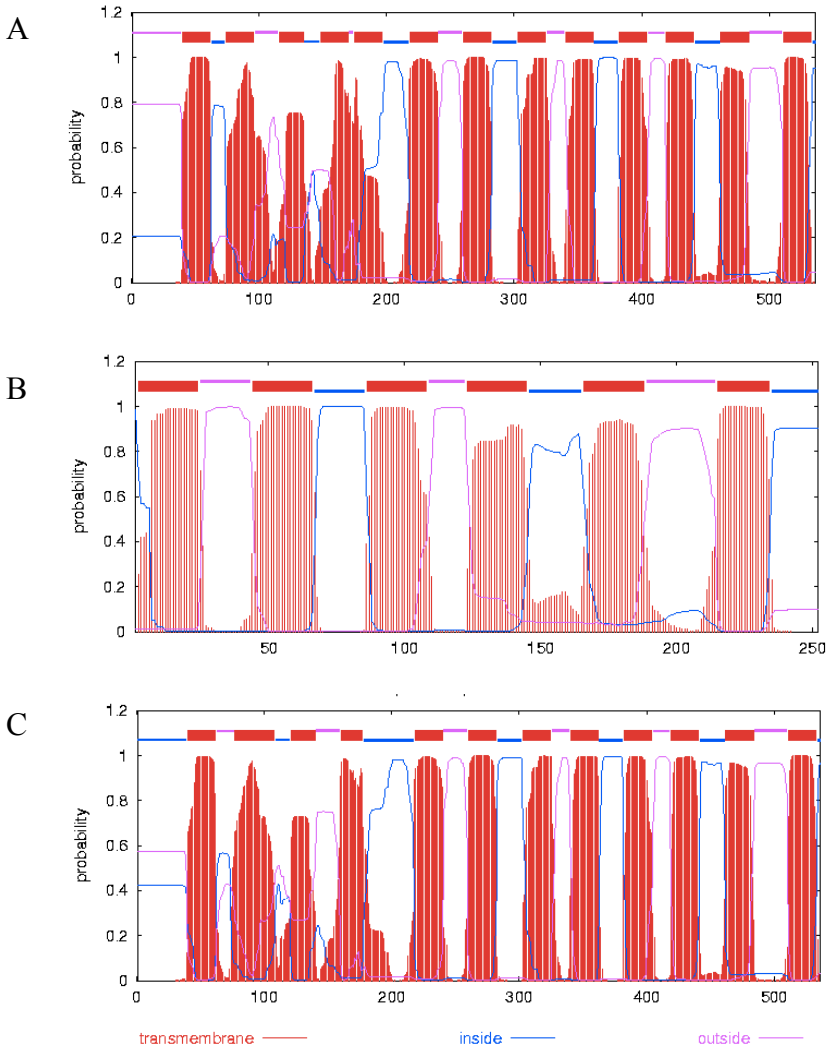


Figure 3.4: Hydrophobicity analysis of *Gm*DT1, *Gm*DT2 and *Gm*DT3.

Hydrophobicity plots obtained in predictions of the transmembrane domains of A) *Gm*DT1, B) *Gm*DT2 and C) *Gm*DT3 using TMHMM. The red curves correspond to predicted transmembrane domains, the blue to intracellular domains and the pink to extracellular domains.

Table 3.6: Similarity matrix for *Gm*DT1, *Gm*DT2, *Gm*DT3 and *Att*DT.

A pairwise protein aligment of *Gm*DT1, *Gm*DT2 and *Gm*DT3 and *Att*DT. The figures are given in percent similarity.

	*Att*DT	*Gm*DT1	*Gm*DT2
*Att*DT	-	-	-
*Gm*DT1	73	-	-
*Gm*DT2	75	88	-
*Gm*DT3	75	95	90

Southern analysis of the GmDT *gene family*

Southern blot analysis was used to estimate the number of genes encoding *Gm*DTs in soybean (Figure 3.5). A Dot blot analysis under stringent conditions confirmed that the *GmDT2* DIG-labelled probe cross-hybridised to *GmDT1* and *GmDT3* as well as *GmDT2* (Figure 3.5 C). Four restriction endonucleases that do not cut within the ORF of *GmDT1*, *GmDT2*, *GmDT3* and partial ORF of *GmDT4* were used. Under low stringency wash conditions the *Eco*RI and *Eco*RV contained five and four fragments, respectively, that hybridised to the *GmDT2* probe. Under the same conditions digestion with *Spe*I and *Xba*I produced two fragments that hybridised to the *GmDT2* probe (Figure 3.5 A and B). In order to confirm that the fragments detected were not due to non-specific hybridisation of the probe to genomic DNA, a second *Eco*RI digest of genomic DNA was hybridised under high stringency conditions. Again under high stringency conditions the *Eco*RI digest contained five fragments that hybridised to the *GmDT2* probe (Figure 3.5 D). This result confirmed that the 5 fragments, detected under low stringency conditions reflected specific labelling of genomic DNA fragments with the *GmDT2* probe. In conclusion the Southern blot analysis showed that a small gene family of not more than five members encodes the *GmDT*s in soybean.

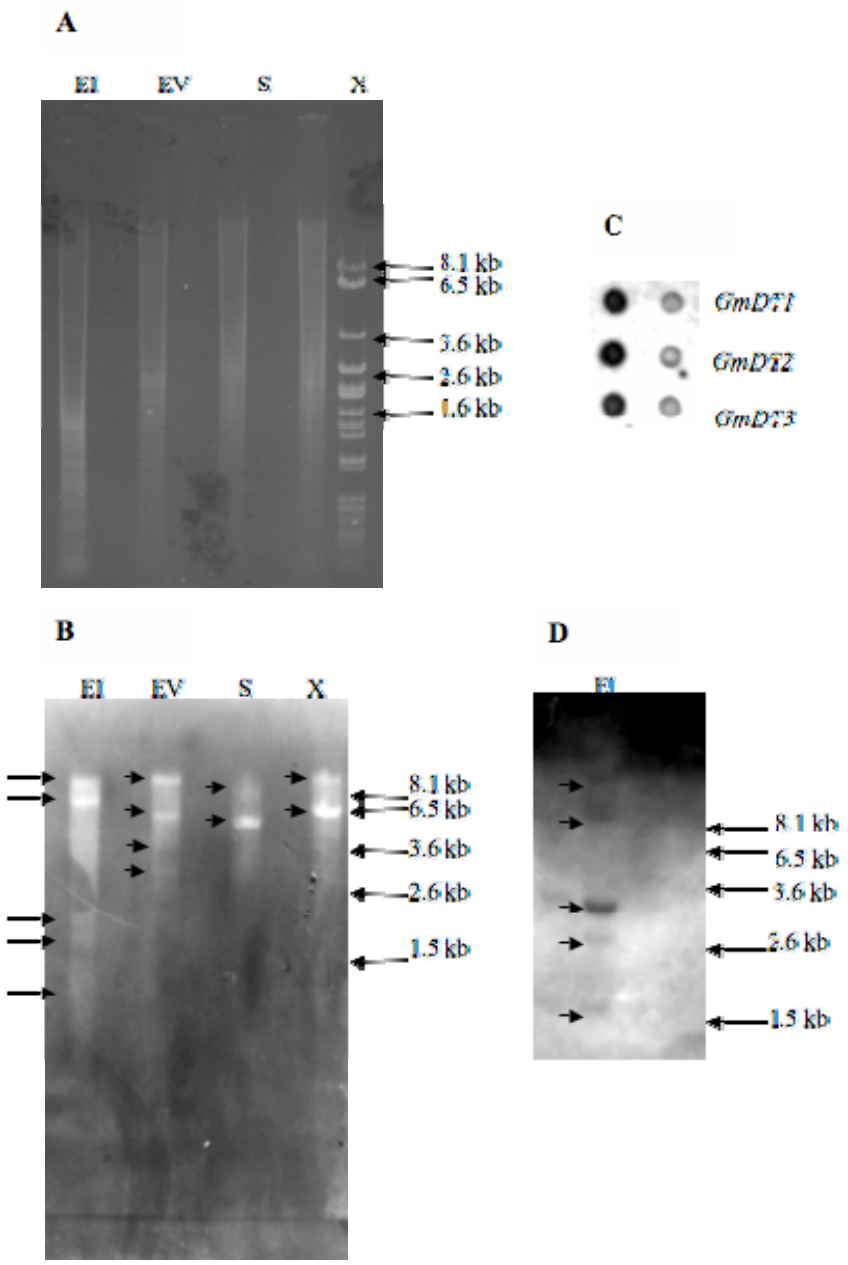


Figure 3.5: Southern blot analysis of the *GmDT* genes.

Soybean genomic DNA (40 µg) was digested with *Eco*RI (EI), *Eco*RV (EV), *Sal*I (S) or *Xba*I (X). (A) Digested genomic DNA was separated by electrophoresis on a 0.8% agarose gel in TAE buffer and stained with ethidium bromide. (B) Digested genomic DNA blotted to a nylon membrane and hybridised under low stringency wash conditions with a DIG-labelled *GmDT2* cDNA fragment corresponding to the ORF of *GmDT2*. The sizes of DNA markers are shown on the right side of each panel. (C) Dot blot of *GmDT1, GmDT2 and GmDT3* PCR products under high stringency wash condition. (D) Southern blot of *Eco*RI digest of genomic DNA under high stringency wash conditions.

Functional complementation of the mae1 *mutant with* GmDT1*,* GmDT2 *and* GmDT3 *cDNAs*

In order to test for functional complementation of the *Sch. pombe mae1* mutant with the full-length soybean cDNAs, the *mae1* mutant was transformed with *GmDT1*::pREP41b, *GmDT2*::pREP41b, *GmDT3*::pREP41b, empty vector pREP41b and mae1::pREP41b (positive control). *Sch. pombe* wild type cells were also transformed with empty vector in order to make it grow on the same plate as the other transformants (Figure 3.6). When grown on MGI (-Leu) medium, the *mae1* mutant transformed with *mae1*::pREP41b was blue in colour similar to when the *Sch. pombe* wild-type cells were grown on the same medium (Figure 3.6, A and B). The blue colour indicates a pH increase in the medium caused by malate uptake and degradation. In contrast when the *mae1* mutant was transformed with either *GmDT1*::pREP41b (Figure 3.6 C), *GmDT2*::pREP41b (Figure 3.6 D), *GmDT3*::pREP41b (Figure 3.6 E) or empty vector, pREP41b (Figure 3.6 F) the the colonies were cream to very light green in colour (same colour as medium). Cream to light green colour of the colonies indicates that the pH was unchanged in the medium, which means that no malate degradation is taking place in these cells. In conclusion, *GmDT1*, *GmDT2* or *GmDT3* do not provide the *mae1* mutant cells with the ability to take up malate.

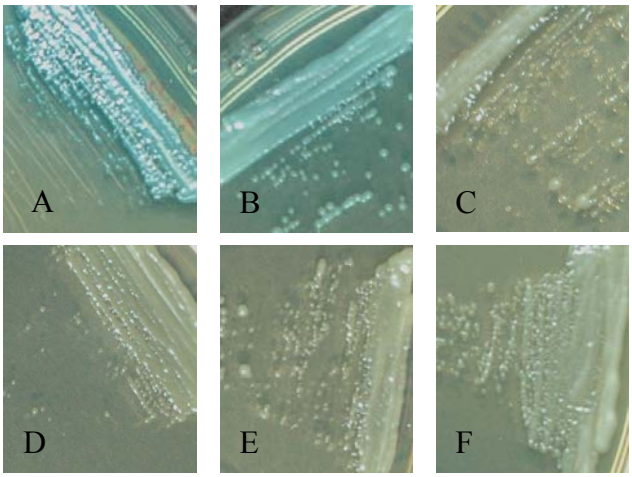


Figure 3.6: Functional analysis of *GmDT1*, *GmDT2* and *GmDT3* cDNA in yeast.

Single colonies of transformed *Sch. pombe* were streaked onto solid MGI (-Leu) medium and incubated for 3 day at 28 ºC. Sections shown are from a single plate. (A) The *mae1* mutant transformed with *mae*1::pREP41b. (B) Wild-type *Sp*.011 transformed with pREP41b. (C), (D), (E) and (F) *mae1* mutant transformed with *GmDT1*::pREP41b, *GmDT2*::pREP41b, *GmDT3*::pREP41b and pREP41b respectively . Blue colour indicates medium pH increase due to malate uptake and subsequent degradation. Cream to light green colour indicates medium pH is unchanged, due to lack of malate degradation.

^{14}C-Malate uptake into yeast cells

^{14}C-malate uptake studies were done in order to test if the *mae1* mutant transformed with any of the *GmDT*s would regain ability to take up malate (Figure 3.7) In addition these studies would serve as confirmation of the results obtained from the functional complementation studies on the MGI (-Leu) medium (Figure 3.6). Uptake was measured after 5 min incubation with ^{14}C malate.

The uptake of ^{14}C malate by the *mae1* mutant cells transformed with *GmDT1* was only 6% of the wild-type cells. The *mae1* mutant cells transformed with *GmDT2* showed a malate uptake of 3.5% relative to the wild-type cells and *mae1* mutant cells transformed with *GmDT3* had a malate uptake that was 4.8% of that of the wild-type cells. The malate uptake of the *mae1* mutant cells alone and the mutant cells transformed with empty vector was 13% and 18% relative to the wild-type cells, respectively. These

results show that none of the *GmDT*s were able to functionally complement the *mae1* mutant's inability to take up malate.

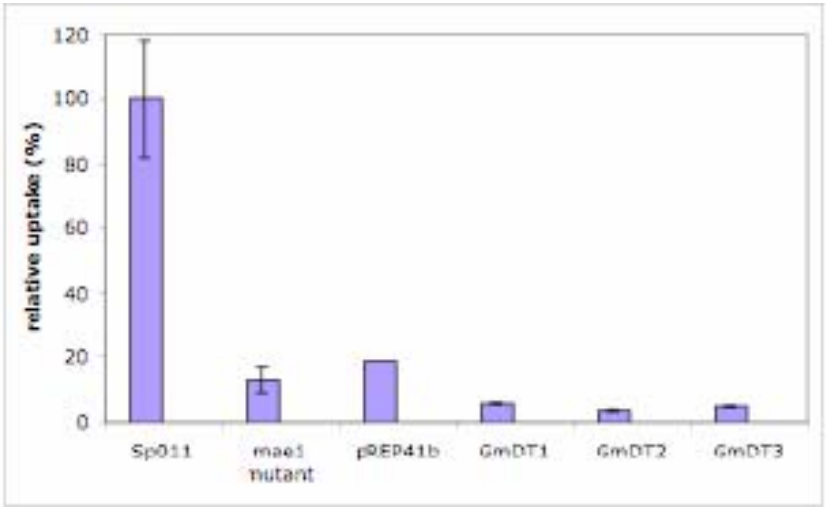


Figure 3.7: Functional analysis of *Gm*DT1, *Gm*DT2 and *Gm*DT3 activity in yeast cells.

^{14}C-malate uptake was measured after 5 min in wild-type strain *Sp*.011, the *mae1* mutant and the *mae1* mutant transformed with empty vector (pREP41b), *GmDT1*::pREP41b (GmDT1), *GmDT2*::pREP41b (GmDT2) or *GmDT3*::pREP41b (GmDT3). Where errors bars are present, data presented are means± SE from n separate experiments otherwise data represent a single experiment with 1 replicate within an experiment. For *Sp*.011 n=5, for *mae1* mutant n=5, for *GmDT1* n=3, for *GmDT2* n=3 and for *GmDT3* n=3.

Expression of GmDTs *in yeast*

The lack of functional complementation in the previous experiment could be due to a failure in the transcription of the *GmDT*s in the heterologous yeast system. To test this, the transformed yeasts were tested for transcription of the *GmDT* genes by RT-PCR experiments (Figure 3.8). Gene specific primers were used (Table 3.3).

RT-PCR from yeast expressing *GmDT1*, *GmDT2* and *GmDT3* produced PCR products of expected sizes 472 bp, 644 bp and 458 bp, respectively to *GmDT1*, *GmDT2* and *GmDT3* (Figure 3.8 A). A partial *kan* gene was amplified from the generated cDNA of each sample and served as a positive control for the RT reaction. Reactions where RT

was omitted failed to produce a product and confirmed that product seen in the presence of RT was due to presence of transcripts and not genomic DNA contamination. These results showed that transcription of *GmDT1*, *GmDT2* and *GmDT3* was occurring in the yeast.

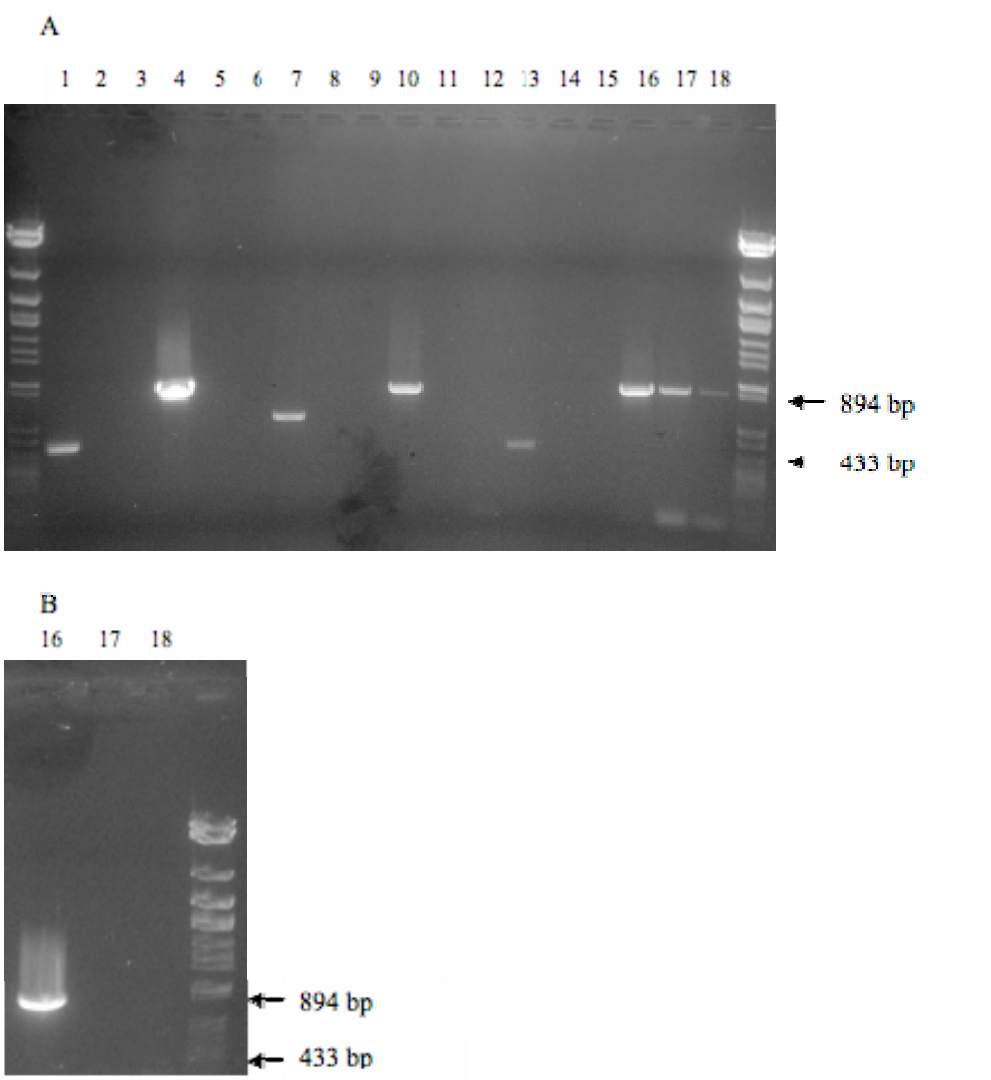


Figure 3.8: RT-PCR showing transcription of GmDT1, GmDT2 and GmDT3 cDNA in the yeast system.

RT-PCR of 1 µg of total RNA extracted from the *mae1* mutant transformed with pREP41b containing *GmDT1*, *GmDT2* or *GmDT3*. PCR products was separated on 1% agarose gel in TAE buffer containing 1µg ml^{-1} ethidium bromide A) Marker: λ-DNA digested with *Ava*II was separated on both sides of the gel. Selected fragment sizes shown on the right. Lanes 1, 7 and 13 show PCR products of 472 bp, 644 bp

and 458 bp from the amplification of *GmDT1*, *GmDT2* and *GmDT3*, respectively (Table 3.3). Lanes 2, 8 and 14 show the corresponding negative control reaction where RT was omitted and lanes 3, 9 and 15 show the corresponding negative control reactions where template was omitted. Lanes 4, 10 and 16 show the primer set KanmaeF1/KanmaeR2 (Table 3.3), which anneals to the *kan* gene in the *mae1* mutant, was used as a positive control and gave a PCR product of 933 bp. Lanes 5, 11 and 17 show the corresponding negative control reaction where RT was omitted lanes and 6, 12 and 18 show the corresponding negative control reactions where template was omitted. B) Marker: λ-DNA digested with *Ava*II was separated on the right. Selected fragment sizes shown on the right. Control reactions shown in (A) lanes 16, 17 and 18 were repeated because of contamination in the negative control reactions. No contamination was then detected.

Functional complementation of E. coli *malate transporter mutant*

The ability of the *GmDT*s to complement the *E. coli* malate transporter mutant CBT315 was tested. *GmDT1*, *GmDT2* and *GmDT3* were cloned (Table 3.2) into the *E. coli* p*Trc*99 expression vector under the control of the *tac* promotor. The *GmDT3* sub-cloning was unsuccessful, due to a mistake in the sequence of the introduced restriction-enzyme site. CBT315 transformed with *GmDT1*::p*Trc*99A, *GmDT1*::p*Trc*99A or p*Trc*99A (negative control), and K12 transformed with p*Trc*99A (positive control) all grew on solid M9 medium containing 1 mM IPTG, 2% glucose, 25 μg ml^{-1} thiamine and 100 μg ml^{-1} ampicillin. However, only the K12 cells transformed with p*Trc*99A grew on M9 medium containing 1 mM IPTG, 25 μg ml^{-1} thiamine, 100 μg ml^{-1} ampicillin and 10 mM malate as sole carbon source (data not shown). These results indicate that *GmDT1* and *GmDT2* did not functionally complement the malate transport inability of the CBT315 cells.

Discussion

In this study, four soybean genes, *GmDT1*, *GmDT2*, *GmDT3* and *GmDT14*, encoding proteins with high similarity to the *Arabidopsis* tonoplast dicarboxylate transporter,

*Att*DT, were identified. Southern blot analysis showed that a small gene family of up to five members encodes the *GmDT* genes in soybean. By contrast, *AttDT* is a single copy gene (Emmerlich *et al.*, 2003). This suggests that the soybean *GmDT* genes may not all encode proteins that are targeted to the vacuolar membrane, as is the target for the *Att*DT protein.

Full-length cDNAs encoding *GmDT1, GmDT2* and *GmDT3* were isolated by RACE techniques. *GmDT1* and *GmDT3* were predicted to encode putative proteins of 537 amino acids in length and containing 13 and 12 transmembrane domains, respectively. No upstream stop codons were found in frame with the proposed first methionine in either *GmDT1* or *GmDT3*. However, due to the very high sequence identity and similar size between these cDNAs and the *AttDT*, it appears unlikely that another methionine further upstream would initiate the synthesis of *Gm*DT1 and *Gm*DT3 proteins. In contrast to *GmDT1* and *GmDT3*, *GmDT2* encodes a putative protein of 252 amino acids and was predicted to contain only 6 membrane-spanning regions. It is unlikely that *GmDT2* is a partial cDNA since stop codons upstream of and in frame with the proposed first methionine was found in cDNAs from nodule, leaf and root tissue. The *Gm*DT2 protein was predicted to be targeted to the secretory pathway with a reliability class (RC) of 2. The RC feature is a useful indication of the level of certainty in the prediction and 93-95% of plant and non-plant sequences with annotated subcellular annotation were correctly predicted (Emanuelsson *et al.*, 2000). When predictions of unannotated data sets were run through the targetP program not all sequences predicted to contain a secretory pathway signal are actually secreted since a subset of them are transmembrane proteins (Emanuelsson *et al.*, 2000). In conclusion, this indicates that *Gm*DT2 has a signal peptide (or anchor protein) that targets the protein to a membrane somewhere in the cell, other than chloroplasts or mitochondria. A signal peptide of 17 amino acids is predicted to occur at the N-terminus in *Gm*DT2. No signal peptides were

predicted for *Gm*DT1 and *Gm*DT3, and these proteins were predicted to be located in any other location than the chloroplasts, mitochondria or the secretory pathway. Like *Gm*DT1 and *Gm*DT3, *At*tDT has not been reported to contain a signal peptide, however it has been localized to the tonoplast (Emmerlich *et al.*, 2003). This demonstrates that targeting prediction programs can yield successful predictions for some proteins and not for others, and that predictions should be taken with caution.

A full length-cDNA for *GmDT4* was not obtained. This was in spite of a PCR fragment of expected size being obtained from a nested 5' RACE PCR for *GmDT4*. Further a PCR with *GmDT4* specific primers using the RACE product gave a product of expected size and the sequence obtained turned out to be unrelated to *GmDT4*. This result can be explained with contamination problems in the latter PCR or that the *GmDT4* specific primers were not specific. Time constraints did not allow further investigation.

On the basis of the high sequence similarity between putative proteins *Gm*DT1, *Gm*DT2, *Gm*DT3 and *At*tDT (73-75%), it is proposed that *GmDT1, GmDT2 and GmDT3* genes encode proteins of similar function to *At*tDT. That is, they transport organic acids. The partial consensus sequence of the *Gm*DT4 protein likewise showed high sequence similarity to *At*tDT (70%), which suggests that *GmDT4* encodes a protein of similar function to *At*tDT.

The NCBI Conserved Domain Database (Marchler-Bauer *et al.*, 2005) was queried for conserved domains in the putative *Gm*DT proteins. The proteins, together with *At*tDT, were found to share similarities with members of the SLC13 gene family. The SLC13 (solute carrier family 13)-type proteins belong to a large superfamily of permeases that are found in all three kingdoms of life (Markovich & Murer, 2004). Common characteristics of the SLC13 genes are that they encode multi-spanning proteins with 8-13 transmembrane domains and have a wide tissue distribution (Markovich & Murer, 2004). They were thought to be DIDS-insensitive sodium-coupled symporters with a

substrate preference for divalent anions such as sulfate and or Krebs cycle intermediates, including mono-, di- and tri-carboxylates. (Markovich & Murer, 2004). Upon identification of the first plant member, *At*tDT, of the SCL13 gene family, it became evident that not all members co-transport sodium as previously reported (Pajor, 2006). The SLC13 gene family now contains members like *At*tDT that, contrary to the mammalian SLC13 transporters that are only found on the plasma membrane, are found on other cell membranes (Pajor, 2006). The *GmDT*s identified in this study are the first legume genes identified within this family.

Functionally important residues of the SLC13 family member NaDC-1 (the renal sodium-dicarboxylate transporter) have been thoroughly investigated in mammals (Pajor, 1999; Pajor, 2000; Pajor, 2001; Pajor & Randolph, 2005; Yao & Pajor, 2002). Some of these amino acids are also conserved in the *Gm*DTs and *At*tDT. The amino acid numbering referred to in the following discussion is for the rabbit NaDC-1. The arginine residue at position 349 (Arg_{349}) is conserved in all members of the SLC gene family, which include both sulphate and dicarboxylate transporters (Pajor *et al.*, 2000). Arg_{349} is also conserved in *Gm*DT1 and *Gm*DT3, but not in *Gm*DT2 or *At*tDT. Arg_{349} has been reported to play a role in substrate recognition and binding however, it is not absolutely required for activity (Yao & Pajor, 2002). Three conserved residues, Ser_{478}, Ala_{481}, and Thr_{482}, have been reported to be conformational sensitive residues (Pajor, 2001). All three residues are conserved in *Gm*DT1, *Gm*DT2 and *Gm*DT3 and *At*tDT. Gln_{475} is likewise conserved in the *Gm*DTs and *At*tDT and has been reported to be involved in determining the affinity for both cations and substrate (Griffith & Pajor, 1999). These findings support the proposed function for the proteins encoded by the *GmDT*s as dicarboxylate transporters. Structure-function studies have identified the C-terminal half SLC13 proteins to be most important for determining function (Pajor *et al.*, 1998); thus although *Gm*DT2 appears to be truncated at its N-terminus compared to

*Gm*DT1 and *Gm*DT3 and other SLC13 family members, it might still function as a dicarboxylate transporter.

Qualitative expression profiles using nodule, leaf and root tissues showed expression of all four *GmDT*s at some level in all the tissues examined. *GmDT1* showed higher expression in all three tissues, indicating that it may encode a protein with a general function such as translocation of malate from shoot to root or pH homeostasis (Martinoia & Rentsch, 1994). *GmDT3* showed lower expression in all three tissues examined. The original *GmDT3* EST originated from seed. Expression of *GmDT3* in seeds was not examined in this study. However, higher expression in seed tissue could suggest a major role for *GmDT3* in malate translocation in seeds. *GmDT2* and *GmDT4* showed highest expression in nodules; however, expression was not specific to nodule tissue. These results exclude these genes from encoding nodulins (nodule-specific proteins); yet they do not exclude the possibility that these genes encode proteins with an important role in the nitrogen fixing process.

Another aim of this part of the project was to functionally characterise the isolated *GmDT*s. Transformation of the *Sch. pombe mae1* malate transporter mutant cells with *GmDT1*, *GmDT2* and *GmDT3*, however, did not increase malate uptake above the levels seen for the untransformed mutant. In order to determine if the lack of complementation of the yeast *mae1* mutant by the *GmDT1* to *3* was due to lack of transcription of the introduced plant genes, an RT-PCR assay was used. It was found that all three *GmDT* genes were transcribed in the heterologous yeast system. These results suggest that post-transcriptional events may be responsible for the lack of functional complementation in the heterologous system. Whether the problem is at the level of translation, protein targeting, or the orientation of the protein within the yeast plasma membrane is unknown. *Sch. pombe* has previously been reported to be an efficient system for functional expression of plant transporters (Taniguchi *et al.*, 2002; Weber *et al.*, 1995).

However, the functional characterisation was carried out with yeast-expressed recombinant proteins reconstituted into liposomes (Taniguchi *et al.*, 2002; Weber *et al.*, 1995). These studies, together with reports that membrane proteins expressed heterologously in the yeast *Saccharomyces cerevisiae*, failed to reach their normal cellular location and instead accumulated in stacked internal membranes or the endoplasmatic reticulum (de Kerchove d'Exaerde *et al.*, 1995; Villalba *et al.*, 1992; Wright *et al.*, 1988; Palmgren & Christensen, 1993; Palmgren & Christensen, 1994) support the view that the problem encountered here is at the post-translational level.

Since functional complementation of *E. coli* transporter mutants by plant genes have been successful (Jeong *et al.*, 2004; Kim *et al.*, 1998), this approach was also taken here. However, transformation of the *E. coli* CBT315 malate transporter mutant with *GmDT1* and *GmDT2* failed to restore malate uptake ability. The lack of growth of CBT315 cells transformed with *GmDT1* and *GmDT2* was not a result of over-expression due to the presence of too much inducer (IPTG), since cell growth on M9 media containing the same amount of IPTG occurred when glucose was the carbon source.

GmDT1, GmDT2 and GmDT3 are now being expressed in *Xenopus* oocytes by our collaborator, Prof. S. D. Tyerman (The University of Adelaide). Expression of plant transporters in *Xenopus* oocytes have proven to be very efficient for studying transport function and the system is well suited for functional measurements including currents, fluxes of electrolytes or non-electrolytes, water permeability and even enzymatic activity (Rentsch, Borer and Frommer, 1998, Zampighi et al, 1995). One of the main advantages of oocytes as a heterologous expression system is that they are self-sufficient in terms of nutrition and therefore provide a system with very low background for foreign transporters expressed in their plasma membrane (Miller & Zhou, 2000). In

addition, electrophysiological measurements of the transport activity of membrane transport proteins are easy to study in oocytes (Miller & Zhou, 2000).

In conclusion, the identification and isolation of the *GmDT*s has brought us a step closer to understanding the mechanisms of malate fluxes in soybean. For the time being the function and physiological role of the different *Gm*DTs can only be hypothesised. *At*tDT has been localised to the tonoplast (Emmerlich *et al.*, 2003) and only recently it has been reported to play a role in pH homeostasis (Hurth *et al.*, 2005). A role in pH homeostasis might, therefore, be expected for at least one of the *Gm*DTs.

Chapter Four: Characterisation of putative amino acid transporters from soybean

Introduction

Nitrogen is an essential macronutrient for plants and often a limiting factor for growth of crops and in most natural environments. Nitrate and ammonium are the most abundant forms of biologically useful inorganic nitrogen (Marschner, 1995). Uptake mechanisms for ammonium and nitrate are therefore vital for the survival of the plant. Subsequent to the uptake of ammonium or nitrate from the soil or acquisition of ammonium from symbiotic nitrogen fixation, the inorganic nitrogen is assimilated into amino acids, a process that occurs either in the root or in the leaf (Marschner, 1995).

Amino acids have major roles in the cell; they are the building blocks of proteins and they are precursors of all other nitrogen-containing compounds, including nucleic acids, growth regulators, photosynthetic pigments, alkaloids, and a wide rage of other essential compounds (Bush, 1999).

Generally amino acids synthesised in the root are transported to mature leaf tissue via the xylem, and amino acids synthesised in the leaf, or arriving from the root in the xylem, are translocated throughout the plant via the phloem (Bush, 1999). Unloading of amino acids from the phloem into heterotrophic sinks such as developing leaves, roots, cortical cells in the stem, seeds and fruits, and apical meristems is a vital process in plant development (Bush, 1999). Amino acid transport is also essential when free amino acids arise from the breakdown of storage proteins in vegetative or reproductive storage organs to supply the developing plant (Fischer *et al.*, 2002). Roots have also been shown to possess uptake systems for amino acids (Frommer *et al.*, 1994b) and amino acids can be taken up directly from the soil. This adaptive strategy is used by lupin for growth in soil low in mineral nitrogen (Hawkins *et al.*, 2005). In many respects, amino acids are the foundation of nitrogen metabolism in the plant and understanding how

amino acids are transported within a plant is a fundamental question in plant biology (Bush, 1999).

Biochemical approaches for identifying and isolating amino acid transport proteins, and transport proteins in general can be problematic (Frommer *et al.*, 1994b). For example one of the major challenges in proteomics is the identification of membrane proteins, and despite considerable progress in two dimensional gel electrophoresis technologies, using Blue Native gel electrophoresis (Wienkoop & Saalbach, 2003) or isoelectric focusing in the first dimension, intrinsic membrane proteins, including amino acid transporters, continue to be difficult to resolve and identify (Ferro *et al.*, 2000). Instead yeast mutants lacking uptake systems for particular amino acids have been used as genetic complementation systems to circumvent biochemical approaches for identifying amino acid transporters (Fischer *et al.*, 1998).

The uptake of amino acids across the plasma membrane in *Saccharomyces cerevisiae* is mediated by 24 different amino acid transporters belonging to the APC superfamily (Wipf *et al.*, 2002). The large number of amino acid transporters in the plasma membrane has made it possible to isolate or create various yeast amino acid transporter mutants. Many Arabidopsis amino acid transporters have been identified using these different yeast mutants (Fischer *et al.*, 2002; Frommer *et al.*, 1994a; Frommer *et al.*, 1993; Frommer & Ninnemann, 1995; Hoyos *et al.*, 2003; Hsu *et al.*, 1993; Rentsch *et al.*, 1995). Amino acid transporters from Arabidopsis are by far the most studied organic nitrogen transporters in plants and more that 50 distinct amino acid transporter genes have been identified in the Arabidopsis genome, indicating that amino acid transport across membranes is a highly complex feature in plants (Wipf *et al.*, 2002). Multiple sets of amino acid transporters are present in plants and by far exceed the number of amino acid transporters in yeast (Wipf *et al.*, 2002). Given that yeast is a

single-cell organism, where long distance translocation is not an issue, this may not be surprising (Su *et al.*, 2004).

Sequence similarity studies have divided the amino acid transporters into two major superfamilies: the amino acid transporter family (ATF) and the amino acid polyamine choline transporter family (APC) (Fischer *et al.*, 1998; Rentsch *et al.*, 1998; Su *et al.*, 2004). The majority of amino acid transporters identified in plants belong to the ATF superfamily (Su *et al.*, 2004). Based on sequence similarity and substrate specificity, the ATF superfamily is further divided into four subfamilies; amino acid permeases (AAPs), proline transporters (ProTs), lysine histidine transporters (LHT-related proteins), and auxin transporters (AUX1-related proteins) (Fischer *et al.*, 1998). The plant APCs can be subdivided into two families, the CATs (cation amino acid transporters) and proteins homologous to the yeast GABA permease-related transporter families (Fischer *et al.*, 1998).

The transport of amino acids between the bacteroids and the infected plant cells of N-fixing legume root nodules is a requirement for nitrogen fixation and the necessary amino acid transporters have been identified in bacteroids (Lodwig *et al.*, 2003). In pea nodules, a model for amino acid cycling between the symbiotic partners has been proposed that involves transport of glutamate, (Glu), or a precursor to it, into bacteroids and aspartate (Asp) or alanine (Ala) transport out of the bacteroids (Lodwig *et al.*, 2003). Aspartate was suggested as the most likely product to be transported out of the bacteroid and across the SM, since Glu stimulated its synthesis in isolated pea symbiosomes (Rosendahl *et al.*, 1992; Lodwig *et al.*, 2003). Although amino acid transport between the plant and the bacteroid in the legume nodule has been shown to be vital for the nitrogen-fixing process, no amino acid transporter has yet been localised to the SM.

In legumes only members of the AAP subfamily of the ATF family have been identified at the molecular level and functionally characterised. Members of the AAP subfamily exhibit a wide specificity for amino acids and are classified either as neutral and acidic amino acid transporters, or as general amino acid transporters (Fischer *et al.*, 1998; Fischer *et al.*, 2002). Using functional complementation of a yeast amino acid transporter mutant, the amino acid transporter from castor bean (*Ricinus communis*), *Rc*AAP1, was the first legume amino acid transporter to be cloned and functionally characterised *in vitro* (Marvier *et al.*, 1998). Five other amino acid transporters have since been identified and functionally characterised *in vitro: Rc*AAP3 from castor bean (Neelam *et al.*, 1999), *Ps*AAP1 from pea (*Pisum sativum*) (Tegeder *et al.*, 2000) and *Vf*AAP1, *Vf*AAP2, *Vf*AAP3 from broad bean (*Vicia faba*) (Miranda *et al.*, 2001; Montamat *et al.*, 1999). Surprisingly, expression of these transporters has not been examined in root nodules.

The NCBI database contains both full-length and partial cDNAs of different *AAP*s from other plant species. Among these is a 1539 bp long soybean gene encoding a putative amino acid transporter (GenBank accession no: AY029352). The function of the encoded protein was presumed based on sequence similarity to other known amino acid transporters. The protein was designated *Gm*AAT-1, however, to emphasise the structural similarity of this sequence to other plant AAPs I will designate it *Gm*AAP1 here.

The aims of the experiments described in this chapter were to: i) identify genes encoding putative amino acid transporters from soybean, ii) use a PCR approach to isolate full-length cDNAs either specific to or with an enhanced expression in the

nodule, and iii) investigate the similarity of identified putative amino acid transporters to known amino acid transporters, mainly from other legumes and Arabidopsis.

Materials and methods

Plant material and growth conditions

Soybean (*Glycine max* (L.) cv. Stevens) plants were grown as described in Chapter 2.

Vectors

The yeast episomal shuttle vector pYES3 (Smith *et al.*, 1995) contains the *AMP* gene for selection in *E. coli* and the *URA3* gene for selection in *ura3* strains, of *S. cerevisiae*. A *GAL1* promoter upstream of the poly-linker controls transcription of DNA inserted into this region. The vector pGEMT was obtained from Promega (Annandale, Australia).

Yeast and bacterial strains

The *E. coli* Epicurian Coli XL10-Gold cells were obtained from Stratagene (La Jolla, USA).

Expression studies in leaf, root and nodule tissue

Complementary DNA was made from total RNA isolated from 3 week-old leaf, root and nodule tissue as described in Chapter 2. Gene specific primers (Table 4.1) were designed and used in PCR assays with *Taq* polymerase (Promega) as described in Chapter 2. The thermocycling program was: 2 min at 94 ºC; 30 cycles of 94 ºC for 15 s, 45 ºC for 30 s, 72 ºC for 1 min; and a final elongation step at 72 ºC for 6 min.

Construction of cDNA libraries

Poly-A+ RNA was isolated as described in Chapter 2 from nodules and leaves of 4-7 week-old and 3 week-old soybean plants, respectively. 5' and 3' BD SMART RACE cDNA (BD Biosciences Clontech) libraries were made as described in Chapter 3.

Table 4.1: Primers used in cloning and expression studies of *GmAAP* genes.

The restriction sites incorporated at the 5' end of the cloning primers are underlined. GGATCC=*Bam*HI, CTCGAG=*Xho*I, GAATTC=*Eco*RI and TCTAGA=*Xba*I.

Primer name	Nucleotide sequence	Specific to
F1	CGGATCCATGTTGCCAAGAAGTAGAAC	*GmAAP1*
R1	CCTCGAGTTAATAACTAGTTTTAAATGGC	*GmAAP1*
F2	CTTCCGCTTAGTATGGAGA	*GmAAP1*
R2	CGGCCTTGATTTAATAAC	*GmAAP1*
F3	AGTGAAACTGCCTTTGCTTCCA	*GmAAP2*
R3	GGCACTTACTATTCCTTCAAG	*GmAAP2*
F4	ATCCATTCATTGTGGGCG	*GmAAP3*
R4	ACAATAAAGCTCAGGCAAC	*GmAAP3*
F5	CTTTCCGAACTGCTTATGTTGC	*GmAAP4*
R5	AAGGCTATCTGAGTCTAATTCC	*GmAAP4*
F6	TATGCTGCTTTCGGAGATGAC	*GmAAP5*
R6	TCATGTGGATTACAATGCAC	*GmAAP5*
F7	GCTTCCATTCTTCAATGAT	*GmAAP6*
R7	GAGGCAAATAGATTATGTGGT	*GmAAP6*
F8	CCACTTTTGTAGCAATGTTGA	*GmAAP7*
R8	GGGTGGATGGAGTTTAATAGT	*GmAAP7*
AW-EcoRI	GGAATTCATGGACGTTGAATTAGCTGCT	*GMAAP5*
AW-XbaI	GCTCTAGACTATTGTTTATACATGAAAAGTTTG	*GMAAP5*

Rapid amplification of cDNA ends (RACE)

The primers used in the RACE PCRs are shown in Table 4.2. The concentration of primers, and the amounts of template and other reagents used in the reactions are as described in Chapter 3. Thermocycling parameters were 2 min at 94 ºC; 5 cycles of 94 ºC for 30 s, 72 ºC for 2 min (omitted for the 5' RACE PCR); 5 cycles of 94 ºC for 30 s, 70 ºC for 30 s, 72 ºC for 2 min; then 30-35 cycles of 94 ºC for 30 s, 65-68 ºC for 30 s, 72 ºC for 2 min. Thermocycling parameters for 3' nested RACE PCR were 2 min at 94 ºC; 5 cycles of 56 ºC for 30 s, 72 ºC for 2 min; then 30 cycles of 94 ºC for 30 s, 50 ºC for 30 s, 72 ºC for 2 min. Thermocycling parameters for 5' nested RACE PCR were 2 min at 94 ºC; then 35 cycles of 65ºC for 30 s, 72 ºC for 2 min. A final extension step at 72 ºC for 6 min was included for all PCR assays.

Gel electrophoresis and purification of PCR products

PCR products were separated by agarose gel electrophoresis and purified from the gel as described in Chapter 2.

Table 4.2: Primers used in RACE PCR.

Primer name	Nucleotide sequence	Specific to
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGT ATCAACGCAGAGT	BD smart II A oligonucleotide
NUP	AAGCAGTGGTATCAACGCAGAGT	BD smart II A oligonucleotide
T7	TAATACGACTCACTATAGGG	Part of UPM primer
RaceAAP6F1	AATGCAAGCAGATGAAAAAGGCCAATGTG	*GmAAP5*
F6	TATGCTGCTTTCGGAGATGAC	*GmAAP5*
RaceAAP6R1	GGGTATTCCTTGTCTATGAAATCTGAACC	*GmAAP5*
RaceAAP6R2	TCAATTATACGAGATAGTGGTTGGGCAAG	*GmAAP5*

Generation of chemically competent cells

E. coli Epicurian Coli XL10-Gold cells were made competent for DNA uptake as described in Chapter 2.

PCR using degenerate oligonucleotide primers

In PCR assays using degenerate primers, the 25 µl reactions contained 0.75- 3 mM $MgCl_2$, 200 µM dNTPs, 3 µM primers, 1.25 U *Taq* polymerase (Promega) and 0.5 µl of a 1:10 dilution of nodule cDNA in water. The thermocycling parameters were 2 min at 94 ºC; 30 cycles of 94 ºC for 30 s, 65-37 ºC for 30 s (touch down) or 50-37 ºC for 30 s (gradient), and 72 ºC for 2 min. A final extension step at 72 ºC for 6 min was included.

Cloning and subcloning

The full-length cDNAs for *GmAAP1* and *GmAAP5* were amplified using primers with incorporated restriction sites at the 5' end of each primer in order to facilitate sub-cloning (Table 4.). PCR assays of 20 - 50 µl contained 1.5 mM $MgCl_2$, 200 µM dNTPs, 300 nM primers, 0.875 U Expand High Fidelity PCR system enzyme mix from Roche (Sydney, Australia) and 1 µl 5' RACE ready cDNA. Thermocycling parameters were 2 min at 94 ºC; 5 cycles of 94 ºC for 30 s, 55 ºC for 30 s, and 72 ºC for 2 min, then 30 cycles of 94 ºC for 30 s, 50 ºC for 30 s, and 72 ºC for 2 min. A final extension step at 72 ºC for 6 min was included. The PCR products were purified and inserted into pGEMT. The integrity of each insert was verified by sequencing before sub-cloning into the *S. cerevisiae* pYES3 expression vector by means of the restriction sites introduced into the PCR primers (Table 4.). After ligation into pYES3, the insert sequence was again verified.

Ligation, bacterial transformation, plasmid isolation and DNA sequencing

Ligation of inserts into vectors, bacterial transformation, plasmid isolation and DNA sequencing were done as described in Chapter 2.

Amino acid sequence analysis

Predictions of transmembrane helices in deduced amino acid sequences were done using the TMHMM 2.0 server (http://www.cbs.dtu.dk/services/TMHMM/). The method for prediction of version 1 is described in Sonnhammer *et al*. (1998). During an evaluation of methods for predicting transmembrane helices by Moller *et al.* (2001), TMHMM was found to be the best performing algorithm

Signal sequence predictions and targeting prediction were done using the SignalP 3.0 and the TargetP 1.1 algorithms, respectively (http://www.cbs.dtu.dk/services/) (Bendtsen *et al.*, 2004; Emanuelsson *et al.*, 2000).

Construction of phylogenetic tree

A blastP search of the GenBank database, using the amino acid sequence deduced from the *At*AAP1 gene as the query sequence, was performed. A total of 73 sequences that showed more than 35% identity to AtAAP1 were chosen for further analysis. Examination of these sequences identified 44 distinct full-length sequences that were aligned using Clustal W (Thompson *et al.*, 1994) and used to produce a phylogenetic tree with the Neighbour Joining Method within the Phylogeny Interference Package (Felsenstein, 1989) of BioManager, a product of ANGIS (http://www.angis.org.au).

Results

Isolation of putative amino acid transporter genes from soybean using a PCR approach

Initially a PCR approach using oligonucleotides with degenerate sequences was taken in order to identify putative amino acid transporters expressed in soybean nodules. A protein alignment (Appendix 9) of 10 plant AAPs was constructed and degenerate primers designed using the CodeHop (Consensus-Degenerate Hybrid Oligonucleotide Primers) program (Rose *et al.*, 1998) of BioManager (http://www.angis.org.au). The AAP subfamily was specifically targeted since its members are known to transport amino acids like Glu, Asp and Ala of particular interest for nitrogen fixation as mentioned in the introduction. Three 5' primers and three 3' primers were designed (Table 4.3). Despite the use of various PCR conditions, no products were amplified (data not shown) and a more specific approach was taken.

Table 4.3: Degenerate primers used in this study.

The underlined sequence in each nucleotide sequence is the degenerate part of the primer. 5' primers indicated with an F, and 3' primers with an R in the primer name. N = A, T, G or C; Y = T or C; R = G or A and D = T, G or A.

Primer name	Nucleotide sequence	Degeneracy
AAdegF1	CCTGTGAACGGAAAGAGAAACTAYACNTAYATG	16
AAdegF2	AGCTACAGTATGATCTCCATCGARATNCARGAY	32
AAdegF3	CCTGTGAACGGAAAGAGAAACTAYACNRAAATG	16
AAdegR1	GATGAAGATCATACTGTAACTNGCRTANGCRAA	64
AAdegR2	GAGCATGTAGGGGTTGTTARAANCCRAA	16
AAdegR3	GGGGACTTTCCTTCTGGATDATRTGCATYTC	12

Identification of soybean ESTs encoding putative amino acid transporters

A tblastn search (Dec., 2003) of the Soybean TIGR Gene Index (www.tigr.org), using the amino acid sequence deduced from the open reading frame of the soybean amino acid transporter gene *GmAAP1* (GenBank accession no: AY029352) as the query sequence, identified a total of 21 soybean expressed sequence tags (ESTs) or tentative consensus (TC) sequences with significant similarity (data not shown). Sequences that showed highest similarity to Arabidopsis cDNAs or that were singletons with < 20% of the EST or TC showing similarity to *GmAAP1* were excluded from further analyses.

Five TC sequences and one singleton were chosen for further investigation because they showed highest identity to other legume cDNAs. In addition the identity (45-65%) to *GmAAP1* was at the 3' end for all five TC sequences (Appendix 10), which increased the likelihood that gene specific primers could be designed. The targets TCs were TC189182, TC200947, TC151206, TC149761 and TC182628 and the target singleton was GenBank accession number AW307506.

Primers for PCR were designed for each of the target sequences. The 5' primer was anchored within the ORF, while the 3' primer was anchored within the 3' untranslated region to increase the probability of producing a gene-specific primer pair or in the 3' end of the ORF (Table 4.1). PCR assays with these primers were used to amplify the target cDNAs from RNA isolated from various soybean tissues. The resulting products were sequenced directly to verify that only one product was obtained and that its sequence matched that of the target TC sequence (data not shown). The positions of the primers designed to be specific for AW307506 were in an area where the AW307506 sequence overlapped only three of the TC sequences and the *GmAAP1* sequence. However, PCR with these primers, using root nodule cDNA as the template followed by direct sequencing of the product, confirmed that the sequence of the PCR product

matched the AW307506 sequence. Each unique sequence obtained by these combined approaches was designated as a separate soybean AAP gene (Table 4.4).

Note that after this work was started, the TC identifiers changed. The original and current TC numbers are shown in Table 4..

Table 4.4: Soybean ESTs with identity to *Gm*AAP1.

Original TC number/GenBank accession number	Percent identity to *Gm*AAP1 amino acid sequence	Current TC number	Designated name (This study)
TC189182	50% over 612 bp	TC203584	GmAAP2
TC200947	53% over 480 bp	TC232997	GmAAP3
TC151206	45 % over 595 bp	TC220062	GmAAP4
AW307506	45% over 576 bp	not applicable	GmAAP5
TC149761	65% over 753 bp	TC218569	GmAAP6
TC182628	55% over 416 bp	TC232413	GmAAP7

Expression profiles of GmAAP*s*

RT-PCR was used to examine the expression profiles in different tissues of *GmAAP1*, *GmAAP2*, *GmAAP3*, *GmAAP4*, *GmAAP5*, *GmAAP6* and *GmAAP7* (Figure 4.1). Gene specific primers for *GmAAP3*, *GmAAP4*, *GmAAP6* and *GmAAP7* (Table 4.1) produced PCR products of expected sizes from leaf, root and nodule cDNA, indicating these genes were expressed in all three tissues. When specific primers for full-length *GmAAP1* were used, a PCR product of expected size was generated from root and nodule cDNA, but not from leaf cDNA. This result was confirmed with primers internal to the *GmAAP1* cDNA. Using *GmAAP2* specific primers, a PCR product of expected size was amplified from leaf and root cDNA but not from nodule cDNA. Primers designed to specifically amplify *GmAAP5* produced a strong PCR product of the expected size from nodule cDNA, but only a very faint product was produced with cDNA from leaf and root. These results show that *GmAAP5* is the only gene whose transcripts are easily amplified from nodule tissue, but are difficult to amplify from leaf and root tissue, making it the strongest candidate for an AAP gene involved in the acquisition of fixed N.

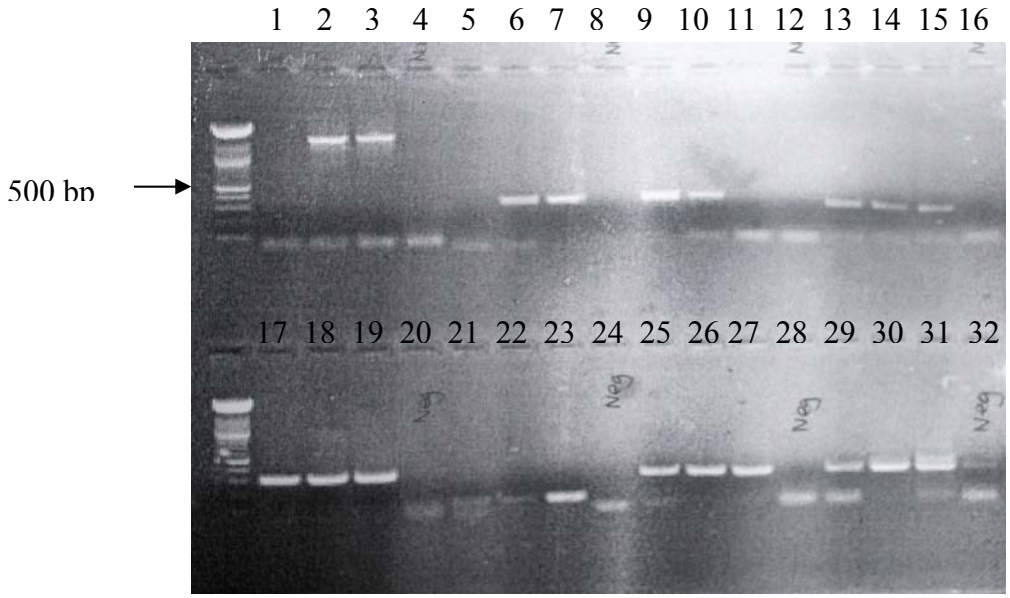


Figure 4.1: Relative expression profiles of *GmAAP* genes in leaf, root and nodule tissue.

Reverse transcription-PCR products produced using specific primer pairs for *GmAAP1* to *GmAAP7* in amplifications of leaf, root and nodule cDNAs. The PCR products were separated on a 1% agarose gel in TAE buffer containing 1 μg ml^{-1} ethidium bromide. Templates used in the reactions were: lanes 1, 5, 9, 13, 17, 21, 25 and 29: leaf cDNA; lanes 2, 6, 10, 14, 18, 22, 26 and 30: root cDNA; lanes 3, 7, 11, 15, 19, 23, 27 and 31: nodule cDNA; lanes 4, 8, 12, 16, 20, 24, 28 and 32: no template controls. Primers used in the reactions were: lanes 1-4: F1/R1 (*GmAAP1*); lanes 5-8: F2/R2 (*GmAAP1*); lanes 9-12: F3/R3 (*GmAAP2*); lanes 13-16: F4/R4 (*GmAAP3*); lanes 17-20: F5/R5 (*GmAAP4*); lanes 21-24: F6/R6 (*GmAAP5*); lanes 25-28: F7/R7 (*GmAAP6*), lanes 29-32: F8/R8 (*GmAAP7*). Molecular weight markers (100 bp ladder), with the 500 bp fragment indicated, are on the left. The band around 100 bp seen in some of the lanes is due to primer dimer formation.

Isolation of a full-length cDNA encoding GmAAP5

Since *GmAAP5* was more highly expressed in nodules than in leaf and root tissues (Figure 4.1), the full-length cDNA encoding this putative soybean amino acid transporter was isolated for further analysis.

In a 3' RACE PCR using the RaceAAP6F1 and the UPM primers (Table 4.2), three products of approximately 850, 750 and 650 bp were generated (data not shown). The three products were designated 6a, 6b and 6c, respectively. The predicted product size

using these primers was 650 bp, based on alignments with *GmAAP1*. PCR products 6a, 6b and 6c were extracted separately from the gel and used as templates in subsequent nested RACE PCRs with primers F6 and T7 (Table 4.2). The resulting products, designated 6A, 6B, and 6C, were inserted into pGEMT and their sequences determined from the 3' end, relative to the contextual direction of the ORF (Appendix 11). While the sequences are incomplete at their 5' ends, analysis of the sequences obtained indicated that the three clones overlapped with the sequence of GenBank accession no. AW307506 and confirmed that they encoded the 3' end of *GmAAP5*. Difference in length of the inserts was observed at the 3' ends. This was due to poly (A) tails located at different positions in the 3' UTR (see later).

A 5' RACE PCR product of the expected size (approximately 1550 bp based on the *GmAAP1* sequence) was produced with the *GmAAP5*-specific primers, RaceAAP6R1 and UPM (Table 4.2). Nested RACE PCR verified the specificity of this 5' RACE product, the nested PCR product was inserted into pGEMT and its sequence determined. The nucleotide sequence overlapped with the sequence of each of the 3' RACE products confirming that it encoded the 5'end of the *GmAAP5*.

Sequence analysis of Gm*AAP1 and* Gm*AAP5*

GmAAP1

The *GmAAP1* ORF encodes a protein of 513 amino acids with a predicted molecular mass of 56,968 Da. Using TMHMM (Sonnhammer *et al.*, 1998) the *Gm*AAP1 protein was predicted to have 9 transmembrane domains (Figure 4.2), with the N-terminus located inside and the C- terminus outside the cell/compartment. The targeting prediction program TargetP (Emanuelsson *et al.*, 2000) predicted *Gm*AAP1 to contain a mitochondrial targeting peptide, with an output score of 0.847 and a reliability class (RC) of 3, indicating not a very strong prediction. A Conserved Domain Database

search for protein classification (Marchler-Bauer *et al.*, 2005; Marchler-Bauer & Bryant, 2004) through the National Center for Biotechnology Information (NCBI) placed *Gm*AAP1 within the amino acid transporter (ATF) superfamily.

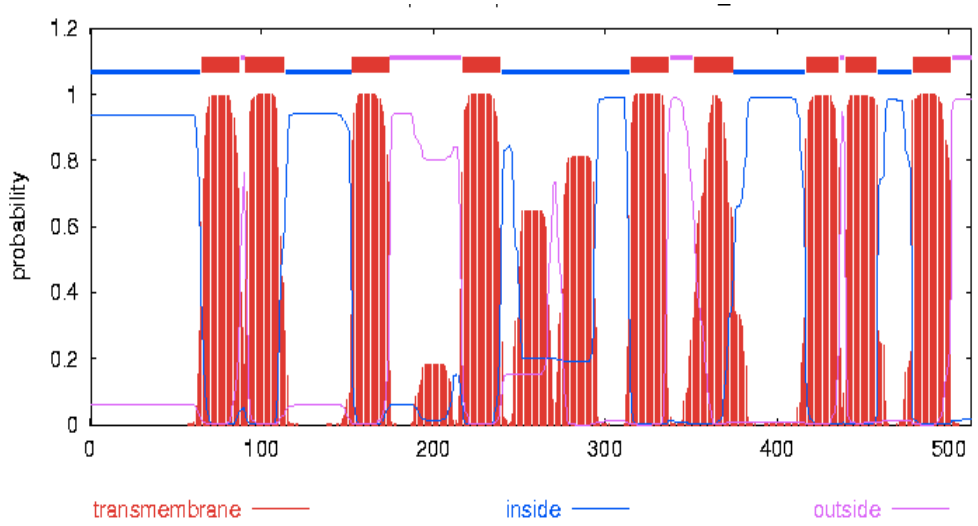


Figure 4.2: Hydrophobicity analysis of *Gm*AAP1.

Hydrophobicity plot obtained in the prediction of the transmembrane domains of *Gm*AAP1 using TMHMM. The red curves correspond to predicted transmembrane domains, the blue curves to predicted intracellular domains and the pink curves to predicted extracellular domains.

GmAAP5

The sequences of the 5' RACE product and the 3' RACE product, 6A, (Appendix 11) overlapped to form a contiguous sequence (contig) of 1843 bp, while the 5' RACE product and the 3' RACE product, 6B, (Appendix 11) generated a single contig of 1781 bp. What appeared to be a poly (A) tail in the shortest 3' RACE product (6C, Appendix 11) was positioned upstream of the predicted stop codon of the contig. In addition no stop codon in frame with the identified ATG in the overlapping 5' sequence was identified; consequently this sequence was excluded from further analysis as a probable artefact of the library construction.

Upon translation of the two complete contig sequences (the 1843 bp and 1781 bp cDNA), an identical ORF of 1413 bp was identified. The ORF was predicted to encode a protein of 470 amino acids with a molecular mass of 51,928 Da. A stop codon

upstream and in frame with the *GmAAP5* ORF suggested that the next in frame ATG codon downstream of this stop encodes the initiating methionine of *Gm*AAP5.

In the 1781 bp cDNA, a poly (A) addition signal, TATAAA, (Wu *et al.*, 1995) was identified 20-26 bp upstream from the poly (A) tail. In addition a GT rich region 55-79 bp upstream of the poly (A) tail was identified. This poly (A) addition signal and GT rich region were also found in the same positions in the longer 1843 bp cDNA, and in addition, a second poly (A) signal, AATAAT, (Wu *et al.*, 1995) and GT-rich region were identified 20-26 bp upstream and 55-68 bp upstream of the poly (A) tail, respectively.

*Gm*AAP5 was predicted to have 11 transmembrane domains using TMHMM (Sonnhammer *et al.*, 1998) and the N- and C-termini of the protein were expected to have intracellular and extracellular locations, respectively (Figure 4.3). A signal anchor (uncleaved signal peptide) was predicted with a probability of 0.947 using the hidden Markov Model in the SignalP 3.0 algorithm (Bendtsen *et al.*, 2004). Using the TargetP 1.1 prediction program (Emanuelsson *et al.*, 2000), the *Gm*AAP5 protein was predicted to have a location other than chloroplast, mitochondrion or secretory pathway, with an output score of 0.985 and an RC of 1, which indicate a strong prediction (Bendtsen *et al.*, 2004; Emanuelsson *et al.*, 2000). A Conserved Domain Database search (Marchler-Bauer *et al.*, 2005; Marchler-Bauer & Bryant, 2004) through NCBI placed *Gm*AAP5, like *Gm*AAP1, within the ATF superfamily.

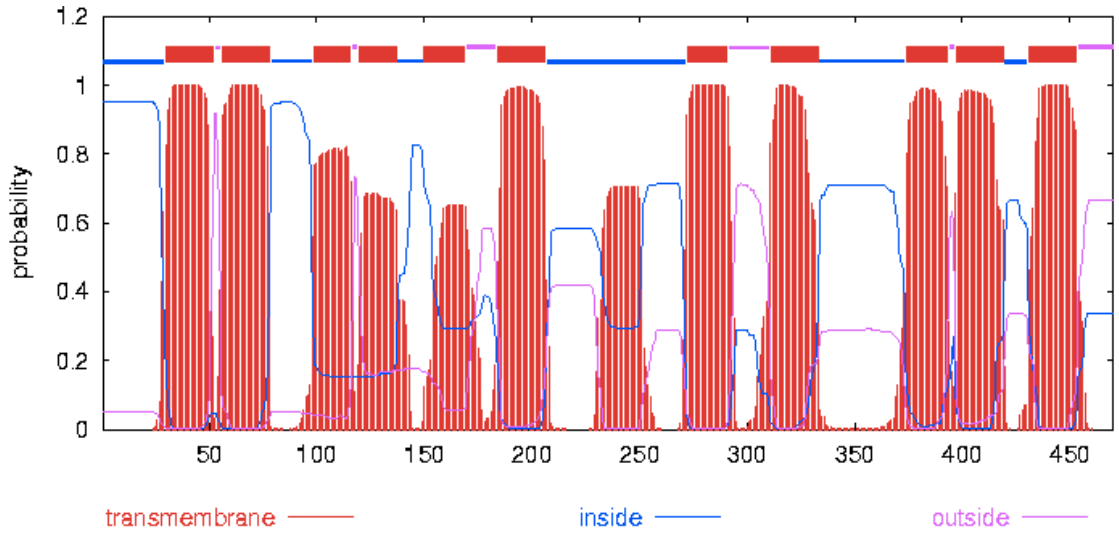


Figure 4.3: Hydrophobicity analysis of *Gm*AAP5.

Hydrophobicity plot obtained in predicting the transmembrane domains of *Gm*AAP5 using TMHMM. The red curves correspond to predicted transmembrane domains, the blue to intracellular domains and the pink to extracellular domains.

Comparison of Gm*AAP1 and* Gm*AAP5 to other plant AAPs*

A comparison of the deduced amino acid sequences of *Gm*AAP1 and *Gm*AAP5 to the functionally characterised AAPs from Arabidopsis and various legume species is shown in Table 4.5 and a sequence alignment is shown in Appendix 12.

Table 4.5: Percent similarity* between the deduced amino acid sequences from *Gm*AAP1, *Gm*AAP5 and functionally characterised AAPs from Arabidopsis and various legume species. Ricinus communis, Rc; Vicia faba, Vf; Pisum sativum, Ps; Glycine max, Gm and Arabidopsis thaliana, At. The numbers indicate the AAP number.

	*Gm*5	*Vf*1	*Vf*2	*Vf*3	*Ps*1	*Rc*1	*Rc*3	*At*1	*At*2	*At*3	*At*4	*At*5	*At*6
*Gm*1	57	75	88	68	74	76	77	65	77	76	79	73	65
*Gm*5	-	61	58	57	58	59	58	61	60	59	59	56	62

*OldDistance in Biomanager by ANGIS (**http://www.angis.org.au**).

The deduced *Gm*AAP1 amino acid sequence showed highest similarity (88%) to the AAP2 sequence from *Vicia faba*. The *Gm*AAP5 deduced amino acid sequence showed

highest similarity (62%) to that of AAP1 from rape (*Brassica napus*). The similarity of the *Gm*AAP1 sequence to those of functionally characterised AAPs from other legumes was between 68% and 88%, while the similarity to *Gm*AAP5 was only 57% (Table 4.5). The sequence similarity of the deduced amino acid sequences of *Gm*AAP5 and those of AAPs from other legumes or Arabidopsis was between 57 and 62 % (Table 4.5). A similar trend was found at the amino acid sequence identity level (data not shown). These results indicate that *Gm*AAP1 has a closer evolutionary relationship with AAPs from other species than with *Gm*AAP5.

A computer-aided analysis of sequence relatedness between 44 distinct full-length plant amino acid transporters was performed by construction of a phylogenetic tree (Appendix 13). A list of the sequences used for constructing this phylogenetic tree is given in Appendix 14. Deletion of the first 66 amino acid positions within the alignment, the least conserved part of the sequences used in the construction of the phylogenetic tree, did not change the clustering of the AAPs in the tree (data not shown).

As the only full-length monocotyledonous plant AAP sequences available in the databases are from rice, and because the rice sequences always grouped together on the same branches of the tree (Appendix 13), the phylogenetic analysis was simplified to include only AAP sequences from dicotyledonous plants. In addition, the two Arabidopsis *AAP7* genes in the NCBI database (accession no: NM_122286 and NM_001036857) are apparently not expressed and were consequently also excluded from further analysis. The resulting phylogenetic tree, which includes the deduced amino acid sequences of all known dicotyledonous AAPs that are expressed, is shown in Figure 4.4.

All the AAP sequences clustered into two major groups, with sequences from both leguminous and non-leguminous plants found in each group. The *Gm*AAP1 sequence

showed the closest relationship to an AAP2 sequence from another legume *Vicia faba*. These sequences were found within a larger cluster containing the non-leguminous sequences *At*AAP2, *At*AAP4 and *Bn*AAP2. Bootstrap analysis indicated these groups are strongly supported. The *Gm*AAP5 sequence clustered within the other major group of AAPs, showing higher sequence similarity to AAPs from non-leguminous plants, namely, *Bn*AAP1, *At*AAP1, *At*AAP8, *Bn*AAP6, *At*AAP6, *St*AAP1 and *Le*AAP1, than to two AAP sequences from *V. faba*. In fact, the *V. faba* AAP sequences formed a sister group to that containing the *Gm*AAP5 sequence. Within this latter group, it is interesting to note that the *Gm*AAP5 sequence is the most divergent. Again, bootstrap values strongly support the clustering within this major group of AAP sequences.

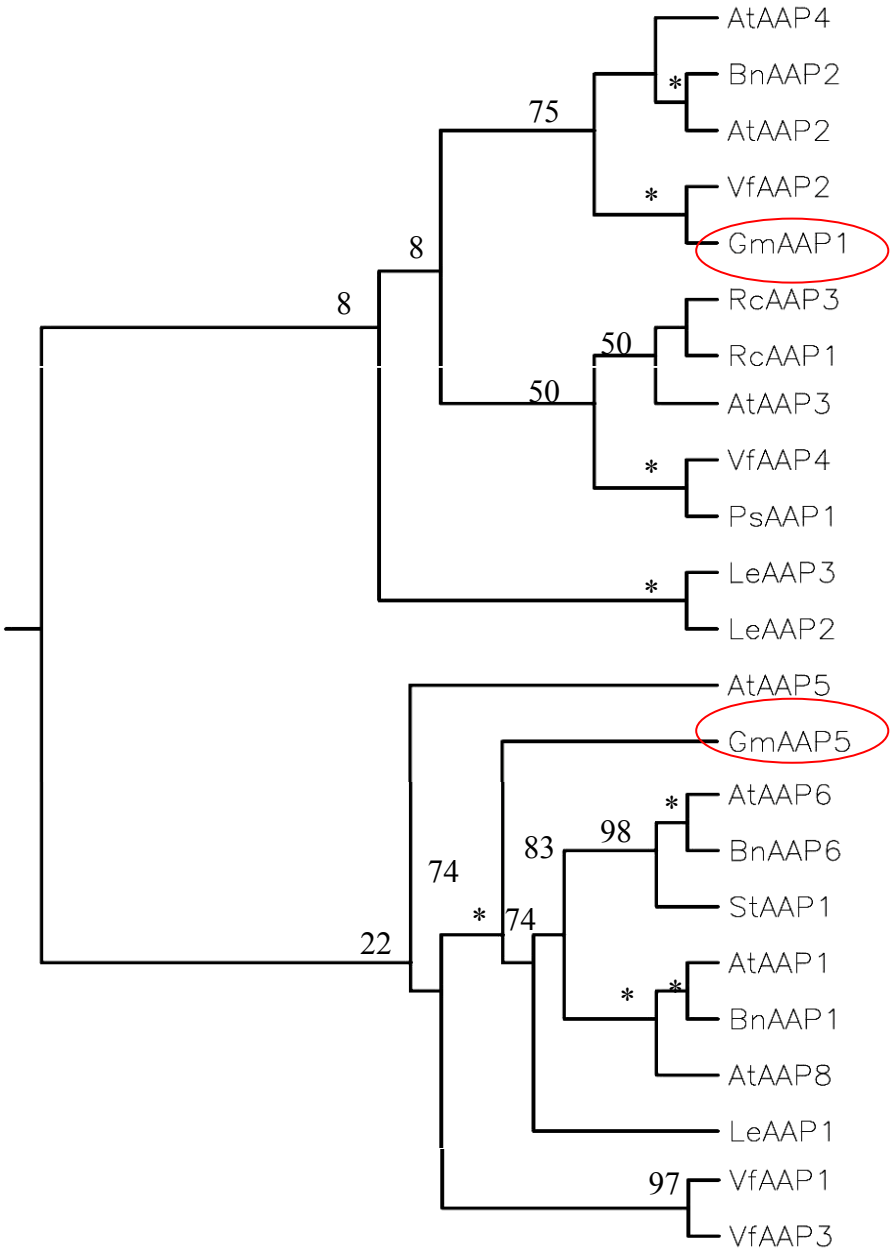


Figure 4.4: Phylogenetic tree showing the relationships of dicotyledonous AAP amino acid sequences.

The tree was constructed from an alignment of the full-length deduced amino acid sequences of AAPs from dicotyledonous plants. This analysis was performed using the Phylogeny Interference Package version 3.2 (Felsenstein, 1989) in BioManager at ANGIS (http://www.angis.org.au). The soybean AAP sequences are circled in red. The numbers indicate the occurrence of a given branch in 100 bootstrap replicates of the given data set, * represents 100. *Ricinus communis*, *Rc*; *Vicia faba*, *Vf*; *Pisum sativum*, *Ps*; *Glycine max*, *Gm* and *Arabidopsis thaliana*, *At*; *Brassica napus*, *Bn*; *Solanum tuberosum*, *St*; *Lycopersicon esculentum*, *Le*. GenBank protein accession numbers are given in Appendix 14.

Discussion

A PCR approach using degenerate primers, designed to amplify AAP-like cDNA sequences from soybean, failed to produce products from nodule cDNA; therefore, a PCR approach using specific primers was taken, which was successful. The expression profile of seven putative soybean amino acid transporter genes (*GmAAP1* to *GmAAP7*) was examined in leaf, root and nodule tissue. The expression of *GmAAP5* was highest in nodule tissue, with very low expression detected in root and leaf tissue. *GmAAP5* was the only gene that showed a nodule-enhanced expression profile and RACE PCR subsequently led to the isolation of the corresponding full-length cDNA.

PCR fragments of three different lengths were generated in 3' RACE reactions of *GmAAP5*. In the shortest fragment (6C), what appeared to be a poly (A) tail was found upstream to where the stop codon was predicted to be in the longer sequences. Fragment 6C was excluded from further analysis, since comparison of the 3' RACE product sequences showed the poly (A) sequence of 6C aligns with a very adenine rich region in the 6A and 6B sequences (Appendix 11). Thus the generation of fragment 6C probably resulted from mis-priming during first strand cDNA synthesis. Two poly (A) addition signals and two GT-rich regions were identified in the 3' UTR of the longest *GmAAP5* 3'-RACE product (6A) while only one poly (A) addition signal and one GT-rich region was found in fragment 6B. It is not unusual to find more than one poly (A) signal in the 3' UTR of plant genes (Wu *et al.*, 1995). The GT-rich region upstream of the poly (A) signal is believed to enhance the recognition of the poly (A) addition signal, and formation of different 3' ends may alter the mode of gene expression (Wu *et al.*, 1995). It is possible that different cell types in the tissue used for construction of the cDNA library contained these different forms of the *GmAAP5* mRNA.

The *GmAAP5* ORF encodes a protein of 470 amino acids and is predicted to contain 11 transmembrane domains. The presence of a signal anchor was predicted with the hidden

Markov model prediction via the SignalP 3.0 server (Bendtsen *et al.*, 2004). The TargetP algorithm (Emanuelsson *et al.*, 2000) predicted that the intracellular location of the *Gm*AAP5 protein was anywhere other than chloroplast, mitochondria or secretory pathway, with an output score of 0.985 and an RC of 1, which indicate a strong prediction (Bendtsen *et al.*, 2004; Emanuelsson *et al.*, 2000). These results clearly indicate that *GmAAP5* encode a protein with intracellular localisation.

*Gm*AAP1 in contrast to *Gm*AAP5 is predicted to have 9 transmembrane domains. Targeting of *Gm*AAP1 is predicted to be to the mitochondria, however, the prediction using the TargetP algorithm (Emanuelsson *et al.*, 2000) was not very strong (RC 3). Proteins of the ATF superfamily, which includes *Gm*AAP1, that have been identified so far are involved in plasma membrane, vesicular and lysosomal amino acid transport (Su *et al.*, 2004). Furthermore, the identified yeast and plant mitochondrial amino acid transporters belong to a very different family of transporters, the mitochondrial transporter family (MCF) (Wipf *et al.*, 2002). It is therefore unlikely that *Gm*AAP1 is targeted to the mitochondria.

Both *Gm*AAP1 and *Gm*AAP5 are members of the ATF superfamily. In addition they both show highest sequence similarity to members of the AAP subfamily. No bacterial proteins have so far been classified to the ATF superfamily (Su *et al*., 2004). All characterised ATF transporters, irrespective of kingdom, are proton coupled, either as symporters or antiporters and have a 9-11 transmembrane domain topology (Russnak *et al.*, 2001; Wipf *et al.*, 2002; Fischer *et al.*, 2002). AAPs transport a wide spectrum of amino acids, including the major transport amino acids glutamine, asparagine and glutamate (Fischer *et al.*, 2002). The individual *At*AAPs differ in their substrate affinity, for example *At*AAP6 has an overall higher substrate affinity and is the only AAP from Arabidopsis that has an affinity for aspartate in the physiological relevant range (Fischer *et al.*, 2002). Electrophysiological analysis of Arabidopsis *AAP*s in *Xenopus* oocytes

showed that these proteins function in a proton-dependent manner, and are energised by a proton gradient and membrane potential across the plasma membrane (Fischer *et al.*, 2002). The secondary structure predicted for *GmAAP1* and *GmAAP5* fit the features of the ATF transporters. In addition *Gm*AAP1 and *Gm*AAP5 both show high sequence identity to Arabidopsis AAPs. It is therefore very likely that *GmAAP1* and *GmAAP5* encode amino acid transporters with similar proton-coupled mechanisms.

A phylogenetic analysis revealed that the *Gm*AAP5 sequence is more closely related to AAPs from non-leguminous plants than to AAPs from other legumes. This relationship is interesting because *GmAAP5* showed relatively high expression in nodule tissue and almost no expression in leaf and root tissues, and a closer relationship to other species capable of developing root nodules would therefore be expected. Of the AAPs that *Gm*AAP5 clustered with, only two, *At*AAP1 and *At*AAP6, have been functionally characterised. Both Arabidopsis AAPs are classified as neutral and acid amino acid transporters, with low and high affinities, respectively (Fischer *et al.*, 1998; Fischer *et al.*, 2002). Therefore it is possible that *Gm*AAP5 is also a neutral and acidic amino acid transporter, and plays a role in the translocation of amino acids derived from fixed N from the root nodule. Clearly however, further characterisation, such as expression of *Gm*AAP5 in a heterologous system and transgenic analyses, are needed to resolve this.

In the phylogenetic analysis, *Gm*AAP1 grouped together with *Vf*AAP2, *At*AAP2, *At*AAP4 and *Bn*AAP2, with the two legume sequences forming a sister group to that containing the Brassica sequences. *VfAAP2* is most strongly expressed in stem and at a lower level in sink leaves, pods and source leaves (Montamat *et al.*, 1999), while expression of *GmAAP1* was detected in root, leaf and nodule tissues. Heterologous expression studies in a yeast amino acid transporter mutant demonstrated that *Vf*AAP2 can transport a broad spectrum of amino acids, and *Vf*AAP2 was classified as a neutral

and acidic amino acid AAP (Montamat *et al.*, 1999), like *At*AAP2 and *At*AAP4 (Fischer *et al.*, 1998). It is therefore very likely that *Gm*AAP1 is also a neutral and acidic amino acid AAP, however, further work is needed to confirm this. Since *Gm*AAP1 is expressed in root, leaf and nodule tissues it would be expected to play a general role in amino acid translocation within soybean plants.

In conclusion, the identification of *GmAAP5,* a gene encoding a novel putative amino acid transporter that shows relatively high expression in soybean nodules compared to leaf and root tissues, has provided new insights into the cycling of amino acids between the two partners in the soybean-rhizobium symbiosis. Future functional characterisation of *GmAAP1* and *GmAAP5*, including expression in heterologous systems, transgenic analyses, and *in situ* hybridisation and immunolocalisation assays, will provide us with a better understanding of the roles of these particular transporters in symbiotic soybeans.

Chapter 5: General discussion

Introduction

Transport of metabolites within the legume root nodule and exchange across the symbiosome membrane (SM) are crucial for optimal nitrogen fixation. Identification of the metabolites that are actually exchanged between the symbiotic partners and understanding the mechanisms behind the exchange of metabolites across the SM, in particular, are key to our understanding how the legume-rhizobia symbiotic relationship is initiated and maintained. Identification of novel genes encoding transport proteins expressed in the nodule adds to our understanding of transport mechanisms in the nodule in particular and our understanding of transport proteins in general.

Putative soybean malate transporters

This study has concentrated on exchange of carbon in the root nodule of soybean. The Arabidopsis tonoplast dicarboxylate transporter (*At*tDT) (Emanuelsson *et al.*, 2000) was considered a possible homolog to the soybean SM malate transporter because of its localisation to the tonoplast, an internal plant membrane like the SM, and its ability to transport malate out of the cytosol. In addition, the uncoupling agent carbonyl cyanide m-chlorophenylhydrazone, that inhibits malate transport across the SM, was found to inhibit the transport of malate by *At*tDT (Emmerlich *et al.*, 2003). The PCR approach employed in this study allowed the identification of four soybean homologs of *At*tDT: *GmDT1*, *GmDT2*, *GmDT3* and *GmDT4*, which encode putative malate transporters. Three full-length cDNAs (*GmDT1*, *GmDT2* and *GmDT3*) and one partial (*GmDT4*) cDNA were isolated and characterised (Chapter 3). These genes were found to encode novel plant members of the SLC13 gene family, and genomic Southern blot analyses showed they make up a small gene family in soybean.

Functional analyses of putative malate transporter genes using heterologous systems

Like plant cells, yeast is eukaryotic, it contains vacuoles, has a cell wall and a large negative electrical potential difference across the plasma membrane and is therefore well suited for functional expression of plant transporters (Ryan *et al.*, 2003). For example, a yeast transport mutant defective in proline uptake was used to isolate the Arabidopsis *AAP1* gene by transforming the mutant with a cDNA library from Arabidopsis seedlings, and screening transformants for complementation on selective media (Frommer *et al.*, 1993). Thus if an appropriate yeast transport mutant is available, plant transporters involved in the uptake of nutrients could be relatively easily identified and isolated using this approach. However this approach is not necessarily ideal when transporters involved in the efflux of nutrients or metabolites are sought (Ryan *et al.*, 2003). In addition this approach may not be straightforward when isolating transporters localised to internal membranes like the tonoplast and the SM. In order for the plant transporter to complement a yeast transport uptake mutant, the plant transporter would need to be localised to the plasma membrane in the yeast system. In addition the plant transporter must be inserted and oriented in the yeast plasma membrane correctly such that uptake into the yeast cell would occur.

In this study a *Sch. pombe* malate transport mutant was generated as a tool to screen isolated genes encoding putative malate transporters and to identify novel malate transporters from legume cDNA libraries (Chapter 2). During the course of this project, several technical difficulties inhibited the isolation of cDNAs encoding novel soybean malate transporters using the *Sch. pombe* mutant (Chapter 2); however, the groundwork has been laid for successful future screening experiments and the mutant will be a valuable tool for these studies.

Interestingly, the inability of the *Sch. pombe* mutant to transport malate was not rescued by the introduction of the *GmDT* cDNAs to the cells, although RT-PCR indicated the cDNAs were expressed in this heterologous system (Chapter 3). Posttranscriptional events such as incorrect protein folding and/or targeting are likely responsible for the lack of complementation and further functional characterisation of the *Gm*DTs is obviously needed to determine if they are in fact malate transporters. This work is under way in collaboration with Prof. S. Tyerman (University of Adelaide); the *GmDT1*, *GmDT2* and *GmDT3* cDNAs were subcloned into vectors (data not shown) for expression in *Xenopus* oocytes.

Heterologous expression in *Sch. pombe* or *E. coli* and subsequent reconstitution of either total protein or recombinant protein, respectively, into phospholipid vesicles led to functional characterisation of the spinach chloroplast 2-oxoglutarate/malate translocator (SoDiT1), and mitochondrial dicarboxylate-tricarboxylate carrier (DTC) from both Arabidopsis and tobacco (Kakhniashvili *et al.*, 1997; Weber *et al.*, 1995). An approach such as this may be useful for the *Gm*DTs and may have advantages over heterologous cell systems, as mis-targeting of the expressed protein should not occur.

Putative soybean amino acid transporters

The cycling of amino acids across the SM has been shown to be important for optimal nitrogen fixation in pea plants (Lodwig *et al.*, 2003), and part of this project was focused on identifying genes encoding soybean nodule amino acid transporters. A novel gene, *GmAAP5*, which showed relatively high expression in soybean nodules compared to minimal expression in leaf and root tissues, was isolated from nodule cDNA (Chapter 4). A molecular characterisation was presented of *GmAAP5* and *GmAAP1* (GenBank Accession no: AY029352), a previously identified putative amino acid transporter gene

(Chapter 4). Both genes were found to be members of the AAP gene subfamily and their products were predicted to be integral membrane proteins with multiple transmembrane regions.

Phylogenetic analysis of the deduced amino acid sequences of known functional AAPs from dicotyledonous plants revealed that *Gm*AAP1 is most closely related to AAP2 from *V. faba*, while *Gm*AAP5 is more closely related to AAPs from non-leguminous plants than from leguminous plants. Based on the functional characterisation of the AAPs with which *Gm*AAP1 and *Gm*AAP5 cluster, it is likely that both transporters are neutral and acidic amino acid transporters within the AAP subfamily.

The *S. cerevisiae* strain 22Δ8AA (Fischer *et al.*, 2002; Jauniaux & Grenson, 1990), which is deficient in the uptake of arginine, aspartate, citrulline, GABA, glutamate or proline, has been used for functional characterisation of several AAPs (Fischer *et al.*, 2002; Tegeder *et al.*, 2000) and would provide a good system for functional complementation and characterisation of *Gm*AAP1 and *Gm*AAP5. The AAPs characterised to date in Arabidopsis and other plants have been shown to be relatively non-specific amino acid transporters (Wipf *et al.*, 2002), and this may be the case for *Gm*AAP1 and *Gm*AAP5 as well.

Expression analyses of putative soybean malate and amino acid transporters

The RT-PCR analyses used in this study for both the *GmDT*s and the *GmAAP*s gave a relative indication of the expression levels in the different tissues examined. RT-PCR assays have been used extensively, and continue to be used, to examine relative gene expression patterns (Su *et al.*, 2004). However, quantitative real-time RT-PCR is a more sensitive technique that allows quantitative comparison of transcript levels in various

plant tissues. Future characterisation of the putative soybean malate and amino acid transporters identified in this study could include the use of these assays during nodule development, giving further insights into the roles of these proteins.

Approaches for identifying transporters located on the SM

The genes encoding soybean SM transporters that have been identified and functionally characterised so far have all shown similarity to other transporters of known function. However, this might not be the case for all SM transporters, which makes identification of genes encoding SM transport proteins more challenging. A clear example of this is the nodule-specific *Alnus* dicarboxylate transporter, which was shown to be a member of the peptide transporter family (Jeong *et al.*, 2004).

As described in Chapter 1, several transport mechanisms on the SM have been physiologically and biochemically characterised. In addition proteomic studies have identified multiple novel proteins, including putative transport proteins, associated with the SM of different legumes (Panter *et al.*, 2000; Saalbach *et al.*, 2002; Wienkoop & Saalbach, 2003) and transcriptome analyses (Colebatch *et al.*, 2004; Fedorova *et al.*, 2002) have identified nodule-enhanced genes encoding putative transporters, some of which might be located on the SM. However, despite the efforts of a number of research groups (Panter *et al.*, 2000; Saalbach *et al.*, 2002; Wienkoop & Saalbach, 2003; Colebatch *et al.*, 2004; Fedorova *et al.*, 2002), only a few SM proteins and the genes encoding them in fact have been identified. This could be due to the amount of SM extractable protein and limitations in the technologies used to identify both genes and proteins.

Soybean transport proteins that have been conclusively identified as SM proteins are the aquaporin, Nodulin 26 (Fortin *et al.*, 1987; Rivers *et al.*, 1997; Weaver *et al.*, 1994), the

zinc transporter, *Gm*ZIP1, (Moreau *et al.*, 2002) and the ferrous iron transporter, *Gm*DMT1 (Kaiser *et al.*, 2003). Nodulin 26 and the cDNA encoding it were identified due to its very high abundance on the SM (Fortin *et al.*, 1987), which is not necessarily the case for all SM transport proteins. For example, the gene encoding *Gm*DMT1 was identified during a 5' RACE PCR experiment designed to amplify the N-terminal sequence of a putative ammonium transporter (Kaiser *et al.*, 2003), and the gene encoding *Gm*ZIP1 was identified and isolated by PCR, based on observed sequence similarity to known zinc transporters from Arabidopsis and pea, and a rice EST (Moreau *et al.*, 2002). All three transport proteins were subsequently localised to the SM by immunolocalisation (Miao *et al.*, 1992; *Kaiser et al.*, 2003; Moreau *et al.*, 2002). *In situ* localisation, in addition to functional characterisation of the proteins encoded by the cDNAs isolated in this study, would provide us with a better understanding of the physiological role of these putative transporters. Immunolocalisation could be one way of determining the intracellular location of the putative malate and amino acid transporters. However, generation of antibodies recognising membrane proteins is not an easy task (Harlow & Lane, 1988).

The use of a transient gene expression system involving the introduction of green fluorescent protein (GFP) fusion constructs into cells such as onion epidermal cells is another system by which the intracellular locations of membrane proteins have been determined (Scott *et al.*, 1999). Recently this technique was used to localise a *Medicago truncatula* zinc transporter (*Mt*ZIP2) to the plasma membrane (Burleigh *et al.*, 2003). Where a protein that is normally targeted to the SM in the infected cells of a soybean nodule would localise in onion epidermal cells is not clear. However when nodulin 26 was expressed in transgenic tobacco, the protein was found in high levels in the tonoplast (Zhang & Roberts, 1995). This could indicate that the symbiosome and the vacuole share a common targeting mechanism. It would be interesting to use the onion

epidermal cell system with the cDNAs identified during this study, particularly with the *Gm*DTs, which all showed some level of expression in root, leaf and nodule tissues and exhibit high sequence similarity to the Arabidopsis vacuolar dicarboxylate transporter (Emmerlich *et al.*, 2003).

In planta analyses of putative soybean malate and amino acid transporters

The generation of mutants that develop nodules with metabolic defects is another approach towards identification of genes encoding SM proteins and is essential for molecular and physiological analyses of suspected transporter gene function *in vivo*. A recent example of this approach is the characterisation of the sulfate transporter SST1, which was identified by map-based cloning of an *L. japonicus* mutant, and is the first transporter gene shown to be crucial for nitrogen fixation in *L. japonicus* nodules (Krusell *et al.*, 2005). The results of this study complemented previous proteomics data that indicated SST1 is located on the SM in *Lotus* nodules (Wienkoop & Saalbach, 2003). The proposed role for SST1 in transport of sulfate across the SM is based on a symbiosis-related phenotype of the SST1 mutant plants, nodule specific expression, and a lower level of the nitrogenase structural protein NifH in *sst1* mutant nodules compared to wild-type nodules (Krusell *et al.*, 2005; Wienkoop & Saalbach, 2003).

Soybean is one of the cultivated legumes that has been difficult to transform using routine *Agrobacterium*-mediated transformation procedures, and only in recent years have the transformation procedures been sufficiently improved (Ko *et al.*, 2003; Olhoft *et al.*, 2003). The hairy root transformation system was recently used to introduce RNA interference-mediated repression of the thioredoxin gene in soybean nodules (Lee *et al.*, 2005). Using the cDNAs isolated in this study in such transformation experiments, and investigating the phenotype of the resulting transgenic nodules, along with their histology and biochemistry, would provide novel information and further insights into

the molecular and physiological role of the putative transporters *in planta*. If, however, the transgenic nodule phenotype showed no obvious phenotypic differences to wild-type nodules, as was reported for pea nodules infected with the rhizobia *bra*/*aap* double mutant (Lodwig *et al.*, 2003), then regeneration of mutant plants may be required to determine the phenotype of the mutant plant.

An *in planta* expression system involving transformation of the Arabidopsis *At*tDT knock-out mutant (Emmerlich *et al.*, 2003) with the *GmDT*s would be another approach towards identification of their malate transport ability.

Other putative malate transporter homologs

*GmDT*2 and *GmDT4* showed higher expression in nodules than in leaf and root tissues (Chapter 3). Therefore they are more likely candidates for genes encoding a SM dicarboxylate transporter than *GmDT1* and *GmDT3*. On the other hand, *GmDT1* and *GmDT3* should not be completely ruled out as SM transporters since, as mentioned above, transporters on the SM might show similarity to transporters with very different functions in other plants that do not form a symbiotic relationship with a nitrogen fixing prokaryote. It may be that the dicarboxylate transporter on the soybean SM is a member of the peptide transporter family as in the case of the *Alnus* dicarboxylate transporter, *Ag*DCAT1 (Jeong *et al.*, 2004). Proteome and transcriptome analyses (Colebatch *et al.*, 2004; Wienkoop & Saalbach, 2003) certainly provide evidence for the presence of peptide transporter (PTR) family members in *Lotus* nodules that show similarity to *Ag*DCAT1 and the nitrate and peptide transporters from other plants. A legume homolog of *Ag*DCAT1 would be a very promising SM dicarboxylate transporter candidate; however, when the *Alnus* work was published, the *GmDT* experiments for this study were already in progress and because of time limitations, a homolog of the *Alnus* dicarboxylate transporter was not pursued.

Molecular and functional characterisation of the aluminium-activated malate transporter (*AlMT1*) from wheat (Sasaki *et al.*, 2004), in addition to the above mentioned plant malate transporters, certainly demonstrate that plants use different mechanisms for transport of malate across membranes. Furthermore *AgDCAT*, *AttDT* and *ALMT1* (Emmerlich *et al.*, 2003; Jeong *et al.*, 2004; Sasaki *et al.*, 2004) encode functional transporters that represent three different transporter families. At present, there is no experimental evidence that rules out homologs of any of the above proteins being responsible for malate transport in soybean.

In summary

The identification and, more importantly, functional characterisation of genes encoding transporters responsible of the exchange of nutrients in the root nodule are vital for our full understanding of the legume-rhizobia symbioses in particular and plant transport processes in general. The identification of four cDNAs encoding putative dicarboxylate transporters and a putative amino acid transporter from symbiotic soybeans are novel contributions to this area of research and brings us closer to the understanding of carbon and nitrogen metabolism and transport in the legume nodule.

Bibliography

Allaway, D., Lodwig, E.M., Crompton, L.A., Wood, M., Parsons, R., Wheeler, T.R. & Poole, P.S. (2000) Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol Microbiol* **36**(2), 508-515.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidmann, J.G., Smith, J.A. & Struhl, K. (1994) Prepararion and analysis of DNA. In *Current Protocols in Molecular Biology*, Vol. 1, p. unit 2.3. Edited by F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmann, J.A. Smith & K. Struhl. John Wiley and Sons. Inc.

Bähler, J. (1998) Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**, 943-951.

Basi, G., Schmid, E. & Maundrell, K. (1993) TATA box mutations in the Schizosaccharomyces pombe nmt1 promoter affect transcription efficiency but not the transcription start point or thiamine repressibility. *Gene* **123**(1), 131-6.

Becana, M. & Rodriguez-Barrueco, C. (1989) Protective mechanisms of nitrogenase against oxygen excess and partially-reduced oxygen intermediates. *Physiol. Plant.* **75**, 429-438.

Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**(4), 783-795.

Bergersen, F.J. (1958) The bacterial component of soybean root nodules; changes in respiratory activity, cell dry weight and nucleic acid content with increasing nodule age. *J Gen Microbiol* **19**(2), 312-323.

Bergersen, F.J. & Turner, G.L. (1967) Nitrogen fixation by the bacteroid fraction of bries of soybean root nodules. *Biochem. Biophys. Acta* **141**, 507-515.

Bettey, M. & Smith, J.A. (1993) Dicarboxylate transport at the vacuolar membrane of the CAM plant Kalanchoe daigremontiana: sensitivity to protein-modifying and sulphydryl reagents. *Biochim Biophys Acta* **1152**(2), 270-279.

Bick, J.A., Neelam, A., Hall, J.L. & Williams, L.E. (1998) Amino acid carriers of Ricinus communis expressed during seedling development: molecular cloning and expression analysis of two putative amino acid transporters, RcAAP1 and RcAAP2. *Plant Mol Biol* **36**(3), 377-385.

Blackford, S., Rea, P.A. & Sanders, D. (1990) Voltage sensitivity of H^+/Ca^{2+} antiport in higher plant tonoplast suggests a role in vacuolar calcium accumulation. *J Biol Chem* **265**(17), 9617-9620.

Blumwald, E., Fortin, M.G., Rea, P.A., Verma, D.P.S. & Poole, P.R. (1985) Presence of host-plasma membrane type H^+-ATPase in the membrane envelope enclosing the bacteroids in soybean root nodules. *Plant Physiol* **78**, 665-672.

Brewin, N.J. (1991) Development of the legume root nodule. *Annu. Rev. Cell. Biol.* **7**, 191-226.

Brown, C.M. & Dilworth, M.J. (1975) Ammonia assimilation by Rhizobium cultures and bacteroids. *J. Gen. Microbiol.* **86**, 39-48.

Burleigh, S.H., Kristensen, B.K. & Bechmann, I.E. (2003) A plasma membrane zinc transporter from Medicago truncatula is up-regulated in roots by Zn fertilization, yet down-regulated by arbuscular mycorrhizal colonization. *Plant Mol Biol* **52**(5), 1077-1088.

Bush, D.R. (1999) Amino acid transport In *Plant amino acids, biochemistry and biotechnology*, pp. 305-318. Edited by B.K. Singh. New York: Marcel Dekker, Inc.

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

Catalano, C.M., Lane, W.S. & Sherrier, D.J. (2004) Biochemical characterization of symbiosome membrane proteins from Medicago truncatula root nodules. *Electrophoresis* **25**(3), 519-531.

Catoni, E., Schwab, R., Hilpert, M., Desimone, M., Schwacke, R., Flugge, U.I., Schumacher, K. & Frommer, W.B. (2003) Identification of an Arabidopsis mitochondrial succinate-fumarate translocator. *FEBS Lett* **534**(1-3), 87-92.

Cerana, F., Giromini, L. & Colombo, R. (1995) Malate-regulted channles permeable to anions in vacuoles of Arabidopsis thaliana. *Aust. J. Plant Physiol.* **22**, 115-121.

Colebatch, G., Desbrosses, G., Ott, T., Krusell, L., Montanari, O., Kloska, S., Kopka, J. & Udvardi, M.K. (2004) Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in Lotus japonicus. *Plant J* **39**(4), 487-512.

Dagert, M. & Ehrlich, S.D. (1979) Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. *Gene* **6**(1), 23-28.

Date, R.A. (1973) Nitrogen, a major limitation in the productivity of natural coomunities, crops and pastures in the pacific area. *Soil Biol. Biochem.* **5**, 5-18.

Day, D.A. & Copeland, L. (1991) Carbon metabolism and compartmentation in nitrogen-fixing legume nodules. *Plant Physiol. Biochem.* **29**, 185-201.

Day, D.A., Kaiser, B.N., Thomson, R., Udvardi, M.K., Monreau, S. & Puppo, A. (2001a) Nutrient transport across symbiotic membranes from legume nodules. *Aust. J. Plant Pyisiol.* **28**, 667-674.

Day, D.A., Poole, P.S., Tyerman, S.D. & Rosendahl, L. (2001b) Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cell Mol Life Sci* **58**(1), 61-71.

Day, D.A., Price, D.G. & Udvardi, M.K. (1989) Membrane interfase of the Bradyrhizobium japonicum-glycine max symbiosis: Peribacteroid units from soybean nodules. *Aust. J. Plant Pyisiol.* **16**, 69-84.

de Kerchove d'Exaerde, A., Supply, P., Dufour, J.P., Bogaerts, P., Thines, D., Goffeau, A. & Boutry, M. (1995) Functional complementation of a null mutation of the yeast Saccharomyces cerevisiae plasma membrane $H(^{+})$-ATPase by a plant $H(^{+})$-ATPase gene. *J Biol Chem* **270**(40), 23828-23837.

Dean, R.M., Rivers, R.L., Zeidel, M.L. & Roberts, D.M. (1999) Purification and functional reconstitution of soybean nodulin 26. An aquaporin with water and glycerol transport properties. *Biochemistry* **38**(1), 347-353.

Delhaize, E., Ryan, P.R., Hebb, D.M., Yamamoto, Y., Sasaki, T. & Matsumoto, H. (2004) Engineering high-level aluminum tolerance in barley with the ALMT1 gene. *Proc Natl Acad Sci U S A* **101**(42), 15249-15254.

Dixon, R.O.D. & Wheeler, C.T. (1986) Nitrogen fixation in plants. pp. 1-44. New York: Chapman and Hall, USA.

Dreyer, I., Horeau, C., Lemaillet, G., Zimmermann, S., Bush, D.R., Rodriguez-Navarro, A., Schachtman, D.P., Spalding, E.P., Sentenac, H. & Gaber, R.F. (1999) Identification and characterization of plant transporters using heterologous expression systems. *Journal of Experimental Botany* **50**, 1073-1087.

Duenas, E., Revuelta, J.L., del Rey, F. & Vazquez de Aldana, C.R. (1999) Disruption and basic phenotypic analysis of six novel genes from the left arm of chromosome XIV of Saccharomyces cerevisiae. *Yeast* **15**(1), 63-72.

Elble, R. (1992) A simple and efficient procedure for transformation of yeasts. *Biotechniques* **13**(1), 18-20.

Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**(4), 1005-1016.

Emmerlich, V., Linka, N., Reinhold, T., Hurth, M.A., Traub, M., Martinoia, E. & Neuhaus, H.E. (2003) The plant homolog to the human sodium/dicarboxylic cotransporter is the vacuolar malate carrier. *Proceedings of the National Academy of Sciences of the United States of America.* **100**(19), 11122-11126.

Engelke, T., Jagadish, M.N. & Puhler, A. (1987) Biochemical and genetical analysis of Rhizobium meliloti mutants defective in C4-dicarboxylate transport. *J. Gen. Microbiol.* **133**, 3019-3029.

Fedorova, E., Thomson, R., Whitehead, L.F., Maudoux, O., Udvardi, M.K. & Day, D.A. (1999) Localization of $H(^{+})$-ATPases in soybean root nodules. *Planta* **209**(1), 25-32.

Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J., Town, C.D., VandenBosch, K.A., Gantt, J.S. & Vance, C.P. (2002) Genome-wide identification of nodule-specific transcripts in the model legume Medicago truncatula. *Plant Physiol* **130**(2), 519-537.

Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (version 3.2) *Cladistics* **5**, 164-166.

Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J. & Joyard, J. (2000) Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *Electrophoresis* **21**(16), 3517-3526.

Finan, T.M., Wood, J.M. & Jordan, D.C. (1983) Symbiotic properties of C4-dicarboxylic acid transport mutants of Rhizobium leguminosarum. *J Bacteriol* **154**(3), 1403-1413.

Fischer, W.N., Andre, B., Rentsch, D., Krolkiewwicz, s., Tegeder, M., Breitkreuz, K. & Frommer, W.B. (1998) Amino acid transport in plants. *Trends in plant science* **3**(5), 188-195.

Fischer, W.N., Loo, D.D., Koch, W., Ludewig, U., Boorer, K.J., Tegeder, M., Rentsch, D., Wright, E.M. & Frommer, W.B. (2002) Low and high affinity amino acid H+-cotransporters for cellular import of neutral and charged amino acids. *Plant J* **29**(6), 717-731.

Fisher, R.F. & Long, S.R. (1992) Rhizobium-plant signal exchange. *Nature* **357**(6380), 655-660.

Fortin, M.G., Morrison, N.A. & Verma, D.P. (1987) Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. *Nucleic Acids Res* **15**(2), 813-824.

Frommer, W.B., Hummel, S. & Rentsch, D. (1994a) Cloning of an Arabidopsis histidine transporting protein related to nitrate and peptide transporters. *FEBS Lett* **347**(2-3), 185-189.

Frommer, W.B., Hummel, S. & Riesmeier, J.W. (1993) Expression cloning in yeast of a cDNA encoding a broad specificity amino acid permease from Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **90**(13), 5944-5948.

Frommer, W.B., Kwart, M., Hirner, B., Fischer, W.N., Hummel, S. & Ninnemann, O. (1994b) Transporters for nitrogenous compounds in plants. *Plant Mol Biol* **26**(5), 1651-1670.

Frommer, W.B. & Ninnemann, O. (1995) Heterologous expression of genes in bacterial, fungal, animal and plant cells. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 419-444.

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

Griffith, D.A. & Pajor, A.M. (1999) Acidic residues involved in cation and substrate interactions in the Na^+/dicarboxylate cotransporter, NaDC-1. *Biochemistry* **38**(23), 7524-7531.

Grobler, J., Bauer, F., Subden, R.E. & Van Vuuren, H.J. (1995) The mae1 gene of Schizosaccharomyces pombe encodes a permease for malate and other C4 dicarboxylic acids. *Yeast* **11**(15), 1485-1491.

Guenther, J.F., Chanmanivone, N., Galetovic, M.P., Wallace, I.S., Cobb, J.A. & Roberts, D.M. (2003) Phosphorylation of soybean nodulin 26 on serine 262 enhances water permeability and is regulated developmentally and by osmotic signals. *Plant Cell* **15**(4), 981-991.

Halbleib, C.M. & Ludden, P.W. (2000) Regulation of biological nitrogen fixation. *J Nutr.* **130**(5), 1081-1084.

Hansen, A.P. (1994) N2 fixation of the leguminosae. In *Symbiotic N_2 fixation of crop legumes*, pp. 1-96. Edited by D.E. Leihner. Weikersheim, Germnay: Margraf Verlag.

Hardarson, G. & Atkins, C. (2003) Optimising biological N_2 fixation by legumes in farming systems. *Plant & Soil* **252**, 41-54.

Hardy, R.W.F. & Havelka, U.D. (1975) Nitrogen fixation research: A key to world food? *Science* **188**, 633-643.

Harlow, E. & Lane, D. (1988) *Antibodies - a laboratory manual.* New York: Cold Spring Harbor Laboratory.

Hawkins, H.J., Wolf, G. & Stock, W.D. (2005) Cluster roots of Leucadendron laureolum (Proteaceae) and Lupinus albus (Fabaceae) take up glycine intact: an adaptive strategy to low mineral nitrogen in soils? *Ann Bot (Lond)* **96**(7), 1275-1282.

Herrada, G., Puppo, A. & Rigaud, J. (1989) Uptake of metabolites by bacteroid-containing vesicles and free bacteroids from French bean nodules. *Journal of General Microbiology* **135**, 3165-3177.

Herrigde, D.J. (1977) Carbon and Nitrogen Nutrition of two Annual Legumes. Perth: University of Western Australia.

Hopkins, W.G. (1995) Nitrogen fixation. In *Introductionto plant physiology*, pp. 101-122. John Wiley & Sons, Inc.

Hosie, A.H., Allaway, D., Galloway, C.S., Dunsby, H.A. & Poole, P.S. (2002) Rhizobium leguminosarum has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J Bacteriol* **184**(15), 4071-4080.

Hoyos, M.E., Palmieri, L., Wertin, T., Arrigoni, R., Polacco, J.C. & Palmieri, F. (2003) Identification of a mitochondrial transporter for basic amino acids in Arabidopsis thaliana by functional reconstitution into liposomes and complementation in yeast. *Plant J* **33**(6), 1027-1035.

Hsu, L.C., Chiou, T.J., Chen, L. & Bush, D.R. (1993) Cloning a plant amino acid transporter by functional complementation of a yeast amino acid transport mutant. *Proc Natl Acad Sci U S A* **90**(16), 7441-7445.

Hurth, M.A., Suh, S.J., Kretzschmar, T., Geis, T., Bregante, M., Gambale, F., Martinoia, E. & Neuhaus, H.E. (2005) Impaired pH homeostasis in Arabidopsis lacking the vacuolar dicarboxylate transporter and analysis of carboxylic acid transport across the tonoplast. *Plant Physiol* **137**(3), 901-910.

Inoue, H., Nojima, H. & Okayama, H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene* **96**(1), 23-28.

Janausch, I.G., Zientz, E., Tran, Q.H., Kroger, A. & Unden, G. (2002) C4-dicarboxylate carriers and sensors in bacteria. *Biochim Biophys Acta* **1553**(1-2), 39-56.

Jauniaux, J.C. & Grenson, M. (1990) GAP1, the general amino acid permease gene in Saccharomyces cerevisiae. Nucleotide sequence, protein similarity with the other bakers yeast amino acid permeases, and nitrogen catabolite repression. *Eur J Biochem* **190**, 39-44.

Jeong, J., Suh, S., Guan, C., Tsay, Y.-F., Moran, N., Oh, C.J., An, C.S., Demchenko, K.N., Pawlowski, K. & Lee, Y. (2004) A nodule-specific dicarboxylate transporter from Alder is a member of the peptide trasnporter family. *Plant Physiology* **134**, 969-978.

Jording, D. & Puhler, A. (1993) The membrane topology of Rhizobium meliloti C4-dicarboxylate permease (DctA) as derived from protein fusions with Escherichia coli K12 alkaline phophatase (PhoA) and β-galactosidase. *Mol Gen Genet* **241**, 106-114.

Jording, D., Sharma, P.K., Schmidt, R., Engelke, T., Uhde, C. & Puhler, A. (1992) Regulatory aspects of the C4-dicarboxylate transport in Rhizobium meliloti: Transcriptional activation and dependence on effective symbiosis. *J. Plant Physiol.* **141**, 18-27.

Jording, D.C., Uhde, C., Schmidt, R. & Puhler, A. (1994) The C4-dicarboxylate transport system of Rhizobium meliloti and its role in nitrogen fixation during symbiosis with alfalfa (medicago sativa). *Experientia* **50**, 875-883.

Journet, E.P., Pichon, M., Dedieu, A., de Billy, F., Truchet, G. & Barker, D.G. (1994) Rhizobium meliloti Nod factors elicit cell-specific transcription of the ENOD12 gene in transgenic alfalfa. *Plant J* **6**(2), 241-249.

Kaiser, B.N., Moreau, S., Castelli, J., Thomson, R., Lambert, A., Bogliolo, S., Puppo, A. & Day, D.A. (2003) The soybean NRAMP homologue, GmDMT1, is a symbiotic divalent metal transporter capable of ferrous iron transport. *Plant J* **35**(3), 295-304.

Kakhniashvili, D., Mayor, J.A., Gremse, D.A., Xu, Y. & Kaplan, R.S. (1997) Identification of a novel gene encoding the yeast mitochondrial dicarboxylate transport protein via overexpression, purification, and characterization of its protein product. *J Biol Chem* **272**(7), 4516-4521.

Kim, E.J., Kwak, J.M., Uozumi, N. & Schroeder, J.I. (1998) AtKUP1: an Arabidopsis gene encoding high-affinity potassium transport activity. *Plant Cell* **10**(1), 51-62.

Ko, T.S., Lee, S., Krasnyanski, S. & Korban, S.S. (2003) Two critical factors are required for efficient transformation of multiple soybean cultivars: Agrobacterium strain and orientation of immature cotyledonary explant. *Theor Appl Genet* **107**(3), 439-447.

Krawchuk, M.D. & Wahls, W.P. (1999) High-efficiency gene targeting in Schizosaccharomyces pombe using a modular, PCR-based approach with long tracts of flanking homology. *Yeast* **15**(13), 1419-1427.

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Kramer, U., Sato, S., Nakamura, Y., Tabata, S., James, E.K., Sandal, N., Stougaard, J., Kawaguchi, M., Miyamoto, A., Suganuma, N. & Udvardi, M.K. (2005) The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in Lotus japonicus root nodules. *Plant Cell* **17**(5), 1625-1636.

Kumar, S., Bourdes, A. & Poole, P. (2005) De novo alanine synthesis by bacteroids of Mesorhizobium loti is not required for nitrogen transfer in the determinate nodules of Lotus corniculatus. *J Bacteriol* **187**(15), 5493-5495.

Larsen, P.B., Degenhardt, J., Tai, C.Y., Stenzler, L.M., Howell, S.H. & Kochian, L.V. (1998) Aluminum-resistant Arabidopsis mutants that exhibit altered patterns of aluminum accumulation and organic acid release from roots. *Plant Physiol* **117**(1), 9-18.

Lee, J.A., Harmer, R. & Ignaciuk, R. (1983) Nitrogen as a limiting factor in plant communities. In *Nitrogen as a ecological factor*, pp. 95-112. Edited by J.A. Lee, S. McNiel & I.H. Rorison. Oxford: Blackwell.

Lee, M.Y., Shin, K.H., Kim, Y.K., Suh, J.Y., Gu, Y.Y., Kim, M.R., Hur, Y.S., Son, O., Kim, J.S., Song, E., Lee, M.S., Nam, K.H., Hwang, K.H., Sung, M.K., Kim, H.J., Chun, J.Y., Park, M., Ahn, T.I., Hong, C.B., Lee, S.H., Park, H.J., Park, J.S., Verma, D.P. & Cheon, C.I. (2005) Induction of thioredoxin is required for nodule development to reduce reactive oxygen species levels in soybean roots. *Plant Physiol* **139**(4), 1881-1889.

Li, Y., Parson, R., Day, D.A. & Bergersen, F.J. (2002) Reassesment of major products of N_2 fixation by bacteroids from soybean root nodules. *Microbiology* **148**, 1959-1966.

Lodwig, E.M., Hosie, A.H., Bourdes, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J.A. & Poole, P.S. (2003) Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. *Nature* **422**(6933), 722-726.

Long, S.R. (1996) Rhizobium symbiosis: nod factors in perspective. *Plant Cell* **8**(10), 1885-1898.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D. & Bryant, S.H. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic. Acids Res.* **33**, 192-196.

Marchler-Bauer, A. & Bryant, S.H. (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, 327-31.

Markovich, D. & Murer, H. (2004) The SLC13 gene family of sodium sulphate/dicarboxylate cotransporters. *Pflugers Arch - Eur J Physiol* **447**, 594-602.

Marschner, H. (1995) Nitrogen fixation AND Function of mineral nutrients: Macronutrients. In *Mineral nutrition of higher plants*, pp. 201-254. Academic Press.

Marston, F.A.O. (1986) The purification of eukaryotic polypeptides synthesised in *E. coli*. *Biochem. J.* **240**, 1-12.

Martinoia, E. & Rentsch, D. (1994) Malate compartmentation - Responses to a complex metabolism. *Annu Rev Plant Physiol. Plant Mol. Biol.* **45**, 447-467.

Martinoia, E., Vogt, E., Rentsch, D. & Amrhein, N. (1991) Functional reconstitution of the malate carrier of barley mesophyll vacuoles in liposomes. *Biochim Biophys Acta* **1062**(2), 271-278.

Marvier, A.C., Neelam, A., Bick, J.A., Hall, J.L. & Williams, L.E. (1998) Cloning of a cDNA coding for an amino acid carrier from Ricinus communis (RcAAP1) by functional complementation in yeast: kinetic analysis, inhibitor sensitivity and substrate specificity. *Biochim Biophys Acta* **1373**(2), 321-331.

Mayer, K. & Temperli, A. (1963) The metabolism of L-malate and other compounds by Schizosaccharomyces pombe. *Arch. Mikrobiol.* **46**, 321-328.

Miao, G.H., Hong, Z. & Verma, D.P. (1992) Topology and phosphorylation of soybean nodulin-26, an intrinsic protein of the peribacteroid membrane. *J Cell Biol* **118**(2), 481-490.

Miao, G.H. & Verma, D.P. (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**(7), 781-794.

Miller, A.J. & Zhou, J.J. (2000) Xenopus oocytes as an expression system for plant transporters. *Biochim Biophys Acta* **1465**(1-2), 343-358.

Miranda, M., Borisjuk, L., Tewes, A., Dietrich, D., Rentsch, D., Weber, H. & Wobus, U. (2003) Peptide and amino acid transporters are differentially regulated during seed development and germination in faba bean. *Plant Physiol* **132**(4), 1950-1960.

Miranda, M., Borisjuk, L., Tewes, A., Heim, U., Sauer, N., Wobus, U. & Weber, H. (2001) Amino acid permeases in developing seeds of Vicia faba L.: expression precedes storage protein synthesis and is regulated by amino acid supply. *Plant J* **28**(1), 61-71.

Moller, S., Croning, M.D. & Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**(7), 646-653.

Montamat, F., Maurousset, L., Tegeder, M., Frommer, W. & Delrot, S. (1999) Cloning and expression of amino acid transporters from broad bean. *Plant Mol Biol* **41**(2), 259-268.

Moreau, S., Thomson, R.M., Kaiser, B.N., Trevaskis, B., Guerinot, M.L., Udvardi, M.K., Puppo, A. & Day, D.A. (2002) GmZIP1 encodes a symbiosis-specific zinc transporter in soybean. *J Biol Chem* **277**(7), 4738-4746.

Mouritzen, P. & Rosendahl, L. (1997) Identification of a Transport Mechanism for NH_4^+ in the Symbiosome Membrane of Pea Root Nodules. *Plant Physiol* **115**(2), 519-526.

Neelam, A., Marvier, A.C., Hall, J.L. & Williams, L.E. (1999) Functional characterization and expression analysis of the amino acid permease RcAAP3 from castor bean. *Plant Physiol* **120**(4), 1049-1056.

Newton, W.E. (2000) Nitrogen fixation: From molecules to crop productivity. In *Nitrogen fixation: from molecules to crop productivity*Edited by F.O. Pedrosa, M. Hungria, G. Yates & W.E. Newton. Dordrecht, The Nederlands: Kluwer Academic Publishers.

Niemietz, C.M. & Tyerman, S.D. (2000) Channel-mediated permeation of ammonia gas through the peribacteroid membrane of soybean nodules. *FEBS Lett* **465**(2-3), 110-114.

Ninnemann, O., Jauniaux, J.C. & Frommer, W.B. (1994) Identification of a high affinity NH_4^+ transporter from plants. *Embo J* **13**(15), 3464-3471.

Okazaki, K., Okazaki, N., Kume, K., Jinno, S., Tanaka, K. & Okayama, H. (1990) High-frequency transformation method and library transducing vectors for

cloning mammalian cDNAs by trans-complementation of Schizosaccharomyces pombe. *Nucleic Acids Res* **18**(22), 6485-6489.

Olhoft, P.M., Flagel, L.E., Donovan, C.M. & Somers, D.A. (2003) Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method. *Planta* **216**(5), 723-735.

Orr-Weaver, T.L., Szostak, J.W. & Rothstein, R.J. (1981) Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A* **78**(10), 6354-6358.

Osothsilp, C. & Subden, R.E. (1986a) Malate transport in Schizosaccharomyces pombe. *J Bacteriol* **168**(3), 1439-1443.

Osothsilp, C. & Subden, R.E. (1986b) Isolation and characterization of *Schizosacccharomyces pombe* mutants with defective NAD-dependent malic enzyme. *Can. J. Micrbiol.* **32**, 481-486.

Ouyang, L.-J., Udvardi, M.K. & Day, D.A. (1990) Spcificity and regulation of the dicarboxylate carrier on the peribacteroid membrane of soybean nodules. *Planta* **182**(3), 437-444.

Ouyang, L.J. & Day, D.A. (1992) Transport properties of symbiosomes isolated from siratro nodules. *Plant Physiology & Biochemistry* **30**(6), 613-623.

Pajor, A.M. (1999) Sodium-coupled transporters for Krebs cycle intermediates. *Annu Rev Physiol* **61**, 663-682.

Pajor, A.M. (2000) Molecular properties of sodium/dicarboxylate cotransporters. *J Membr Biol* **175**(1), 1-8.

Pajor, A.M. (2001) Conformationally sensitive residues in transmembrane domain 9 of the Na^+/dicarboxylate co-transporter. *J Biol Chem* **276**(32), 29961-29968.

Pajor, A.M. (2006) Molecular properties of the SLC13 family of dicarboxylate and sulfate transporters. *Pflugers Arch* **451**(5), 597-605.

Pajor, A.M., Kahn, E.S. & Gangula, R. (2000) Role of cationic amino acids in the Na+/dicarboxylate co-transporter NaDC-1. *Biochem J* **350 Pt 3**, 677-683.

Pajor, A.M. & Randolph, K.M. (2005) Conformationally sensitive residues in extracellular loop 5 of the Na^+/dicarboxylate co-transporter. *J Biol Chem* **280**(19), 18728-18735.

Pajor, A.M., Sun, N., Bai, L., Markovich, D. & Sule, P. (1998) The substrate recognition domain in the Na^+/dicarboylate and Na^+/sulphate cotransporters is located in the carboxy-terminal portion of the protein. *Biochimica et Biophysica Acta* **1370**(1), 98-106.

Palmgren, M.G. & Christensen, G. (1993) Complementation in situ of the yeast plasma membrane $H(^+)$-ATPase gene pma1 by an $H(^+)$-ATPase gene from a heterologous species. *FEBS Lett* **317**(3), 216-222.

Palmgren, M.G. & Christensen, G. (1994) Functional comparisons between plant plasma membrane $H(^+)$-ATPase isoforms expressed in yeast. *J Biol Chem* **269**(4), 3027-3033.

Panter, S., Thomson, R., de Bruxelles, G., Laver, D., Trevaskis, B. & Udvardi, M. (2000) Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. *Mol Plant Microbe Interact* **13**(3), 325-333.

Pantoja, O. & Smith, J.A. (2002) Sensitivity of the plant vacuolar malate channel to pH, Ca^{2+} and anion-channel blockers. *J Membr Biol* **186**(1), 31-42.

Pawlowski, K. & Bisseling, T. (1996) Rhizobial and Acthinorhizal Symbioses: What are tha shared features. *The Plant Cell* **8**, 1899-1913.

Peiter, E., Yan, F. & Schubert, S. (2004) Amino acid export from infected cells of vicia faba root nodules: Evidence for an apoplastic step in the infected zone. *Physiologia plantarum* **122**, 107-114.

Pelissier, H.C., Frerich, A., Desimone, M., Schumancher, K. & Tegeder, M. (2004) PvUPS1, an allantoin transporter in nodulated roots of French *Plant Physiol* **134**, 664-675.

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

Picault, N., Palmieri, L., Pisano, I., Hodges, M. & Palmieri, F. (2002) Identification of a novel transporter for dicarboxylates and tricarboxylates in plant mitochondria. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J Biol Chem* **277**(27), 24204-24211.

Poole, P.S., Franklin, M., Glenn, A.R. & Dilworth, M.J. (1985) The transport of L-glutamate by Rhizobium leguminosarum involves a common amino acid carrier. *J. Gen. Microbiol.* **131**(1441-1448).

Prell, J. & Poole, P. (2006) Metabolic changes of rhizobia in legume nodules. *Trends Microbiol* **14**(4), 161-168.

Price, D.G., Day, D.A. & Gresshoff, P.M. (1987) Rapid Isolation of Intact Peribacteroid Envelops from Soybean Nodules and Demonstartion of Selective Permability to Metabolites. *J. Plant Physiol.* **130**, 157-164.

Rastogi, V., Labes, M., Finan, T. & Watson, R. (1992) Overexpression of the dctA gene in Rhizobium meliloti: effect on transport of C4 dicarboxylates and symbiotic nitrogen fixation. *Can J Microbiol* **38**(6), 555-562.

Rentsch, D., Boorer, K.J. & Frommer, W.B. (1998) Structure and function of plasma membrane amino acid, oligopeptide and sucrose transporters from higher plants. *J Membr Biol* **162**(3), 177-190.

Rentsch, D., Laloi, M., Rouhara, I., Schmelzer, E., Delrot, S. & Frommer, W.B. (1995) NTR1 encodes a high affinity oligopeptide transporter in Arabidopsis. *FEBS Lett* **370**(3), 264-268.

Rivers, R.L., Dean, R.M., Chandy, G., Hall, J.E., Roberts, D.M. & Zeidel, M.L. (1997) Functional analysis of nodulin 26, an aquaporin in soybean root nodule symbiosomes. *J Biol Chem* **272**(26), 16256-16261.

Roberts, D.M. & Tyerman, S.D. (2002) Voltage-dependent cation channels permeable to NH_4^+, K^+, and Ca^{2+} in the symbiosome membrane of the model legume Lotus japonicus. *Plant Physiol* **128**(2), 370-378.

Robertson, J.G. & Lyttleton, P. (1984) Division of peribateroid mambranes in root nodules of white clover. *J. Cell. Sci.* **69**, 147-157.

Rodriquez, S.B. & Thorton, R.J. (1990) Factors influencing the utilisation of L-malate by yeasts. *FEMS Microbiol. Letters* **72**, 17-22.

Rolfe, B.G. & Gresshoff, P.M. (1988) Genetic analysis of legume nodule initiation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**, 297-319.

Rolletschek, H., Hosein, F., Miranda, M., Heim, U., Gotz, K.P., Schlereth, A., Borisjuk, L., Saalbach, I., Wobus, U. & Weber, H. (2005) Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. *Plant Physiol* **137**(4), 1236-1249.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M. & Henikoff, S. (1998) Consesus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. *Nucleic Acids Res* **26**(7), 1628-1635.

Rosendahl, L., Dilworth, M. & AR, Glenn. (1992) Exchange of Metabolites Across the Peribacteroid Membrane in Pea Root Nodules *J. Plant Physiol.* **139**, 635-638.

Rosendahl, L., Mouritzen, P. & Rudbeck, A. (2001) Nitrogen transfer in the interface between the symbionts in pea root nodules. *Plant and Soil* **230**, 31-37.

Roth, L.E., Jeon, K.W. & Stacey, G. (1988) Homology in endosymbiotic systems: The term " symbiosome". In *Molecular genetics of Plant-microbe interactions*, pp. 220-225. Edited by R. Palacios & D.P.S. Verma. St. Paul, Minnesota: APS Press.

Rothstein, R.J. (1983) One-step Gene Disruption in Yeast. *Methods in Enzymology* **101**, 202-211.

Rudbeck, A., Mouritzen, P. & Rosendahl, L. (1999) Characterization of aspartate transport across the symbiosome membrane in pea root nodules. *J. Plant Physiol.* **155**, 576-583.

Russnak, R., Konczal, D. & McIntire, S.L. (2001) A family of yeast proteins mediating bidirectional vacuolar amino acid transport. *J Biol Chem* **276**(26), 23849-23857.

Ryan, P.R., Delhaize, E. & Randall, P.J. (1995) Characterisation of Al-stimulated efflux of malate from the apices of Al-tolerant wheat roots. *Planta* **196**, 103-110.

Ryan, P.R., Dong, B., Watt, M., Kataoka, T. & Delhaize, E. (2003) Strategies to isolate transporters that facilitate organic anion efflux from plant roots. *Plant and Soil* **248**, 61-69.

Saalbach, G., Erik, P. & Wienkoop, S. (2002) Characterisation by proteomics of peribacteroid space and peribacteroid membrane preparations from pea (Pisum sativum) symbiosomes. *Proteomics* **2**(3), 325-337.

Saleki, R., Jia, Z., Karagiannis, J. & Young, P.G. (1997) Tolerance of low pH in Schizosaccharomyces pombe requires a functioning pub1 ubiquitin ligase. *Mol Gen Genet* **254**(5), 520-528.

Salmon, J.M. (1987) L-Malic-acid permeation in resting cells of anaerobically grown Saccharomyces cerevisiae. *Biochimica et Biophysica Acta* **901**, 30-34.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular cloning. A laboratory manual.* New York: Cold Spring Harbor Laboratory Press.

Sasaki, T., Yamamoto, Y., Ezaki, B., Katsuhara, M., Ahn, S.J., Ryan, P.R., Delhaize, E. & Matsumoto, H. (2004) A wheat gene encoding an aluminum-activated malate transporter. *Plant J* **37**(5), 645-653.

Schertler, G.F.X. (1992) Overproduction of membrane proteins. *Curr. Opin. Struc. Biol.* **2**, 534-544.

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

Scott, A., Wyatt, S., Tsou, P.L., Robertson, D. & Allen, N.S. (1999) Model system for plant cell biology: GFP imaging in living onion epidermal cells. *Biotechniques* **26**(6), 1125-1132.

Seron, K., Blondel, M.O. & Haguenauer-Tsapis, R. (1999) Disruption of six novel genes from the left arm of chromosome XV of Saccharomyces cerevisiae and basic phenotypic analysis of the generated mutants. *Yeast* **15**(1), 73-79.

Smith, F.W., Ealing, P.M., Hawkesford, M.J. & Clarkson, D.T. (1995) Plant members of a family of sulfate transporters reveal functional subtypes. *Proc Natl Acad Sci U S A* **92**(20), 9373-9377.

Snapp, S.S. & Vance, C.P. (1986) Asparagine biosynthesis in alfalfa (Medicago sativa L.) root nodules. *Plant Physiol* **82**, 390-395.

Sonnhammer, E.L., von Heijne, G. & Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**, 175-182.

Sousa, M.J., Mota, M. & Leao, C. (1992) Transport of malic acid in the yeast Schizosaccharomyces pombe: evidence for a proton-dicarboxylate symport. *Yeast* **8**(12), 1025-1031.

Streeter, J.G. (1991) Transport and metabolism of carbon and nitrogen in legume nodules. In *Advances in botanical research*, Vol. 18, pp. 129-187. Edited by J.A. Callow. London: Harcourt Brace Jovanovich, Publishers.

Stumpf, D.K. & Burris, R.H. (1979) A micromethod for the purification and quantification of organic acids of the tricarboxylic acid cycle in plant tissues. *Anal Biochem* **95**(1), 311-315.

Su, Y.H., Frommer, W.B. & Ludewig, U. (2004) Molecular and functional characterization of a family of amino acid transporters from Arabidopsis. *Plant Physiol* **136**(2), 3104-3113.

Szczyglowski, K., Kapranov, P., Hamburger, D. & de Bruijn, F.J. (1998) The Lotus japonicus LjNOD70 nodulin gene encodes a protein with similarities to transporters. *Plant Mol Biol* **37**(4), 651-661.

Taniguchi, M., Taniguchi, Y., Kawasaki, M., Takeda, S., Kato, T., Sato, S., Tabata, S., Miyake, H. & Sugiyama, T. (2002) Identifying and characterizing Plastidic 2-oxoglutarate/malate and dicarboxylate transporters in Arabidopsis thalinia *Plant Cell Physiol* **43**(7), 706-717.

Taniguchi, Y., Nagasaki, J., Kawasaki, M., Miyake, H., Sugiyama, T. & Taniguchi, M. (2004) Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize. *Plant Cell Physiol* **45**(2), 187-200.

Tate, R., Riccio, A., Merrick, M. & Patriarca, E.J. (1998) The Rhizobium etli amtB gene coding for an NH_4^+ transporter is down-regulated early during bacteroid differentiation. *Mol Plant Microbe Interact* **11**(3), 188-198.

Tegeder, M., Offler, C.E., Frommer, W.B. & Patrick, J.W. (2000) Amino acid transporters are localized to transfer cells of developing pea seeds. *Plant Physiol* **122**(2), 319-326.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Toya, M., Motegi, F., Nakano, K., Mabuchi, I. & Yamamoto, M. (2001) Identification and functional analysis of the gene for type I myosin in fission yeast. *Genes Cells* **6**(3), 187-199.

Tyerman, S.D., Whitehead, L.F. & Day, D.A. (1995) A channel-like transporter for NH_4^+ on the symbiotic interface of N_2-fixating plants. *Nature* **378**, 629-632.

Udvardi, M.K. & Day, D.A. (1989) Electrogenic ATPase Activity on the peribacteroid membrane of soybean (glycine max L.) root nodules. *Plant Physiol* **1989**(90), 982-987.

Udvardi, M.K. & Day, D.A. (1997) Metabolite Transport across Symbiotic Membranes of Legume Nodules. *Annu Rev Plant Physiol Plant Mol Biol* **48**, 493-523.

Udvardi, M.K., Lister, D.L. & Day, D.A. (1991) ATPase activity and anion transport across the peribacteroid membrane of isolated soybean symbiosomes. *Arch Microbiol* **156**, 362-366.

Udvardi, M.K., Price, G.D., Gresshoff, P.M. & Day, D.A. (1988a) A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. *FEBS Letters* **231**(1), 36-40.

Vance, C. (2002) Root-bacteria interactions: Symbiotic N_2 fixation. In *Plant roots, the hidden half*, pp. 839-868. Edited by W. Y, E. A & K. U. New York: Marcel Dekker, Inc.

Vance, C.P., P.H., G. & Allan, D.L. (2000) Biological nitrogen fixation - A critical future need? . In *Nitrogen fixation: From molecule to crop productivity*, pp. 509-514. Edited by F.O. Pedrosa, M. Hungria, G.M. Yates & W.E. Newton. Dordrecht, The Netherlands: Kluwer academic publishers.

Verma, D.P. & Hong, Z. (1996) Biogenesis of the peribacteroid membrane in root nodules. *Trends Microbiol* **4**(9), 364-368.

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

Villalba, J.M., Palmgren, M.G., Berberian, G.E., Ferguson, C. & Serrano, R. (1992) Functional expression of plant plasma membrane $H(^{+})$-ATPase in yeast endoplasmic reticulum. *J Biol Chem* **267**(17), 12341-12349.

Vincill, E.D., Szczyglowski, K. & Roberts, D.M. (2005) GmN70 and LjN70. Anion transporters of the symbiosome membrane of nodules with a transport preference for nitrate. *Plant Physiol* **137**(4), 1435-1444.

Volschenk, H., van Vuuren, H.J. & Viljoen-Bloom, M. (2003) Malo-ethanolic fermentation in Saccharomyces and Schizosaccharomyces. *Curr Genet* **43**(6), 379-391.

Wach, A. (1996) PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in S. cerevisiae. *Yeast* **12**(3), 259-265.

Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. (1994) New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10**(13), 1793-1808.

Walshaw, D.L. & Poole, P.S. (1996) The general L-amino acid permease of Rhizobium leguminosarum is an ABC uptake system that also influences efflux of solutes. *Mol Microbiol* **21**(6), 1239-1252.

Walshaw, D.L., Wilkinson, A., Mundy, M., Smith, M. & Poole, P.S. (1997) Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in Rhizobium leguminosarum. *Microbiology* **143**, 2209-2221.

Watson, R.J. (1990) Analysis of the C4-dicarboxylate transport genes of Rhizobium meliloti: nucleotide sequence and deduced products of dctA, dctB, and dctD. *Mol Plant Microbe Interact* **3**(3), 174-181.

Weaver, C.D., Shomer, N.H., Louis, C.F. & Roberts, D.M. (1994) Nodulin 26, a nodule-specific symbiosome membrane protein from soybean, is an ion channel. *J Biol Chem* **269**(27), 17858-17862.

Weber, A., Menzlaff, E., Arbinger, B., Gutensohn, M., Eckerskorn, C. & Flügge, U. (1995) The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: Molecular cloning of a translocator containing a 12-helix motif and expression of the functional protein in yeast cells. *Biochemisrty* **34**, 2621-2627.

Whitehead, L.F. & Day, D.A. (1997) The peribacteroid membrane. *Physiol. Plant.* **100**, 30-44.

Wienkoop, S. & Saalbach, G. (2003) Proteome analysis. Novel proteins identified at the peribacteroid membrane from Lotus japonicus root nodules. *Plant Physiol* **131**(3), 1080-1090.

Wipf, D., Ludewig, U., Tegeder, M., Rentsch, D., Koch, W. & Frommer, W.B. (2002) Conservation of amino acid transporters in fungi, plants and animals. *Trends Biochem Sci* **27**(3), 139-147.

Wright, R., Basson, M., D'Ari, L. & Rine, J. (1988) Increased amounts of HMG-CoA reductase induce "karmellae": a proliferation of stacked membrane pairs surrounding the yeast nucleus. *J Cell Biol* **107**(1), 101-114.

Wu, L., Ueda, T. & Messing, J. (1995) The formation of mRNA 3'-ends in plants. *Plant J* **8**(3), 323-329.

Yamaguchi, M., Sasaki, T., Sivaguru, M., Yamamoto, Y., Osawa, H., Ahn, S.J. & Matsumoto, H. (2005) Evidence for the plasma membrane localization of Al-activated malate transporter (ALMT1). *Plant Cell Physiol* **46**(5), 812-816.

Yao, X. & Pajor, A.M. (2002) Arginine-349 and aspartate-373 of the $Na(^{+})$/dicarboxylate cotransporter are conformationally sensitive residues. *Biochemistry* **41**(3), 1083-1090.

Yarosh, O.K., Charles, T.C. & Finan, T.M. (1989) Analysis of C4-dicarboxylate transport genes in Rhizobium meliloti. *Mol Microbiol* **3**(6), 813-823.

Zhang, Y. & Roberts, D.M. (1995) Expression of soybean nodulin 26 in transgenic tobacco. Targeting to the vacuolar membrane and effects on floral and seed development. *Mol Biol Cell* **6**(1), 109-117.

Appendices

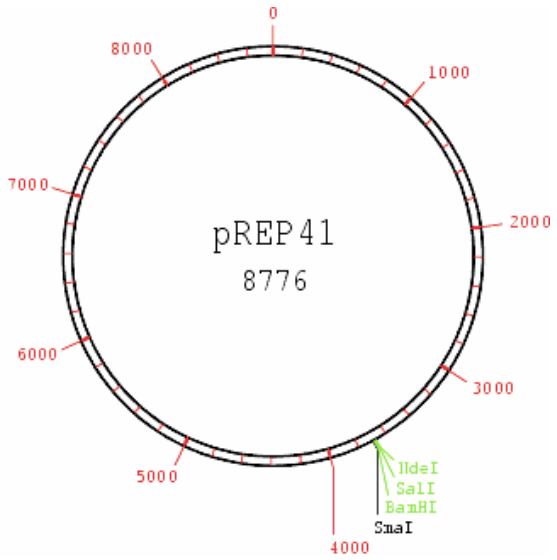
Appendix 1: Chapter 2 - oligonucleotide list

All oligonucleotide used were synthesized and desalted by Sigma or Proligo.

Restriction enzymes sites, introduced to enable sub-cloning are underlined. Where the primer was designed for full-length amplification and if the stat and stop codon are represented in the primer it is indicated in bold in the nucleotide sequence. Sequences indicated with asterisk were phosphorylated at the 5' end.

Name	Nucleotide sequence
mae1-NdeI	GCCATATGGGTGAACTCAAGGAAATC
mae1-BamHI	GCGGATCCTTAAACGCTTTCATGTTCAC
LjNOD70-NdeI	GCCATATGATGTTTTGTGCACCACATCC
LjNOD70-BamHI	GCGGATCCTGGTGTCGTGAGGTTATC
pREP-MCS-F	GGAATTCCATATGTGAGGATCCATCAGTAACGGCCG
pREP-MCS-R	TCCCCCGGGCTCGAGCGGCCGCCAGTGTC
MCS-pREP41b-A	*TATGGGATCCGAATTCGCGGCCGCC
MCS-pREP41b-A	*ACCCTAGGCTTAAGCGCCGGCGGAGCT
mae1iIII	GCCGTCAATGTCTTACGGTT
mae1iIV	AACCTGAGAAAGCTGGTGGA
mae1iVI	GAGAAACGCACGGACCATTA
Mae1-usF	CTTTTACTTGTTTGCTACAC
Mae1-exR	GTGTGCTGCTGATAAGAGTAG
kanmaeF1	GATTGTATGGGAAGCCCGATGCG
kanmaeR2	GATGGCGGCGTTAGTATCGA
kanmaeR3	CACCTGAATCAGGATATTCTTC

Appendix 2: Construction of pREP41b



pREP41a
The *Nde*I/*Sma*I fragment was excised from the poly-linker of pREP41 and replaced with the fragment below:

T*ATG* TGA GGATCC ----pYES3 poly-linker----CTCGAG CCC
GTATAC ACT CCTAGG-----PYES3 poly-linker----GAGCTC GGG
*Nde*I *Bam*HI *Xho*I *Sma*I partial

This introduced a stop codon (bold) in frame with the start codon (italics) in the *Nde*I site and additional restriction sites of the pYES3 poly-linker. This vector was designated pREP41a.

pREP41b
The NdeI/XhoI fragment was excised from the poly-linker of pREP41a and replaced with the fragment below:

TATG GGATCC GAATTC GCGGCCGC C
AC CCTAGG CTTAAG CGCCGGCG GAGCT
*Bam*HI *Eco*RI *Not*I *Xho*I partial

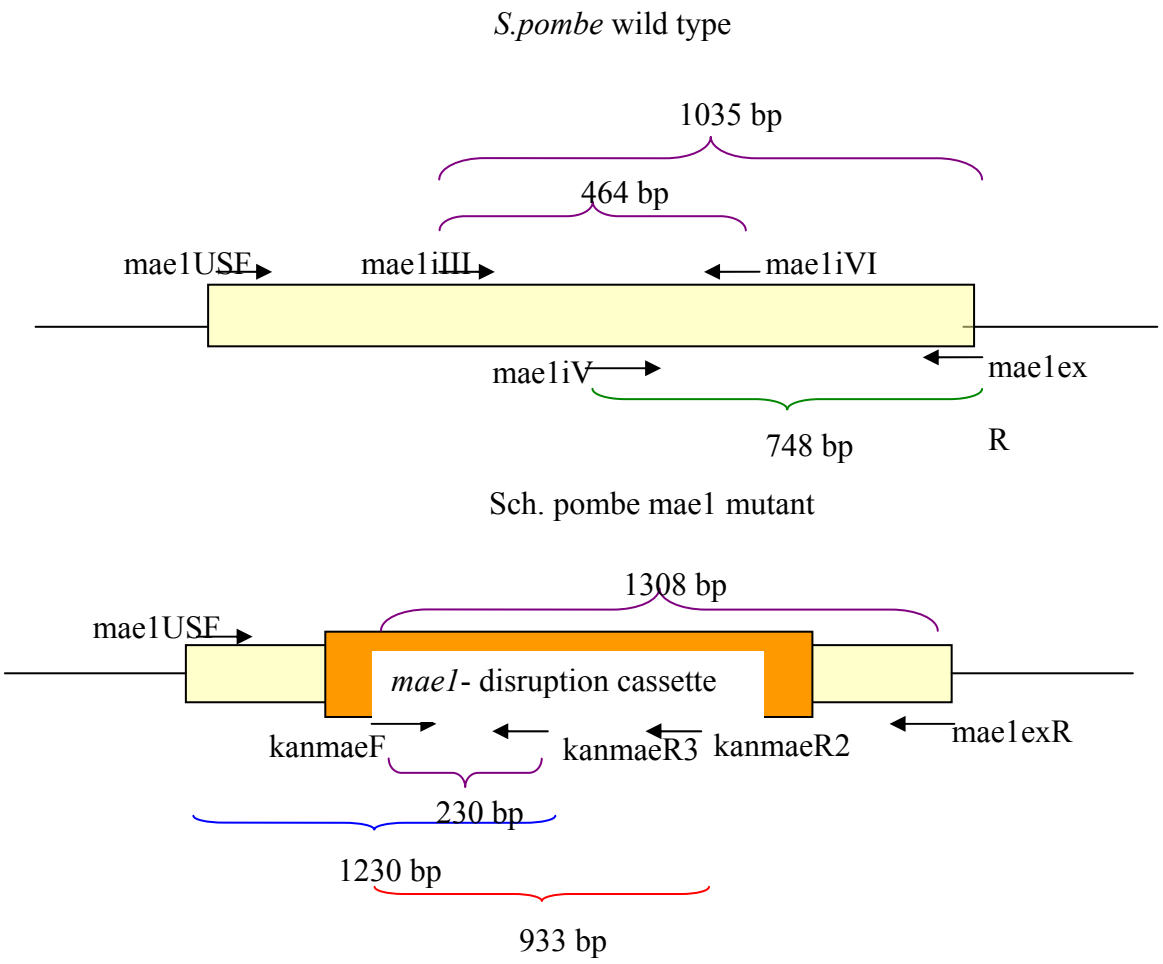
This resulted in a poly-linker with the restriction enzyme sites; *Nde*I, *Bam*HI, *Eco*RI, *Not*I, *Xho*I and *Sma*I. This vector was designated pREP41b.

Appendix 3: Sequence of mae1 disruption cassette

Sequence parts shown in red and blue represents the *mae1* untranslated homologous regions 5' and 3' of *mae1*, respectively. Sequence regions shown underlined are priming sites, used when the mae1 disruption cassette was constructed. Start and stop codon of the ORF of the *E. coli kan*^{r}. For more details on the pFA6-kanMX6 plasmid see Wach (1996) and Wach *et al.,* (1994).

AGCTTATTTGTTGCTGCACTAGACTTTTTGTTTGATTTCTCATCCTACTTCTGTATCGGCAGT
TTGCTCATTTACTAAGACTAGCAACAGCCAGTCATTCATTTTTTACACTCTCTATCATTTTTT
ATTTTCATCACGATAACTAACATGTGCGATTAGACTCACAGATAAATTGCTAGCAATTGGTT
GTCTCTTTCCTTCCTCCGTCTTTTTCGTACGCTGCAGGTCGACGGATCCCCGGGTTAATTA
AGGCGCGCCAGATCTGTTTAGCTTGCCTCGTCCCCGCCGGGTCACCCGGCCAGCGACAT
GGAGGCCCAGAATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGA
CTGTCGCCCGTACATTTAGCCCATACATCCCCATGTATAATCATTTGCATCCATACATTTTG
ATGGCCGCACGGCGCGAAGCAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGA
AACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAATATA
AAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTCATATACTTCCTTTTAAAATCTTGCTAGGA
TACAGTTCTCACATCACATCCGAACATAAACAACC**ATG**GGTAAGGAAAAGACTCACGTTTC
GAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGAT
AATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAG
TTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGAC
TAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGAT
GATGCATGGTTACTCACCACTGCGATCCCCGGCAAAACAGCATTCCAGGTATTAGAAGAAT
ATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTC
GATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAAT
CACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCC
TGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTC
ACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTAT
TGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTG
CCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATC
CTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTC**TAA**TCAGTACTGACA
ATAAAAAGATTCTTGTTTTCAAGAACTTGTCATTTGTATAGTTTTTTTATATTGTAGTTGTTCT
ATTTTAATCAAATGTTAGCGTGATTTATATTTTTTTTCGCCTCGACATCATCTGCCCAGATGC
GAAGTTAAGTGCGCAGAAAGTAATATCATGCGTCAATCGTATGTGAATGCTGGTCGCTATA
CTGCTGTCGATTCGATACTAACGCCGCCATCCAGTTTAAACGAGCTCGAATTCATCGATAA
TCTGTGTGCCCTGCTCTTAATACCATTATAGATTAATCATTTTGAATCATTCTGTATCTTTATT
GTACTACTGGTACTAATTTTGCTTAGACATTTTTGCTCCTTCTTCTTCTTTTTGTTTAAATTAT
ACATACCAAAATTTTGGACTTTGAATAATGGTAATTTTTGGTTGTCGTAGTGTTAAATATGTA
TGCGTCTTGCATATGAATC

Appendix 4: Schematic explanation of Table 2.2.



Appendix 5: *GmDT1* cDNA

Complete cDNA of *GmDT1*. The start and the stop codon of the ORF of *GmDT1* are indicated with red circle around the codons. Poly(A) signals (AATGAA) in the 3'UTR are indicated by black circle.

Appendix 6: *GmDT2* cDNA

Complete cDNA of *GmDT2*. The start and the stop codon of the ORF of *GmDT2* are indicated with red circle around the codons. Stop codons upstream and in frame with the start codon are indicated by blue circle.

```
   1  ACGCGGGGGGTGTTTTAGTTTATGGAAAAAAAATCATCGATAAAAGGGCA
  51  GCACTCAGTCTTTACTTCTTCCAGCCTAAAGTCCCCCCTTCTTCCAATTC
 101  ACCAACAAACACAGAGCTCCTCCCCCTTCACCCTGAAATCCCTTTTGTAT
 151  GTGATCATAGGACCCTTCTTGTCCCTCGTAATATGTCTCTTTGTCAACCT
 201  CGATGCTTGCCGTGATTGCTTGGGTCTATTCTTGGTGGGTCACCGCCGCC
 251  GTGCCTCTTCCGGTGACCTCCATGTGTCCCAAGTTCAACAGAGCGGTGGT
 301  TCTGACGGTGGTGTACGCCACCCCCATTGGAGGGATCAGCACTTCACAAT
 351  TTGCTCTGATACAATTTTTAGAATGTAAGTTTAGGCCTAATTCAACCCAA
 401  AAGCTAGCTCTAAGTATTCTAACTTGGCTTTACTTTTAGCGGATGTGACT
 451  TGGGTTTTTCTCAACATGGTTCTTCTTCGGTTTTCCTGTTGCCATTCTCT
 501  TCCTCTTCTCTTTGGGTGCATCGTCTGTTTCCTCTATGTTCCAAAGGACT
 551  CCTCTCGGGCCTTGTCTTCTTTATCATTTGACATTCCCGGATGGGGAGCT
 601  TGCTTCCATGGTCTTGTTGGTGATGGAAGTGTTAGTGTTATGGTGGCTGT
 651  GTTATTGTTCATAATCCCAAACAAGAAACAAGAGGGGGAGAAGCTAATGG
 701  ATTGGAATGAACGCAAGAAGTTACCTTGGAACCTTATTTTGCTTCTTGGA
 751  GCAGGGTTTGCTTTAGCTGATGGAGTCCAATCTAGTGGCCTAGCAGACGT
 801  GCTTTCCAGAGCCTTGGATTTCTTGGAAGATACACCATACTGGGCCATTG
 851  CTCCTGCTGTTAGTCTTATGAGTAGCATTATTACAGAGTTCATCACATCT
 901  AACGATGCCACTGCCACTGCCACACTTCTAATCCCACTTCTATATAACAT
 951  AGCCAAGACTATGCATGTGCATCCTCTTCTTCTTATGGTACCTGGAGCAG
1001  TGGCAACTGCTTTCTTACTTCCAACTTCAACTCCATCAAATGTAGTGGGA
1051  TTTGCCACTGGCCATATTGAAATCCTAGACATGCTCAAAATTGGCGTGCC
1101  GCTTAAGGTTGCAGGGATAGCTGTGCTCTCAATTTTCATGCCAACTCTAG
1151  GAGCTATTGTTTTTGGAACAGATGACTATATTCAATGGATGCCAAACATA
1201  AACTCTCCTTGGTTGAGAAATTGATTTTTTCTCCAAATTCCACCCTGCTT
1251  TACAAATTCCTGGTGTTCTCTATCATTGACAATCCACACGACCTCCCTTA
1301  TTACTATGTAGATCAANAGCATGGCAATTGTGGAAACCCAAAAAAAAAAA
1351  AAAAAAAAAAAAAAA
```

Appendix 7: *GmDT3* cDNA

Complete cDNA of *GmDT3*. The start and the stop codon of the ORF of *GmDT3* are indicated with red circle around the codons. Poly(A) signals (AATGAA) in the 3'UTR are indicated by black circle.

```
   1  CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCG
  51  GGGACAAAAACCTTCTCCCGCAAACTCATTCTGGGCGGCGATAGTTTTAT
 101  CGTTCTCTATCTCCAAGATGGTCGGAGAACACGGTACGATGTCACTTGCC
 151  TCCGATGATCAGAAAACACCACTTCTTCCCGTGCACGAGCCAATGGAAGG
 201  AACACGCACCAGCGTCTCCTTCAGCATCAAAACAATTCTAACACTGGAAA
 251  ACTTGTACGTGATTCTGGGACCTATGCTGTCTCTTTTCATATGCCTCTTC
 301  GTGAAGCTCGACGCGCCACCCACAAGCCAGAAAATGCTTGGTGTGATTGC
 351  GTGGGTCTTCGCTTGGTGGGTCACGCAGGCCGTGCCACTTCCCGTGACTT
 401  CCATGTGTCCTCTGTTTTTGTTCCCAATTTTCGGGATAGCTTCAGCTGAT
 451  AGTGTGGCGCACTCTTACATGGATGACGTCATCACCCTCGTTTTGGGAAG
 501  CTTCATTCTCGCCCTCGCCGTCGAACGCTACAACGTTCACCGGAGATTGG
 551  CCTTGAATGTAACACTGCTGTTTTGTGGCGACCCGGTTAATCCAGCGCTG
 601  TTACTGCTGGGGCTGTGCGCGACGATATTCTTCGTGAGCATGTGGCTGCA
 651  CAACGTGGCGACGGCGGTGATGATGATGCCGGTGGCGACGGGGATCGTGC
 701  AACGGCTACCGCCGGTGCACGAGCAGTCGGAGGCGGTGAACAAGTTCTGC
 751  CGCGCGGTGATTCTGACGGTGGTCTACGCCACGCCCATCGGAGGAATAAG
 801  CACACTAACGGGAACGGGTGTGAACTTGATAATAATCGGAATGTGGAAGA
 851  GCCTCTTCCCTGAGGCAAAACCAATAAGCTTCAACACGTGGTTCTTCTAT
 901  GGTTTCCCTGTTGCTGTTCTCATCCTCATTTGTTTTTGGTGCATCATTTG
 951  CCTCCTCTATGTGCCAAAAGGTTCGGGTCGGGCTTTGTCTGCTTACTTGG
1001  ATAGATCTCACTTGAAGAGAGACCTCGAAGCTCTTGGTCCCATGGCTTTT
1051  GCGGAGAAGATGGTCTTGTCTGTGTTTGGGTTGCTGATCATACTTTGGAT
1101  GACAAGAAGAATAACAGATGACATTCCTGGATGGGGATCTTTATTCCATG
1151  GCCTAGTTGGAGATGGAAGTGTTAGTGTTATGGTGGCTGTTTTATTGTTC
1201  ATAATCCCAAACATGAAGCAAGAGGGGGAGAAGCTAATGAGTTGGAATGA
1251  CTGCAAAAAGTTACCTTGGAACCTCATTTTGCTTCTGGGAGCTGGTTTTG
1301  CCATAGCTGATGGAGTACAATCTAGTGGCCTAGCAGATGTTTTGTCAAGA
1351  GCCTTGGATTTCTTGGAAGATGCTCCATACTTGGCAATTGTTCCTGCTGT
1401  TAGTCTAATATGTAGCATCATCACTGAGTTCATCACCTCCAATGATGCTA
1451  CTGCCACCCTTCTTGTCCCACTTCTGTATCACATAGCCAGAACTATGCAT
1501  GTGCACCCTCTTCTTCTTATGGTCCCTGGAGGAATAGCAACTGAGTTTGC
1551  CTTCTGGCTTCCAACTTCAACACCATCAAATGTAGTTGGCTTTGCCCCTG
1601  GACACATAGAAATTAAAGACATGCTCAAAGTTGGTGTGCCACTCAAGGTT
1651  GCTGGGATTGTTGTGCTGTCTCTTCTCATGCCATCACTAGGAACTATTGT
1701  TTTTGGAGCGGGCACTGGTACTCAATAAGTGATATTGCCGAAGGAAAACT
1751  CTTGTTGGTTAGGAAATTGATTGATTTCTTCTCCAAATTGCAGAATCTTG
1801  TTTGTCTAGGAGATCAATGAAATTTTTAATATGCAGGGCCTTATGCTATG
1851  TTTAGGTAATGAGGATATATCAACAGGGCAAAGGAGAAGGAGAGAAAGAT
1901  CTCCATCCTTTTCTTGCTTTTATTTTGTATTTCTTTGTTGATTTGTTCAT
1951  AATTTTTTTACAGGGTTATATTATAAAGTATAGAATGAATTTCCAGCAGT
2001  TGAACTTATAAAAAAAAAAAAAAAAAAAAAAAAAA
```

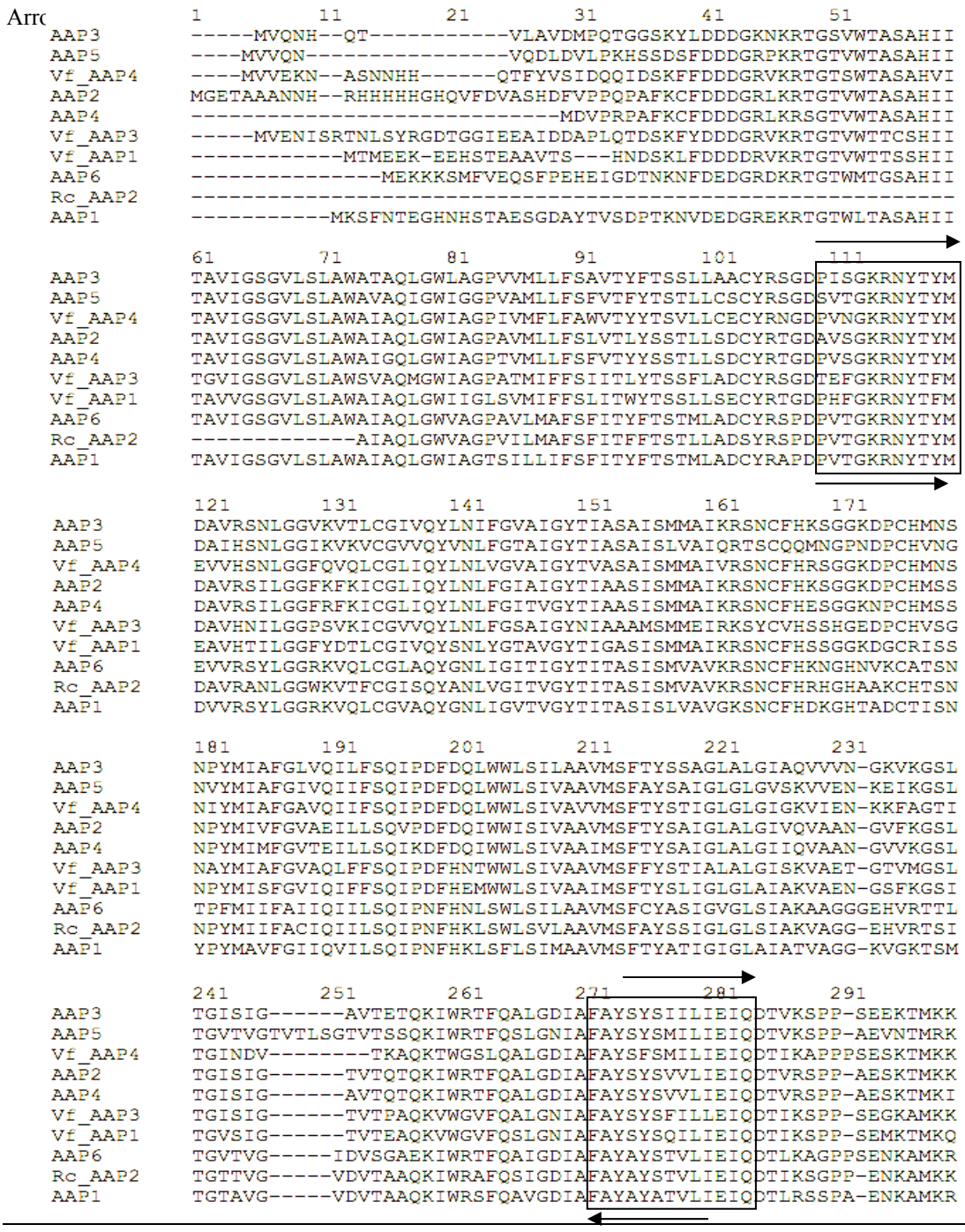
Appendix 8: *GmDT4* cDNA

Partial cDNA of *GmDT4*. The stop codon of the partial ORF of *GmDT4* are indicated with red circle around the codons. Poly(A) signals (AATGAA) in the 3'UTR are indicated by black circle.

```
  1  CTGTGTTTGGGTTGCTGATCATACTTTGGATGACAAGAAGAATAACAGAT
 51  GACATTCCTGGATGGGGATCTTTATTCCATGGCCTAGTTGGAGATGGAAG
101  TGTTAGTGTTATGGTGGCTGTTTTATTGTTCATAATCCCAAACATGAAGC
151  AAGAGGGGGAGAAGCTAATGAGTTGGAATGACTGCAAAAAGTTACCTTGG
201  AACCTCATTTTGCTTCTGGGAGCTGGTTTTGCCATAGCTGATGGAGTACA
251  ATCTAGTGGCCTAGCAGATGTTTTGTCAAGAGCCTTGGATTTCTTGGAAG
301  ATGCTCCATACTTGGCAATTGTTCCTGCTGTTAGTCTAATATGTAGCATT
351  ATCACTGAGTTCATCACCTCCAATGATGCTACTGCCACCCTTCTTGTCCC
401  ACTTCTGTATCACATAGCCAGAACTATGCATGTGCACCCTCTTCTTCTTA
451  TGGTCCCTGGAGGAATAGCAACTGAGTTTGCCTTCTGGCTTCCAACTTCA
501  ACACCATCAAATGTAGTTGGCTTTGCCACTGGACACATAGAAATTAAAGA
551  CATGCTCAAAGTTGGTGTGCCACTCAAGGTTGCTGGGATTGTTGTGCTGT
601  CTCTTCTCATGCCATCACTAGGAACTATTGTTTTTGGAGCGGGCACTGGT
651  ACTCAATAAGTGATATTGCCGAAGGAAAACTCTTGTTGGTTAGGAAATTG
701  ATTGATTTCTTCTCCAAATTGCAGAATCTTGTTTGTCTAGGAGATGAATG
751  AAATTTTTAATATGCAGGGCCTTATGCTATGTTTAGGTAATGAGGATATA
801  TCAACAGGGCAAAGGAGAAGGAGAGAAAGATCTCCATCCTTTTCTTGCTT
851  TTATTTTGTATTTCTTTGTTGATTTGTTCATAATTTTTTTACAGGGTTAT
901  ATTATAAAGTATACAATGAATTTCCAGCAGTTGAACTTTATATCAGTAAT
951  GGATGAATGGAATTGATTTTAAAAAAAAAAAAAAAA
```

Appendix 9: Protein sequence alignment of plant amino acid permeases for degenerate primer design.

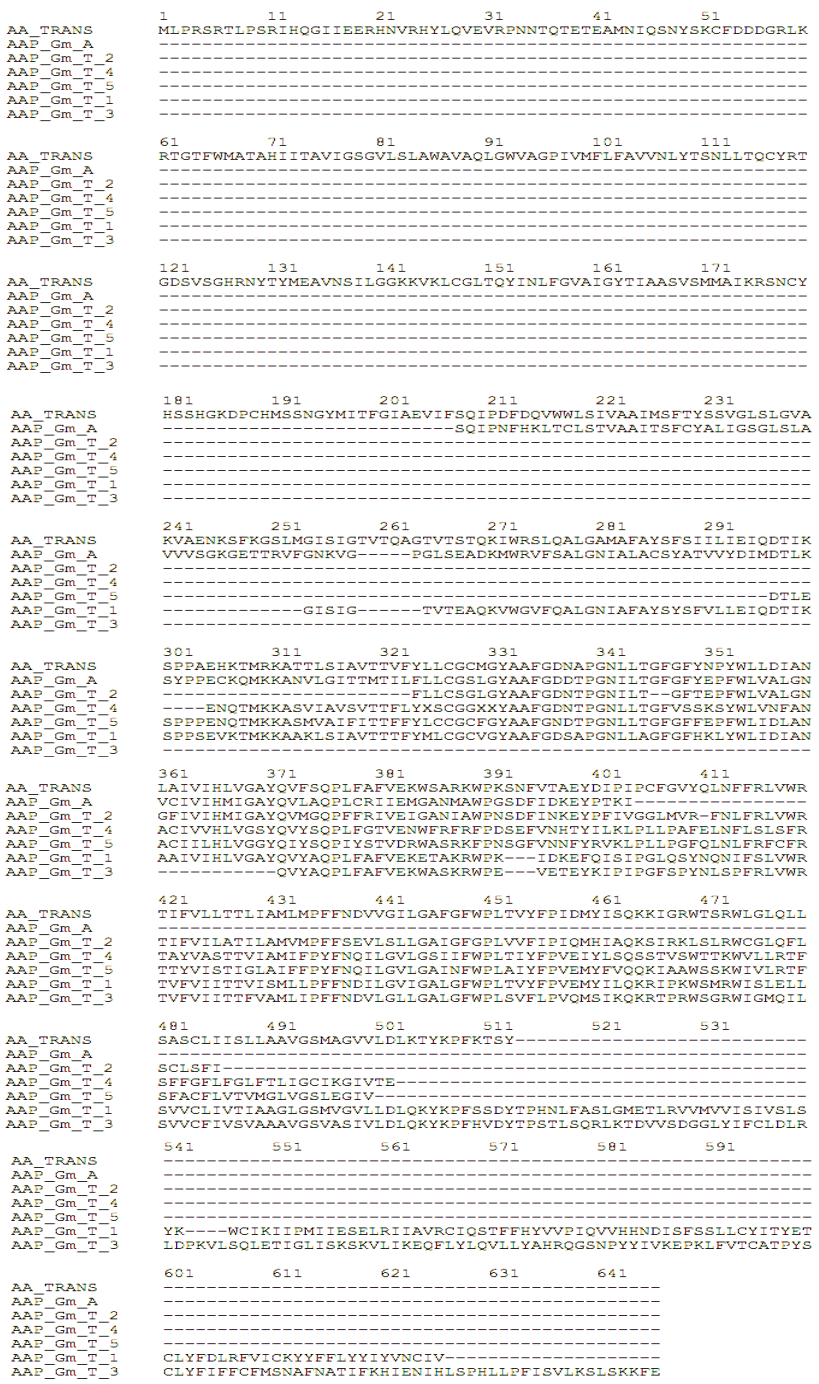
Arabidopsis amino acid permease (AAP) numbers 1-6 are indicated by AAP followed by number (Accession numbers: NP_176132, CAC05448, NP_177862, NP-201190, NP_175076, NP_199774, respectively). *Vicia faba* AAP1, AAP3 and AAP4 are indicated by Vf_AAP followed by number (Accession numbers: CAC_51423, CAC_51424, CAC_51425, respectively). *Ricinus communis* AAP2 (Rc; Accession number: CAA72006). The boxes indicate regions to which the primers were designed.



Appendix 9, continued.

Appendix 10

Protein sequence alignment of *Gm*AAP1 and the TC and GenBank sequences translated in the frame that aligned to *Gm*AAP1. AA_Trans = *Gm*AAP1, AAP_GM_A = *Gm*AAP5, AAP_GM_T1 = TC149761, AAP_GM_T2 = TC200947, AAP_GM_T3 = TC182628, AAP_GM_T4 = TC121506, AAP_GM_T5 = TC189182.



Appendix 11: Sequence alignment of *Gm*AAP5

ClustalW alignment of the consensus sequence (cons_seq) generated from the 5' RACE product and the 3' RACE product 6A, and the sequence of the three 3' RACE products (6A, 6B and 6C). The 5' end of RACE product 6A, 6B and 6C was not sequenced, the high quality 3' sequence is shown in this alignment. Start and stop codons are boxed in black. Stop codons in frame and upstream of the start codon are underlined. Regions with identity to consensus poly(A) signals (TATAAA and TAATAA) are boxed in red. Blue boxes indicate the GT-rich regions.

```
                1         11        21        31        41        51
cons_seq        CCGAACACTGCGTTTGCTGGCTTTGATGAAAAAGAATAAGAGTGATTGTGAAGGGAAGGC
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                61        71        81        91        101       111
cons_seq        CGTACTTGGCACAAAGCATTCTGGTTTGAAACACTCGAATATCACTGAAGAGAGAAGAAA
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                121       131       141       151       161       171
cons_seq        CATTTTAGTACAGAAAATTAACATAGTATTCTGGTTCGAAACTAACAGGTGCTACCCTTT
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                181       191       201       211       221       231
cons_seq        TAAGCTATCAGATTCCTTCAATTTGTAGGATTGTGAGTTGCAAGCAGGGACCTCTTGAAT
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                241       251       261       271       281       291
cons_seq        CTTCATGCACGTTGAATTAGCTGCTAAGAGTGTTTCTAGGAGTGAGGAATTGGATGATGA
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                301       311       321       331       341       351
cons_seq        TGGCAGAATCAAAAGAACCGGGAACGTGTTCACCGCTAGCATTCACATAGTAACGGTGGT
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                361       371       381       391       401       411
cons_seq        GGTGGGTGCAGGAGTGCTGGCTCTAGCATGGGCCATGGCCCAGCTCGGATGGATACCTGG
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                421       431       441       451       461       471
cons_seq        CTTAGCCACTATGATCATCTTTGCATGCATTTCCATTTATACTTACAATCTTGTAGCTGA
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                481       491       501       511       521       531
cons_seq        TTGCTATAGATACCCTGACCCAATCAATGGCAAGAGGAACTACACTTACATGCAAGCTGT
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                541       551       561       571       581       591
cons_seq        CGATGCATACCTTGGTGGAACAATGCACGTGTTTTGCGGATTAATTCAATATGGGAAGCT
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------

                601       611       621       631       641       651
cons_seq        TGCTGGGCTTACAGTGGGCTACACTATAACTTCGTCTACAAGCTTGGTGGCTATAAAGAA
6A              ------------------------------------------------------------
6C              ------------------------------------------------------------
6B              ------------------------------------------------------------
```

Appendix 11, continued

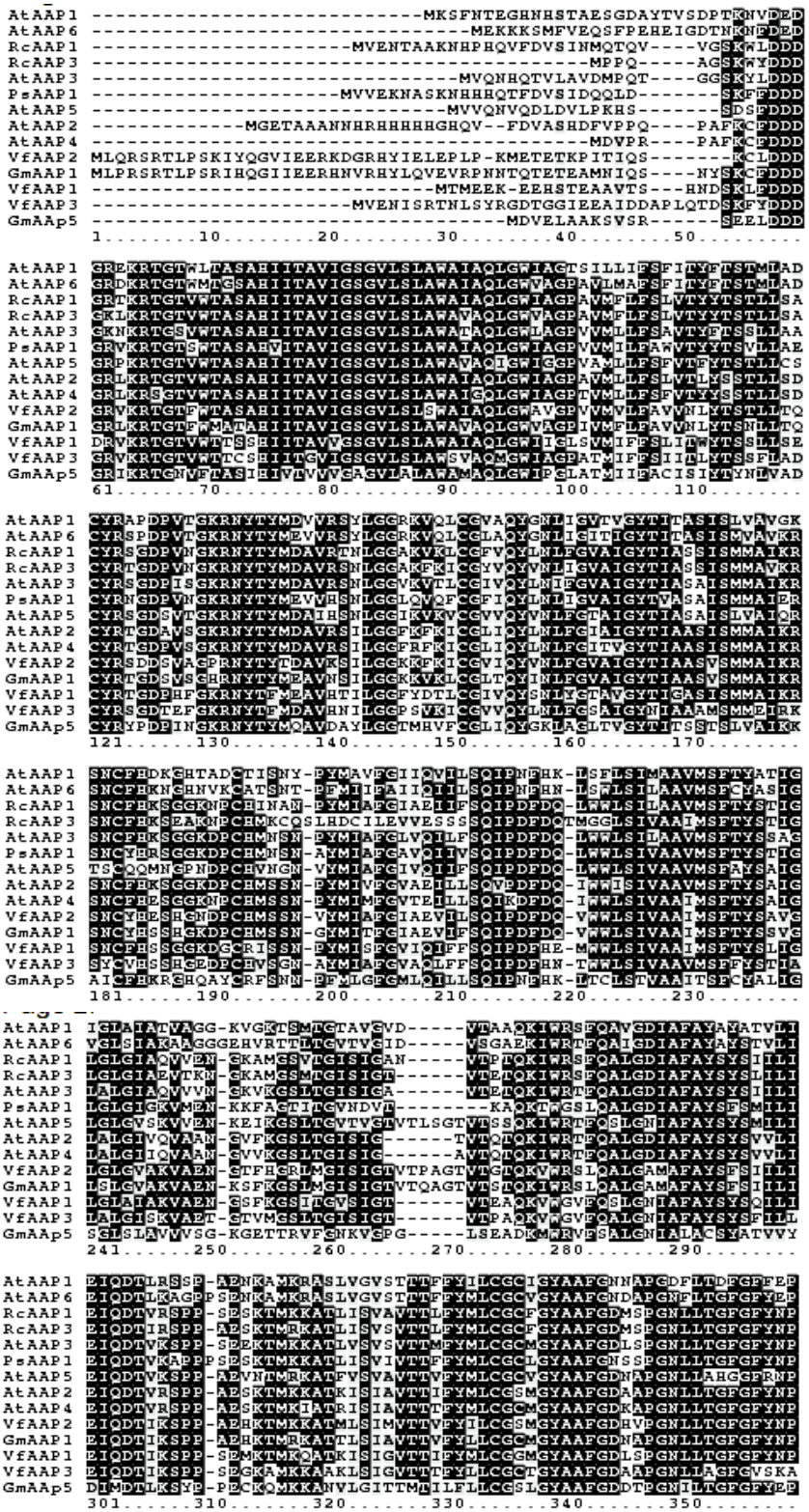
```
            661       671       681       691       701       711
cons_seq    GGCCATTTGCTTTCACAAAAGAGGCCACCAAGCTTATTGCAGGTTTTCAAATAATCCCTT
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            721       731       741       751       761       771
cons_seq    TATGCTCGGTTTTGGGATGTTGCAAATTTTATTGTCTCAAATCCCAAACTTCCACAAGTT
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            781       791       801       811       821       831
cons_seq    AACATGCCTTTCAACCGTTGCTGCTATTACCTCTTTTTGTTATGCATTAATTGGAAGCGG
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            841       851       861       871       881       891
cons_seq    GCTTTCTCTGGCAGTAGTTGTCTCAGGTAAAGGAGAAACAACCCGTGTATTTGGAAACAA
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            901       911       921       931       941       951
cons_seq    AGTAGGGCCAGGACTATCTGAAGCAGATAAAATGTGGAGGGTTTTCAGTGCTTTGGGAAA
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            961       971       981       991       1001      1011
cons_seq    CATTGCACTTGCTTGCTCTTATGCTACTGTTGTGTATGATATAATGGACACGTTGAAGTC
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            1021      1031      1041      1051      1061      1071
cons_seq    ATATCCACCAGAATGCAAGCAGATGAAAAAGGCCAATGTGTTAGGAATCACAACAATGAC
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            1081      1091      1101      1111      1121      1131
cons_seq    AATACTTTTCCTGCTATGTGGTAGCCTTGGCTATGCTGCTTTCGGAGATGACACCCCGGG
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ------------------------------------------------------------

            1141      1151      1161      1171      1181      1191
cons_seq    AAACATCCTCACTGGCTTTGGATTTTACGAGCCATTCTGGTTGGTTGCCCTTGGGAATGT
6A          ------------------------------------------------------------
6C          ------------------------------------------------------------
6B          ----------------------------------------TTGGTTGCCCTTGGGAATGT

            1201      1211      1221      1231      1241      1251
cons_seq    GTGCATTGTAATCCACATGATTGGAGCATATCAGGTGCTTGCCCAACCACTATTTCGTAT
6A          ------------------------------------------------------------CGTAT
6C          ------------------------------------------------------------
6B          GTGCATTGTAATCCACATGATTGGAGCATATCAGGTGCTTGCCCAACCACTATTTCGTAT

            1261      1271      1281      1291      1301      1311
cons_seq    AATTGAGATGGGTGCAAACATGGCATGGCCAGGTTCAGATTTCATAAACAAGGAATACCC
6A          AATTGAGATGGGTGCAAACATGGCATGGCCAGGTTCAGATTTCATAAACAAGGAATACCC
6C          ------------------------------------------------------------
6B          AATTGAGATGGGTGCAAACATGGCATGGCCAGGTTCAGATTTCATAAACAAGGAATACCC

            1321      1331      1341      1351      1361      1371
cons_seq    AACCAAAATAGGCTCCCTAACATTCAGTTTCAACTTGTTTAGGCTAATTTGGAGGACAAT
6A          AACCAAAATAGGCTCCCTAACATTCAGTTTCAACTTGTTTAGGCTAATTTGGAGGACAAT
6C          ----------------TAACATTCAGTTTCAACTTGTTTAGGCTAATTTGGAGGACAAT
6B          AACCAAAATAGGCTCCCTAACATTCAGTTTCAACTTGTTTAGGCTAATTTGGAGGACAAT
```

Appendix 11, continued

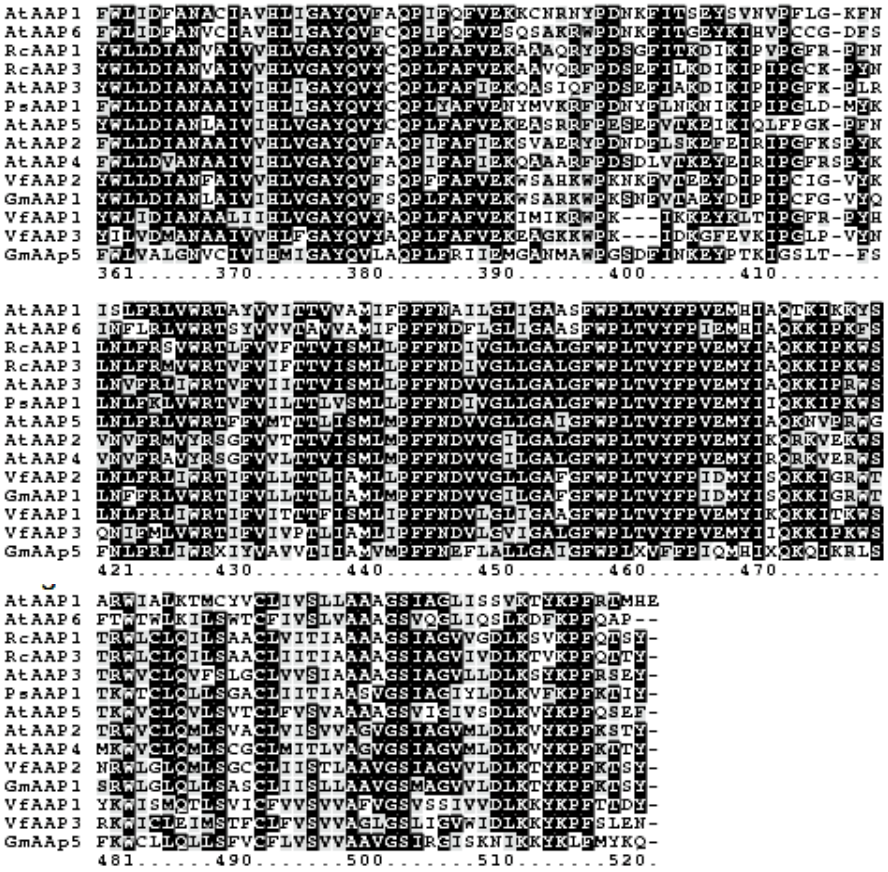
```
            1381      1391      1401      1411      1421      1431
cons_seq    ATACGTGGCAGTGGTCACAATCATTGCCATGGTGATGCCGTTTTTCAACGAGTTTCTTGC
6A          ATACGTGGCAGTGGTCACAATCATTGCCATGGTGATGCCGTTTTTCAACGAGTTTCTTGC
6C          ATACGTGGCAGTGGTCACAATCATTGCCATGGTGATGCCGTTTTTCAACGAGTTTCTTGC
6B          ATACGTGGCAGTGGTCACAATCATTGCCATGGTGATGCCGTTTTTCAACGAGTTTCTTGC

            1441      1451      1461      1471      1481      1491
cons_seq    CCTGCTTGGAGCAATTGGGTTTTGGCCTCTCATCGTCTTTTTCCCTATACAAATGCACAT
6A          CCTGCTTGGAGCAATTGGGTTTTGGCCTCTCATCGTCTTTTTCCCTATACAAATGCACAT
6C          CCTGCTTGGAGCAATTGGGTTTTGGCCTCTCATCGTTTTTTTCCCTATACAAATGCACAT
6B          CCTGCTTGGAGCAATTGGGTTTTGGCCTCTCATCGTCTTTTTCCCTATACAAATGCACAT

            1501      1511      1521      1531      1541      1551
cons_seq    TGCTCAGAAACAGATCAAAAGATTATCATTTAAATGGTGTCTGCTCCAGCTATTGAGCTT
6A          TGCTCAGAAACAGATCAAAAGATTATCATTTAAATGGTGTCTGCTCCAGCTATTGAGCTT
6C          TGCTCAGAAACAGATCAAAAGATTATCATTTAAATGGTGTCTGCTCCAGCTATTGAGCTT
6B          TGCTCAGAAACAGATCAAAAGATTATCATTTAAATGGTGTCTGCTCCAGCTATTGAGCTT

            1561      1571      1581      1591      1601      1611
cons_seq    TGTGTGCTTCCTTGTTTCGGTAGTCGCCGCAGTTGGTTCCATTCGTGGAATTAGCAAGAA
6A          TGTGTGCTTCCTTGTTTCGGTAGTCGCCGCAGTTGGTTCCATTCGTGGAATTAGCAAGAA
6C          TGTGTGCTTCCTTGTTTCGGTAGTCGCCGCAGTTGGTTCCATTCGTGGAATTAGCAAAAA
6B          TGTGTGCTTCCTTGTTTCGGTAGTCGCCGCAGTTGGTTCCATTCGTGGAATTAGCAAGAA

            1621      1631      1641      1651      1661      1671
cons_seq    TATCAAAAAATACAAACTTTTCATGTATAAACAATAGGCCATTTTGCGCATTACTTACCA
6A          TATCAAAAAATACAAACTTTTCATGTATAAACAATAGGCCATTTTGCGCATTACTTACCA
6C          AAAAAAAAAAAAAAAAAAAAAAAA----------------------------------
6B          TATCAAAAAATACAAACTTTTCATGTATAAACAATAGGCCATTTTGCGCATTACTTACCA

            1681      1691      1701      1711      1721      1731
cons_seq    TCTTTCGTGTACGTGTACGTTTAAGATAAGTGAATATAAACGGACAGAAAATGACTTGTA
6A          TCTTTCGTGTACGTGTACGTTTAAGATAAGTGAATATAAACGGACAGAAAATGACTTGTA
6C          ----------------------------------------------------------
6B          TCTTTCGTGTACGTGTACGTTTAAGATAAGTGAATATAAACGGACAGAAAATGACTTGTA

            1741      1751      1761      1771      1781      1791
cons_seq    AACCCGTTAAAATTGAGCAGTGAATTATACAGATGTAACACTGTTCAACTCATTAATAAT
6A          AACCCGTTAAAATTGAGCAGTGAATTATACAGATGTAACACTGTTCAACTCATTAATAAT
6C          ----------------------------------------------------------
6B          AAAAAAAAAAAAAAAAAAAAAAAAAAAA------------------------------

            1801      1811      1821      1831      1841
cons_seq    GGGTTAGAAGTTAGATAGTGAAAAAAAAAAAAAAAAAAAAAAAA
6A          GGGTTAGAAGTTAGATAGTGAAAAAAAAAAAAAAAAAAAAAAAA
6C          ----------------------------------------------
6B          ----------------------------------------------
```

Appendix 12:

ClustalW sequence alignment of deduced amino acid sequences from *Gm*AAP1, *Gm*AAP5 and functionally characterised AAPs from Arabidopsis and different various legumes. *Ricinus communis*, *Rc*; *Vicia faba*, *Vf*; *Pisum sativum*, *Ps*; *Glycine max*, *Gm* and *A*



Appendix 12, continued



Appendix 13: Phylogenetic tree of 44 full-length AAP amino acid sequences.

The tree was constructed from alignments of full-length amino acid sequences deduced from each AAP nucleotide sequence. This analysis was performed using the Phylogeny Interference Package version 3.2 (Felsenstein, 1989) in BioManager by ANGIS (http://www.angis.org.au). *Ricinus communis*, *Rc*; *Vicia faba*, *Vf*; *Pisum sativum*, *Ps*; *Glycine max*, *Gm* and *Arabidopsis thaliana*, *At*; *Brassica napus*, *Bn*; *Solanum tuberosum*, *St*; *Lycopersicon esculentum*, *Le*. The soybean AAPs are indicated by are red circles. Black circles indicate clusters containing sequences from legume species other than soybean. See Appendix 14 for the list of AAP sequences and protein accession numbers used to construct the tree.

Appendix 14: List of sequences used for construction of the AAP phylogenetic tree.

The two first letters in the left column, which are followed by the AAP gene number, indicate the species names as shown in Appendix 13. If no number has been given to the putative AAP gene, the protein accession number is given. Protein accession numbers: *At*AAP1, NP_176132; *At*AAP2, NP_196484; *At*AAP3, NP_177862; *At*AAP4, NP_201190; *At*AAP5, NP_175076; *At*AAP6, NP_199774; *At*AAP8, NP_172472; *Ps*AAP1, AAX56951, *Gm*AAP1, AAK33098; *Vf*AAP1,CAC51423; *Vf*AAP2, CAA70778; *Vf*AAP3, CAC51424; *Vf*AAP4, CAC51425 *Rc*AAP1, CAA07563; *Rc*AAP3, CAA10608; *Bn*AAP1, CAD92449; *Bn*AAP2, CAF22024; *Bn*AAP6, CAD92450; *St*AAP1, CAA70968; *Le*AAP1, AAO013688; *Le*AAP2,AAO13689; *Le*AAP3, AAO13687.

protein sequence	Sequence/ Description				
BnAAP1	MKSFNTDQHG	...	485	...	TYKPFRTIHE
AtAAP8	MDAYNNPSAV	...	475	...	KSYKPFKNLD
BnAAP6	MEKKSMFIEQ	...	481	...	KDFKPFQAPE
StAAP1	MAPEFQKNAM	...	469	...	SCWIDPGTCH
LeAAP1	MGSGAIDDSS	...	465	...	QEFEPFQSRS
OsAAP_BAD53554	MAKDVEMAVR	...	487	...	KHYVPFKTKS
LeAAP2	MADKPHQVFE	...	471	...	DLKVYRPFQS
OsAAP_XP_476538	MGMERPQEKV	...	487	...	AHYKPFSVSS
OsAAP_AAX96051	MGKAAAMEVS	...	476	...	ALKVYRPFSG
OsAAP_XP_476539	MAAAGRTLGC	...	460	...	AHYKPFSVSS
BnAAP2	MGETAAANHH	...	487	...	KVYKPFQSTY
LeAAP3	MGDSTNFAAK	...	476	...	LKVYKPFKFT
OsAAP_ABA95951	MASGQKVVKP	...	475	...	ALKVYRPFSG
OsAAP_AAV24773	MGENGVVASK	...	496	...	KVYKPFATTY
OsAAP_ABA95955	MASGQKVVKP	...	475	...	ALKVYRPFSG
OsAAP_XP_463528	MGENVVGTYY	...	488	...	LKVYRPFKGY
VfAAP4	MVVEKNASNN	...	481	...	KVFKPFKTIY
OsAAP_BAD53557	MDVYLPRTQG	...	459	...	KTYVPFKTRS
OsAAP_BAD37473	MASVDLELGR	...	484	...	AHYVPFKSKL
OsAAP_XP_463785	MLPRSRTLPP	...	518	...	KSYRPFRSTY
OsAAP_ABA96629	MVQIEPLEVS	...	468	...	DFMKFRPFSG
OsAAP_BAD81895	MDVEKVERKE	...	466	...	RHVTIFQTQL
OsAAP_XP_472988	MSLADDLAAV	...	488	...	KAHNPFCWTC
OsAAP_NP_915026	MNKNAAPEDV	...	465	...	KVATPFKTVS
OsAAP_ABA96081	MDRRAVVYDA	...	468	...	KAAAPFKTVN
OsAAP_NP_915029	MDVEKVERKE	...	531	...	RHVTIFQTQL
OsAAP_XP_472826	MALGDGDDGA	...	466	...	IQGLISQKLG
OsAAP_XP_474206	MAPQLPLEVA	...	469	...	VEGVIRKRLG
AtAAP7_NP_197770	MDIKEDDESR	...	467	...	IYGLVGAKFG
AtAAP7_NP_001031934	MDIKEDDESR	...	361	...	VHILENKLPF
AtAAP5	MVVQNVQDLD	...	480	...	KVYKPFQSEF
AtAAP1	MKSFNTEGHN	...	485	...	TYKPFRTMHE
AtAAP3	MVQNHQTVLA	...	476	...	KSYKPFRSEY
AtAAP2	MGETAAANNH	...	493	...	KVYKPFKSTY
AtAAP6	MEKKKSMFVE	...	481	...	LKDFKPFQAP
AtAAP4	MDVPRPAFKC	...	466	...	KVYKPFKTTY
VfAAP1	MTMEEKEEHS	...	475	...	KKYKPFTTDY
VfAAP3	MVENISRTNL	...	486	...	KKYKPFSLEN
PsAAP1	MVVEKNASKN	...	482	...	KVFKPFKTIY
RcAAP1	MVENTAAKNH	...	486	...	KSVKPFQTSY
GmAAP5	MDVELAAKSV	...	470	...	KKYKLFMYKC
RcAAP3	MPPQAGSKWY	...	466	...	KTVKPFQTTY
VfAAP2	MLQRSRTLPS	...	509	...	KTYKPFKTSY
GmAAP1	MLPRSRTLPS	...	513	...	KTYKPFKTSY