ELUCIDATION OF A PERIBACTEROID MEMBRANE-BOUND bHLH TRANSCRIPTION FACTOR REQUIRED FOR LEGUME NITROGEN FIXATION

A THESIS SUBMITTED BY

PATRICK CHARLES LOUGHLIN

SCHOOL OF AGRICULTURE, FOOD AND WINE

UNIVERSITY OF ADELAIDE

NOVEMBER 2007

1. Introduction

1.1 INTRODUCTION

Despite being one of the most abundant elements on earth, nitrogen is the most commonly limiting nutrient for plant growth. Most plants acquire nitrogen through their roots from soil ammonium (NH_4^+) and nitrate (NO_3^-) pools (Williams and Miller, 2001). Alternatively, many plant species from the family *Leguminosae* (legumes) form a symbiosis with a group of nitrogen-fixing bacteria, known collectively as rhizobia, from the genera *Rhizobium, Bradyrhizobium, Sinorhizobium, Mesorhizobium* and *Azorhizobium*. The rhizobia typically form a non-pathogenic infection in the legume root cortical cells, in a novel organ called the nodule. The plant-bacteria relationship is mutually beneficial, with the plant providing nutrients including photosynthate to the rhizobia in exchange for fixed nitrogen. This nutrient exchange occurs across the plant-derived peribacteroid membrane (PBM). Atmospheric N_2 makes up 79% (v/v) of air and diffuses into the nodule where the rhizobia reduce the N_2 to NH_3 as follows:

$$N_2 + 8H^+ + 8e^- + 16ATP \xrightarrow{\quad nitrogenase \quad} 2NH_3 + H_2 + 16ADP + 16P_i$$
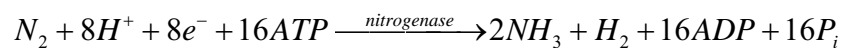
The energy requirement for this N_2 reduction is provided by photosynthate from the host plant and is catalysed by the oxygen labile rhizobium-encoded nitrogenase enzyme.

The industrial production of reduced N fertiliser using the Haber process is both economically and environmentally costly. As oil reserves diminish and demand for N fertiliser increases, production costs will also increase. Environmental concerns also need to be addressed; N-fertiliser production contributes to air and water pollution, and over-application of the fertiliser causes nitrification of the soil and waterways (Vance, 2001). Extending the symbiosis forming ability of nitrogen fixing rhizobia to non-legumes, particularly the agriculturally important *Graminaceae* e.g., rice, wheat and maize, may help alleviate these problems. There are many examples of mutualistic plant-nitrogen fixing microbe associations (Hirsch et al., 2001; Yanni et al., 2001), including interactions between actinorhizal plant/*Frankia* and cycad/cyanobacteria (Vessey et al., 2005) however the tropical tree *Parasponia* (originally misidentified as *Trema*) is the only known non-legume which forms a symbiosis with rhizobia.(Trinick, 1973). Before even the first steps toward design of a rhizobium-*Graminaceae* symbiosis are possible, a better understanding of the interaction between the rhizobium and legume is required.

1.2 NODULATION

1.2.1 SIGNAL EXCHANGE AND NODULE DEVELOPMENT

Nodulation begins with an intricate signal exchange between the legume and bacteria. The legume roots secrete flavonoids and other signalling molecules, into the rhizosphere which attract rhizobia (Peters et al., 1986; Stougaard, 2000) and induces rhizobial expression of nodulation genes (Nod genes), eliciting synthesis and

secretion of specific lipo-chitooligosaccharides or 'Nod factors' (Lerouge et al., 1990). These Nod factors are recognised by 'Nod factor' receptors on legume root hair cells in the susceptible zone behind the root tip (see below). In general, particular species or strains of rhizobia will only form a symbiosis with specific legume species, due in part to legume flavonoid/rhizobia Nod factor recognition (Stougaard, 2001). However there are examples of exceptionally promiscuous strains of rhizobia (Pueppke and Broughton, 1999).

In recent years, research into the molecular identity of this Nod factor receptor(s) and downstream signalling elements in the legume have intensified and with the adoption of the model legume species *Lotus japonicus* and *Medicago truncatula* and their genomic sequencing projects (Young et al., 2005), the signalling pathway is rapidly being elucidated. Numerous nodulation mutants have been identified with arrested nodule formation at various points in development. Initial Nod factor recognition is thought to be mediated by two LysM-type serine/threonine receptor kinases, NFR1 (putatively MtLYK3 in *M. truncatula*) and NFR5 in *L. japonicus* (Madsen et al., 2003; Radutoiu et al., 2003; Smit et al., 2007). Interestingly, although MtLYK3 is necessary for the earliest responses to rhizobial infection, including root hair curling, it is not required for induction of the nodulation associated putative transcription factor NIN (Smit et al., 2007). It is likely that rather than a linear molecular/biochemical signal pathway from Nod factor recognition to nodule development, a number of pathways exist, culminating in the formation of a functional nodule. Recently it has been shown that expression of the NFR1 and NFR5 receptors from *L. japonicus* in *M. truncatula* and *Lotus filicaulis* extends their host range to include rhizobia which do not normally infect these plants (Radutoiu et al.,

2007). Furthermore the authors demonstrate that the LysM domains from NFR1 and NFR5 mediate Nod factor recognition. These transmembrane kinases most likely form heterodimers in the plasma membrane and upon Nod factor recognition, initiate a phosphorylation cascade which in turn stimulates a calcium signalling pathway.

Within seconds of Nod factor recognition, the plasma membrane of the receptive root hair cell depolarises and intracellular calcium levels fluctuate (called 'calcium spiking') (Felle et al., 1998; Cardenas et al., 2000; Shaw and Long, 2003). Calcium is released from intracellular stores to the cytoplasm and nucleus. It was initially believed the ion-channel-like protein DMI1 from *M. truncatula* was responsible for this calcium release (Ana et al., 2004) however it now appears unlikely that DMI1 is directly responsible for mediating the calcium oscillations, but may indirectly regulate channel(s) which do mediate this calcium flux (Peiter et al., 2007). This calcium signalling is thought to be transduced into a transcriptional output via the nuclear localised (Kalo et al., 2005) calcium- calmodulin- dependent protein kinase (CCaMK) DMI3 in *M. truncatula* (Mitra et al., 2004). Recently, the central role of Ca^{2+} signalling in nodule formation has unequivocally been demonstrated, with the identification of autoactive mutant forms of CCaMKs in both *M. truncatula* and *L. japonicus* leading to spontaneous nodulation independent of rhizobia or Nod factor application (Gleason et al., 2006; Tirichine et al., 2006). In *M. truncatula*, downstream of this calcium signalling, two members of the GRAS family of transcriptional regulators NSP1 and NSP2 have been identified (Kalo et al., 2005; Smit et al., 2005), and more recently an ERF transcription factor ERN has also been characterised (Middleton et al., 2007).

This signalling pathway induces curling of a susceptible root hair around the rhizobia, and the de-differentiation of adjacent root cortical cells. As the root hair curls around the rhizobia, a cell wall ingrowth called the infection thread forms. The rhizobia enter the nodule primordium via the infection thread, which fuses with the plasma membrane of a target cell. The bacteria are released into the plant cytosol enclosed in a plant-derived membrane called the peribacteroid membrane (PBM). Inside this membrane, the rhizobia differentiate into obligate symbionts called bacteroids and begin N_2 fixation. The organelle-like structure consisting of the bacteroid, surrounded by the PBM, along with the intervening matrix (the peribacteroid space or PBS), is referred to as the symbiosome (Figure 1; Roth et al., 1988). Prior to differentiation into bacteroids, the rhizobia proliferate in the infected cell which requires co-ordinated PBM synthesis by the host plant. Depending on the legume species, each symbiosome may contain 1 to 20 mature bacteroids, and up to 20,000 symbiosomes per cell (Whitehead and Day, 1997). The central vacuole of the infected cell breaks down to accommodate the proliferating bacteroids, and the cell expands to at least twice the size of uninfected nodule cells.

1.3 THE PERIBACTEROID MEMBRANE

1.3.1 BIOGENESIS OF THE PERIBACTEROID MEMBRANE

The peribacteroid membrane (PBM) serves a unique purpose in the plant cell, regulating the exchange of nutrients between the plant cytosol and the peribacteroid space (PBS). The PBM is originally derived from the end of the infection thread, as the rhizobia are endocytosed into the cell, essentially being an 'inside-out' plasma

membrane. However with the proliferation of the bacteroids within the infected cell, a co-ordinated synthesis of phospholipid membrane and associated proteins is initiated and the PBM becomes a biochemically distinct membrane. The area of new membrane required to accommodate the bacteroids in a single cell is estimated to be 21,500 μm^2, compared with 2,800 μm^2 for the plasma membrane (Roth and Stacey, 1989). Just how the plant cell targets the trafficking of the unique protein and phospholipid complement to the PBM is unclear, however a number of studies are beginning to shed light on this phenomenon. For example, two nodule expressed GTP-binding Rab proteins, Rab1 and Rab7 have been implicated in the trafficking of phospholipid vesicles to the PBM (Cheon et al., 1993). Homologues of these Rab proteins from yeast are required for correct vesicle trafficking and normal ER and Golgi apparatus structure (Spang, 2004). Antisense lines of *rab1* (from soybean) and *rab7* (from *Vigna acontifolium*), show abnormal PBM development (Cheon et al., 1993). Recently proteomic studies have identified various putative vesicle trafficking-related proteins on the PBM of pea and *L. japonicus* (Saalbach et al., 2002; Wienkoop and Saalbach, 2003), and the syntaxin, MtSYP132, from the PBM of *M. truncatula* (Catalano et al., 2004), which has more recently been localised using fluorescence microscopy to both the PBM and the infection thread membrane (Catalano et al., 2007). Syntaxins are commonly situated at a 'target' membrane, conferring a biochemical identity to the membrane, allowing correct targeting of cargo vesicles to their destination. It will be interesting to see whether MtSYP132 is essential for correct PBM biogenesis in *M. truncatula* and orthologous syntaxins in other legumes.

1.3.2 PROTEIN TRAFFICKING TO THE PBM

Prior to bacterial release, PBM-specific proteins are expressed in the nodule primordium (Morrison and Verma, 1987; Mitra et al., 2004), presumably in preparation for insertion into the PBM. Proteins are thought to be trafficked to the PBM in a number of ways. Some proteins are translated on cytosolic ribosomes and then trafficked to the PBM (Simonsen and Rosendahl, 1999). Other proteins on the PBM appear to be of bacteroid origin (Simonsen and Rosendahl, 1999; Catalano et al., 2004), however it is difficult to be certain this result is not due to bacteroid contamination. Most research has examined the role of vesicle trafficking of PBM proteins, through the ER and Golgi (see Whitehead and Day, 1997, and papers therein), however a direct ER to PBM route for vesicle trafficking has also been proposed (Catalano et al., 2004). A cleavable, N-terminal signal sequence and ER vesicle-mediated trafficking to the soybean PBM have been determined for the PBM protein NOD24 (Cheon et al., 1994). However this signal sequence is clearly not universal for PBM proteins, as the PBM protein NOD26 has no homologous signal sequence. Interestingly though, when expressed in tobacco, soybean NOD26 is trafficked to the vacuole (Zhang and Roberts, 1995). An N-terminal signal peptide is also predicted for the PBM zinc transporter GmZIP1 from soybean (Moreau et al., 2002).

1.4 FUNCTIONALLY CHARACTERISED PBM ENZYMES AND CARRIERS

1.4.1 H^+-ATPASE AND OTHER PBM-ASSOCIATED ENZYMES

Biochemical and biophysical studies indicate the activity of a number of enzymes on the PBM including a calcium-dependent protein kinase (Weaver et al., 1991), Fe(III)-

chelate reductase (LeVier et al., 1996), and H^+-ATPase (Blumwald et al., 1985). This Mg^{2+}-dependent H^+-ATPase is one of the best characterised PBM-associated enzymes (Fedorova et al., 1999). The ATPase pumps protons across the PBM, maintaining an acidic environment in the PBS. This is thought to provide energy for the secondary transport of nutrients across the membrane. Vanadate inhibition sensitivity, stimulation by cations, and similar immunological characteristics all suggest the PBM ATPase is closely related to plasma membrane ATPases (Blumwald et al., 1985; Udvardi and Day, 1989; Fedorova et al., 1999). Recent proteome studies support the presence of P-type ATPases on the PBM, however, in contrast to Fedorova et al., (1999) and others, also demonstrate the presence of V-type ATPases (Saalbach et al., 2002; Wienkoop and Saalbach, 2003).

1.4.2 NOD26

NOD26 was one of the first PBM-specific proteins to be cloned from soybean (Fortin et al., 1987). It is a member of the aquaporin or major intrinsic protein (MIP) family, members of which have been implicated in the transport of a range of molecules including H_2O, glycerol, CO_2, and boron (Tyerman et al., 2002). It has been speculated that NOD26 functions as an NH_3/NH_4^+ channel, a malate channel, or an osmoregulator of the PBM. The high density of NOD26 on the PBM agrees well with it being responsible for the NH_4^+ channel activity described by Tyerman et al., (1995) (see below). Although the physiological significance of NH_3 permeation across the PBM is uncertain, (the acidic environment of the PBS favours NH_4^+ in the $NH_4^+ \sqrt{}$ NH_3 equilibrium, with the pKa of $NH_4^+ = 9.24$), a reduction in NH_3 permeation across the PBM is observed in the presence of the aquaporin inhibitor $HgCl_2$ (Niemietz and

Tyerman, 2000). The possibility of NOD26 functioning as a channel for malate, the probable reduced carbon source exported to the symbiosome (see below), has also been explored. When reconstituted in artificial lipid bilayers, NOD26 mediates transport of small anions (including malate) and cations, with a slight preference for the former (Weaver et al., 1994) and phosphorylation of NOD26 leads to an increase in malate uptake by intact symbiosomes (Ouyang et al., 1991). In contrast to artificial bilayer investigations, *Xenopus* oocyte expression studies demonstrate NOD26 mediates transport of H_2O and small neutral molecules (Dean et al., 1999). Furthermore, phosphorylation of NOD26 *in vivo* is regulated by osmotic stress (Guenther et al., 2003). It may be that NOD26 has a multi-functional role in the PBM.

1.4.3 PBM ION TRANSPORT

PBM-enclosed bacteroids are reliant on their legume host for almost all nutrients, including metallic ions. PBM iron transport has received a lot of attention owing to the fact that iron is an essential constituent in a number of bacteroid enzymes including nitrogenase and electron transport chain proteins. Both ferric (Fe^{3+}) (Moreau et al., 1995; LeVier et al., 1996) and ferrous (Fe^{2+}) iron (Moreau et al., 1998) are transported across the PBM. A nodule-specific ferrous iron transporter (GmDMT1) located on the soybean PBM has been cloned (Kaiser et al., 2003) and is a probable candidate for ferrous iron transport detected previously (Moreau et al., 1998). In a similar study, a PBM-specific Zn^{2+} transporter has also been identified in soybean (GmZIP1) and data suggest that it is at least partially responsible for symbiosome Zn^{2+} uptake (Day and Copeland, 1991; Moreau et al., 2002).

The homologous inorganic anion transporters LjNod70 and GmNod70 from *L. japonicus* and *G. max* respectively show expression predominantly on the PBM and transport studies in *Xenopus* oocytes suggest they preferentially transport nitrate and nitrite (Szczyglowski et al., 1998; Vincill et al., 2005). The authors suggest the presence of this anion/nitrate transporter on the PBM may have a role in collapsing the $\Delta\Psi$ generated by the H^+/ATPase across the PBM. This in turn would reduce dicarboxylate and ammonium transport and eventually stop the symbiotic interaction if the nitrate concentration reached a certain threshold in the plant cytosol (Vincill et al., 2005). A sulfate transporter identified in a proteomics study of the PBM of *L. japonicus* (Wienkoop and Saalbach, 2003) called LjSST1 has been shown to be essential for an effective symbiosis to occur, suggesting a role for this transporter in providing sulfur to the bacteroid (Krusell et al., 2005).

1.4.4 EXCHANGE OF FIXED CARBON AND NITROGEN ACROSS THE PBM

Fixed carbon-nitrogen exchange between symbionts is the major nutrient exchange across the PBM. Consequently, many PBM studies have focussed firstly on identifying the transported forms of fixed carbon and nitrogen, and secondly, characterising the mechanism of this transport across the PBM.

1.4.4.1 CARBON FLUX ACROSS THE PBM

Bacteroids require large amounts of energy and reducing power for the reduction of dinitrogen to ammonium. This means a relatively high rate of transport of fixed carbon across the PBM and bacteroid membranes is required. Possible candidates for the transported carbon species include hexose sugars, amino acids, and

dicarboxylates. In soybean, neutral hexose sugars diffuse slowly across the PBM, at an insufficient rate to support the rapidly respiring bacteroid (Day and Copeland, 1991). Soybean studies demonstrate the PBM is relatively impermeable to glutamate (Udvardi et al., 1988), however this may or may not be the case in pea (see Appels and Haaker 1991; Rosendahl et al., 1992). Three lines of evidence support the role of dicarboxylates and in particular malate, as the major transported carbon species across the PBM. Firstly, malate concentration in nodules is high (4-7 μmol/gfw) and nodule enhanced enzymes involved in malate synthesis have been identified (Miller et al., 1998). Secondly, dicarboxylates are transported across both the peribacteroid and bacteroid membranes of all symbiosomes examined (Herrada et al., 1989; Yang et al., 1990; Ouyang and Day, 1992) and a carrier with a preference for dicarboxylates has been characterised on soybean PBM (Udvardi et al., 1988). Finally, rhizobia mutants incapable of utilising hexose sugars form a normal symbiosis, however dicarboxylate uptake mutants do not (Finan et al., 1983; Vance and Heichel, 1991). More recently a member of the peptide transporter family, AgDCAT1, which is localised to the symbiotic interface between *Alnus glutinosa* and its actinomycete symbiont Frankia, has been shown to transport dicarboxylates including malate (Jeong et al., 2004). Whether a similar transporter is responsible for malate transport across the PBM remains to be determined.

1.4.4.2 NITROGEN FLUX ACROSS THE PBM

The bacteroid provides the host plant with fixed nitrogen in exchange for plant derived nutrients. Ammonium is probably the major transported nitrogenous compound across the PBM (Udvardi and Day, 1997), although it should be noted that

some studies put forward other nitrogenous compounds including alanine (Waters et al., 1998) or aspartate (Appels and Haaker, 1991). NH_3 is produced by the reduction of N_2 by nitrogenase and diffuses through the bacteroid membrane into the acidic PBS and is protonated. Acidity of the PBS is maintained by an ATPase proton pump (see above) and the bacteroid electron transport chain. Re-uptake of the NH_4^+ by bacteroids is prevented by repression of its ammonium transport system (Tate et al., 1998). NH_4^+, and possibly some NH_3, is then transported across the PBM to the plant cytosol where it is rapidly assimilated by glutamine synthetase (Day et al., 2001). As a consequence of this rapid assimilation and the action of the PBM ATPase pump, a chemical and voltage gradient across the PBM is generated, allowing the facilitated passive transport of NH_4^+ from the PBS to the plant cytosol.

1.4.4.3 CHARACTERISATION OF THE PBM AMMONIUM CARRIER

Using a patch clamp technique, an ammonium channel has been described in intact soybean symbiosomes (Tyerman et al., 1995; Obermeyer and Tyerman, 2005) and also in pea symbiosome membrane vesicles (Mouritzen and Rosendahl, 1997). The channel is selective for small monovalent cations, and in particular NH_4^+, with a $K_m=$ 37.5 mM, which agrees with the predicted NH_4^+ concentration in the PBS (Streeter, 1989). Unidirectional channel activity is dependent on Mg^{2+} and Ca^{2+} concentration, whereby high concentration of either cation on a particular side of the PBM inhibits NH_4^+ flow from that side (Whitehead et al., 1998). The authors suggest that the cytosolic Mg^{2+} concentration is such that the Mg^{2+} blocks back flow of NH_4^+ to the PBS, and that Ca^{2+} in the PBS may fine tune the regulation of NH_4^+ flux. Ca^{2+} and Mg^{2+} also inhibit NH_4^+ permeability of a similar channel described on the *L. japonicus*

PBM, however this channel is also slightly permeable to Ca^{2+}, suggesting a secondary role as a Ca^{2+} channel (Roberts and Tyerman, 2002). It has been suggested that in soybean, NOD26 (see above) or another non-selective cation channel (Whitehead et al., 2001) may be the NH_4^+ channel described here. Another possible candidate, GmSAT1, is described below.

1.5 GmSAT1

1.5.1 FUNCTIONAL CHARACTERISATION OF GmSAT1

GmSAT1 (*Glycine max* symbiotic ammonium transporter 1) was isolated from a soybean nodule cDNA library and initially characterised as an ammonium transporter (Kaiser et al., 1998). It was isolated based on its ability to complement growth of the yeast ammonium transport mutant 26972c on 1 mM ammonium media. Upon GmSAT1 expression, enhanced methylammonium uptake and a time dependent cation flux was measured in yeast and yeast spheroplasts respectively. GmSAT1 is located on the soybean PBM (Kaiser et al., 1998) and similar cation transport kinetics and response to external Mg^{2+} and Ca^{2+} regulation when expressed in yeast make it a possible candidate for the previously described NH_4^+ channel (Tyerman et al., 1995). GmSAT1 does not share homology with any known ammonium transporter (Howitt and Udvardi, 2000). It is predicted to have a single transmembrane domain, and a transcription factor motif which raises speculation about whether GmSAT1 is an ammonium transporter as such.

1.5.2 GmSAT1 AND YEAST AMMONIUM TRANSPORT

S. cerevisiae have three high affinity ammonium transporters; Mep1, Mep2 and Mep3, with ammonium binding constants of 5-10 μM, 1-2 μM and 1.4-2.1 mM respectively (Marini et al., 1994; Marini et al., 1997). The function of any one of these transporters allows yeast growth on 1 mM ammonium media. Yeast strain 26972c, which was used to isolate Gmsat1, has a mep2 deletion (Δmep2) and a point mutation causing a Gly→Asp substitution in the C-terminal tail of Mep1 which knocks out its function. This *mep1-1* mutant protein also post-translationally inhibits MEP3, however over-expression of MEP3 overcomes this inhibition (Marini et al., 2000). Complementation of strain 26972c was previously used to isolate the first plant ammonium transporter, AtAMT1;1, from *Arabidopsis* (Ninnemann et al., 1994). Members of the AMT family, for example LjAMT1;1, LjAMT1;2, and LjAMT1;3 from *L. japonicus*, complement the growth of a triple null MEP yeast mutant 31019b (Δ*mep1*Δ*mep2*Δ*mep3*) on 1 mM ammonium media (D'Apuzzo et al., 2004). In contrast, GmSAT1 does not complement growth of this yeast mutant or a MEP3 deletion mutant in a 26972c background, suggesting that GmSAT1 cannot, by itself, transport NH_4^+ (Marini et al., 2000).

1.5.3 IS GmSAT1 A TRANSCRIPTION FACTOR?

GmSAT1 contains a domain homologous to members of the basic helix-loop-helix (bHLH) transcription factor family. Plant members of this large transcription factor family are involved in various processes including flavonoid synthesis, phytochrome signalling, ER structure, photomorphogenesis and cell fate determination of

epidermal, trichome, and stomatal cells (Payne et al., 2000; Fujimori et al., 2004; Matsushima et al., 2004; Broun, 2005; Duek and Fankhauser, 2005; MacAlister et al., 2007; Pillitteri et al., 2007). Phylogenetic analysis of the full length GmSAT1 protein places it in a small subfamily of putative bHLH transcription factors including members from *Arabidopsis*, rice, *M. truncatula*, *V. vinifera* and *M. crystallinum* (Buck and Atchley, 2003; Figure 2-8B). The three *Arabidopsis* homologues are located in tandem on chromosome 2. QTL (quantitative trait loci) analysis has implicated this locus in the response of *Arabidopsis* to growth under varying nitrogen conditions (Rauh et al., 2002). More recently however, it has been demonstrated that one of these *Arabidopsis* GmSAT1 homologues, AtNAI1 (At2g22770) plays a critical role in the formation of ER-associated structures called ER bodies (Matsushima et al., 2004).

Marini *et al.,* (2000) demonstrated that in 26972c, GmSAT1 expression increases MEP3 protein level, overcoming inhibition by *mep1-1*, but suggest it is unlikely GmSAT1 is acting as a transcription factor as it is localised on the plasma membrane of the yeast (Kaiser et al., 1998). However there are many examples of membrane-bound transcription factors (Hoppe et al., 2001) and it has been hypothesised that GmSAT1 may upregulate *mep3* expression in yeast (Dommelen et al., 2001) however this is yet to be tested.

1.6 MEMBRANE-BOUND TRANSCRIPTION FACTORS

Transcription factors are potent regulators of cellular activities and as such are highly regulated themselves. As DNA-binding transcription factors need to be in the nucleus

to function a common form of post-translational regulation is the sequestration of an inactive form of the protein outside of the nucleus. Transcription factors may be anchored in the endomembrane system of a cell, cleaved from the membrane upon a particular signal and translocated into the nucleus to activate or repress target gene transcription (Hoppe et al., 2001). The location of a membrane-bound transcription factor is commonly relevant for its functional role. For example membrane-bound transcription factors which upregulate genes involved in the ER stress response are localised to the ER membrane (Ye et al., 2000; see below). The mammalian sterol responsive element binding protein (SREBP) is the best characterised membrane-bound transcription factor, whose activity is dependent on its cleavage from the Golgi apparatus membrane (Brown and Goldstein, 1999). The SREBP is an ER resident protein with an N-terminal bHLH/leucine zipper (LZ) transcription factor domain and two central transmembrane domains. It is inserted into the membrane such that both its C- and N-termini are cytoplasmic. In response to sterol depletion, SREBP is translocated in COPII vesicles to the Golgi, where it undergoes two stepwise proteolytic cleavage events. Firstly, a Golgi membrane associated subtilisin-like serine protease, termed the site-1 protease (S1P) cleaves after the leucine in an RXXL recognition site in the luminal loop. The N-terminal half of the protein becomes substrate for the zinc metallo-protease termed site-2 protease (S2P), which cleaves within the transmembrane domain of the protein, releasing it from the membrane (reviewed in Espenshade, 2006). S1P/S2P proteolysis has also been shown to be necessary for the activation of the type II membrane-associated basic-ZIP (bZIP) transcription factors ATF6 (Ye et al., 2000; Shen et al., 2002), OASIS (Murakami et al., 2006) and BBF2H7 (Kondo et al., 2007), which are involved in ER stress in mammals. Additionally, it has recently been demonstrated that the ER-stress inducing

agent thapsigargin also induces S1P/S2P proteolysis of SREBP in mammalian cell culture (Colgan et al., 2007).

1.6.2 PLANT MEMBRANE-BOUND TRANSCRIPTION FACTORS

The first evidence that regulated proteolysis of membrane-bound proteins may also occur in plants came with the identification and characterisation of the rhomboid intramembrane serine protease AtRBL2 in *Arabidopsis* (Kanaoka et al., 2005). Furthermore, an ER stress-responsive bZIP transcription factor with a putative C-terminal transmembrane domain, AtbZIP60, has been identified in *Arabidopsis* (Iwata and Koizumi, 2005). Its expression is sufficient to activate expression of ER chaperone proteins such as BiP and calnexin, which alleviate ER stress by enhancing correct protein folding (Iwata and Koizumi, 2005). The localisation of AtbZIP60 to the ER membrane is yet to be definitively proven, however if it is ER membrane-localised it will be interesting to determine how AtbZIP60 is liberated from the membrane and imported into the nucleus. An *Arabidopsis* orthologue of S1P, AtS1P, has been found to cleave an ER-membrane associated bZIP transcription factor, AtbZIP17 under salt stress conditions (Liu et al., 2007). This cleavage allows translocation of the N-terminal part of the protein, which includes the bZIP domain, to the nucleus. The authors point out that the salt stress may in fact be inducing ER stress in the plant, and that this is causing cleavage of AtbZIP17 by AtS1P, in a system analogous to that observed in mammalian cells exposed to ER stress (see above). Interestingly, the GmSAT1 homologue from *M. crystallinum* (accession AF097665) was isolated from a root cDNA library of plants under salinity stress. Recently the *Arabidopsis* membrane localised NAC transcription factor AtNTM1

(NAC with transmembrane motif 1) was also characterised. AtNTM1 was shown to be membrane-associated and inhibitor studies suggest it may be cleaved from the membrane by the proteolytic activity of a calpain protease (Kim et al., 2006). In a recent review of *Arabidopsis* transcription factors, at least thirteen NAC, six bZIP and five bHLH transcription factors were predicted to contain transmembrane domains (Kim et al., 2007). It is becoming apparent that the liberation of the soluble, DNA-binding domain of a transcription factor from its membrane anchor may be a relatively common event in plants.

1.7 CHARACTERISED NODULE TRANSCRIPTION FACTORS

Structurally, GmSAT1 shares similarity with many of these known membrane bound transcription factors – having an N-terminal transcription factor domain and a putative C-terminal membrane anchor. It is feasible to hypothesise that GmSAT1, which has been shown to localise to membrane structures in soybean nodules and yeast (Kaiser et al., 1998) is being cleaved from the membrane and imported into the nucleus to activate gene expression. Its expression in the nodules (and nodulated root tissue) of soybean suggests a role for GmSAT1 in either nodule development or maintenance. To date, few transcription factors have been identified as involved in the nodulation process, despite the obvious importance of novel transcription factors in this process. LjNIN (nodule inception) is predicted to be a membrane-bound transcriptional regulator which is required for early events in *L. japonicus* nodule development (Schauser et al., 1999). NIN knockouts in both *L. japonicus* and pea do not form rhizobia infections and root hairs become hyper-curled forming 'shepherd's crook'-like structures when exposed to Nod factors (Schauser et al., 1999; Borisov et al.,

2003). GRAS family transcriptional regulators NSP1 and NSP2 (Kalo et al., 2005; Smit et al., 2005), have been shown to be important in the initial development of the nodule as has the ERF transcription factor ERN (Middleton et al., 2007). The *L. japonicus* 'astray' mutant, which has a deletion in its bZIP transcription factor LjBzf, has both symbiotic and non-symbiotic phenotypes, with defects in photo- and gravi-tropic responses, and an increased zone of nodulation, leading to around twice the normal number of nodules (Nishimura et al., 2002). Reducing the nodule expression of two *ndx* homeobox genes in *L. japonicus* causes structural defects in the nodules, which although able to actively fix nitrogen, do not provide sufficient fixed nitrogen for plant growth (Gronlund et al., 2003). Described in this thesis is the initial characterisation of a novel soybean nodule transcription factor, GmSAT1.

1.8 PROJECT AIMS

GmSAT1 has previously been localised to the PBM and other internal membranes of soybean infected cells and the plasma membrane in yeast expressing GmSAT1 (Kaiser et al., 1998). It was characterised as an ammonium transporter in the yeast expression system, however further work suggests it probably does not act directly as a transport protein (Kaiser et al., 1998; Marini et al., 2000). Work described in this thesis includes a re-examination of the functional role of GmSAT1 in both yeast and soybean nodules. It has been speculated that the homology of GmSAT1 with known basic Helix-Loop-Helix transcription factors suggests it may act as a DNA binding protein. In this study, mutation of key amino acid residues in the bHLH domain of GmSAT1 did disrupt its function, providing evidence that GmSAT1 acts as a transcription factor. It was also necessary to reconcile the membrane localisation of

GmSAT1 with its putative activity as a DNA-binding protein. A number of approaches were undertaken to examine the localisation of the GmSAT1 protein, both in yeast and soybean. GmSAT1 was tagged with both GFP and an artificial transcription factor to demonstrate the protein is localised to two distinct locations in the yeast cell - punctate bodies around the periphery of cell and the nucleus. Immunogold labelling of rhizobia-infected soybean nodule cells also localised GmSAT1 to the nucleus. Western blotting indicated that rather than the full-length GmSAT1 protein being targeted to both the membrane and nucleus, a proteolytic event takes place to release the soluble N-terminal part of the protein which is imported into the nucleus. Finally, silencing of GmSAT1 in soybean roots causes the development of small, ineffective nodules, with few symbiosomes housed within the infected cells. This suggests GmSAT1 is critical for the development and/or maintenance of the legume-rhizobia symbiosis.
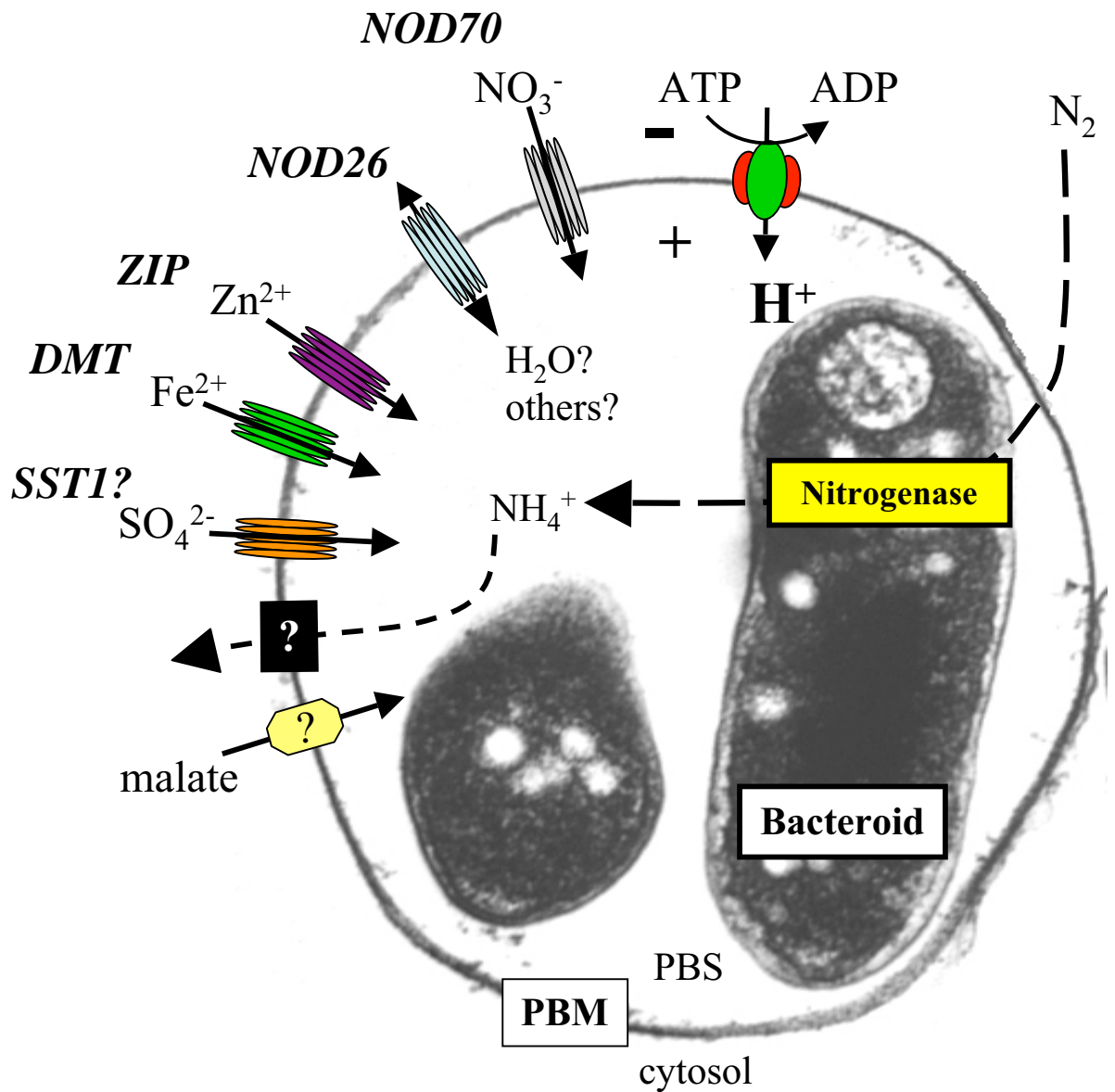
Figure 1-1. The soybean symbiosome. The nitrogen fixing bacteroid is surrounded by the acidic peribacteroid space (PBS), enclosed in the plant-derived peribacteroid membrane (PBM). Characterised transport systems found on the peribacteroid membrane are highlighted. Electron micrograph courtesy of Dr. Dean Price (ANU).

2. A Functional Analysis of GmSAT1 using Site-Directed Mutagenesis

2.1 INTRODUCTION

2.1.1 THE BASIC HELIX-LOOP-HELIX TRANSCRIPTION FACTOR FAMILY

The basic Helix-Loop-Helix (bHLH) family is a well-characterised transcription factor family, with members identified in organisms as diverse as mammals, plants and yeast. In particular mammalian members have been studied intensively, due to their functional importance in processes such as haematopoiesis, neurogenesis and oncogenesis (Massari and Murre, 2000). In plants, bHLH members have been implicated in diverse cellular processes including flavonoid synthesis, phytochrome signalling, ER structure and a number of cellular differentiation processes including epidermal, trichome, and stomatal cell fate (Payne et al., 2000; Matsushima et al., 2004; Broun, 2005; Duek and Fankhauser, 2005; MacAlister et al., 2007; Pillitteri et al., 2007). The bHLH motif comprises a sequence of 10-15 predominantly basic amino acids, which recognise and bind its cognate DNA (Voronova and Baltimore, 1990), followed by a pair of parallel α-helices interrupted by a loop of variable length. These helices are involved in the homo- and/or hetero-dimerisation of these proteins, which is a prerequisite for DNA binding.

2.1.2 THE FUNCTION OF GmSAT1 AND STRUCTURAL ASPECTS OF ITS bHLH MOTIF

GmSAT1 was initially isolated from a soybean nodule cDNA library based on its ability to complement growth of the ammonium transport deficient yeast strain

26972c on media where 1 mM ammonium was the sole nitrogen source (Kaiser et al., 1998). Similar to ammonium transporters of the AMT/MEP family, GmSAT1 also induced toxicity in 26972c when grown in the presence of 100 mM methylammonium (MA), a toxic ammonium analogue, which is caused by a measurable enhancement in the accumulation of MA (Kaiser et al., 1998). Unlike other high affinity ammonium transporters, which typically have 10-12 transmembrane domains, GmSAT1 has a single predicted C-terminal transmembrane domain. Despite its unusual structure, its apparent capacity for ammonium flux and its plasma membrane localisation (Kaiser et al., 1998) suggested a direct role for GmSAT1 in ammonium uptake. GmSAT1 also has a predicted bHLH motif, indicative of it being a transcription factor. As such, I examined this conserved bHLH domain of the protein through site-directed mutagenesis.

The crystal structures of six bHLH transcription factors have been solved, five mammalian members and one, Pho4p, from *S. cerevisiae* (Ferredamare et al., 1993; Ellenberger et al., 1994; Ferredamare et al., 1994; Ma et al., 1994; Brownlie et al., 1997; Shimizu et al., 1997; Parraga et al., 1998). These crystal structures of bHLH transcription factors, from a number of sub-families, suggest that the general 3D structure of the motif is conserved across the family. Threading the predicted bHLH primary sequence of GmSAT1 onto known 3D structures of bHLH members (Figure 2-1) along with the high conservation of particular amino acid residues (Figure 2-2) provided a guide for identifying functionally important amino acids to target for substitution in this study. Atchley and Fitch (1997) have devised a nomenclature system for residues in the bHLH motif, which is used in this study when describing mutations made in the bHLH region of GmSAT1 (Figure 2-2). bHLH proteins which

have been shown to bind DNA have a glutamate residue at position 9 of the bHLH motif (Figure 2-2). This glutamate contacts the C and A of the canonical bHLH recognition motif, CANNTG and is stabilised by the presence of an arginine residue at position 12. Other amino acids have been shown to be important for binding specificity within the basic region. For example, GmSAT1 has the sequence H-E-R at positions 5-9-13 of the basic region, which are major determinants for the recognition of the palindromic CACGTG DNA sequence (Figure 2-2; Atchley and Fitch, 1997). Amino acids which determine DNA binding specificity outside of this hexanucleotide-binding motif are less well characterised. The helices of bHLH proteins have a number of highly conserved hydrophobic amino acids which participate in stabilising the structure of the helices by forming a hydrophobic core, both within the monomer and through interactions with its binding partner. Amino acid residues involved in dimer partner specificity are not as well characterised as those important for DNA binding. In some instances domains such as the Per-Arnt-Sim homology (PAS) or Leucine Zipper (LZ) adjacent to the bHLH domain are responsible for choice of binding partner (Chapman-Smith and Whitelaw, 2006), however in other bHLH proteins, including GmSAT1, no recognisable dimerisation motif outside of the bHLH is present. Although the loop region of the bHLH domain is relatively variable and has little secondary structure, loop replacement and mutagenesis studies have shown that the sequence is functionally important (Winston and Gottesfeld, 2000; Ma et al., 2007).

Described in this chapter are a number of amino acid substitutions in the bHLH (Figure 2-2) and transmembrane domains of GmSAT1 (Figure 2-3C), which provide further evidence that GmSAT1 acts as a membrane-bound transcription factor.

Functional activity of GmSAT1 mutants was determined through examination of the growth of 26972c expressing the mutant proteins on 1 mM ammonium and 100 mM MA media. The ability of these mutant GmSAT1 to enhance accumulation of MA in the yeast was determined by ^{14}C-MA tracer experiments, as ^{14}C-labelled MA is a convenient ammonium transport analogue and is commonly used in ammonium transport studies (Hackette et al., 1970). A brief study of the overexpression of endogenous yeast bHLH transcription factors and three nitrogen metabolism-associated proteins is also described in this chapter.

2.2 RESULTS

2.2.1 MUTATIONS IN THE TRANSMEMBRANE DOMAIN OF GmSAT1 HAVE MIXED EFFECTS ON ITS FUNCTION

GmSAT1 is predicted to have a single C-terminal transmembrane domain (TMD) (Figure 2-3A & B), which is at odds with other characterised ammonium transporters of the AMT/MEP family which generally have 11-12 TMDs (Howitt and Udvardi, 2000). It has been suggested that similar to the well characterised single-pass transmembrane cation channel M-2 from the influenza virus (Pinto et al., 1997), GmSAT1 may homooligomerise to form a hydrophilic pore to allow transport of substrate. Supporting this is the presence of a hydrophilic face on the otherwise hydrophobic TMD of GmSAT1 (Figure 2-3B). Prior to work described here, I had determined that GmSAT1 lacking its TMD was non-functional (Section 3.2.1). In a more subtle approach, substitution of two amino acids in the transmembrane domain of GmSAT1 was also performed to examine their role and the functional importance of the transmembrane domain. Aspartate 317 (D317) was substituted for asparagine

and alanine and isoleucine 318 (I318) was mutated to glutamine (Figure 2-3C). D317 was targeted for substitution as the presence of a charged amino acid in the predicted transmembrane domain is energetically unfavourable and so may be functionally important, potentially lining a hydrophilic pore. Mutation of D317→N, which would disrupt the formation of any salt bridges mediated by this residue but would still allow hydrogen bonding, did not affect the ability of GmSAT1 to complement 26972c on low NH_4^+ or accumulate toxic levels of MA (Figure 2-4A & B). The D317→A mutation did disrupt GmSAT1 function (Figure 2-4A & B), however this was due to a reduction in protein translation or stability, as GmSAT1 protein was not detected in yeast expressing the D317A mutant (Figure 2-4C). Ile318 is well conserved in the SAT family (Figure 2-3C) and was also chosen for mutagenesis. Interestingly, mutation of this I318→Q abolished the enhanced accumulation of methylammonium and its associated toxicity, but did not affect complementation of 26972c on low ammonium (Figure 2-4A & 4B). Additionally this mutation reduced the presence of the smaller, 35 kDa GmSAT1-associated product when analysed by immunoblotting (Figure 2-4C).

2.2.2 THE BASIC REGION OF THE bHLH DOMAIN OF GmSAT1 IS ESSENTIAL FOR FUNCTION

Three amino acids, at positions 5, 9 and 13 in the basic region of the bHLH motif are directly involved in the recognition and binding of the DNA sequence CANNTG, with others involved in stabilising the helical structure of this region (Figure 2-2; (Atchley and Fitch, 1997). Initially the highly conserved glutamate (E) at position 9 (E177) of the bHLH domain of GmSAT1 was targeted for mutagenesis (Figure 2-2,

Figure 2-5C). This residue is essentially invariant in bHLH transcription factors as it hydrogen bonds with the C and A residues of the canonical E-box CANNTG DNA sequence. E177 was replaced by three amino acids 1) glutamine (Q), which is of similar size to glutamate and is still able to hydrogen bond 2) aspartate (D) which is one carbon smaller than glutamate but retains a negative charge and 3) alanine (A), which is a small hydrophobic amino acid unable to hydrogen bond. Mutation of E177 to any of these other amino acids completely abolished the ability of GmSAT1 to enhance uptake/accumulation of ammonium and MA, as shown by growth studies and ^{14}C-MA flux experiments in Figures 2-5A & 2-5B. Interestingly the E177A mutation not only abolished complementation of 26972c on low ammonium, but also retained its toxicity on MA medium (Figure 2-5A). It is probable that this mutant form of GmSAT1 is toxic to the yeast, as expression of the E177A mutant caused cell death on media where proline was supplied as a nitrogen source (Fig 2-5A, first column) and no enhancement of methylammonium flux was detected in ^{14}C-methylammonium uptake assays (Figure 2-5B). No GmSAT1 protein was detected by Western blotting in cells harbouring the E177D and E177Q mutants (Figure 2-5D). The conservative mutation of Arg180→Lys (R180K) at position 12 of the GmSAT1 bHLH motif (Figure 2-2 & 2-5C) also abolished the function of GmSAT1 (Figure 2-5A & 5B). Crystal structures of other bHLH proteins demonstrate this Arg hydrogen bonds with the DNA backbone, and helps stabilise the position of Glu 9 (Ma et al., 1994). Unlike the other conservative mutations of E177, the R180K mutation did not disrupt the stability of GmSAT1 (Figure 2-5D), suggesting that this residue is functionally important for GmSAT1.

2.2.3 MUTATION OF HYDROPHOBIC RESIDUES IN THE HELICES OF THE bHLH DOMAIN ALSO AFFECT GmSAT1 FUNCTION

Contiguous with the basic region of bHLH proteins is the first helix, which is often broken by a proline residue at position 28, marking the beginning of the variable loop region. Helix 2 begins with a well-conserved lysine residue in many bHLH members (arrow heads, Figure 2-2). Both Helix 1 and 2 are amphipathic, with dimerisation of bHLH proteins at least in part being mediated by the hydrophobic core formed through interactions of the hydrophobic faces of the helices. Within helix 1, leucine (L) at position 23 (L191 in GmSAT1) is completely conserved across the *Arabidopsis* bHLH family (Toledo-Ortiz et al., 2003) and in identified animal members, conservation is also very high (Atchley et al., 1999). In this study, L191 was replaced by three hydrophobic residues – isoleucine (I), valine (V) and alanine (A) (Figure 2-2 & 2-6C). Isoleucine is of similar size to leucine, with valine being one carbon smaller and alanine being three carbons smaller than leucine. Mutation of L191 had some unexpected consequences for the function of GmSAT1. Replacement with either alanine (L191A) or valine (L191V) disrupted the ability of GmSAT1 to complement 26972c on 1 mM NH_4^+ and abrogated sensitivity of the yeast to MA (Figure 2-6A). The mutation of L191A decreased the stability of the protein as Western blots demonstrate some degradation of GmSAT1 with this mutation, however mutation L191V did not affect stability of the protein (Figure 2-6D). Mutation L191I led to a perplexing phenotype whereby complementation of 26972c on low ammonium media was not affected however the toxic effect of MA was considerably reduced (Figure 2-6A). ^{14}C-MA uptake data suggest that yeast expressing the L191I mutant GmSAT1 does not accumulate MA at 1 mM as observed with wild-type GmSAT1 (Figure 2-6B). This separation of the ammonium

and MA phenotypes, similar to that observed in the I191→Q mutation in the TMD (see section 2.2.1) was somewhat surprising and suggests two distinct mechanisms causing the observed phenotypes.

In helix 2 of the bHLH domain, two amino acids were chosen for substitution – leucine at position 54 (L207) and methionine at position 64 (M217) (Figure 2-2 & 2-7C). Conservative replacement of Leu207→Val (L207V) completely abolished GmSAT1 function (Figure 2-7A & 7B), without affecting protein stability (Figure 2-7D). This result reinforces the importance of specific hydrophobic packing between helices, as valine is only one carbon smaller than leucine. In a survey of *Arabidopsis* bHLH members, Toledo-Oritz et.al., (2003) found that 93% of bHLH proteins had a leucine at position 64. GmSAT1 has a Met residue at this position (M217), which is conserved in two of the SAT members identified in the legume species *M. truncatula,* but not in SAT homologues identified in non-legume species (Figure 2-8). This residue was also found to be of functional significance in GmSAT1, with mutation of this methionine to leucine (M217L), abolishing GmSAT1 activity (Figure 2-7A & 7B). However this is most likely due to reduced protein stability (Figure 2-7D)

2.2.4 OVEREXPRESSION OF ENDOGENOUS YEAST bHLH TRANSCRIPTION FACTORS AND NITROGEN METABOLISM-RELATED PROTEINS DO NOT MIMIC GmSAT1 EXPRESSION

One interpretation of the functional sensitivity of GmSAT1 to mutations in its predicted bHLH domain is that it acts as a bHLH transcription factor in the yeast 26972c to allow growth on 1 mM NH_4^+ and induce toxicity on 100 mM MA. *S. cerevisiae* have eight endogenous bHLH transcription factors which have been

characterised to various degrees (Robinson and Lopes, 2000; Chen and Lopes, 2007), as well as the ScNoc3 protein, which shares homology with bHLH proteins but appears to be involved in rRNA processing (Milkereit et al., 2001) and the initiation of genomic DNA replication (Zhang et al., 2002) and the dubious open reading frame (ORF) YGR290W (Chen and Lopes, 2007). It is clear that although the functions of these yeast bHLH proteins are not redundant and that target genes have distinct promoter elements (e.g., Fisher and Goding, 1992), there are also networks of these factors heterodimerising and working in concert (Robinson et al., 2000; Chen and Lopes, 2007). Although there is little overall homology between GmSAT1 and yeast bHLH proteins, within the bHLH domain there is some similarity (Figure 2-9).

To examine the possibility that the overexpression of other bHLH transcription factors may present a similar phenotype to GmSAT1, 26972c was transformed with vectors from a yeast ORF collection (Open Biosystems), harbouring four known yeast bHLH transcription factors, ScCbf1, ScIno4, ScPho4, and ScRtg3, as well as the bHLH-like protein ScNoc3 and the dubious ORF YGR290W, expressed under the Gal1 promoter (Figure 2-10). Overexpression of tagged ScPho4, ScCbf1, ScRtg3 and ScNoc3 was toxic to the yeast when grown on minimal media with proline as a nitrogen source (Figure 2-11). Overexpression of ScIno4p and YGR290W on the other hand, were not toxic to the yeast and although their overexpression was not able to complement 26972c growth on low ammonium, both induced a MA associated toxicity (Figure 2-11). This result was explored further through ^{14}C-MA uptake assays of 26972c yeast expressing ScIno4 under a Gal1 promoter. Contrary to growth studies on 100 mM MA media, yeast overexpressing ScIno4 exhibited

similar levels of MA accumulation as empty vector control yeast, when exposed to 1 mM MA for 10 min (Figure 2-11, last column).

GmSAT1 may act indirectly to allow growth of 26972c on low ammonium and induce toxicity on MA. As well as mimicking other endogenous bHLH transcription factors, GmSAT1 may potentially interact or mimic other transcription factors, the most likely of which being those involved in maintaining nitrogen balance within the yeast cell. Nitrogen catabolite repression (NCR) refers to the ability of *S. cerevisiae* to repress expression of genes involved in the uptake and catabolism of non-preferred nitrogen sources e.g., urea or proline, when grown in the presence of a rich nitrogen source e.g., high concentration ammonium or glutamine (Magasanik and Kaiser, 2002). ScGln3 and ScGat1 are both zinc-finger transcription factors, which are positive regulators of NCR-sensitive genes (Magasanik and Kaiser, 2002). Overexpression of these transcription factors, as well as the inositol/choline-repressible (Santiago and Ben Mamoun, 2003), NCR-sensitive urea/polyamine transporter ScDur3 (Uemura et al., 2007) was examined as for the yeast bHLH proteins described above. Overexpression of ScGln3 was toxic to yeast grown on minimal media when supplied with proline as a nitrogen source (Figure 2-11) and both ScDur3 and ScGat1 were partially toxic under these conditions, and completely toxic when grown in the presence of 100 mM MA (Figure 2-11). Similar to the yeast bHLH protein ScIno4 described above, yeast overexpressing ScGat1 exhibited no enhancement of MA accumulation in ^{14}C-MA uptake experiments performed at 1 mM MA concentration (Figure 2-11, last column).

2.3 DISCUSSION

GmSAT1 is an intriguing protein, having both a putative DNA binding domain as well as a predicted transmembrane spanning domain. Initially GmSAT1 was characterised as an ammonium transporter (Kaiser et al., 1998), however this was brought into question due to the inability of GmSAT1 to complement an ammonium transport yeast mutant 31019b, which has all three high affinity ammonium transporters deleted (Marini et al., 2000). Furthermore the unusual structure of GmSAT1 suggests it may have a transcriptional regulation role, rather than direct ammonium transport activity. Described in this chapter are a number of amino acid substitutions made in the predicted DNA binding and dimerisation domains of the bHLH region and in the C-terminal TMD of the protein.

Although the exact mechanism of how GmSAT1 modifies the ammonium transport activity in 26972c remains unclear its potential as a putative transcription factor is tenable based on its conserved bHLH motif. Results presented in this chapter suggest that the putative DNA binding domain of GmSAT1 is essential for its activity in complementing 26972c with the C-terminal transmembrane domain also being important for its function. All mutations, even conservative ones, in the basic region of GmSAT1 abolished function either through reducing the stability of the protein as is the case for E177D and E177Q, or through other unresolved means in the case of E177A and R180K. The mutation of E177A was the only mutant form of GmSAT1 which was toxic to the yeast when grown on minimal medium supplemented with proline as a nitrogen source. Potentially the other two mutations

of Glu177 (E177D and E177Q) may also have produced a toxic protein, however these proteins were unstable (Figure 2-5D).

The importance of the helix-loop-helix motif of GmSAT1, which potentially mediates dimerisation of the protein was also demonstrated here. Mutation of the conserved hydrophobic residues in the two helices of the bHLH domain caused disruptions in the function of GmSAT1. Of particular interest was the mutation of L191I. Leucine at this position has been shown to participate in the dimerisation of the mammalian bHLH protein Max (Brownlie et al., 1997). Functionality of the GmSAT1 protein was measured based on its ability to allow growth of 26972 on 1 mM ammonium and to induce toxicity on 100 mM of the toxic ammonium analogue MA. The L191I mutation of GmSAT1 disrupted the MA toxicity phenotype without affecting growth of the yeast on 1 mM ammonium (Figure 2-6A). This phenotype may be due to a number of reasons. If GmSAT1 is heterodimerising with a number of different partners, it may be that this mutation prevents GmSAT1 from interacting with one of these partners without affecting the other. This may have downstream consequences, leading to the retention of one phenotype and the loss of the other. There is the possibility that this mutation is affecting some other process that the dimerisation helices are involved in, mediating interaction with proteins at the membrane for example, which may induce the MA transport and toxicity. It is interesting to note that mutation of I318, in the transmembrane domain of GmSAT1 has a similar phenotype to the L191I mutation. The results may also demonstrate an intermediate phenotype. The L191I mutation may destabilise the protein enough to disrupt its ability to induce toxicity on MA, but not enough such that it prevents ammonium complementation. Supporting this is the fact that there is a intermediate

growth inhibition of cells expressing the L191I mutant GmSAT1 grown on 100 mM MA when compared to the empty vector control (Figure 2-6A) although the level of MA accumulation measured in ^{14}C-MA uptakes failed to show anything different to the empty vector control (Figure 2-6B). The L191I mutation does highlight an important observation that the mechanism most likely involved in the ammonium complementation does not contribute to the ^{14}C-MA uptake, which is routinely observed with the expression of GmSAT1. This result is compelling in that it demonstrates that the complementation of growth on 1 mM ammonium by GmSAT1 and the mechanisms involved in this process are most likely not responsible for accumulation of ^{14}C-MA or the observed toxicity phenotypes on 100 mM MA. It is unclear what mechanism may be at play, such that mutations in different domains of the protein i.e., the bHLH and transmembrane domains, cause this phenotype.

At position 64 in the second helix of bHLH proteins, 93% of *Arabidopsis* members have leucine, including the 3 closest orthologues of GmSAT1, At2g22750, At2g22760 (AtNai1) and At2g22770. In GmSAT1 and two *M. truncatula* homologues, MtSAT1;1 and MtSAT1;2, this leucine is replaced by methionine. Mutation of this methionine in GmSAT1 to the more common leucine introduced some apparent instability in the protein and abolished its function (Figure 2-7). This was somewhat surprising as Met and Leu side chains have similar physicochemical properties. In fact, the stability of the protein would be expected to increase with a Met→Leu substitution of a buried Met residue (Ann-Lipscomb et al., 1998), however steric interference may explain the loss of stability and function of the M217L mutant. Indeed, the exquisite sensitivity of GmSAT1 to even conservative mutations of hydrophobic residues in these helices has been repeatedly demonstrated

in this study, for example the L191V and L207V mutations. Interestingly, expression of At2g22750, the closest *Arabidopsis* orthologue of GmSAT1, is unable to complement 26972c growth on 1 mM ammonium, or induce toxicity on 100 mM MA (data not shown). The bHLH motifs of GmSAT1 and At2g22750 are 94% similar (see alignment, Figure 2-8), however with the removal of the last 5 amino acids of helix 2, which includes M217, the similarity rises to 98%. This raises the possibility that the differences in this region of the HLH may be responsible for the inability of At2g22750 to complement 26972c on 1 mM ammonium. Alternatively the difference in complementation may be unrelated to the bHLH region of the two proteins as the full length proteins share only ~50% similarity (Figure 2-8B).

One possible mechanism to explain the action of GmSAT1 is that its overexpression is mimicking the action of another endogenous bHLH transcription factor. This possibility was explored in this study with the overexpression of four yeast bHLH proteins, which share modest homology with GmSAT1 within the bHLH domain. When overexpressed, three of the four bHLH proteins, ScCbf1, ScPho4 and ScRTG3 were toxic to the yeast when grown on proline media, with the exception being ScIno4 which grew but also induced a toxicity phenotype like GmSAT1 on 100 mM MA. ScIno4p is a transcriptional regulator of phospholipid synthesis, responding to limiting phospholipid precursors inositol and choline (Schwank et al., 1997). Whether or not ScIno4 contains a transcriptional activation domain is unclear (Schwank et al., 1995; Robinson et al., 2000), however it is dependent on forming heterodimers with ScIno2 to mediate transcriptional regulation of structural genes involved in phospholipid synthesis (Schwank et al., 1995). How overexpression of ScIno4 induces toxicity in the presence of 100 mM MA was not determined in this

study, however as there was no enhancement of MA accumulation measurable in ^{14}C-MA flux experiments and no complementation of 26972c on 1 mM ammonium, its mode of action is most likely distinct from that of GmSAT1. Overexpression of ScGAT1, one of the positive transcriptional regulators involved in NCR in yeast was similarly toxic to the yeast when grown in the presence of 100 mM MA, and did not enhance MA accumulation over a 10min uptake period. Interestingly, a recent yeast transcriptome study identified the expression of the ammonium transporter Mep2 and the polyamine/urea transporter Dur3 as being repressed in the presence of inositol and choline (Santiago and Ben Mamoun, 2003), suggesting a possible connection between the regulation of nitrogen uptake/assimilation and phospholipid synthesis.

2.4 METHODS

2.4.1 PCR AMPLIFICATION OF MUTANT *GmSAT1* cDNA

Standard DNA manipulations were carried out essentially as described by Sambrook and Russell (2001). Mutant GmSAT1 were generated by PCR of pYES3-GmSAT1 using mutagenic primers as listed in table 1. The high fidelity polymerase *Pfu* (Promega) was used according to the manufacturer's instructions, using the following PCR conditions: 95°C 1 min; 30 cycles of 95°C 45 sec, 50°C 45 sec, 72°C 16 min.

Reactions were spiked with 1 µl of DpnI, a methylated DNA-specific restriction endonuclease, and incubated at 37°C for at least 2 hours to remove contaminating

template plasmid DNA. DH5α *E.coli* were transformed with 5 µl of the PCR and plated on LB supplemented with ampicillin (100 µg/ml).

2.4.2 SELECTION OF MUTANT *GmSAT1*

E. coli harbouring mutant *GmSAT1* were selected using restriction enzyme sites introduced or removed through the mutagenesis PCR, as indicated in table 1. The full length of the mutant *GmSAT1* cDNAs were then sequenced to verify the presence of the desired mutation and that no spurious mutations were introduced during the mutagenesis PCR. Sequencing primers specific to the pYES3 plasmid were designed 5' and 3' of the *GmSAT1* cDNA (pYES3_449_FW 5'-AACCCCGGA TCGGACTACTA-3' and pYES3_674_RV 5'-CTTTTCGGTTAGAGCGGATG-3' respectively), and primers for mutagenesis were used for internal sequencing. Sequencing reactions were performed using Big Dye Terminator v3.1 (ABI), according to the manufacturers protocol. Reactions were isopropanol precipitated and sent to the Institute of Medical and Veterinary Sciences (Adelaide, SA) for analysis.

2.4.3 YEAST TRANSFORMATION

Yeast were transformed using a modified lithium acetate method as described by Gietz and Woods, 2002. A 10 mL YPAD (1% (w/v) yeast extract; 2% (w/v) peptone; 2% (w/v) glucose; and 0.004% (w/v) adenine hemisulfate) culture was spiked with yeast and grown overnight at 28°C, shaking 200 rpm. The culture was diluted with fresh YPAD to approximately $OD_{600}= 0.2$, and grown further at 28°C, shaken at 200 rpm, for approximately 2 rounds of cell division (3 hours). 1.5 mL of cells were

pipetted into a microfuge tube and the cells pelleted by centrifugation in a bench top centrifuge at 4000 rpm, 5sec. The supernatant was discarded and the cells were gently resuspended by vortexing in the transformation mix (240 µl 50% (w/v) PEG 3350; 25 µl 2mg/ml ssDNA (pre-boiled, then chilled on ice); 36 µl 1 M lithium acetate) plus 200 ng plasmid DNA. Cells were incubated at 42°C for 40 min before being pelleted by centrifugation at 8000 rpm, 5sec. The yeast pellet was gently resuspended in 200 µl 0.9% (w/v) NaCl by pipetting up and down with a wide bore pipette. 100-200 µl was plated on uracil deficient selective minimal media (YNB w/o amino acids (Difco) supplemented with 2% (w/v) glucose and 1.5% (w/v) agar and incubated at 30°C for 2-3 days.

2.4.4 GROWTH STUDIES OF 26972C EXPRESSING MUTANT GmSAT1

26972c yeast harbouring constructs of interest were grown to mid log phase in YNB (w/o amino acids) with 2% (w/v) glucose. Yeast were pelleted, washed with sterile ddH$_2$O and resuspended at an OD$_{600}$= 0.3. A 10 fold dilution series of the yeast was created and 5 µL spots were carefully pipetted onto solid minimal media (Grenson 1966) supplemented with 2% (w/v) galactose or 2% (w/v) glucose as a carbon source and either 1 mM NH$_4$Cl, 0.1% (w/v) proline or 0.1% (w/v) proline/100 mM methylammonium chloride. As a loading control, yeast were also spotted onto YNB (w/o amino acids, Difco) plus 2% (w/v) glucose plates. Yeast were allowed to grow for 2-5 days at 28°C.

2.4.5 ^{14}C-METHYLAMMONIUM FLUX

Yeast were grown overnight in YNB (w/o amino acids), supplemented with 2% (w/v) glucose to mid log phase then washed twice with sterile H_2O and used to inoculate 25 mL Grenson's minimal medium (Grenson, 1966) supplemented with 0.1% (w/v) proline, 100 mM methylammonium chloride and 2% (w/v) glucose at an $OD_{600} = 0.2$. Yeast were grown overnight to mid-log phase in this medium to prevent the reversion of the 26972c yeast strain to a wild type-like strain with respect to ammonium transport. Yeast were then washed twice with sterile H_2O and grown overnight in Grenson's medium with 1 mM NH_4Cl and 2% (w/v) galactose to induce GmSAT1 expression. Mid-log phase yeast were washed with sterile H_2O, concentrated to $OD_{600} = 4.0$ and incubated in a 20 mM phosphate buffer (pH 6.2) plus 2% (w/v) galactose, with 1 mM ^{14}C-methylammonium (Amersham) for 10 min before filtration through 0.45 µm nitrocellulose membranes (Sigma). 10 mL of ice-cold phosphate buffer was immediately washed through the membrane to prevent further ^{14}C-methylammonium uptake. Membranes plus cells with accumulated ^{14}C-methylammonium were placed into scintillation vials to which 3 ml of StarScint scintillant (Perkin-Elmer) was added. β-decay was measured using a Tri-Carb 2100 scintillation counter (Beckman). Uptake results were normalised using total protein of the yeast determined using a modified Lowry method (Peterson 1977).

2.4.6 TOTAL YEAST PROTEIN EXTRACTION FOR SDS/PAGE

Yeast were grown to mid-log phase in YNB (w/o amino acids) supplemented with 2% (w/v) galactose at 30°C with shaking. Cells were pelleted (4000 rpm, 3 min) and the pellet washed once with 1 mM EDTA, pH 8.0. The pellet was then resuspended in 200 µl 2 M NaOH and transferred to a 1.5ml microfuge tube and incubated for 10

min on ice. 200 µl 50% (w/v) trichloroacetic acid was added, mixed well and the mixture left on ice for at least 1 hour. Samples were spun at 14,000 rpm at 4°C for 20 min in a benchtop microfuge to sediment the precipitated protein. The pelleted protein was washed once with 100% acetone before being resuspended in 5% (w/v) SDS. Depending on the size of the protein pellet between 50-200 µl 5% (w/v) SDS was used. An equal volume of 2x SDS sample buffer (25 mM Tris-HCl (pH 6.8), 9 M urea, 1 mM EDTA, 1% (w/v) SDS, 0.7 M β-mercaptoethanol, 10% (v/v) glycerol) was then added to the resuspended protein. The pH of the protein sample was neutralised with 2 µl of 1 M Tris-HCl pH 8.8 as required. Protein concentration was estimated by Coomassie blue R250 stain (Sigma) of an SDS/PAGE gel.

2.4.7 SODIUM DODECYL SULFATE/ POLYACRYLAMIDE GEL ELECTROPHORESIS

Standard SDS/PAGE in a Tris-Glycine running buffer (0.384 M glycine, 0.1% (w/v) SDS, 0.1 M Tris-HCl) was performed on protein samples. Polyacrylamide gels (12% (w/v) bis-acrylamide) were prepared using the MiniProtean system (BioRad). Total protein was examined using Coomassie blue R250 stain (Sigma).

2.4.8 WESTERN BLOTTING

After separation of proteins by SDS/PAGE, proteins were transferred to nitrocellulose C^+ membranes (Schleicher & Schuell) using a semi-dry method. The transfer was run at 65 mA for 90-120 min. The membrane stained with Ponceau (Sigma) and the ladder marked on the membrane. The Ponceau stain was rinsed from the membrane using Tris/saline (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) plus 0.1% (v/v) Tween-20 (Sigma), and the membrane blocked with blocking solution

(Tris/saline plus 5% (w/v) skim milk powder) for at least 1 hour at room temperature. The membrane was then incubated in a 1:4000 dilution of the polyclonal GmSAT1-specific antibody in blocking solution either for 1 hour at room temperature or overnight at 4°C. The membrane was washed for 2 x 10 min in Tris/saline plus 0.1% (v/v) Tween-20 (Sigma), followed by 2 x 10 min washes with blocking solution, and then incubated in a 1:10,000 dilution of a horseradish peroxidase-conjugated anti-rabbit antibody (Roche) in blocking buffer for 30-60 min at room temperature. The membrane was washed 4 x 15 min in Tris/saline plus 0.1% (v/v) Tween-20 and incubated for 1 min in the horseradish peroxidase substrate according to the manufacturer's instructions (Roche). Luminescence was imaged using a ChemiDoc XRS (Biorad) camera system.

NAME	SEQUENCE (5' - 3')	Restriction site change
E177A_FW	CCACATAATAGCTGCAAGGAAGCG	SatI introduced
E177A_RV	CGCTTCCTTGCAGCTATTATGTGG	
E177D_FW	CACATAATAGCTGATCGGAGGCGAAGAG	DpnI introduced
E177D_RV	CTCTTCGCCTCCGATCAGCTATTATGTG	
E177Q_FW	CACATAATAGCGCAGCGGAAGCGAAGAG	SatI introduced
E177Q_RV	CTCTTCGCTTCCGCTGCGCTATTATGTG	
R180K_FW	GAAAGGAAGAAAAGAGAGAAGC	Eam1104I introduced
R180K_RV	GTTCTCTCTTTTCTTCCTTTC	
L191A_FW	AACGGTTCATTGCTGCCTCTGCTCTTGTTC	MnlI introduced
L191A_RV	GAACAAGAGCAGAGGCAGCAATGAACCGTT	
L191I_FW	GCCAACGGTTCATTGCGATATCTGCTCTTGTTCCGGGC	EcoRV introduced
L191I_RV	GCCCGGAACAAGAGCAGATATCGCAATGAACCGTTGGC	
L191V_FW	AACGGTTCATAGCTGTATCTGCTCTTGTTC	AluI introduced
L191V_RV	GAACAAGAGCAGATACAGCTATGAACCGTT	
L207V_FW	GATGGACAAAGCGTCGGTTGTTGGAGAAGC	HindIII deleted
L207V_RV	GCTTCTCCAACAACCGACGCTTTGTCCATC	
M217L_FW	ATTTGAAGCAACTGCAGGAGAAAGTGAGTG	PstI introduced
M217L_RV	CACTCACTTTCTCCTGCAGTTGCTTCAAAT	
D317A_FW	GTTTCATCCTGGCCATAACCATTATTGC	HaeIII introduced
D317A_RV	GCAATAATGGTTATGGCCAGGATGAAAC	
D317N_FW	GTTTCATCCTGAATATAACCATTATTGC	MvaI deleted
D317N_RV	GCAATAATGGTTATATTCAGGATGAAAG	
I318Q_FW	CATCCTGGATCAAACCATTATTGCTC	DpnI introduced
I318Q_RV	GAGCAATAATGGTTTGATCCAGGATG	

Table 2-1. Primers used to introduce point mutations into the bHLH and TMD of GmSAT1.
Amino acids substituted are indicated in the first column. Red letters in the primer sequence correspond to mutated nucleotides to introduce the amino acid substitutions as well as silent mutations to delete/introduce restriction enzyme sites designated in column 3. These restriction enzyme sites were used to distinguish mutant GmSAT1 cDNAs from wild type GmSAT1.
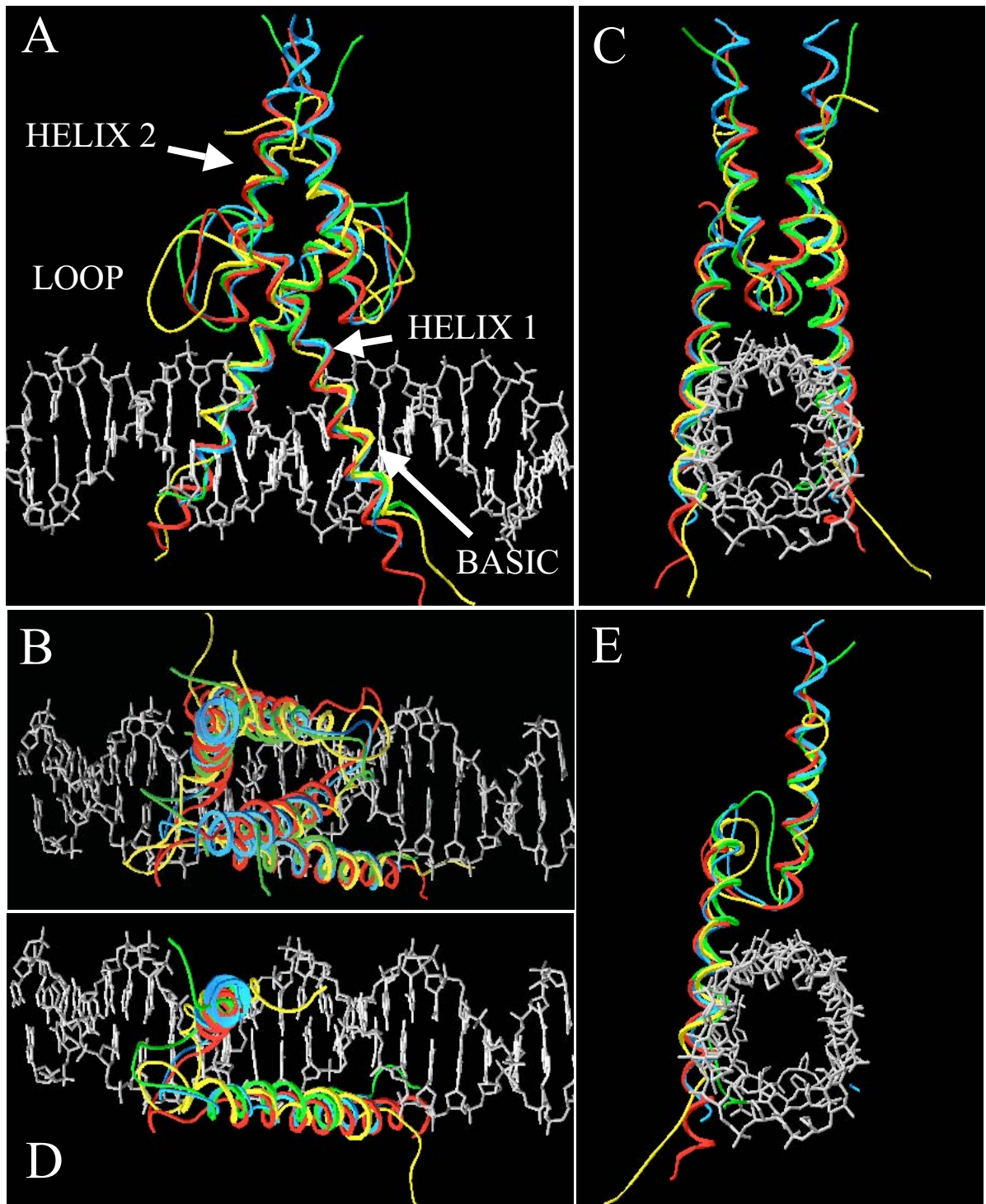
Figure 2-1. A model of the putative homodimeric bHLH domain of GmSAT1 bound to DNA. The amino acid sequence of GmSAT1 (Pro165-Ala223) was threaded onto the three dimensional structures of the bHLH domains of 4 bHLH proteins using the Swiss-Pdb Viewer (http://www.expasy.org/spdbv/). **(A)** & **(B)** Two views of the full length putative GmSAT1 homodimer. **(C)** For clarity the variable loop region (Pro196-Lys203) of the protein was removed. **(D)** & **(E)** Two views of monomeric GmSAT1. Grey = DNA, Blue = HsMax (Ferredamare et al., 1993), yellow = HsUSF **(**Ferredamare et al., 1994), red = HsMyoD (Ma et al., 1994), and green = ScPho4 (Shimizu et al., 1997).

```
                    BASIC ↓          ↓              HELIX 1        ↓
              ┌──────────────┬───────────────┬─────────────────────────────┐
              │      *        *        *      │/////////////////////////////│
GmSAT1      Q P Q D  H I I A E R K R R  E K L S Q R F I A L S A L V
MtSAT1;2    Q A H D  H I V T E R K R R  E K L S Q R F I A L S A L V
VvSAT1      Q T Q D  H I I A E R K R R  E K L S Q R F I A L S A I V
At2g22750   N A Q D  H I L A E R K R R  E K L T Q R F V A L S A L I
ScPho4      K R E S  H K H A E Q A R R  N R L A V A L H E L A S L I
HsMyc       K R R T  H N V L E R Q R R  N E L K R S F F A L R D Q I
HsSREBP     K R T A  H N A I E K R Y R  S S I N D K I I E L K D L V
DmMlx-int   R R A G  H I H A E Q K R R  Y N I K N G F D T L H A L I
                     H I . A E R K R R    . L   . . F . A L   A L .
            1 2 3 4  5 6 7 8 9 0 1 2 3  4 5 6 7 8 9 0 1 2 3 4 5 6 7
                                  10                      20

          ▼              LOOP          ▼             HELIX 2      ↓                 ↓
          ┌───────────────────────────┬──────────────────────────────────────────┐
          │:::::::::::::::::::::::::::::│////////////////////////////////////////////│
GmSAT1      P G L K - - - - - - - K M D  K A S V L G E A I K Y L K Q M
MtSAT1;2    P N L K - - - - - - - K M D  K A S V L G E A I R Y L K Q M
VvSAT1      P G L K - - - - - - - K M D  K A S V L G D A I K Y L K Q L
At2g22750   P G L K - - - - - - - K M D  K A S V L G D A I K H I K Y L
ScPho4      P A E W K Q Q N V S A A P S  K A T T V E A A C R Y I R H L
HsMyc       P E L E N - - - - N E K A P  K V V I L K K A T A Y I L S V
HsSREBP     V G T E A - - - - - - K L N  K S A V L R K A I D Y I R F L
DmMlx-int   P Q L Q L - - N P N A K L S  K A A M L Q K G A D H I K Q L
            P . L     Q Q N   N A K      K A . V L   A I . Y I K   L
            8 9 0 1 2 3 4 5 6 7 8 9 0 //  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4
                30                40   50                 60
```
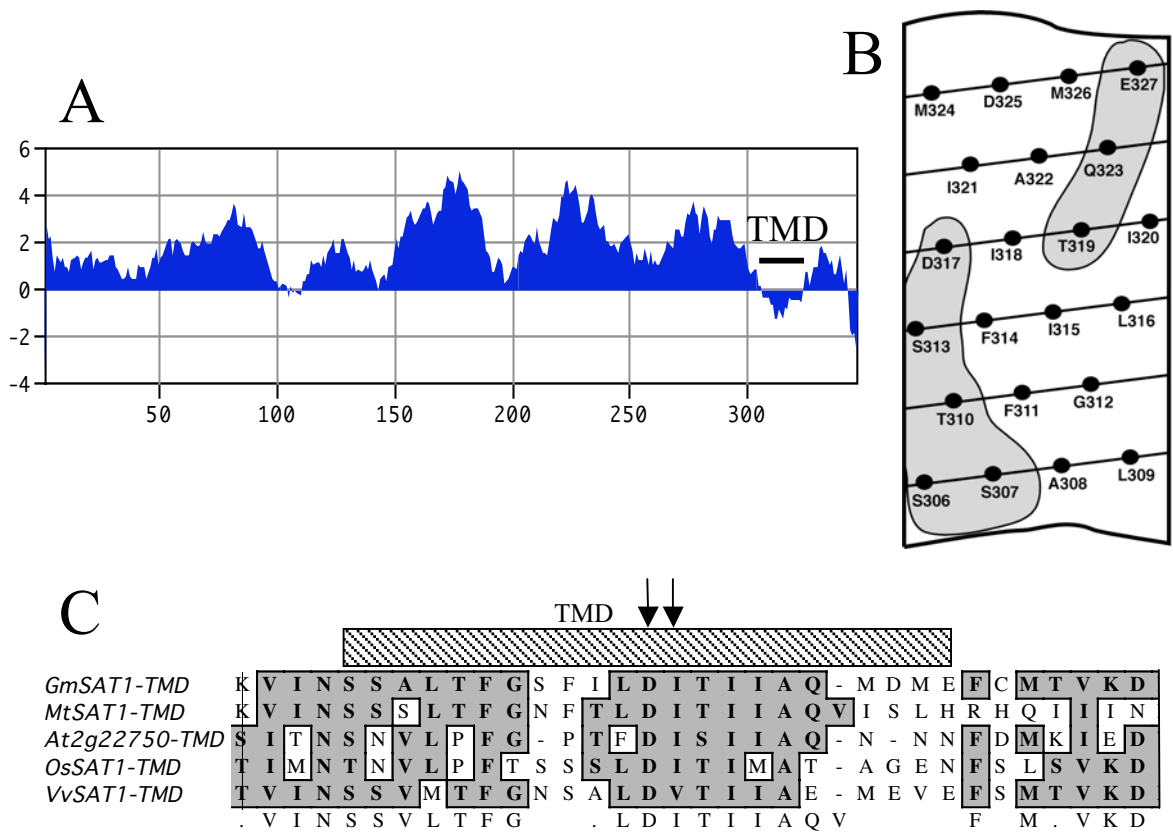
Figure 2-2. Sequence alignments of the bHLH domain of GmSAT1 (Q169 → M217) and selected bHLH proteins. The basic, helix 1 and 2 and variable loop are marked above the sequences. Residues substituted in this study are marked with arrows. The H-E-R motif which is the primary determinant for the binding of bHLH proteins to the palindromic CACGTG DNA sequence is marked with asterisks (*). Arrow heads indicate the conserved Pro and Lys residues which commonly mark the beginning and end of the variable loop region. The numbering system for bHLH amino acid residues devised by Atchley et al., 1997 is indicated below the sequence. // indicates gap in numbering in the variable loop region. Gm - *Glycine max,* Mt - *Medicago truncatula*, Vv - *Vitis vinifera*, At - *Arabidopsis thaliana*, Hs - *Homo sapiens*, Dm - *Drosophila melanogaster.* Accessions (in order): AAC32828, AC130800, CAN75020, NP_001077944, P07270, NP_002458, NP_004590, NP_724328

Figure 2-3. GmSAT1 has a predicted C-terminal transmembrane domain (TMD).
(A) The predicted TMD of GmSAT1 is marked on the GES hydrophilicity plot
(window size = 19). **(B)** A helical plot of the predicted TMD of GmSAT1. The shaded
area represents a predicted hydrophilic face of the TMD. **(C)** An alignment of the
predicted TMD of GmSAT1 and SAT-like proteins from *M. truncatula*, *A. thaliana*,
O.sativa, and *V. vinifera*. Amino acids mutated in this study are marked with arrows.
Accessions (in order): AAC32828, AC130800, NP_001077944, NP_001051131
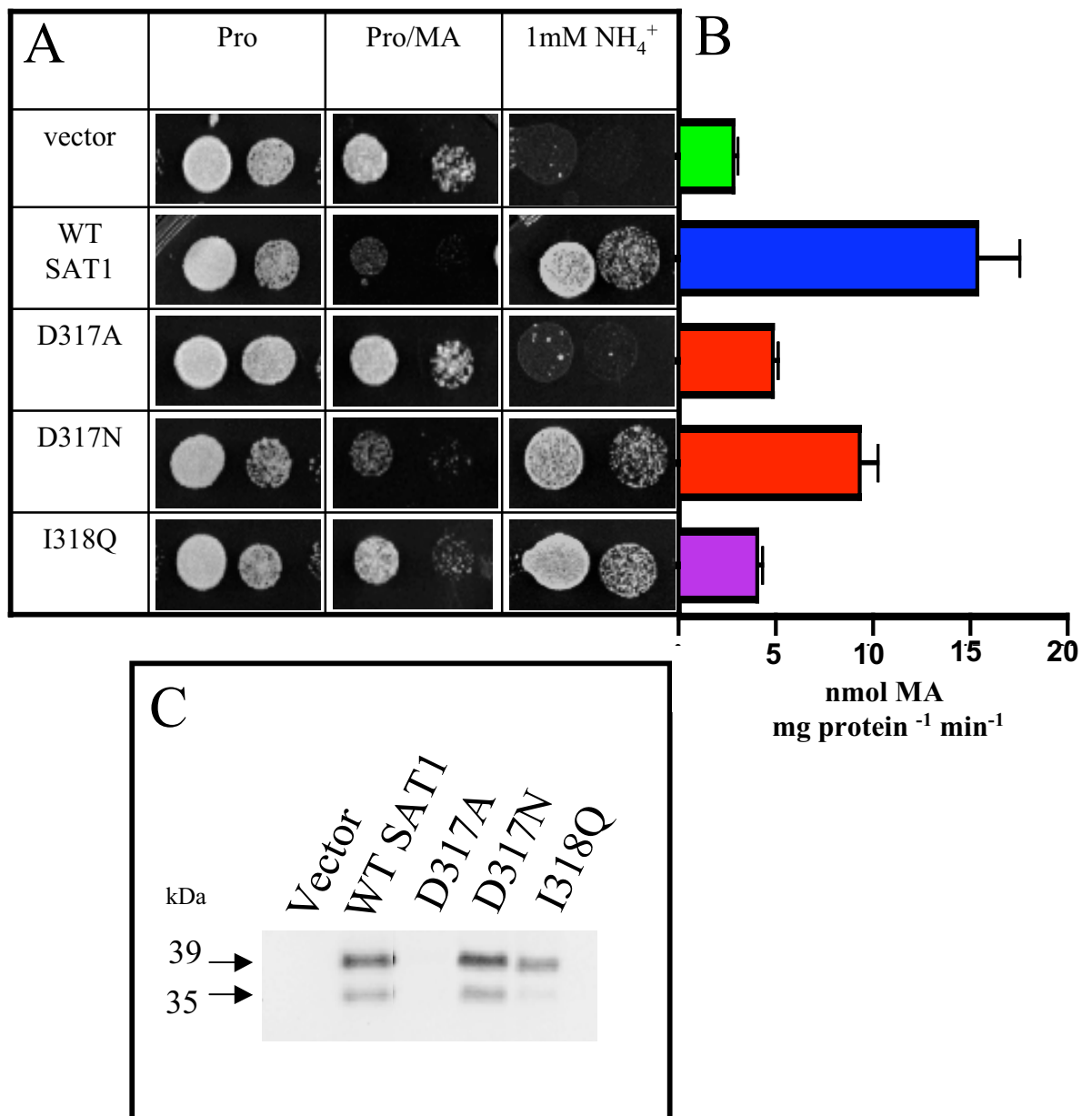CAN75020. Thanks to Dr. Susan Howitt for modeling GmSAT1 in Figure B.

Figure 2-4. Effects of amino acid substitutions in the predicted C-terminal transmembrane domain of GmSAT1. (A) 26972c yeast harbouring the empty pYES3 vector (vector), the wild-type GmSAT1 cDNA in pYES3 (WT SAT1), or with the amino acid substitutions D317A, D317N or I318Q were spotted onto solid minimal media plus 2% (w/v) galactose, supplemented with 0.1% proline (Pro), 100mM methylammonium chloride/ 0.1% (w/v) proline (Pro/MA) or 1 mM NH_4Cl (1mM NH^+). **(B)** ^{14}C-MA uptake experiments of 26972c yeast expressing wild-type and mutant GmSAT1 were performed as described in section 4.4.5. Values are the means of 3 independent experiments of 4 reps each ± SEM. **(C)** Total protein from 26972c cells expressing the GmSAT1 mutants were subjected to SDS-PAGE, western blotted and probed with anti-GMSAT1 serum. GmSAT1 consistently ran as two bands, approximately 39 kDa and 35 kDa in size.
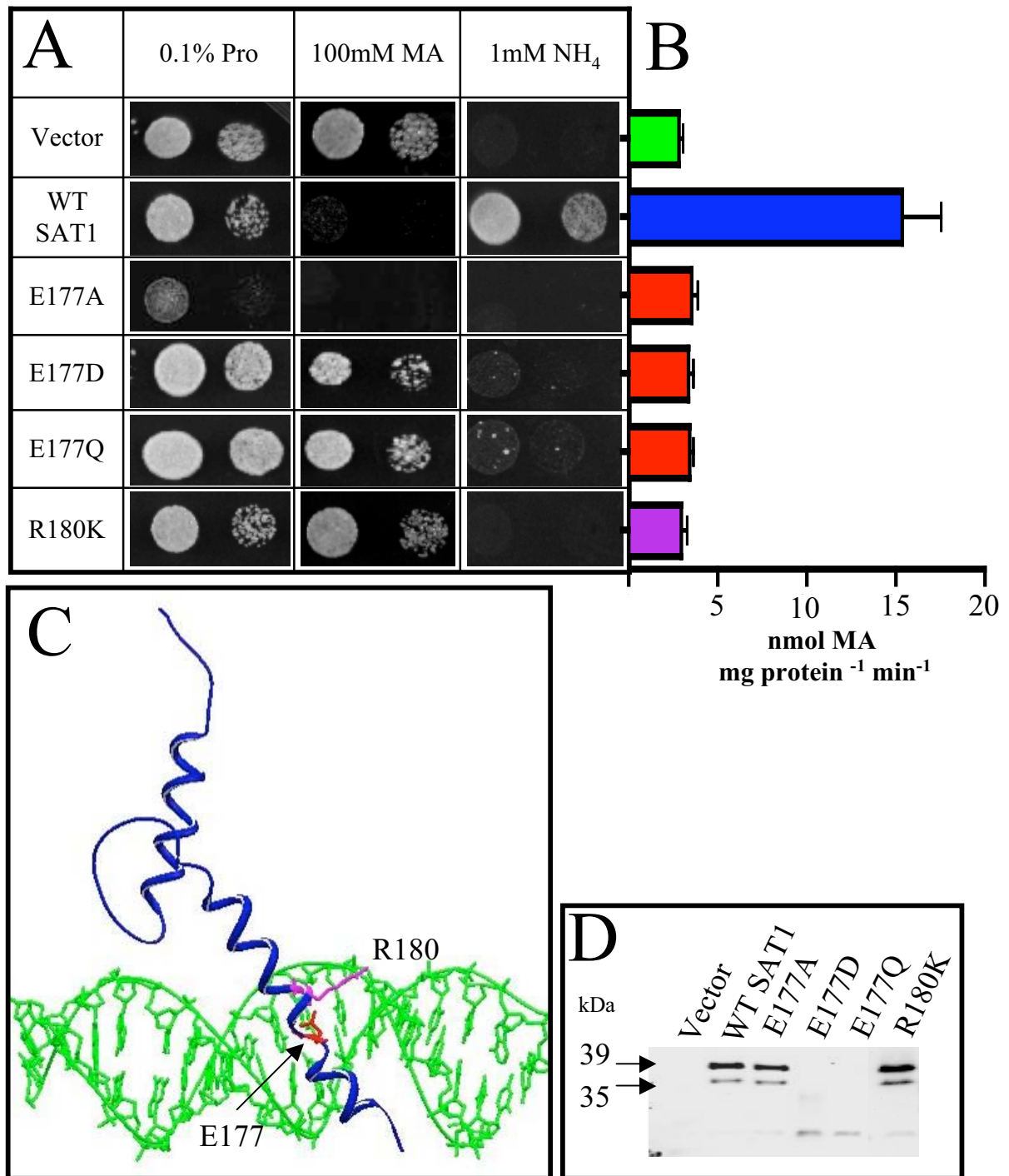
Figure 2-5. The effects of mutations in the basic region of the GmSAT1 bHLH domain. (A) 26972c yeast, harboring the empty pYES3 vector (vector), the wild-type GmSAT1 cDNA in pYES3 (WT SAT1), or with the amino acid substitutions indicated were spotted onto solid minimal media plus 2% (w/v) galactose, supplemented with 0.1% proline (Pro), 100mM methylammonium chloride/ 0.1% (w/v) proline (Pro/MA) or 1mM NH_4Cl (1mM NH_4). **(B)** ^{14}C-MA uptake experiments of 26972c yeast expressing wild-type and mutant GmSAT1 were performed as described in section 4.4.5. Values are the means of 3 independent experiments of 4 reps each ± SEM. **(C)** The primary sequence of the bHLH domain of GmSAT1 was threaded onto the crystal structure of ScPho4 (Shimizu et al., 1997) bound to its cognate DNA sequence (CACGTG) using the Swiss-Pdp Viewer (http://www.expasy.org/spdbv/). Pictured is the predicted backbone structure of monomeric GmSAT1, with the side-chains of the selected mutated amino acid indicated. **(D)** Total protein from 26972c cells expressing the GmSAT1 mutants were subjected to SDS-PAGE, western blotted and probed with anti-GMSAT1 serum. When stable, GmSAT1 consistently ran as two bands, approximately 39 kDa and 35 kDa in size.
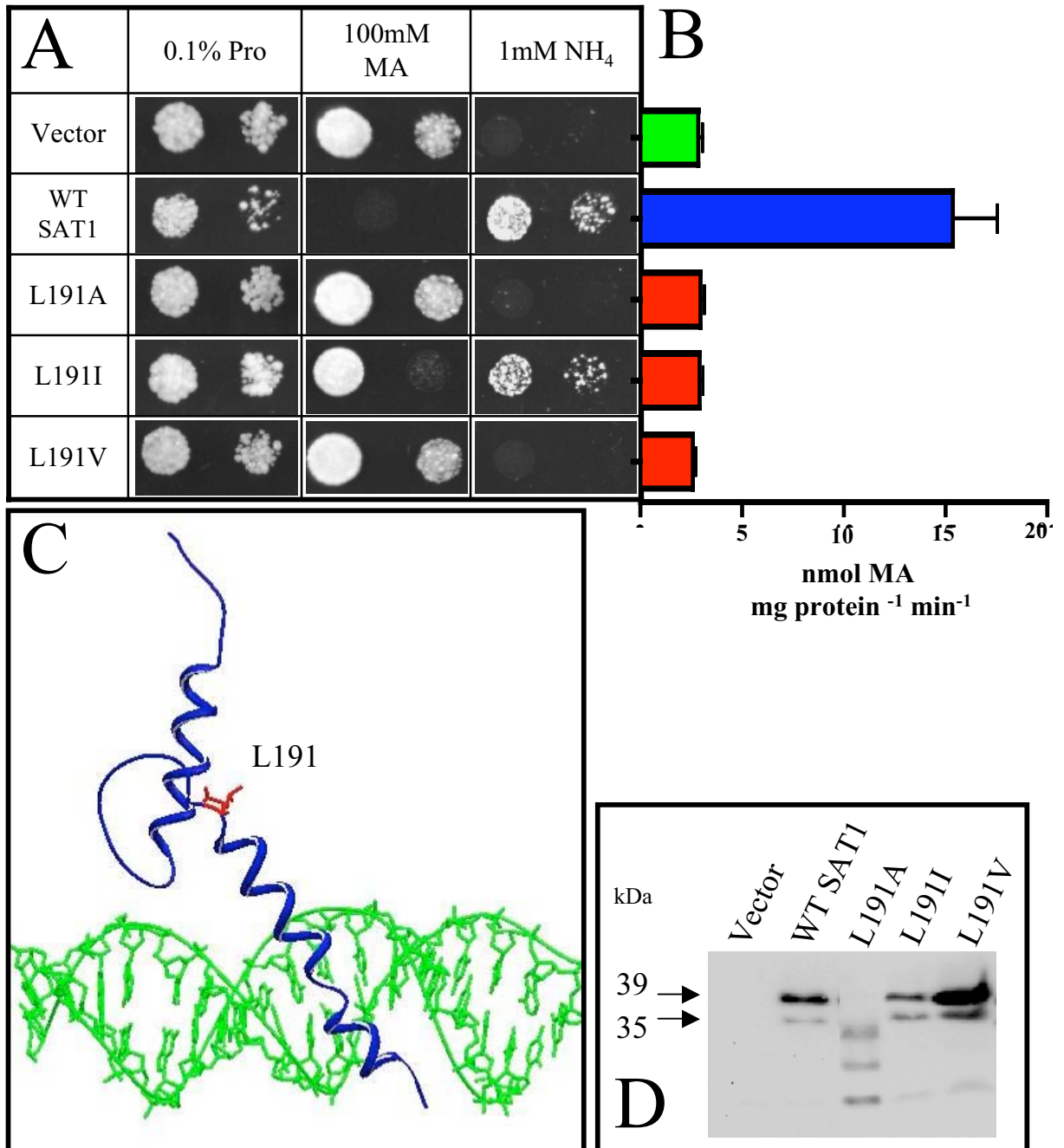
2-26

Figure 2-6. The effects of mutations in helix 1 of the GmSAT1 bHLH domain. (A)
26972c yeast, harboring the empty pYES3 vector (vector), the wild-type GmSAT1
cDNA in pYES3 (WT SAT1), or with the amino acid substitutions indicated were
spotted onto solid minimal media plus 2% (w/v) galactose, supplemented with 0.1%
proline (Pro), 100mM methylammonium chloride/ 0.1% (w/v) proline (Pro/MA) or
1mM NH_4Cl (1mM NH_4). **(B)** ^{14}C-MA uptake experiments of 26972c yeast expressing
wild-type and mutant GmSAT1 were performed as described in section 4.4.5. Values
are the means of 3 independent experiments of 4 reps each ± SEM. **(C)** The primary
sequence of the bHLH domain of GmSAT1 was threaded onto the crystal structure of
ScPho4 (Shimizu et al., 1997) bound to its cognate DNA sequence (CACGTG) using
the Swiss-Pdp Viewer (http://www.expasy.org/spdbv/). Pictured is the predicted
backbone structure of monomeric GmSAT1, with the side-chains of the selected
mutated amino acid indicated. **(D)** Total protein from 26972c cells expressing the
GmSAT1 mutants were subjected to SDS-PAGE, western blotted and probed with anti-
GMSAT1 serum. When stable, GmSAT1 consistently ran as two bands, approximately
39 kDa and 35 kDa in size.

Figure 2-7. The effects of mutations in helix 2 of the GmSAT1 bHLH domain. (A)
26972c yeast, harboring the empty pYES3 vector (vector), the wild-type GmSAT1 cDNA
in pYES3 (WT SAT1), or with the amino acid substitutions indicated were spotted onto
solid minimal media plus 2% (w/v) galactose, supplemented with 0.1% proline (Pro),
100mM methylammonium chloride/ 0.1% (w/v) proline (Pro/MA) or 1mM NH$_4$Cl (1mM
NH$_4$). **(B)** ^{14}C-MA uptake experiments of 26972c yeast expressing wild-type and mutant
GmSAT1 were performed as described in section 4.4.5. Values are the means of 3
independent experiments of 4 reps each ± SEM. **(C)** The primary sequence of the bHLH
domain of GmSAT1 was threaded onto the crystal structure of ScPho4 (Shimizu et al.,
1997) bound to its cognate DNA sequence (CACGTG) using the Swiss-Pdp Viewer
(http://www.expasy.org/spdbv/). Pictured is the predicted backbone structure of
monomeric GmSAT1, with the side-chains of the selected mutated amino acid indicated.
(D) Total protein from 26972c cells expressing the GmSAT1 mutants were subjected to
SDS-PAGE, western blotted and probed with anti-GMSAT1 serum. When stable,
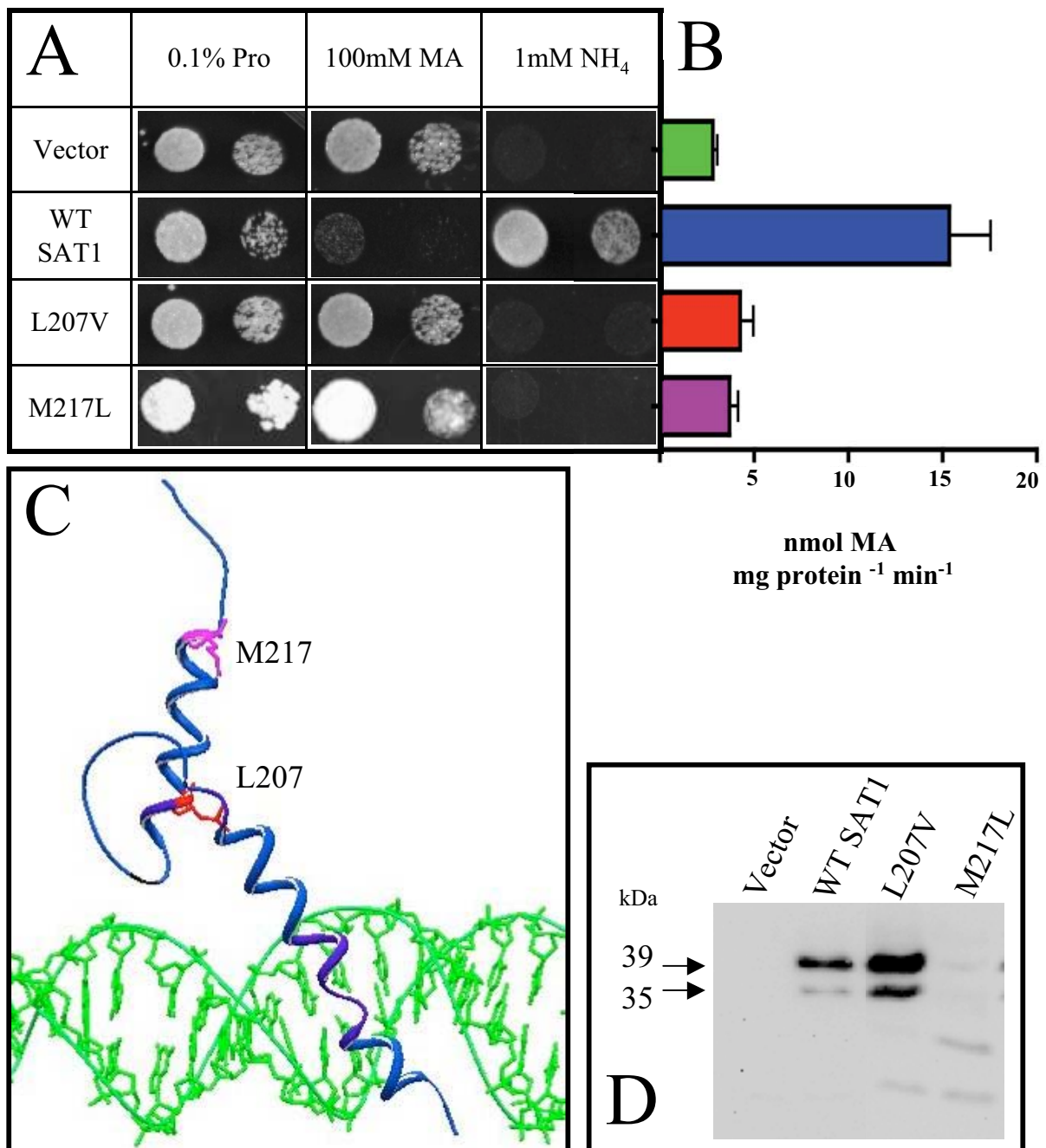GmSAT1 consistently ran as two bands, approximately 39 kDa and 35 kDa in size.

Figure 2-8. Sequence alignment (A) and phylogenetic tree (B) of the basic Helix-Loop-Helix domain of plant members of the SAT family. An asterisk marks the position of the highly conserved Leu at position 64 (Toledo-Oritz et.al., 2003) which is replaced by Met in GmSAT1 (M217) and MtSAT1;2 and MtSAT1;1. Neighbour joining; best tree. Distance: absolute. Gm - *Glycine max*, Mt - *Medicago truncatula*, Vv - *Vitis vinefera*, Os - *Oryza sativa*, At - *Arabidopsis thaliana*, Mc - *Mesembryantheum crystallinum.* Accessions (in order): AAC32828, AC130800, AC130800, AC124214, CAN75020, NP_001051131 , NP_001077944, NP_179861.2, NP_850031.2, AAD11428

BASIC · HELIX

```
           *
GmSAT1   Q P Q D - H I I A E R K R R E K L S Q R F I A L S A L V P G - -
ScCbf1   Q R K D S H K E V E R R R R E N I N T A I N V L S D L L P - - -
ScIno4   Q I R I N H V S S E K K R R E L E R A I F D E L V A V V P D L Q
ScPho4   D K R E S H K H A E Q A R R N R L A V A L H E L A S L I P A E W
ScRtg3   R K R E F H N A V E R R R R E L I K Q K I K E L G Q L V P P S L
ScNoc3   E Q A V S A E E R E R N Q S E I L K I V F T I Y L N I L K N N A
YGR290W  N W A E L V V N S K W S S S K A L S F F P S S K S S V L K T S C
           .   . . H   . E R . R R E . L   .       L . L . P
```

LOOP · HELIX

```
GmSAT1   - - - - - - - - - - - L K K M D K A S V L G E A I K - - Y L K Q M
ScCbf1   - - - - - - - - - - - V R E S S K A A I L A R A A E - - Y I Q K L
ScIno4   - - - - - - - - - - P Q E S R S E L I I Y L K S L S - - Y L S W L
ScPho4   K Q Q N - - - - - - V S A A P S K A T T V E A A C R - - Y I R H L
ScRtg3   L N Y D D - - - L G K Q I K P N K G I I L D R T V E - - Y L Q Y L
ScNoc3   K T L I G S V L E G L T K F G N M A N F D L L G D - - F L E V M
YGR290W  W F T - - - - - - - - I S L E D S E M F L S S F F E A A F F L L T
             S V L   .       K . . L     . . . A A Y L           L
```
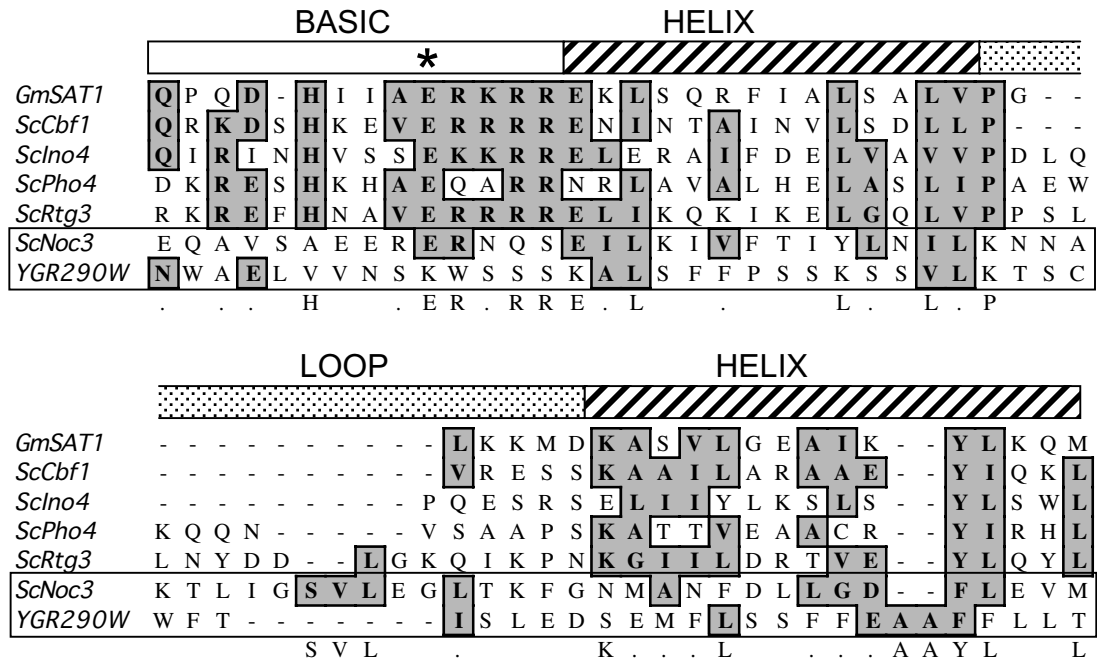
Figure 2-9. Sequence alignment of GmSAT1 with the yeast bHLH and bHLH-like proteins (boxed) used in this study. The basic Helix-Loop-Helix region of the proteins are indicated above the sequences. Note that YGR290W does not have glutamate at position 9 (asterisk) and probably does not bind DNA. The putative bHLH domain of ScNoc3 is also divergent from the other yeast bHLH transcription factors and GmSAT1. Accessions (in order): AAC32828, P17106, EDN63767, NP_116692, EDN64520, Q07896, P53339
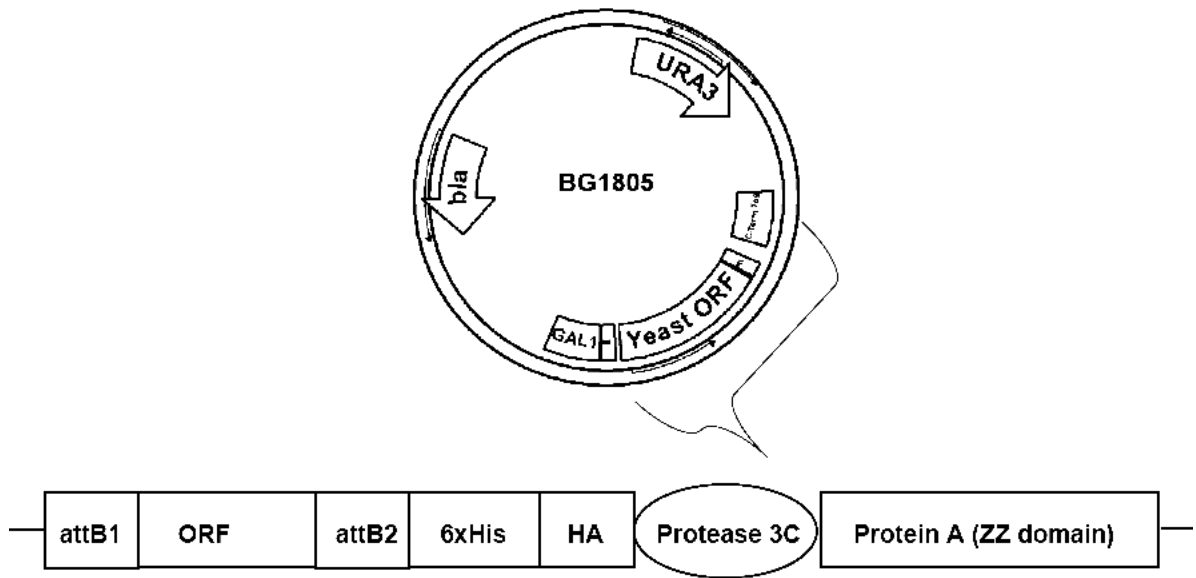
Figure 2-10. Map of yeast ORF overexpression shuttle vector BG1805 (Open Biosystems). URA3 and bla are uracil and ampicillin selectable markers for yeast and *E. coli* respectively. The yeast ORF is inserted in the attB1/attB2 Gateway® recombination sites, fused in-frame to the 6xHIS, HA epitope, Protein A C-terminal protein tag. Over-expression of the tagged ORF is driven by the yeast Gal1 promoter.

	YNB	Pro	Pro/MA	1mM NH_4^+	nmol MA mg protein^{-1} min^{-1}
vector					3.4±0.5
GmSAT1					24.5±2.1
ScCBF1					n.d.
ScINO4					3.9±0.7
ScPHO4					n.d.
ScRTG3					n.d.
ScNOC3					n.d.
YGR290W					n.d.
ScDUR3					n.d.
ScGLN3					n.d.
ScGAT1					3.6±0.4

(Rows ScCBF1 through YGR290W are grouped as "Yeast bHLH"; rows ScDUR3 through ScGAT1 are grouped as "Yeast NCR")

Figure 2-11. Growth studies of 26972c yeast overexpressing endogenous bHLH and nitrogen catabolite repression- (NCR) associated genes. 26972c yeast overexpressing various bHLH genes and NCR-associated genes from the BG1805 vector under the galactose inducible Gal1 promoter (Figure 10) were spotted onto either YNB (w/o amino acids) plus 2% (w/v) glucose (YNB) or minimal medium plus 2% (w/v) galactose supplemented with 0.1% (w/v) proline (Pro), 100mM methylammonium chloride/0.1% (w/v) proline (Pro/MA), or 1mM NH_4Cl (1mM NH_4^+). ^{14}C-MA uptake experiments of 26972c yeast (last column) were performed as described in section 4.4.5. For yeast bHLH accessions, refer to Figure 2-9. Accessions for ScDUR3, ScGLN3 and ScGat1 are NP_011847, NP_010958 and NP_116632 respectively