UNIVERSITY OF READING
SCHOOL OF ANIMAL AND MICROBIAL SCIENCES

ADAPTATION OF *RHIZOBIUM* TO ENVIRONMENTAL STRESS

by
Marc A. Fox

Submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy September 2005

I declare that this is my own account of my research and that this work has not been submitted for a degree at any other university. However, I would like to acknowledge that certain vectors and strains were constructed by members of the laboratory, as described in the text. I also acknowledge the help I received from the undergraduate project students Claire Vernazza, Lara Clewes-Garner and David Stead in long and arduous task of screening the LB3 library, under my joint supervision with Professor Philip Poole.

Marc Fox

ACKNOWLEDGEMENTS

Firstly, I would like to thank Philip Poole for supervising me throughout this project. His guidance over the years has proved fundamental to this research.

Thanks too should go to my colleagues in Lab 160. James White for assistance with radioactive assays, Tim Mauchline for development of colony PCR, Karunakaran for the creation of vectors, Alexandré Bourdes for aid given with plants, Laura Fooks, Alex Pudney and Elham Moslehi-Mohebi for help with protein work and Arthur Hosie, Mary Leonard, Michelle Barr, Alison East, Jon Seaman and Emma Lodwig for their assistance in general. Thanks are also due to the staff of Central Science Services, AMS, especially Jane Clarke for the thousands of plates she prepared for me.

I would also like to thank the University of Reading and the BBSRC for awarding me the funding to carry out this research.

Thanks to the other labs I visited whilst carrying out this project, whose staff were both welcoming and helpful. Thanks to Alan Williams, Martin Krehenbrink and Allan Downie from the John Innes Centre, Norwich, for allowing me use of their Tn*5* library; and to Laurence Dupont, Geneviève Alloing and Daniel LeRudulier from INRA-CNRS, Université de Nice Sophia-Antipolis, for allowing me to carry out proline betaine transport assays with them.

Finally, a big thank-you to my parents and my sister Sarah, for their continued support throughout all my academic studies and also to John, Emma, Jenny and Rach for helping me get through this.

ABSTRACT

A previously created promoter probe library of *Rhizobium leguminosarum* 3841, LB3, was investigated to identify genes that are induced under stressful conditions. Each bacterium in the library contains a plasmid with a random chromosomal insert, upstream of a promoterless *gfp*UV reporter. If the insert contains a promoter that responds to a stress it will activate production of green fluorescent protein (GFP) and colonies will fluoresce bright green when examined under UV light.

Over 30,000 colonies were screened on various media designed to reproduce hyper-osmotic stress, acidic stress and metal toxicity and 32 were induced. The release of the preliminary genome of 3841 allowed the genes, or operons, associated with each of the isolated stress-induced fusions from LB3 to be identified.

Mutations were made in ten of the genes selected from LB3 that are upregulated by hyper-osmosis. The mutants were then tested to see how they would grow in standard and stressed conditions, and if the way which they interacted with pea plants was altered. This led to the discovery of a two-component response regulator system (RL1156 and RL1157) responsible for controlling the transcription of RL1155 in response to low pH and hyper-osmosis.

One of the genes isolated from LB3 was upregulated by hyper-osmosis and is part of an operon for an ABC transporter that shares sequence identity to the well characterised glycine betaine transporter (ProU). This led to the identification of five other ABC systems that shared a significantly similar sequence identity to this transporter. One of these transporters (termed QAT1 in this work) appears to be the homologue of the Cho system in *S. meliloti* as it is induced by choline and is responsible for its uptake. Studies also demonstrated that hyper-osmosis temporarily inactivates solute uptake via ABC transporters (but not secondary permeases).

LIST OF ABBREVIATIONS USED

aa	Amino acids
ABC	ATP-binding cassette
AIB	2-Amino-isobutyric acid
ALA	δ-Aminolevulinic acid
$AlCl_3$	Aluminium chloride
AMA	Acid minimal agar
Amp	Ampicillin
AMS	Acid minimal salts
ASP	Acid shock protein
ATP	Adenosine triphosphate
BAP	Bacterial alkaline phosphatase
bp	Base pair
cfu	Colony forming units
CSP	Cold shock protein
$CuCl_2$	Copper chloride
DFI	Differential fluorescence induction
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharide
et al.	*et alii*
Fix	Fixation
GABA	γ-amino-n-butryic Acid
Gen	Gentamycin
GFP	Green fluorescent protein
glc	Glucose
GDW	Glass distilled water
H_2O_2	Hydrogen peroxide
HSP	Heat shock protein
IMP	Integral membrane permease
IPTG	Isopropyl β-D-thiogalactoside
IS50	Insertion sequence 50
Kan	Kanamycin

Kb(p)	Kilobase (pairs)
KCl	Potassium chloride
LA	Luria-Bertani agar
LB	Luria-Bertani broth
LPS	Lipopolysaccharide
MES	2-Morpholinoethanesulfonic acid
MFS	Major facilitator superfamily
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MGT	Mean generation time
MIC	Minimal induction concentration
MOPS	3-[N-morpholino]propanesulfonic acid
N-free	Nitrogen free
NaCl	Sodium chloride
Nal	Naladixic acid
Neo	Neomycin
nH_2O	Nanopure water
NH_4	Ammonium
Nod	Nodulation
Nys	Nystatin
OD	Optical density
OEP	Outer membrane efflux protein
ORF	Open reading frame
PCR	Polymerase chain reaction
PEG	Polyethylene-glycerol
PHB	Polyhydroxybutyrate
QAC	Quaternary amine compound
QAT	Quaternary amine transporter
^{r}	Resistant
RBS	Ribosome binding site
RMS	*Rhizobium* minimal salts
rpm	Revolutions per minute
^{s}	Sensitive
SBP	Solute binding protein
Spc	Spectinomycin

Str	Streptomycin
TAE	Tris acetate EDTA
TCA	Tricarboxylic acid
Tet	Tetracyclin
Tn*5*	Kanamycin/Neomycin resistant transposon
TY	Tryptone-Yeast media
UV	Ultraviolet
VS	Vincent's sucrose
wt	Wild-type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
X-Glc-A	5-bromo-4-chloro-3-indolyl-β-D-glucuronide
$ZnCl_2$	Zinc chloride

CONTENTS

CHAPTER 1: INTRODUCTION

1.1. *Rhizobium*

1.1.1. Taxonomy

Rhizobia is the common name given to a group of small, rod-shaped, Gram-negative bacteria that collectively have the ability to produce nodules on the roots of leguminous plants and belong to the family *Rhizobiaceae*, which are part of the α-proteobacteria. In early studies, the taxonomy of rhizobia was based on the rate of growth of isolates on laboratory media and their selective interaction with their plant hosts. It was soon established that no strain could nodulate all plants, but that each could nodulate some legumes though not others (Long, 1989). This led to the concept of cross-inoculation groups, with organisms grouped according to the hosts they nodulated. Within the genus *Rhizobium* several strains nodulate a common host, but are distinct according to genetic and/or phenotypic properties and are therefore classified as distinct species (e.g. *R. tropici* and *R. etli*). However some strains cannot be distinguished other than by their host range, therefore the species is further classified into biovars (bv.) (e.g. *R. leguminosarum* is split into three biovars that nodulate clover, peas and beans) (Table 1.1).

For a time this was the basis on which rhizobia were identified. However, developments in molecular biology and advances in bacterial taxonomy (Graham, *et al.*, 1991) in have resulted in a rhizobial taxonomy based on a wide range of characteristics and to the distinction of new genera and species. Currently six genera and at least 42 species have been distinguished (Table 1.1), but a number of these remain in question (Tighe *et al.*, 2000; Willems, *et al.*, 2003; Young, 2003). These new classifications have corroborated previous divisions, e.g. the genus *Bradyrhizobium* is made up of the strains that took the longest to grow in laboratory conditions.

Table 1.1. Examples of the Genera and Species of *Rhizobium*. The major host for each species is shown in bold. This is not a complete list. (Updated from Zakhia & de Lajudie, 2001.)

Genera/Species	Host(s)	Reference
Allorhizobium		
A. undicola	***Neptunia natans,*** *Acacia*, *Faidherbia*, *Lotus*	de Lajudie *et al*., 1998a
Azorhizobium		
A. caulinodans	***Sesbania rostrata***	Dreyfus *et al*., 1988
Bradyrhizobium		
B. elkanii	***Glycine max***	Kuykendall *et al*., 1992
B. japonicum	***Glycine max***	Jordan, 1984
B. liaoningense	***Glycine max***	Xu *et al*., 1995
B. yuanmingense	***Lespedeza,*** *Medicago*, *Melilotus*	Yao *et al*., 2002
Mesorhizobium		
M. amorphae	***Amorpha fruticosa***	Wang *et al*., 1999b
M. chacoense	***Prosopis alba***	Velasquez *et al*., 1998
M. ciceri	***Cicer arietinum***	Nour *et al*., 1994
M. huakuii	***Astragalus sinicus,*** *Acacia*	Chen *et al*., 1991; Jarvis *et al*., 1997
M. loti	***Lotus corniculatus***	Jarvis *et al*., 1982; Jarvis *et al*., 1997
M. mediterraneum	***Cicer arietinum***	Nour *et al*., 1995; Jarvis *et al*., 1997
M. plurifarium	***Acacia senegal, Prosopis juriflora,*** *Leucaena*	de Lajudie *et al*., 1998b
M. septentrionale	*Astragalus adsurgens*	Goa *et al*., 2003
M. temperatum	*Astragalus adsurgens*	Goa *et al*., 2003
M. tianshanense	***Glycyrrhiza pallidflora,*** *Swansonia*, *Glycine*, *Caragana*, *Sophora*	Chen *et al*., 1995
Rhizobium		
R. etli	***Phaseolus vulgaris,*** *Mimosa affinis*	Segovia *et al*., 1993; Wang *et al*., 1999a
R. galegae	***Galega orientalis, G.officinalis***	Lindstrom, 1989;
R. gallicum	***Phaseolus vulgaris,*** *Leucaena*, *Macroptilium*, *Onobrychis*	Amarger *et al*., 1997
R. giardini	***Phaseolus vulgaris,*** *Leucaena*, *Macroptilium*	Amarger *et al*., 1997
R. hainanense	***Desmodium sinuatum,*** *Stylosanthes*, *Vigna*, *Arachis*, *Centrosema*	Chen *et al*., 1997
R. huautlense	***Sesbania herbacea***	Wang *et al*., 1998
R. indigoferae	***Indigofera***	Wei *et al*., 2002

R. leguminosarum • *bv. trifolii* • *bv. viciae* • *bv. phaseoli*	• ***Trifolium*** • ***Lathyrus, Lens, Pisum,*** *and* ***Vicia*** • ***Phaseolus vulgaris***	Dangeard, 1926; Jordan, 1984
R. mongolense	***Medicago ruthenica,*** *Phaseolus vulgaris*	van Berkum, *et al*., 1998
R. sullae	***Hedysarum coronarium***	Squartini, *et al*., 2002
R. tropici	***Phaseolus vulgaris,*** *Dalea*, *Leucaena*, *Macroptilium*, *Onobrychis*	Martinez-Romero *et al*., 1991
R. yanglingense	***Amphicarpaea trisperma, Coronilla varia, Gueldenstaedtia multiflora***	Tan *et al*., 2001
Sinorhizobium		
S. abri	***Abrus precatorius***	Ogasawara *et al*., 2003
S. adhaerens	unknown	Willems *et al*., 2003; Young, 2003
S. americanus	***Acacia spp.***	Toledo *et al*., 2003
S. arboris	***Acacia senegal, Prosopis chilensis***	Nick *et al*., 1999
S. fredii	***Glycine max***	Scholla *et al*., 1984; Chen *et al*., 1988
S. indiaense	***Sesbania rostrata***	Ogasawara *et al*., 2003
S. kostiense	***Acacia senegal, Prosopis chilensis***	Nick *et al*., 1999
S. kummerowiae	***Kummerowia stipulacea***	Wei *et al*., 2002
S. medicae	***Medicago truncatula, M. polymorpha, M.orbicularis***	Rome *et al*., 1996
S. meliloti	***Medicago, Melilotus, Trigonella***	Dangeard, 1926; de Lajudie *et al*., 1994
S. morelense	***Leucaena leucocephala***	Wang *et al*., 2002
S. sahelense	***Acacia, Sesbania***	de Lajudie *et al*., 1994; Boivin & Giraud, 1999
S. terangae	***Acacia, Sesbania***	de Lajudie *et al*., 1994; Lortet *et al*., 1996
S. xinjiangense	***Glycine max***	Peng *et al*., 2002

As mentioned above, rhizobia are all member of the α-subset of proteobacteria, but some species of β-Proteobacteria, such as *Burkholderia* and *Ralstonia* have been found to nodulate legumes (Moulin et al., 2001; Chen et al., 2003). More recently a member of the γ-proteobacteria has been found that also nodulates legumes (Benhizia *et al*., 2004).

There has been extensive study of the *Rhizobium*-legume symbiosis, identifying many of the rhizobial genes required for nodulation and nitrogen fixation. However, the genes allowing growth and survival of free-living *Rhizobium* in the soil remain largely unknown. Identifying molecules that have effects on bacterial growth in the rhizosphere

and determining the genes that are involved in responding to these factors is vital to understanding how the bacteria develop in this environment. This could also lead to a greater understanding into the *Rhizobium*-plant symbiosis and the relationship between the two distinct growth states (as a free-living saprophyte in the soil and in a symbiotic relationship with leguminous plants).

1.1.2. Symbiosis

As mentioned above, rhizobia are taxonomically diverse members of the α-sub-division of the proteobacteria and can exist in two states: as a free-living saprophyte in the soil and in a symbiotic relationship with leguminous plants. The latter interaction begins with a specific molecular signal exchange between the legume and the free-living *Rhizobium*. Plant roots secrete many different organic compounds into the soil, some of which allow microorganisms to grow in the rhizosphere and include carbohydrates, amino acids, organic acids, vitamins and phenolic derivatives. In terms of symbiosis, flavonoids are the most important of these compounds, as they trigger the induction of bacterial nodulation (*nod*) genes (Redmond *et al*., 1986), although oxygen limitation also plays a key role in symbiotic gene expression (Soupène *et al*., 1995).

1.1.2.1. The *nod* Genes

The *nod* genes can be divided into three classes known as common, host specific and *nodD*.

The first *nod* gene involved in nodulation is the only one that is constitutively expressed, *nodD* (Long, 1989). The protein it encodes for, NodD, is a member of the LysR family of transcriptional activators (Schell, 1993) and causes the transcription of the other *nod* genes, when activated in response to specific plant stimuli. As well as activating the transcription of other *nod* genes, it also regulates its own expression in *R. leguminosarum* (Rossen *et al*., 1985). The N-terminus of NodD is highly conserved, indicating a role in DNA binding. The *nod* genes/operons induced by NodD all contain a highly conserved sequence termed the '*nod* box' where it is believed the N-terminus of NodD binds and initiates transcription of the genes/operons (Hong *et al*., 1987). NodD's C-terminus is more variable and it may have a function involving flavonoid binding (Shearman *et al*., 1986). As shown above (Table 1.1), each *Rhizobium* is able to inoculate only certain legumes. Different *Rhizobium* have different NodD proteins, which respond to different flavonoids specific for different legume types. The ability of NodD to react to specific flavones is a key part that determines the range of plants each species of *Rhizobium* can nodulate; either

broad range, nodulating many different plants or narrow range, nodulating one or few hosts. *R. leguminosarum* bv. *viciae* responds to hesperitin (Laeremans & Vanderleyden, 1998), which is released by pea and vetch roots, whereas *S. meliloti* contains three *nodD* genes, allowing it to respond to a wider array of flavonoids and hence leguminous plants (Honma *et al*., 1990). Mutations in *nodD* can lead rhizobia to respond to a wider range of plant-derived compounds (Burn *et al*., 1987).

The common *nod* genes are *nodABC* and a mutation in any of these prevents the formation of nodules on inoculated plant roots (Nod$^-$ phenotype) (Debruijn & Downie, 1991). The proteins encoded by *nodABC*, NodA (acyl-transferase), NodB (deacetylase) and NodC (*N*-acetylglucosaminyltransferase or chitin synthase) function together to catalyze the synthesis of the monoacylated tetrameric or pentameric chitin core structure required in nodule formation (Spaink, 1996). They are found across the range of *Rhizobium* strains, have no effect on plant specificity and so as such are functionally interchangeable between strains (Kondorosi *et al*., 1984; Djordjevic *et al*., 1985; Fisher *et al*., 1985). The *nodIJ* genes are often considered to be common *nod* genes as they are found in many rhizobial species, including *R. leguminosarum* bv. *viciae*, bv. *trifolii*, *R. etli* and *S. meliloti*. Their products, NodI and NodJ, are involved in the transport of Nod factors and are believed to be part of an ATP-binding cassette (ABC) transporter (Evans & Downie, 1986; Cardenas *et al*., 1996).

Additional *nod* genes appear to affect the nodulation efficiency on a given plant host and also control host-plant sensitivity (Downie & Johnston, 1988); e.g. the main factor that determines host specificity in *R. leguminosarum* is *nodE* (Spaink *et al*., 1989; Spaink *et al*., 1991), whereas in *S. meliloti nodH* and *nodPQ* are responsible for specifying the nodulation of alfalfa (Faucher *et al*., 1989, Roche *et al*., 1991). By extensive genetic and complementational analysis, thirteen different *nod* genes have been identified in *R. leguminosarum* biovar *viciae* and are organised into five operons; *nodABCIJ*, *nodD*, *nodFEL*, *nodMNT and nodO* (Downie & Surin, 1990; van Rhijn & Vanderleyden, 1995). The organization of *nod* genes differs between different species, although *nodDABCIJ* are normally clustered into one organizational unit.

Together these *nod* genes synthesize molecules known as Nod factors, which initiate nodule formation in the plant.

1.1.2.2. Nodule Formation

On production of Nod factors, the bacteria then surround and attach to the root, causing the root to start to curl (Yao & Vincent, 1969). Rhizobia trapped in a curled hair, or

between a hair and another cell, proliferate and begin to infect the outer plant cells, which in turn stimulates plant cells to produce infection threads (Callaham & Torrey, 1981). Bacteria released from infection threads into the cytoplasm of plant cells are surrounded by the plant plasma membrane and then briefly replicate their DNA and divide before stopping both processes (Robertson *et al*., 1978). Finally, the endosymbiotic forms of the bacteria (referred to as bacteroids) make up a new organ of the plant on the root (called the root nodule) and begin to fix nitrogen by the action of the enzyme nitrogenase (Xi *et al*., 2000).

As mentioned, the *Rhizobium*-legume symbiosis is very specific between both the species of rhizobia and the species of legume involved (Long, 1989) and certain bacterial genes will only activate under symbiotic conditions (Long, 1989; Cabanes *et al*., 2000). Nodules formed on different plants by different bacteria nonetheless display striking developmental similarities.

1.1.2.3. Nitrogen Fixation

Once the rhizobia are in the root nodules and have differentiated into bacteroids, most *nod* genes are no longer expressed (Schlaman *et al*., 1991), probably due to the fact that large quantities of Nod factors have been shown to bring about plant defence reactions (Savouré *et al*., 1997), and the bacteroids express nitrogen fixing genes instead. Many species of the family *Rhizobiaceae* possess the ability to fix atmospheric nitrogen, a mechanism that is exclusive to prokaryotes (Long, 1989).

The bacterial genes for nitrogen fixation fall into two broad categories. Those that have homologies amongst organisms (e.g. *Klebsiella* spp.) that can fix nitrogen in the free living state are known as *nif*, whilst those that are unique to symbiotic nitrogen fixation are known as *fix* (Arnold *et al*., 1988; Long, 1989). Mutations within these genes result in *Rhizobium* that are still able to undergo nodulation with their legume hosts, but are unable to fix nitrogen (Nod^+ Fix^- phenotype). While *Rhizobium* fix nitrogen within the nodules produced by their host, a few have been shown to exhibit this property when growing in pure culture; *Azorhizobium caulinodans* and some *Bradyrhizobium* strains exhibit low levels of nitrogen fixation activity in older cultures (Dreyfus, *et al*., 1988).

Nitrogenase is the two-component enzyme complex responsible for the process of nitrogen fixation and is structurally highly conserved throughout nitrogen-fixing bacteria (Dean & Jacobson, 1992). The α and β subunits of component I (the dinitrogenase or MoFe protein) are encoded for by *nifD* and *nifK* respectively; component II (the dinitrogenase reductase or Fe protein) is encoded for by *nifH* (Halbleib & Ludden, 2000). Component I requires a co-factor, encoded by *nifB*, *nifE* and *nifN* genes, which is believed to be the site

of substrate binding and reduction (Dean *et al*., 1993; Kim et al., 1995). The genes *nifH*, *nifM*, *nifQ* and *nifV* are also required for synthesis and maturation of the active enzyme complex (Filler *et al*., 1986; Howard *et al*., 1986; Imperial *et al*., 1984). The enzyme is slow in its action, large in size and can account for up to 30% of the protein present in bacteroids (Haaker & Klugkist, 1987).

Just as there are rhizobial genes that are only expressed in roots as part of symbiosis, there are also plant genes that only are expressed in these conditions; these are called nodulins, which are only transcribed in nodules and include those responsible for the production of leghaemoglobin (Fuller *et al*., 1983). Leghaemoglobin binds oxygen but releases it when the local concentration of O_2 drops below a certain level, thus providing a high flux for the bacteroid to use in respiration, but an environment with low free oxygen that is also required, as nitrogenase is irreversibly inactivated by oxygen (Appleby, 1984). It is the pigment in leghaemoglobin that gives healthy nodules their pink/red colour. Oxygen concentration is the trigger for nitrogen fixation and in *S. meliloti* it is the oxygen sensing *fixLJK* that regulates this process (see section 1.2.2.3).

Nitrogenase made within these bacteroids converts N_2 into NH_3 by reduction of di-nitrogen gas and protons as indicated in the following reaction (Bergersen, 1965).

$$8H^+ + N_2 + 8e^- \rightarrow 2NH_3 + H_2$$

This is a very energy intensive process, as nitrogen is highly inert at normal atmospheric temperature and pressure, and requires a minimum of 16 ATP molecules per molecule of nitrogen reduced, although it has been estimated that the energy requirements under certain circumstances may be as high as 42 ATP molecules per molecule of nitrogen fixed (O'Brian, 1996). Therefore, bacteroids need to respire at a high rate to generate the ATP required for nitrogen fixation, but it is believed that part of their metabolism is shut down on entering symbiosis with the plant (Copeland *et al*., 1989). It has always been believed that the plant provided carbon to the rhizobia (in the form of dicarboxylates) with which to respire, in return for fixed nitrogen (in the form of ammonium); this was later revised to fixed nitrogen (ammonium) and alanine (Allaway *et al*., 2000). However, recent work has shown that the *Rhizobium*-legume symbiosis and the exchange between the two organisms, is more complex (Lodwig *et al*., 2003).

The plant provides the bacteria with an environment with controlled amounts of oxygen, dicarboxylates (taken in by rhizobia via the dicarboxylate transport (DCT) system) and glutamate (or glutamine), which is then used for respiration in the *Rhizobium* via the tricarboxylic acid (TCA), generating the ATP required for nitrogen fixation. In return the bacteria supply the plant with ammonia, aspartate and alanine. Aspartate is converted into

asparagine that is used by the plant, whilst the ammonia serves to replace the glutamate that was donated to the bacteroid (Lodwig *et al*., 2003). This dependence on amino acid cycling between rhizobia and legume has significant consequences on their symbiosis. The plant provides amino acids to the bacteroids, allowing them to shut down ammonium assimilation; likewise the bacteroids must export ammonia to the plant in order to obtain amino acids. The plant cannot dominate the relationship by restricting amino acid availability though, as the bacteroids act as plant organelles and as such are responsible for the aspargine synthesis of the plant. This provides a selective pressure for mutualism between the two organisms, rather than dominance on either side.

This process continues until the plant dies hence releasing its nitrogen into the biomass and contributing to the nitrogen cycle. *Rhizobium*-legume symbiosis is the primary source of fixed nitrogen in land-based systems, providing well over half of the biological source (Zahran, 1999).

1.2. Stress Response

1.2.1. What is a Stress Response?

When studying any living organism, it is important to know how each species grows and responds to certain conditions that can be found in their natural environment. Ascertaining how bacteria respond to environmental signals, or stressful conditions, is a vital part to understanding how those microbes live, thrive and survive. Every bacterium has optimum conditions that make this process easier, however in order to survive in a changing environment (or some other form of stress) the bacteria must be able to adapt. This adaptation is a stress response.

Two types of stress responses operate in microorganisms: the general stress response and specific stress responses. The general stress response is normally controlled by a single, or a few master regulators (Bremer & Krämer, 2000) and provides cross-protection against a wide variety of environmental cues, regardless of the initial stimulant (Hecker *et al*., 1996; Hecker & Völker, 1998). This response is effective in allowing the cell to survive, but it may not be enough to let the cell grow under the stressful conditions (Bremer & Krämer, 2000). Under prolonged stress conditions cells employ specific stress responses, which utilise highly integrated networks of genetic and physiological adaptation mechanisms (Bremer & Krämer, 2000). Usually, there is also a complex relationship between cellular response systems and global regulators, adding another level of control to the cell's emergency stress response and long-term survival reactions (Hengge-Aronis, 1999). Although the above description is usually what happens, not all general responses

occur immediately on stressful stimuli as some activate on entry into stationary phase; likewise some specific stress response are induced as soon as stress is detected.

Stress can take many forms, as shown below.

1.2.2. Examples of Stress Response in *Rhizobium*

1.2.2.1. Osmotic Stress

Water, and its availability, is one of the most vital environmental factors to affect the growth and survival of micro-organisms (Potts, 1994). A change in the external osmolality immediately causes water to be moved along the osmotic gradient, which could result in a cell swelling and bursting (in hypotonic environments – hypo-osmosis), or plasmolysis and dehydration (in hypertonic environments – hyper-osmosis). In general, cells respond more rapidly to hypo-osmosis, than to hyper-osmosis as the risk of bursting is more severe than that of dehydration (Wood, 1999). Maintenance of cell turgor is vital for almost any form of life, as it provides the mechanical force for expansion of the cell wall (van der Heide *et al*., 2001). Cells prevent these two possible outcomes by using active countermeasures to retain a level of cytoplasmic water (Galinski & Trüper, 1994; Miller & Wood, 1996; Poolman & Glaasker, 1998).

Bacteria may detect a change in osmotic pressure by many different ways, including: a change in cell turgor, deformation of cell membrane and changes in the hydration state of membrane proteins; but the key signal is believed to be a change in intercellular ionic solutes (Poolman *et al*., 2002). Potassium ions (K^+) are rapidly transported into cells and accumulated immediately after an osmotic upshift, although they have no known function within bacteria other than to act as a secondary messenger to activate other hyper-osmotic stress responses (Miller & Wood, 1996; Wood, 1999).

Two distinct mechanisms are responsible for initialising the movement of water across a cell membrane under osmotic stress. Simple diffusion is usually adequate in balancing solute levels under low osmotic conditions; however, a much faster transfer of water is achieved through water-specific channels (aquaporins) (Bremer & Krämer, 2000). Aquaporins facilitate rapid water movement across a cell membrane, are abundant in animal and plant cells (Agre *et al*., 1995) and are also present in *Saccharomyces* (Bonhivers *et al*., 1998) and in several bacterial species (Bremer & Krämer, 2000). The *E. coli* aquaporin (AqpZ) (Calamita *et al*., 1995) serves as a model for bacterial water channels and has been shown to mediate rapid and large water fluxes, both into and out of a cell, in response to an osmotic up- or downshift (Delamarche *et al*., 1999). This shows that aquaporins can play an important role in the survival of bacteria under osmotic stress.

A more flexible and versatile osmotic stress response is used by bacteria that generally inhabit environments of varying salinity or water activity (Bremer & Krämer, 2000). This group of bacteria, which include rhizobia, utilise osmoprotectants and compatible solutes. Osmoprotectants are exogenous solutes that stimulate bacterial growth in an environment with high osmolality, whilst compatible solutes are specific organic osmolytes that accumulate in high amounts within a cell to counter a hyper-osmotic gradient, but do not conflict with cellular functions (Miller & Wood 1996). Several compatible solutes have also been shown to stabilise enzyme stability in cells under stressful conditions (Poolman *et al*., 2002). Some compounds can function as osmoprotectants and compatible solutes, whilst some can only function as one of these groups. Many osmprotectants are transported into the cytoplasm where they act as, or are converted into compatible solutes. Compatible solutes can be collected in high concentrations (several moles per litre) (Bremer & Krämer, 2000). Since only a limited number of compounds meet the required criteria, the same compatible solutes are employed against hyper-osmosis throughout various bacteria (Braun, 1997). Different compatible solutes work more effectively then others within their bacteria; e.g. glycine betaine is more effective in *S. meliloti* and *E. coli* than it is in *Bacillus subtilis* (Botsford & Lewis, 1990); whilst proline is a compatible solute in *E. coli* but not in rhizobia (Gloux & LeRudulier, 1989). Also, the strength of hyper-osmolarity can determine how the bacteria respond and what osmoprotectants are used (Breedveld *et al*., 1990; Gouffi *et al*., 2000). In a similar way, the compound used to bring about hyper-osmosis can stimulate a stronger stress response compared to others; e.g. generally sodium chloride (NaCl) induced hyper-osmosis causes a stronger stress response then sucrose induced hyper-osmosis, due to the ionic nature of NaCl (Gloux & Le Rudulier, 1989). Compatible solutes can either by synthesised *de novo*, when required by the bacteria, or they are accumulated from the environment, depending on the situation. Under conditions where osmotic upshift is severe and immediate, cells do not have the time required to synthesise compatible solutes and so must acquire them from their environment. In general, rhizobia do not, or cannot, synthesise their own solutes so use uptake systems to accumulate them (Gloux & Le Rudulier, 1989).

The two most studied systems used to transport compatible solutes are ProP and ProU in *E. coli*. ProP is a secondary transporter that is predominately controlled post-translationally and is strongly activated by an osmotic upshift; however, transcription of the gene that encodes for it (*proP*) is also enhanced under hyper-osmosis but only two- to five-fold (Csonka & Epstein, 1996). ProQ is known to be required post-translationally for the optimum functionality of ProP, but the actual function ProQ has is unknown (Smith *et al*.,

2004). ProU is an ABC transporter encoded by the *proU* operon (*proVWX*) that is transcriptionally induced more than 100-fold under hyper-osmosis (Csonka & Epstein, 1999). ABC transporters have been identified in eukaryotes and prokaryotes. They can transport a wide variety of substrates including amino acids, sugars, inorganic ions, polysaccharides and peptides (Walshaw, 1995). In eukaryotes, ABC transporters have been linked to cystic fibrosis and multidrug resistance (Higgins & Linton, 2004). ABC transporters are made up of four domains; consisting of two integral membrane permeases (IMP) and two ATP-binding cassette (ABC) domains that energise the transport (Fig. 1.1). Both the IMPs and ABCs can be homodimers or heterodimers. As well as the four core domains, prokaryotic ABC transporters involved in solute uptake use a substrate-binding protein (SBP), which is found in the cell's periplasm.

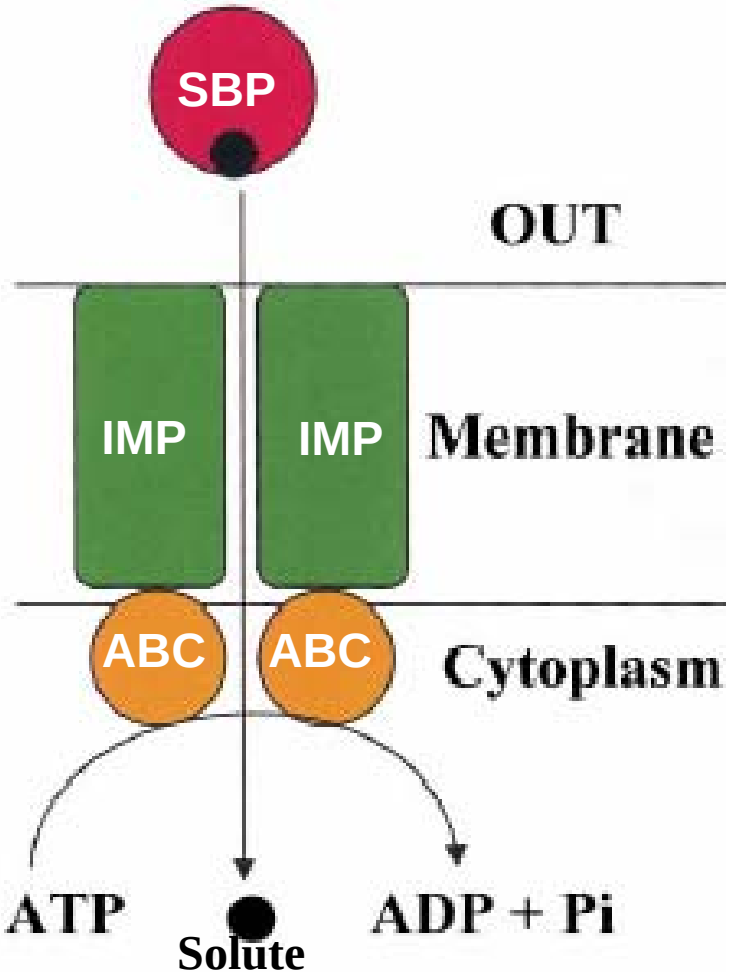


Figure 1.1. Prokaryotic ABC Transporter Schematic.

SBPs are essential for optimum uptake via the transporter with which they are associated, as shown by spheroplast studies (Heppel, 1969); however, some mutants show limited uptake in the absence of SBPs (Petronilli & Ames, 1991). Spheroplasts are

essentially cells with their outer membrane and cell wall removed, so all periplasmic components escape. This means spheroplasts contain no SBPs and so transport via ABC transporters is no longer possible. Spheroplasts can still maintain solute exchange with their environment through secondary transporters, as these do not require SBPs. This ability is used to check that spheroplasts are still viable and had not burst during the hyper-osmotic conditions (20% sucrose) used in spheroplast generation (Hosie *et al*., 2002b). SBPs tend to be highly solute specific, although in some cases ABC transporters can interact with more then one SBP (Higgins & Ames, 1981). Generally, the number of ABC transporters in an organism is proportional to its genome size, i.e. the larger the genome, the more ABC transporters. However, α-proteobacteria often have a disproportionately increased number of ABC transporters (Konstantinidis & Tiedje, 2004). It is theorised that as α-proteobacteria are so ecologically versatile, they need a wider range of ABC transporters in order to deal with the various conditions they may encounter.

Neither ProP or ProU respond to osmotic upshift directly, as both require K^+ to be activated (Csonka & Epstein, 1996). ProP and ProU were both named after proline, a strong osmoprotectant of *E. coli* that both systems are responsible for transporting, however, as mentioned above, proline does not act as an osmoprotectant in rhizobia (Gloux & LeRudulier, 1989). *Rhizobium* can use proline betaine, as well as glycine betaine and other betaines (or quaternary amine compounds – QACs) (Bernard *et al*., 1986). QACs are N-methylated compounds and other methylated compounds can also act as compatible solutes, such as some S-methylated compounds, which have been shown to be effective in *E. coli*. The protection provided by S-methylated compounds has been shown to be limited to 3-dimethylsulfoniopropionate (DMSP) in *S. meliloti*, as 2-dimethylsulfonionacetate (DMSA) is in fact toxic (Pichereau *et al*., 1998).

Given the models provided by the ProP and ProU systems, transporters that may import compatible solutes into rhizobia have been investigated and there are potentially two homologous systems in *S. meliloti*. The SBP component of an ABC transporter has been found that is induced on osmotic upshift and specific to glycine betaine (Talibart *et al*., 1990; LeRudulier *et al*., 1991), indicating the presence of a ProU-like system. Whilst a secondary transporter, BetS, is present that is analogous to ProP in that it is also is consitiutively expressed but activated rapidly by hyper-osmosis and transports proline betaine and glycine betaine into stressed cells (Boscari *et al*., 2002). Choline is also collected by *S. meliloti* when under hyper-osmosis, however, it is not accumulated and so is an osmprotectant and not a compatible solute (Brhada *et al*., 2001). This is because the *Rhizobium* has the ability to break choline (or choline-O-sulphate) down into glycine

betaine, converting it into the compatible solute. This is accomplished by the *betICBA* operon and its products; a choline sulfatase (*betC*), a choline dehydrogenase (*betA*), a betaine aldehyde dehydrogenase (*betB*) and the regulator of this system (*betI*) (Mandon *et al*., 2003). There are three choline transport systems in *S. meliloti* but none of these are induced by an osmotic upshift (Dupont at *al*., 2004). Hyper-osmosis also effects bacteroids in root nodules and is detrimental to the nitrogen-fixing process; many genes involved in osmotic upshift stress response in free-living bacteria are also required for efficient symbiosis and nitrogen fixation (Nogales *et al*., 2002; Djordjevic *et al*., 2003). Choline, glycine betaine and proline betaine uptake have been observed in *S. meliloti* bacteroids, alleviating stress and restoring nitrogen fixation (Fougère & LeRudulier, 1990a & b). Choline and other betaines are readily found in plants, providing a source for nodules under stress (Fougère & LeRudulier, 1990b; Pichereau *et al*., 1998). Rhizobia, unlike *E. coli*, can use choline, glycine betaine and proline betaine as carbon and/or nitrogen sources, although the metabolic pathways associated with these compounds are all repressed when cells are growing under hyper-osmotic stress (Miller & Wood, 1996).

Trehalose is another compatible solute used by rhizobia, however, this compound is normally synthesised by the stressed bacteria instead of being transported into the cells (Breedveld *et al*., 1990) although some import does occur (Miller & Wood, 1996). Like choline and the betaines, trehalose can be used by *Rhizobium* as a carbon source, although genes involved in this metabolic role are repressed under an osmotic upshift (Breedveld *et al*., 1993). Glutamate is very similar to trehalose: it can be used as a compatible solute in rhizobia, it is synthesised in stressed cells instead of transported in and is a carbon source, but again these metabolic genes are repressed under hyper-osmosis (Botsfold & Lewis, 1990; Breedveld *et al*., 1990; Gonzalez-Gonzalez *et al*. 1990). Glutamate appears to be accumulated at a lower osmotic threshold to trehalose though, showing that bacteria use different solutes in response to different degrees of stress (Miller & Wood, 1996). N-acetylglutaminylglutamine amide (NAGGN) is also synthesised by *S. meliloti* under osmotic upshift but in higher concentrations than trehalose, however, NAGGN cannot be used as a carbon or nitrogen source (Smith & Smith, 1989). The biosynthetic pathway of NAGGN has not been fully characterised though an N-actetylglutaminylglutamine sythetase has been identified that is transcriptionally induced under hyper-osmosis and is stimulated by the presence of K^+ (Miller & Wood, 1996). The role of NAGGN as a compatible solute appears to be limited to *S. meliloti* as it is not accumulated in other rhizobia (including *R. leguminosarum* and *R. fredii*) (Smith & Smith, 1989).

Pipecolic acid (PIP), an imino acid, has been shown to act as an osmoprotectant in *S. meliloti* as its presence promotes the restoration of growth in bacteria under severe hyper-osmosis. Interestingly, both the $_{D}$- and $_{L}$- isomers of PIP must be supplied together to be effective; it is believed that only $_{D}$-PIP accumulates to relieve osmotic pressure whereas $_{L}$-PIP participates in the synthesis of glutamate and NAGGN (Gouffi *et al*., 2000). Ectoine, a tetrahydropyrimidine, is another osmoprotectant in several rhizobial species (and *E. coli*) and has been shown to be as effective as proline betaine in improving the growth of *S. meliloti* under an osmotic upshift (Talibart *et al*., 1994). It is believed to be imported into cells by an ABC transporter (and by a separate system to that of glycine betaine), but it does not accumulate in *Rhizobium* and instead stimulates the synthesis of trehalose, glutamate and NAGGN (Talibart *et al*., 1994). Ectoine can also be used by *Rhizobium* as a carbon and/or nitrogen source (Miller & Wood, 1996).

Disaccharides have also been recognised as osmoprotectants; however, this function has not been investigated extensively as disaccharides are commonly used as growth substrates and also as a means to induce hyper-osmosis experimentally (Gouffi *et al*., 1999). These sugars do not accumulate in the cell and instead act as a driving force for the *Rhizobium* to grow and divide, employed if compatible solutes begin to deplete. Interestingly, only certain disaccharides can function in this way; e.g. sucrose can whereas lactose cannot (Gouffi *et al*., 1999). Further to this, a glucose-6-phosphate dehydrogenase (encoded by the *zwf* gene) is required for sucrose (and trehalose) to be efficient osmoprotectants, but not for ectoine or glycine betaine (Barra *et al*., 2003). It is believed that the glucose-6-phosphate dehydrogenase involvement in the hyper-osmotic stress response results from the production of reactive oxygen species that may have been produced during the osmotic upshift. This suggests a cross over between hyper-osmosis and oxidative stress response.

Although most work has been conducted on *S. meliloti*, there are similarities between the compatible solutes used by this and by *R. leguminosarum*, *R. tropici*, *S. fredii*, *R. galegae*, *Agrobacterium tumefaciens*, *M. loti*, *M. huakuii*, *Agrobacterium rhizogenes*, *R. etli* and *B. japonicum* (Boncompagni *et al*., 1999). This further emphasises the fact that as only a limited number of compounds meet the criteria and that the same compatible solutes are employed against hyper-osmotic stress throughout many bacteria (Braun, 1997). When the surrounding environment drops below hyper-osmotic, the bacteria need to be able to dispose of the acquired compatible solutes quickly or suffer from hypo-osmosis. This is done by either, initiating an efflux system (Wood, 1999), or by the active catabolism of the compounds (Fougère & Le Rudulier, 1990b).

Table 1.2 summarises the above osmoprotectant/compatible solute data specifically for *S. meliloti* as that is the most studied organism, although as mentioned above the solutes and their effects have been recorded in other species of rhizobia.

Table 1.2. Use of Osmoregulatory Solutes by *S. meliloti*. Table shows if solutes are accumulated as compatible solutes or not, how they get into cells (by syntheis or uptake) and if they can be used as a carbon/nitrogen source.

Solute	Accumulated	Synthesized	Transported	C/N
Proline	-	-	-	+
Proline betaine	+	-	+	+
Glycine betaine	+	-	+	+
Choline	-	-	+	+
Trehalose	+	+	+	+
Glutamate	+	+	-	+
NAGGN	+	+	-	-
PIP	-	-	+	+
Ectoine	-	-	+	+
Disaccharides	-	+	+	+

Other changes recorded in rhizobia experiencing hyper-osmosis include changes in the synthesis of extracellular, capsular polysaccharides and lipopolysaccharides (LPS). This alteration in the cell's membrane can impair the *Rhizobium*-legume interaction; LPS is especially important for nodule development (Zahran, 1999). Entry into stationary phase caused by nutrient limitation can also protect cells from osmotic upshift (Thorne & Williams, 1997) and cross over between acid stress, osmotic stress and the responses they induce is very common as a change is osmotic gradient can lead to a change in pH gradient and visa versa (Fujihara & Yoneyama, 1993; Leyer & Johnson, 1993). Hyper-osmotic stress can also lead to nutrient stress (see below).

A study using a genome-wide DNA microarray to monitor the gene expression of *S. meliloti* under osmotic upshift generated by NaCl has recently been conducted and revealed the induction and repression of many genes (Rüberg *et al*., 2003). A decreased expression of flagellum genes (*flaA*, *flaB*, *flaC*, *flaD*) and chemotaxis genes (*mcpZ*, *mcpX*, *cheY1*, *cheW3*) was observed suggesting *S. meliloti* can shut down flagella synthesis (a process requiring large amounts of energy) under adverse conditions, which may help to save more energy for survival; *E. coli* employs a similar strategy (Shi *et al*., 1993). The repression of genes involved in cysteine (*cysK2*), proline (*smc*03253), serine (*serA*, *serC*) and thiamine (*thiC*, *thiE*, *thiG*) biosynthesis as well as those related to iron uptake was also seen. The latter included genes involved in the synthesis and regulation of the siderophore rhizobactin 1021 (*rhbA*, *rhbC*, *rhbD*, *rhbE*, *rhbF*, *rhrA*, *rhtA* and *sma*2339), genes connected to siderophore-type iron transporters (*exbD*, *exbB*), genes encoding a haem compound

transporter (*hmuT*, *hmuS*) and other genes related to iron uptake (*smc*02726, *smb*21431, *smb*21432, *smc*00784, *fhuA1*, *fhuA2*).

In contrast, 14 genes involved in transport of small molecules like amino acids, amines and peptides (*smb*20476, *smb*21572, *dppA2*, *smc*03124, *smc*04293, *smc*04439), anions (*phoD*, *phoE*, *phoT*) and alcohols (*smc*02774) were induced under the osmotic upshift. These genes are most likely involved in the accumulation of compatible solutes (as mentioned above). Genes which are involved in surface polysaccharide biosynthesis and regulation were also found to be induced in response to salt stress (*smb*20825, *exoY*, *exoN*, *exsI*). This supports the data that the synthesis of extracellular, capsular polysaccharides and LPS are altered under hyper-osmosis (see above).

All of the above examples (with the exception of aquaporins) deal with *Rhizobium* encountering hyper-osmosis, as this is more common and the most studied form of water stress, however, cells may also have to deal with hypo-osmosis. Under these conditions bacteria can use mechanosensitive channels that detect tension in the cell membrane and open, allowing water and solutes to escape with little discrimination except for size (Poolman *et al*., 2002). *E. coli* has three main mechanosensitive channels, MscL (mechanosensitive channel of large conductance), MscS (small conductance) and MscM (mini conductance). These proteins are constitutively expressed and open at different membrane tensions, with more tension required to gate the channels with larger conductance, providing the cells with another degree of control to the response they use (Li *et al*., 2002). Recently another mechanosensitive channel (MscK) has been found that is regulated by K^+ and appears to have more of a physiological role than the other channels (Li *et al*., 2002). Initial studies showed no genes with significant sequence identity to any of the *msc* genes in the preliminary genomic sequence of 3841.

1.2.2.2. pH Stress

One of the most important factors that affects the efficiency of symbiosis between rhizobia and plants is the pH of the soil in which they interact (Glenn & Dilworth, 1994). The host plant to any symbiotic *Rhizobium* appears to be the limiting factor for growth in extreme pH, as most legumes require a neutral or slightly acidic soil for growth especially when they depend on symbiotic nitrogen fixation (Zahran, 1999). Every bacterium has its own optimum conditions, under which it grows at its best. Although neutral conditions are generally optimum for bacteria, different species of *Rhizobium* display varying degrees of pH resistance as measured by their ability to grow (not just survive) (Glenn & Dilworth, 1994). Some mutants of *R. leguminosarum* have been reported to be able to grow at a pH as

low as 4.5 (Chen *et al*., 1993), *S. meliloti* is viable only down to pH 5.5 (Foster, 2000), *S. fredii* can grow well between pH 4 – 9.5 but *B. japonicum* cannot grow at the extremes of that range (Fujihara & Yoneyama, 1993). These values are the extremes, when the rhizobia can no longer grow; their growth starts to be impeded between 1 and 2 pH units before those figures, as does their ability to successfully nodulate (Richardson & Simpson, 1989).

The more common, and characterised, pH stress found in soil is acidic as opposed to basic, though defence mechanisms are similar (Fujihara & Yoneyama, 1993). Many Gram-negative and Gram-positive neutralophiles utilise different, and in several cases overlapping, approaches for coping with acid stress. Some inducible systems raise the internal pH of the bacterium, in order to counter any intruding acidic molecules or protonated species. These systems employ ABC systems (see above) and other transport mechanisms to either move acidic molecules out of the cell, or import basic ones (Foster, 2000; Priefer *et al*., 2001). This process is only usually successful if the difference between internal and external pH is of approximately 1 pH unit (Foster, 2000).

Another common response to acid shock is for the bacteria to produce acid shock proteins (ASPs). These contribute to acid tolerance by conferring acid protection on the bacteria but do not alter the internal pH of the cell (Foster, 1993). Some ASPs are induced by the internal pH, whilst others are induced by the external pH (Foster, 2000). There are two main types of ASPs: chaperones and proteases. Chaperones are proteins that either bind to other proteins, preventing them from misfolding under stress; some can also repair proteins that have already misfolded as a result of the acidic conditions (Foster, 1993 & 2000). Proteases are enzymes that break down any misfolded proteins that the chaperones cannot save (Foster, 1993 & 2000). This response generally takes over from the previously described 'pump' mechanism when external pH gets too acidic.

At least twenty genes have been identified in *R. leguminosarum* that are specific to acid stress response in rhizobia and are termed *act* genes (acid tolerance) (Kurchak *et al*., 2001). In order to bring about an acid shock response the bacteria and/or root nodule must have some form of sensing mechanism (Glenn & Dilworth, 1994). Such systems for environmental sensing and response are generally made up of two components: a sensor and a regulator, and one has been found in *S. meliloti*; the genes *actR* and *actS* encode for the regulator and sensor respectively (Tiwari *et al*., 1996b). ActS is the membrane bound product of *actS* that, on detection of external acidity, activates ActR (product of *actR*) via phosphorylation. ActR then goes on to activate the transcription of other acid response genes within the bacterium (Tiwari *et al*., 1996b). Research on *S. meliloti* has shown that calcium can also play a key role in acid tolerance (Tiwari *et al*., 1996a). It has been shown

that some tolerance mechanisms can function under greater stress (i.e. increasing acidity) on addition of increasing amounts of calcium (Tiwari *et al*., 1996a). Although how calcium facilitates this longevity is unknown, it has led to a new means of grouping acid stress response, either calcium repairable or not (Tiwari *et al*., 1996a). In a similar way, glutathione has been shown to be involved in acid tolerance (as well as other stresses) in *Rhizobium tropici*, though it is not known how (Riccillo *et al*., 2000). Perhaps the thiol forms a complex with the reactive protonated species, thus removing their effect over the bacterial cells.

TypA is also required for growth at low pH and is believed to act as a regulator by controlling the phosphorylation of proteins (Kiss *et al*., 2004a). Acid shock has also been shown to induce the pH regulated repressor (PhrR) protein (Reeve *et al*., 1998). It was suggested that exopolysaccharides (EPS) may have a protective role, as *Rhizobium* that produce greater amounts of EPS are able to survive in acidic conditions more successfully than *Rhizobium* that can only produce smaller amounts (Cunningham & Munns, 1984). Potassium and phosphorus are also known to increase in concentration in *R. leguminosarum* cells exposed to acid stress, though the role they play is unknown – possibly secondary messengers (c.f. potassium in hyper-osmotic stress) (Watkin *et al*., 2003).

The responses outlined above are all initiated by the stress, however, some genes are constitutively expressed that function under stress conditions; e.g. *actA* in *S. meliloti*, the first rhizobial acid tolerance gene to be found (Tiwari *et al*., 1996a). The membrane bound product of *actA* is basic and responsible for maintaining internal pH at around 7, when the external pH drops below 6.5 (Tiwari *et al*., 1996a). Mutants defective in this gene are unable to maintain intracellular pH and cannot grow at a pH lower then 6. Although it is know to be expressed, it is unknown what the function of this gene, or its product, is under neutral conditions (Tiwari *et al*., 1996a).

Entry into stationary phase caused by nutrient limitation can protect against acid stress (Thorne & Williams, 1997) and cross over between acid stress, osmotic stress and the responses they induce is very common as a change is osmotic gradient can lead to a change in pH gradient and visa versa (Fujihara & Yoneyama, 1993; Leyer & Johnson, 1993). Acidic stress can also lead to metal stress and nutrient stress (see below).

High pH can also prevent *Rhizobium* from growing and undergoing nodulation, although *R. leguminosarum* bv. *trifolii* has been reported to colonise soil at a higher rate and produce nodulates at a higher frequency in alkaline conditions; it is also known to grow unaffected at pH 11.5 (Zahran, 1999). Homospermidine, a polyamine present in high

concentrations in root nodule bacteria, is also known to accumulate in *B. japonicum* in alkaline conditions, although its function is unknown (Fujihara & Yoneyama, 1993).

1.2.2.3. Oxygen/Oxidative Stress

Given the prominent role of oxygen in the critical function of energy generation as well as in the generation of oxidative stress, it is not surprising that many organisms sense and adapt to changing oxygen concentrations in their environment (Patschkowski *et al*., 2000). Such adaptive strategies are well illustrated in the lifestyles of many bacteria, where oxygen tension serves as an important environmental cue to initiate major changes in gene expression. The oxygen sensitive assimilatory process of nitrogen fixation in rhizobia-legume symbiosis but the need for oxygen in order for cells to respire is an example of the balance that must be carefully controlled (Fischer, 1994).

The root nodules formed as part of symbiosis between bacteria and plant, as described above, provide the ideal and essential microaerobic environment for the nitrogen-fixing bacteria. However, not all genes induced (or repressed) by low oxygen concentrations are involved within a root nodule or in nitrogen fixation (and visa versa), so there is a clear difference between genes regulated by limited oxygen and by the symbiosis process (Becker *et al*., 2004). In *S. meliloti* a two-component regulatory system, encoded by the genes *fixL* (sensor) and *fixJ* (regulator), is responsible for sensing microaerobic conditions and controls the expression of at least 11 other loci, also induced by low oxygen concentrations (Trzebiatowski *et al*., 2001). FixJ controls the expression of many symbiosis specific genes via the activation of the FixK, which is otherwise repressed by FixT (Batut *et al*., 1989; Foussard *et al*., 1997). FixK is the microaerobic regulatory protein for the *fixNOQP* operon, which is essential for symbiotic nitrogen fixation (Lopez *et al*., 2001). Although the FixL/FixJ system is responsible for the regulation of most of the genes in *S. meliloti* under low oxygen conditions, mutational analysis has found some genes/operons that can activate independently of the sensor/regulator, indicating the presence of at least one other regulatory system or level of control (Trzebiatowski *et al*., 2001). While the *S. meliloti* requires FixLJK to regulate nitrogen fixation, the mechanism is much more complex in *R. leguminosarum*; e.g. *R. leguminosarum* bv. *viciae* VF39 has two FixK/Fnr-like genes but no FixJ, and a FixL homologue that is a hybrid of FixL and FixJ that performs the functions of both proteins (Patschkowski *et al*., 1996; Lopez *et al*., 2001).

Oxidative stress is quite different to changes brought about by oxygen gas; it is caused by increased levels of superoxide anions ($O_2\bullet^-$), hydrogen peroxide (H_2O_2) or hydroxyl radicals (HO•) (Storz & Zheng, 2000). These reactive species, which can be

generated by exposure to radiation, metals and redox-active drugs, can lead to the damage of all cellular components by a similar mode of action to protonated species generated acidic stress (see above). In addition, animals, plants and microorganisms all possess mechanisms to specifically generate oxidants as a defence against bacterial invasion (Storz & Zheng, 2000). As legumes produce this defensive response regardless to the bacteria 'attacking' it, *Rhizobium* must overcome this stress in order to undergo symbiosis (Santos *et al*., 2001). The general response for bacteria against oxidative stress is to produce reductases and other compounds, such as catalases. These counter the oxidative nature of the reactive species, preventing them from damaging the cell (Storz & Zheng, 2000).

S. meliloti contains three genes that encode for catalases, *katA* (induced by H_2O_2), *katB* (constitutive) and *katC* (induced on entry to stationary phase) (Sigaud *et al*., 1999). KatA is involved with protecting free-living cells from oxidative stress, whilst KatB and KatC are required for cells to successfully by-pass plant defence systems and undergo the nodulation process (Jamet *et al*., 2003). Catalase activity in *R. leguminosarum* bv. *phaseoli* is dependent on growth phase, as stationary phase cells were more resistant to lethal concentrations (3mM) of H_2O_2 than those in exponential phase. Also cells that were exposed to low-levels (200μM) of H_2O_2 were more resistant to later exposure to 3mM H_2O_2 (Crockford *et al*., 1995). Oxidative shock has also been shown to induce the PhrR repressor protein (Reeve *et al*., 1998).

Rhizobium cells have been shown to be resistant to oxidative shock in response to other stresses, as part of a cross-protection, and by the NolR regulator (Thorne & Williams, 1997; Chen *et al*., 2000). Glutathione has also been shown to contribute to the oxidative stress response in *R. tropici*, in the same way as it does in acid tolerance, though it is unknown how (Riccillo *et al*., 2000). Perhaps the thiol forms a complex with the reactive oxygen species, thus removing their effect over the bacterial cells.

1.2.2.4. Metal Stress

Many of the transitional elements function as essential cofactors in metabolic pathways and are required for microbial growth. However, when in excess these, and other metal ions, can lead to harmful effects in bacteria, including enzyme inhibition, biopolymer hydrolysis and uncontrolled redox reactions within the cell (Outten *et al*., 2000). Characterising the minimum and maximum concentrations of each metal is imperative in determining the difference between a standard and a stress response. Stress response genes are induced as metal ion concentrations increase from starvation to toxic levels.

Metal ions are known to cause oxidative stress by the Fenton reaction and whilst there is some knowledge as to how rhizobia counter oxidative stress (see above), there is little known on that caused by heavy metals (Balestrasse *et al*., 2001). It has been shown that there are genes that are expressed under general metal stress (Outten *et al*., 2000) and genes expressed to a specific metal, such as nickel (Singh *et al*., 2001). Responses to some of these metals have been characterised; e.g. high intercellular carbohydrates and large cell inclusions increase the resistance of *R. leguminosarum* to cadmium, copper, nickel and zinc, whilst production of thiols has also been shown to counter heavy metal-induced oxidation (Balestrasse *et al*., 2001; Singh *et al*., 2001). Thiols bind to the metal ions, forming a complex and preventing any cell damage by inactivating the ion's redox potential and have been shown to be effective against cadmium, gold, mercury and lead toxicity (Singh *et al*., 2001).

Some responses are not as well understood; e.g. the previously described acid tolerance gene *actA* seems to be required in *S. meliloti* to bring about copper and zinc resistance, though it is not known why (Tiwari *et al*., 1996a). Furthermore, mutations in the acid-induced genes *actA*, *actR* or *actS* are sensitive to copper and zinc, although this phenotype is calcium repairable (Reeve *et al*., 2002). A connection between acidity and metal toxicity has previously been identified (Keyser & Munns, 1979; Dilworth *et al*., 2001) and an acid-induced copper pump, ActP, has also been found in *S. meliloti* that is controlled by a heavy metal-responsive regulator (HmrR) (Reeve *et al*., 2002). Copper and zinc also bring about the activation of the PhrR repressor (Reeve *et al*., 1998).

In *Rhizobium*-legume symbiosis, it is usually the plant that is the limiting factor with regard to tolerance to metal toxicity. This has been illustrated with aluminium, copper, iron and cadmium (Richardson *et al*., 1988; Balestrasse *et al*., 2001) and can sometimes be the case with other stresses as well (see pH and oxygen stress, above). Nodules can help plants survive because the bacteroids counter metal stress (by thiol inactivation as outlined above), further supporting the fact that symbiosis is mutually beneficial to legume and rhizobia (Balestrasse et al., 2001).

1.2.2.5. Temperature Stress

As has been already mentioned, every bacterium has its own optimum conditions, under which it grows at its best. For most rhizobia, the optimum temperature range for growth is 28 – 31°C, and many are unable to grow at 37°C (Zahran, 1999). Not only do the bacteria themselves have an optimum temperature range, but the processes within them do as well. Temperature affects root hair infection, bacteroid differentiation, nodule structure,

the functioning of the legume root nodule and nitrogen fixation. These processes usually function over a range of ~5°C, but this differs between legumes and is obviously dependant on the environment the rhizobia naturally occupy (Zahran, 1999). Temperature stress is generally divided into two classes: heat shock and cold shock.

Bacterial heat shock is the more characterised of the two (Phadtare *et al*., 2000). The heat shock response is very similar to the acid stress response, in that many proteins with a similar mode of action are synthesised. Heat shock proteins (HSPs) contribute to heat tolerance by conferring heat protection on the bacteria but do not alter the internal temperature of the cell (Yura *et al*., 2000). Like ASPs, there are two main types of HSPs: chaperones and proteases. These work in the same way as the ASPs, as outlined in pH stress above. HSPs, and their regulation, structure and function, have been studied in great detail. Their function appears to be highly conserved between both prokaryotes and eukaryotes (Netzer & Hartl, 1998). Some of these proteins are also vital under normal (non-heat shock) growth conditions (Münchbach *et al*., 1999).

Most bacteria only have a small number of HSPs but *Rhizobium* seem to be an exception to this observation (Michiels *et al*., 1994; Wallington & Lund, 1994); e.g. research has shown that *R. leguminosarum* contains at least three copies of the HSP gene *cpn60* that encode for Cpn60 (or GroEL) (Wallington & Lund, 1994). The Cpn60 protein interacts with another protein called Cpn10 (or GroES) encoded by *cpn10* and a copy of a *cpn10* gene is upstream of at least two of the *cpn60* genes (Wallington & Lund, 1994). A superfamily of at least six small HSPs, one of which is essential for symbiosis, has also been located throughout the *Rhizobium*, though initially in *B. japonicum* (Münchbach *et al*., 1999; Natera *et al*., 2000). It is unclear why rhizobia possess so many HSPs in comparison to other bacteria, it may be so they can bring about an immediate response in times of heat stress, minimising damage caused. It may also be that the genome of *Rhizobium* contains many copies of many genes that could be homologues or paralogues; this indicates that a high level of redundancy in some systems may be present.

Like acid tolerance, heat tolerance can also induce cross protection against other stresses, indicating it is a can be part of a general stress response. However, the rhizobial superfamily of small HSPs mentioned above have not been shown to offer any cross protection indicating that a specific response to heat stress is also present (Münchbach *et al*., 1999).

Cold shock is essentially the opposite of heat shock. Instead of proteins misfolding and denaturing, cells undergoing cold shock have to contend with a loss of membrane and cytosol fluidity and with the stabilisation of secondary structures of RNA/DNA (Phadtare *et*

al., 2000). RNA/DNA stabilisation leads to a decrease in the efficiency of translation, transcription and replication. Bacterial cold shock response is an immediate and transient response to the temperature downshift. This is followed by low temperature adaptation that allows continued growth at low temperatures (Panoff *et al*., 1997). Generally, bacteria overcome loss of fluidity by increasing the amount of unsaturated fatty acids in the membrane phospholipids (Phadtare *et al*., 2000).

Cold shock response also leads to the production of many cold shock proteins (CSPs). Just like ASPs and HSPS, these too are mainly chaperones and proteases (Phadtare *et al*., 2000). However, instead of protecting against the misfolding of proteins, the CSP chaperones are primarily used to bind to RNA/DNA to prevent stabilisation and allow translation and transcription to proceed as usual (Phadtare *et al*., 2000). CspA is an RNA chaperone and a major CSP found in many bacteria (Jiang *et al*., 1997). A CspA homologue is present in *S. meliloti* and is induced following a temperature downshift from 30 to 15°C, along with the three rRNA (*rrn*) operons. It is unknown what function the genes and products of the *rrn* operons or CspA have in response to cold shock, as mutations made in these genes showed no change in cell phenotype at 15°C compared to the wild-type (O'Connell *et al*., 2000; Gustafson *et al*., 2002). TypA is also required for growth at low temperatures and is believed to act as a regulator by controlling the phosphorylation of proteins (Kiss *et al*., 2004a).

Both HSPs and CSPs have been shown to be induced by other stresses, as part of a cross-protection, and by the NolR regulator, which is more associated with the nodulation process (Thorne & Williams, 1997; Chen *et al*., 2000).

1.2.2.6. Starvation Stress

When in their natural environment, rhizobia are rarely in conditions with a constant nutrient supply, sometimes, albeit even more rarely, nutrients are in abundance. More often than not, the bacteria are starving with no, or only sub-optimal levels of, nutrients present. When growing at sub-optimal levels of nutrients bacteria express appropriate cellular responses and many different things can be termed as nutrients for bacteria, including energy, carbon, nitrogen, phosphorus, sulphur and other trace compounds (Ferenci, 2001). From this list carbon, nitrogen and phosphorus limitation are the best studied within rhizobia. Non-growth is ordinarily the rule as opposed to the exception in most natural environments with the majority of bacteria being in a nutrient-limited stationary phase (the stringent response) (Wells & Long, 2002). *Rhizobium* have therefore developed a number of mechanisms that allow them to survive even long-term nutrient starvation and then to

resume growth once conditions are favourable again (Thorne & Williams, 1997; Djordjevic *et al*., 2003).

Spore formation is a strategy used by some bacteria (*Bacilli*, *Clostridia*, *Myxococcus* and *Azospirilli*) to survive these periods, however most bacteria, which include the *Rhizobiaceae*, lack this survival mechanism (Thorne & Williams, 1997; Summers *et al*., 1998; Davey & de Bruijn, 2000). In the majority of cases, nutrient starvation is not a rapid occurrence, so *Rhizobium* are able to enter stationary phase on detection of the early stages of nutrient limitation. If nutrients are removed quickly, *R. leguminosarum* in exponential growth phase are unable to adapt and only cells that are in stationary phase will survive (Thorne & Williams, 1997). Many mechanisms exist which detect and regulate the entry into stationary phase in nutrient deprived bacteria. These include a novel two-component sensor-regulator system pairing a TspO homologue (regulator) and the microaerobic sensor FixL, in *S. meliloti*. This mechanism is required for full expression of a nutrient-deprivation induced (*ndi*) locus, although it is unknown what function the genes, and products, of the *ndi* locus have in response to starvation. The *ndi* locus is also activated under osmotic stress and oxygen limitation, which further illustrates the cross-induction and overlapping responses that can occur in rhizobia (Davey & de Bruijn, 2000).

On entering stationary phase, cells undergo many changes. The most important one is that cell metabolism slows to an almost halt. This is due to the obvious fact that the bacteria have little or no nutrients available to metabolise. The other important process that occurs is the global control of mRNA in the *Rhizobium*. The mRNA pool is stabilised and its turnover is retarded as a means of maintaining gene expression. This limits the production of novel proteins and other compounds so that cells can then stabilise their biomass in order to survive the nutrient limitation (Thorne & Williams, 1997; Summers *et al*., 1998). Cells have been shown to survive nutrient-limited for up to two months and exit from the stringent response is rapid, taking as few as five hours for *R. leguminosarum* to enter exponential phase growth on availability of nutrients (Thorne & Williams, 1997)

Nutrient-limited conditions have been shown to have an effect on other rhizobial processes. *B. japonicum* has been shown to have improved root association with soybean under nitrogen-starved conditions (López-García *et al*., 2001). Osmotic stress and soil acidity can also lead to nutrient-limited conditions, as the demands for calcium and phosphorus increase under these conditions. This in turn can lead to *Rhizobium* being unable to attach to root hairs during the nodulation process as the calcium-dependant cell surface components can not function (Zahran, 1999). Also, the stringent response has shown to cause cross protection in *Rhizobium*, as they are also tolerant to pH, heat, osmotic

and oxidative stress (Thorne & Williams, 1997; Summers *et al*., 1998) and has also been shown to affect many factors involved with symbiosis in *S. meliloti* (Wells & Long, 2002).

1.3. Research Objectives

1) The primary goal of this project is to identify markers of stress induction. This involves determining the conditions that would qualify as stressful to *R. leguminosarum* 3841. As mentioned above, there can be a fine line between a stress response and a 'normal' response in bacteria. Ideally markers can be isolated that induce under one or two of the common stresses encounter by rhizobia.

2) Are these markers cross induced? Do they respond to one stress alone, as part of a specific stress response, or do they respond to many stresses, as part of a general stress response? Determining when in the bacterial cell cycle these genes activate will also aid in the determination of the type of stress response. When available, the genomic sequence of *R. leguminosarum* 3841 will allow the identification of the genes, associated with each reporter plasmid. This will allow potential operons to be recognised and may reveal how stress-induced genes are arranged within the genome of *R. leguminosarum*. Identification of stress-induced genes/operons will also allow similar genes/operons to be discovered, which may then be investigated in conjunction with those found with the reporter plasmids.

3) How are these markers regulated? The potential regulatory systems employed by 3841 to bring about and control stress response may be identified using the stress-induced markers. Determining if there is a global regulator of stress response in 3841 would be a major development in the understanding of rhizobia. It is already known that *Rhizobium* do not contain an *rpoS* (Galibert *et al*., 2001) and so cannot produce the stress-related sigma factor for which it encodes (σ^S). Consequently, it is unknown how the bacteria control their stress response; whether there is a one regulator that takes the place of σ^S, or many separate regulators responsible for different stresses.

4) How vital are the stress-induced genes? Once identified, the importance of stress-induced genes to the growth and survival of 3841 can be investigated with mutational studies.

CHAPTER 2: METHODS & MATERIALS

2.1. List of Strains

Strains are described in Table 2.1.

Table 2.1. Strains Used

Strains	Details	Reference
Rhizobium leguminosarum		
3841	Wild-type (wt) *R. leguminosarum*, str^r	Glenn *et al*., 1980
LB3	pOT1 off library in 3841, str^r gen^r	Allaway *et al*., 2001
3841::Tn*5*	3841 mutant library, str^r kan/neo^r	This work
RU1158	Clone isolated from LB3 library with constitutive low GFP expression, str^r gen^r	Allaway *et al*., 2001
RU1159	Clone isolated from LB3 library with constitutive high GFP expression, str^r gen^r	Allaway *et al*., 2001
RU1160	Clone isolated from LB3 library with no GFP expression, str^r gen^r	Allaway *et al*., 2001
RU1505	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1506	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1507	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1508	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1509	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1510	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1511	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1512	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1513	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1514	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1515	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work

RU1516	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1517	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1518	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1519	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1520	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1521	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1522	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1523	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1524	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1525	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1526	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100mM sucrose, str^r gen^r	This Work
RU1527	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1528	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1529	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1530	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1531	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1532	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work

RU1533	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1534	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) buffered at pH 5.75, str^r gen^r	This Work
RU1642	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 100μM aluminium, str^r gen^r	This Work
RU1643	Clone isolated from LB3 library that expresses GFP on AMA (10mM glc, 10mM NH_4) + 30μM copper, str^r gen^r	This Work
RU1848	3841 with pRU1216 that expresses GFP on AMA (10mM glc, 10mM NH4) + 100mM sucrose, str^r gen^r	This Work
RU2184	Insertion mutant of a two component response regulator gene generated with pRU1451, str^r kan/neo^r	This Work
RU2185	Insertion mutant of a carboxypeptidase-related protein gene generated with pRU1336, str^r kan/neo^r	This Work
RU2186	Insertion mutant of a fatty aldehyde dehydrogenase gene generated with pRU1337, str^r kan/neo^r	This Work
RU2187	Insertion mutant of a hypothetical gene generated with pRU1338, str^r kan/neo^r	This Work
RU2188	Insertion mutant of a hypothetical gene generated with pRU1339, str^r kan/neo^r	This Work
RU2189	Insertion mutant of a hypothetical gene generated with pRU1340, str^r kan/neo^r	This Work
RU2190	Insertion mutant of a major facilitator superfamily transporter gene generated with pRU1341, str^r kan/neo^r	This Work
RU2191	Insertion mutant of a *nodT* homologue generated with pRU1342, str^r kan/neo^r	This Work
RU2192	Insertion mutant of the QAT6 ABC gene generated with pRU1343, str^r kan/neo^r	This Work
RU2193	Insertion mutant of a *gntR* orthologue generated with pRU1189, str^r kan/neo^r	This Work
RU2248	RU1736::Tn*5* mutant that grew on AMA (10mM glc, 10mM NH_4) but not on AMA (10mM glc, 10mM NH_4) + 300mM sucrose AMA, str^r kan/neo^r spc^r tet^r	This Work
RU2283	3841::Tn*5* mutant that grew on AMA (10mM glc, 10mM NH_4) + 300mM sucrose but not on AMA (10mM glc, 10mM NH_4), str^r kan/neo^r	This Work
RU2300	3841 with pRU1614, str^r gen^r	This Work
RU2358	RU2184 with pRU862 and pRU1645, str^r gen^r tet^r	This Work
RU2359	RU2184 with pRU862 and pRU1646, str^r gen^r tet^r	This Work

RU2360	RU2184 with pRU862 and pRU1647, str^r gen^r tet^r	This Work
RU2361	RU2184 with pRU862 and pRU1683, str^r gen^r tet^r	This Work
RU2385	3841 with pRU1700, str^r gen^r	This Work
RU2410	Tn*5* mutant of QAT1 operon, str^r kan/neo^r	This Work
RU2411	Tn*5* mutant of QAT2 operon, str^r kan/neo^r	This Work
RU2412	Tn*5* mutant of QAT5 operon, str^r kan/neo^r	This Work
RU2415	Transductant of RU2248 in 3841 background, str^r kan/neo^r spc^s tet^s	This Work
RU2416	Transductant of RU2372 in 3841 background, str^r kan/neo^r tet^s	This Work
RU2422	Transductant of RU2283 in 3841 background, str^r kan/neo^r	This Work
RU2423	3841 with pRU1758, str^r gen^r	This Work
RU2424	3841 with pRU1759, str^r gen^r	This Work
RU2425	3841 with pRU1760, str^r gen^r	This Work
RU2426	3841 with pRU1761, str^r gen^r	This Work
RU2427	3841 with pRU1762, str^r gen^r	This Work
RU2428	3841 with pRU1763, str^r gen^r	This Work
RU2429	3841 with pRU1764, str^r gen^r	This Work
RU2430	3841 with pRU1765, str^r gen^r	This Work
RU2431	3841 with pRU1766, str^r gen^r	This Work
RU2496	Insertion mutant in QAT3 operon generated with pRU1800, str^r kan/neo^r	This Work
RU2497	Insertion mutant in QAT4 operon generated with pRU1801, str^r kan/neo^r	This Work
Escherichia coli		
DH5α	*Escherichia coli*, nal^r	Sambrook et al., 1989
DH5α T1	*Escherichia coli*, nal^r	Invitrogen
TOP10	*Escherichia coli*, str^r	Invitrogen

2.2. List of Plasmids/Cosmids

Plasmids/cosmids are described in Table 2.2.

Table 2.2. Plasmids/Cosmids Used

Plasmid/ Cosmid	Details	Reference
pOT1	Promoter probe vector containing promoterless *gfpuv* and a polylinker between two transcriptional terminators, gen^r	Allaway *et al*., 2001
pRK2013	Helper plasmid required to allow DH5α to conjugate with 3841, kan/neo^r	Figurski & Helinski, 1979
pCR® 2.1-TOPO®	TA PCR Cloning vector containing *lacZ*, amp^r kan/neo^r	Invitrogen
pBluescript® II SK$^-$	pUC19 derivative containing *lacZ*, amp^r	Stratagene
pK19mob	Used to generate insertion mutants in 3841 containing *lacZ*, kan/neo^r	Schäfer *et al*., 1994
pRK415	Broad host-range cloning cosmid vector, tet^r	Keen *et al*., 1988
pSUP202-1::Tn*5*	Contains *mob* site, used for Tn*5* mutagenesis, amp^r kan/neo^r	Simon *et al*., 1983
pRU843	pOT derivative isolated from RU1507, gen^r	This work
pRU844	pOT derivative isolated from RU1508, gen^r	This work
pRU845	pOT derivative isolated from RU1509, gen^r	This work
pRU846	pOT derivative isolated from RU1510, gen^r	This work
pRU847	pOT derivative isolated from RU1511, gen^r	This work
pRU848	pOT derivative isolated from RU1512, gen^r	This work
pRU849	pOT derivative isolated from RU1513, gen^r	This work
pRU850	pOT derivative isolated from RU1514, gen^r	This work
pRU851	pOT derivative isolated from RU1515, gen^r	This work
pRU852	pOT derivative isolated from RU1516, gen^r	This work
pRU853	pOT derivative isolated from RU1517, gen^r	This work
pRU854	pOT derivative isolated from RU1518, gen^r	This work
pRU855	pOT derivative isolated from RU1519, gen^r	This work
pRU856	pOT derivative isolated from RU1520, gen^r	This work
pRU857	pOT derivative isolated from RU1521, gen^r	This work
pRU858	pOT derivative isolated from RU1522, gen^r	This work
pRU859	pOT derivative isolated from RU1506, gen^r	This work
pRU860	pOT derivative isolated from RU1505, gen^r	This work
pRU861	pOT derivative isolated from RU1523, gen^r	This work
pRU862	pOT derivative isolated from RU1524, gen^r	This work
pRU863	pOT derivative isolated from RU1525, gen^r	This work
pRU864	pOT derivative isolated from RU1526, gen^r	This work
pRU865	pOT derivative isolated from RU1527, gen^r	This work
pRU866	pOT derivative isolated from RU1528, gen^r	This work
pRU867	pOT derivative isolated from RU1529, gen^r	This work
pRU868	pOT derivative isolated from RU1530, gen^r	This work
pRU869	pOT derivative isolated from RU1531, gen^r	This work
pRU870	pOT derivative isolated from RU1532, gen^r	This work
pRU871	pOT derivative isolated from RU1533, gen^r	This work

pRU872	pOT derivative isolated from RU1534, gen^r	This work
pRU1059	pOT derivative isolated from RU1642, gen^r	This work
pRU1060	pOT derivative isolated from RU1643, gen^r	This work
pRU1064	pJP2 based reporter vector containing a promoterless *gusA* and *gfpuv*, tet^r	Karunakaran, unpublished
pRU1097/D-TOPO®	Reporter vector containing a promoterless *gfp* mut3.1, gen^r	Invitrogen
pRU1181	p430 & p431 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1189	Internal region of a *gntR* orthologue extracted from pRU1181 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1195	p453 & p454 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1196	p455 & p456 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1197	p457 & p458 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1198	p459 & p460 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1199	p461 & p462 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1200	p463 & p464 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1201	p465 & p466 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1202	p467 & p468 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1203	p469 & p470 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1216	pRU857 with fragment excised via *Pst*I to remove 500bp from the 3' end of the insert, gen^r	This work
pRU1336	Internal region of a gene encoding a carboxypeptidase-related protein extracted from pRU1196 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1337	Internal region of a gene encoding a fatty aldehyde dehydrogenase extracted from pRU1197 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1338	Internal region of a gene encoding a hypothetical protein extracted from pRU1198 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1339	Internal region of a gene encoding a hypothetical protein extracted from pRU1199 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work

pRU1340	Internal region of a gene encoding a hypothetical protein extracted from pRU1200 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1341	Internal region of a gene encoding a major facilitator superfamily transporter extracted from pRU1201 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1342	Internal region of a *nodT* homologue extracted from pRU1202 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1343	Internal region of the QAT6 ABC gene extracted from pRU1203 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1451	Internal region of a gene encoding a two component response regulator extracted from pRU1195 via *Eco*RI and ligated into pK19mob, kan/neo^r	This work
pRU1600	Genomic region of RU2248 extracted via *Eco*RI and ligated into pBluescript® II SK$^-$, amp^r kan/neo^r	This work
pRU1601	pRU1600 with 3kb region excised via *Bam*HI to remove an IS50R region, amp^r	This work
pRU1611	p527 & p637 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1613	p637 & p638 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1614	p623 & p624 PCR product from pRU843 DNA inserted into pRU1097/D-TOPO, gen^r	This work
pRU1645	Two component response regulator gene excised from pRU1611 via *Kpn*I and ligated into pRK415	This work
pRU1646	Two component response regulator gene excised from pRU1611 via *Kpn*I and ligated into pRK415 (in opposite orientation to pRU1645), tet^r	This work
pRU1647	Two component response regulator and kinase genes excised from pRU1613 via *Kpn*I and ligated into pRK415, tet^r	This work
pRU1683	Two component response regulator and kinase genes excised from pRU1613 via *Kpn*I and ligated into pRK415(in opposite orientation to pRU1647), tet^r	This work
pRU1700	Probable promoter region of QAT1 operon (p663 & p664 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1758	Probable promoter region of QAT2 operon (p694 & p695 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work

pRU1759	Probable promoter region of a *lysR* orthologue within QAT2 operon (p696 & p697 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1760	Probable promoter region of QAT3 operon (p698 & p699 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1761	Probable promoter region of QAT3 operon (p700 & p701 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1762	Probable promoter region of QAT4 operon (p702 & p703 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1763	Probable promoter region of QAT4 operon (p704 & p705 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1764	Probable promoter region of QAT5 operon (p706 & p707 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1765	Probable promoter region of of a *lysR* orthologue within QAT5 operon (p708 & p709 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1766	Probable promoter region of QAT5 operon (p710 & p711 PCR product) directionally inserted into pRU1097/D-TOPO, gen^r	This work
pRU1784	p718 & p719 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1785	p793 & p794 PCR product from 3841 DNA inserted into pCR TOPO 2.1, amp^r kan/neo^r	This work
pRU1800	Internal region of a gene encoding the QAT3 IMP gene extracted from pRU1784 via *Hin*dIII/*Xba*I and ligated into pK19mob, kan/neo^r	This work
pRU1801	Internal region of a gene encoding the QAT4 ABC gene extracted from pRU1785 via *Hin*dIII/*Xba*I and ligated into pK19mob, kan/neo^r	This work

2.3. Primers Used

Primers are described in Table 2.3.

Table 2.3. Primers Used

Name	Sequence 5' – 3'	Target
IS50R	AGGTCACATGGAAGTCAGATC	Sequencing primer used for Tn*5* mutants
IS50 Downie	GAACGTTACCATGTTAGGAGGT	Primer used for screening Tn*5* mutants
pOTfor	CGGTTTACAAGCATAAAGC	Sequencing primer used for the 5' end of inserts in the pOT1 vector
pOTfor_far	GACCTTTTGAATGACCTTTA	Sequencing primer used for the 5' end of inserts in the pOT1 vector
pOTrev	CATTTTTTCTTCCTCCACTAGTG	Sequencing primer used for the 3' end of inserts in the pOT1 vector
pOTrev_gfp	GAAAATTTGTGCCCATTAAC	Sequencing primer used for the 3' end of inserts in the pOT1 vector
pK19/18A	ATCAGATCTTGATCCCCTGC	Antisense primer used for PCR amplification for region of pK19/18mob
pK19/18B	GCACGAGGGAGCTTCCAGGG	Sense primer used for PCR amplification for region of pK19/18mob
p430	GTCGGGATCGCCGGTTTCGAT	Antisense primer used for PCR amplification of an internal region of a predicted *gntR*-like gene
p431	GCCATGCTGTCCGTCAGCCGC	Sense primer used for PCR amplification of an internal region of a *gntR* orthologue
p453	GGGACGGGCTGTTTCAGGCG	Sense primer used for PCR amplification of an internal region of a predicted two component response regulator gene
p454	TATGGGTCTCGACGACGCTGGT	Antisense primer used for PCR amplification of an internal region of a predicted two component response regulator gene
p455	CCGATTCCGTCACCGAGCAT	Sense primer used for PCR amplification of an internal region of a predicted carboxypeptidase-related gene
p456	GAGTTCGTTGCGGGCGTAAT	Antisense primer used for PCR amplification of an internal region of a predicted carboxypeptidase-related gene
p457	GTTCCGAAGCCGTTGACCAG	Antisense primer used for PCR amplification of an internal region of a predicted fatty aldehyde dehydrogenase gene

p458	CGAGCGATTGACGAGACTGG	Sense primer used for PCR amplification of an internal region of predicted fatty aldehyde dehydrogenase gene
p459	CCTTCCAGTGTTCCTCCACG	Antisense primer used for PCR amplification of an internal region of a predicted hypothetical gene
p460	TTTCAGGGCGGTGGTGCTCT	Sense primer used for PCR amplification of an internal region of a predicted hypothetical gene
p461	GTTCAATGGTTCCGACACAAG G	Antisense primer used for PCR amplification of a region of a predicted hypothetical gene
p462	CGACGAATGGCGATGGCTTC	Sense primer used for PCR amplification of an internal region of predicted hypothetical protein gene
p463	CTGGATCTGGGAACAGGGAT	Antisense primer used for PCR amplification of an internal region of a predicted hypothetical gene
p464	TCGAATGGAACGCCTGCTGG	Sense primer used for PCR amplification of an internal region of a predicted hypothetical gene
p465	CGCACGCTGCTTTTGACCCTG A	Sense primer used for PCR amplification of an internal region of a predicted MFS gene
p466	GAAGGCGGAATGGTTGGACG	Antisense primer used for PCR amplification of an internal region of a MFS gne
p467	ATTCCTCAGCCGTCTGCACT	Antisense primer used for PCR amplification of an internal region of a predicted *nodT*-like gene
p468	GCCTGCCGAGCCTCGATGTC	Sense primer used for PCR amplification of an internal region of a predicted *nodT*-like gene
p469	GATCATGCCGTGATAGGTCT	Antisense primer used for PCR amplification of an internal region of the QAT6 ABC gene
p470	AATTCATGCAGCAACGGGCT	Sense primer used for PCR amplification of an internal region of the QAT6 ABC gene
p473	AGAAAGTATCCATCATGGCT	Sense primer used for PCR amplification of an internal region of Tn*5*
p474	AATTCGTTCTGTATCAGGCG	Antisense primer used for PCR amplification of an internal region of Tn*5*
p496	TAATTAAGTCGACCCTTCACC	Sense primer used for PCR amplification of a region upstream of the insertion site of pRU1097/D-TOPO®

p519	TGCGGTAGAGCGGCGATCCG	Primer used for PCR amplification of region downstream of a predicted carboxypeptidase-related gene
p520	GGTCAACGCATCAATCGAGG	Primer used for PCR amplification of a region upstream of a predicted fatty aldehyde dehydrogenase gene
p521	CTTTCGAGGATTTTCGCCTC	Primer used for PCR amplification of region downstream of a predicted hypothetical gene
p522	TGCATGTCGAGCGGGCAGAG	Primer used for PCR amplification of region downstream of a predicted hypothetical gene
p523	GGATGTCTCGACCAGCCTCT	Primer used for PCR amplification of region downstream of a predicted *nodT*-like gene
p524	GAAGGCGTCGTCGCTGTGTT	Primer used for PCR amplification of a region upstream of the QAT6 ABC gene
p525	CGGCGAAATCCCGCTTCACC	Primer used for PCR amplification for region downstream of a predicted *gntR*-like gene
p527	CGCCCGCCTGCATCCGTCAG	Primer used for PCR amplification for region downstream of a predicted *gntR*-like gene
p528	CCGCAGGAGTGCTGGTAGCG	Primer used for PCR amplification for region upstream of a predicted hypothetical gene
p529	CAATGCGTCGGCTACCTGCT	Primer used for PCR amplification of region upstream of a predicted carboxypeptidase-related gene
p530	CTCTCGCCAGGTCTAGTCGA	Primer used for PCR amplification of a region downstream of a predicted fatty aldehyde dehydrogenase gene
p531	TGGTCGAACGGAGTAGCAAG	Primer used for PCR amplification of region upstream of a predicted hypothetical gene
p532	GGTTCAACTTGGCGGCGACT	Primer used for PCR amplification of region upstream of a predicted hypothetical gene
p533	CTGAACCGAGATGTGCGACG	Primer used for PCR amplification of region upstream of a predicted *nodT*-like gene

p534	CCAGAGGGCGATGTGATTGA	Primer used for PCR amplification of a region downstream of the QAT6 ABC gene
p535	TCAGCAACAGAACGAAAGGA	Primer used for PCR amplification for region upstream of a predicted *gntR*-like gene
p537	AAGACCTTCCACAAAAGGCT	Primer used for PCR amplification for region upstream of a predicted two component response regulator gene
p538	ATCGTCTCGGTCGCCGATAG	Primer used for PCR amplification for region downstream of a predicted hypothetical gene
p545	TCGAAGCGACGCTGACTTAC	Primer used for PCR amplification for region downstream of a predicted MFS gene
p546	TCCTGACAAAGGGCAGAAAT	Primer used for PCR amplification for region upstream of a predicted MFS gene
p564	TCTTGTAGTTCCCGTCATCT	Antisense Primer used for PCR ampliication of an internal region of the *gfpuv* reporter
p623	**CACC**CGTTGTGAAACCTTACT ACG	Primer used for PCR amplification upstream of intergenic region before the QAT6 ABC gene capped with CACC (in bold)
p624	AA**TCTAGA**CGGATGCTCGCC GAAGACTT	Primer used for PCR amplification downstream of intergenic region before the QAT6 ABC gene capped with *Xba*I (in bold)
p637	**GGTACC**TTGCAGTGTGCAGA GGTAGC	Primer used for PCR amplification upstream of a predicted two component response regulator gene & kinase capped with *Kpn*I (in bold)
p638	AAGGCAGCAGGCTGGATTGC	Primer used for PCR amplification downstream of a predicted two component response regulator gene & kinase
p663	**CACC**GAAGGCTGCGATCAGT TGCA	Primer used for PCR amplification upstream of intergenic region before the QAT1 operon capped with CACC (in bold)

p664	CC**TCTAGA**AAGCGGACGGTAGTGCAGTT	Primer used for PCR amplification downstream of intergenic region before the QAT1 operon capped with *XbaI* (in bold)
p694	**CACC**AACGTCATCGCCAGTTCGGT	Primer used for PCR amplification upstream of intergenic region before the QAT2 operon capped with CACC (in bold)
p695	CCC**AAGCTT**CGTTGCCGAGCAGCGTGCCT	Primer used for PCR amplification downstream of intergenic region before the QAT2 operon capped with *Hin*dIII (in bold)
p696	**CACC**ACGTTGCCGAGCAGCGTGCC	Primer used for PCR amplification upstream of intergenic region before a *lysR* orthologue in the QAT2 operon capped with CACC (in bold)
p697	CCC**AAGCTT**AACGTCATCGCCAGTTCGGT	Primer used for PCR amplification downstream of intergenic region before a *lysR* orthologue in the QAT2 operon capped with *Hin*dIII (in bold)
p698	**CACC**TTGAAACCTTTGTCGGCTAT	Primer used for PCR amplification upstream of intergenic region before the QAT3 operon capped with CACC (in bold)
p699	CCC**AAGCTT**ACCTCCGCACTCTGCCAGTT	Primer used for PCR amplification downstream of intergenic region before the QAT3 operon capped with *Hin*dIII (in bold)
p700	**CACC**CCTGCCGTCATCGCCAAGGC	Primer used for PCR amplification upstream of intergenic region before the QAT3 operon capped with CACC (in bold)
p701	CCC**AAGCTT**AACCAGTAGAAACTCGTGCG	Primer used for PCR amplification downstream of intergenic region before the QAT3 operon capped with *Hin*dIII (in bold)
p702	**CACC**AGATGGATGCTGTCAGGGCG	Primer used for PCR amplification upstream of intergenic region before the QAT4 operon capped with CACC (in bold)

p703	CCC**AAGCTT**CGTACTTTTCGC TCAAACTG	Primer used for PCR amplification downstream of intergenic region before the QAT4 operon capped with *Hin*dIII (in bold)
p704	**CACC**AGCCCCGAACTCGACC GATA	Primer used for PCR amplification upstream of intergenic region before the QAT4 operon capped with CACC (in bold)
p705	CCC**AAGCTT**CGTCAGGTCCTC CCACGACA	Primer used for PCR amplification downstream of intergenic region before the QAT4 operon capped with *Hin*dIII (in bold)
p706	**CACC**GACGCCATAGGAGGTC TCGA	Primer used for PCR amplification upstream of intergenic region before the QAT5 operon capped with CACC (in bold)
p707	TTT**AAGCTT**GCCAGGAAACG CTCCTCGAC	Primer used for PCR amplification downstream of intergenic region before the QAT5 operon capped with *Hin*dIII (in bold)
p708	**CACC**GCCAGGAAACGCTCCT CGAC	Primer used for PCR amplification upstream of intergenic region before a *lysR* orthologue in the QAT5 operon capped with CACC (in bold)
p709	TTT**AAGCTT**GACGCCATAGGA GGTCTCGA	Primer used for PCR amplification downstream of intergenic region before a *lysR* orthologue in the QAT5 operon capped with *Hin*dIII (in bold)
p710	**CACC**CGAATTGGGCAAAGGA TCGG	Primer used for PCR amplification upstream of intergenic region before the QAT5 operon capped with CACC (in bold)
p711	TTT**AAGCTT**GGCTCGATGACG GTTTCGCC	Primer used for PCR amplification downstream of intergenic region before the QAT5 operon capped with *Hin*dIII (in bold)
p718	CGAAAGCCTCTTCTCCCCGC	Sense primer used for PCR amplification of an internal region of the QAT3 IMP gene
p719	CCGATCATCGAAGCGACGAC	Antisense primer used for PCR amplification of an internal region of the QAT3 IMP gene

p793	TTCAACTGCGTCGTCGGCGTGTCC	Sense primer used for PCR amplification of an internal region of the QAT4 ABC gene
p794	CAGCGAGGCGATGTCGGAATTCG	Antisense primer used for PCR amplification of an internal region of the QAT4 ABC gene
p839	AGACGGTTGATGTGGCGGAT	Primer used for PCR amplification of region downstream of QAT3 IMP
p840	GGGTTTCGCCTGAGGCAGCC	Primer used for PCR amplification of region upstream of QAT3 IMP
p841	GACCAGGTAGAAAGGCGGAA	Primer used for PCR amplification of region downstream of QAT4 ABC
p842	CGGCCACAAGGGCTTCTCCG	Primer used for PCR amplification of region upstream of QAT4 ABC
pRL120515rh	CCTGTTCTAAGTTGAATAGTG	Primer used for PCR amplification of region upstream of QAT3 IMP
pRL120532rh	GCGACCAGGTAGAAAGGC	Primer used for PCR amplification of region downstream of QAT4 ABC
pRL120753lh	CGCTCCGCTATTGGTTGC	Primer used for PCR amplification of region upstream of QAT5 IMP
RL0511lh	CCAGCCGGTTGATCATCC	Primer used for PCR amplification of region downstream of QAT2 IMP
RL3534lh	TGGGCAAGATTCTCAACGAC	Primer used for PCR amplification of region upstream of QAT1 IMP

2.4. Media & Growth Conditions Used

Media are described in Table 2.4.

Table 2.4. Media Used

Media	Reference	Adaptations (if any)
Tryptone-Yeast (TY)	Beringer, 1974	None
Acid Minimal Salts/Agar (AMS/A)	Poole *et al*., 1994	None
Acidic AMS/A	This Work	20mM MES replaces MOPS (pH 5.75) (2% agar required for solid media)
Sucrose AMS/A	This Work	AMS/A + 100mM Sucrose
NaCl AMS/A	This Work	AMS/A + 100mM NaCl
Mannitol AMA	This Work	AMA + 100mM Mannitol
Aluminium AMA	This Work	AMA + 100μM $AlCl_3$
Copper AMA	This Work	AMA + 30μM $CuCl_2$
Zinc AMA	This Work	AMA + 100μM $ZnCl_2$
Hydrogen Peroxide AMA	This Work	AMA + 1mM H_2O_2
Paraquat AMA	This Work	AMA + 250μM Paraquat dichloride
Rhizobium Minimal Salts (RMS)	Brown & Dilworth, 1975	None
Luria Agar/Broth (LA/B)	Miller, 1972	None
N-free Rooting Solution	Allaway, *et al*., 2000	None

AMS consists of 0.5ml 1M K_2HPO_4; 0.5g $MgSO_4.7H_2O$; 0.2g NaCl; 4.19g MOPS Buffer; 1ml *Rhizobium* solution A (15g EDTA-Na_2; 0.16g $ZnSO_4.7H_2O$; 0.2g $NaMoO_4.2H_2O$; 0.25g H_3BO_3; 0.2g $MnSO_4.4H_2O$; 0.02g $CuSO_4.5H_2O$; 1mg $CoCl_2.H_2O$; made up to a litre with GDW); 1ml *Rhizobium* solution B (12.8g $CaCl_2$; 3.3g $FeSO_4$; made up to a litre with GDW); buffered to pH 7.0 with 1M NaOH; made up to a litre with GDW. 1ml *Rhizobium* solution C (1g Thiamine HCl; 2g D-Pantothenic acid Ca salt; 1mg Biotin; made up to a litre with GDW) is added after autoclaving along with any antibiotics and carbon and nitrogen sources.

R. leguminosarum strains were grown at 27°C on either Tryptone-Yeast (TY) media or on acid minimal salts (AMS) with any adaptations in text. AMS was supplemented with 10mM glucose (glc) as a carbon source and 10mM ammonium (NH_4) as a nitrogen source, unless otherwise stated in text. *E. coli* strains were grown at 37°C in Luria-Bertani (LB) broth.

Strains were routinely stored at -20°C and -80°C in 15% glycerol after snap freezing in liquid nitrogen.

2.5. Antibiotics Used

Antibiotics, fungicide and stains are described in Table 2.5.

Table 2.5. Antibiotics, Fungicide and Stains Used

Antibiotic	Concentration (μg/ml)	
	***E*. coli**	***R. leguminosarum***
Ampicillin	50	-
Gentamycin	10	20
Kanamycin	-	40
Nalidixic Acid	20	-
Neomycin	20	80
Spectinomycin	-	100
Streptomycin	-	500
Tetracycline	10	2 (in AMA) 5 (in TY)
Fungicide		
Nystatin	-	50
Stains		
IPTG	40	-
X-Gal	40	-
X-Glc-A	-	125

2.6. Molecular Techniques

2.6.1. DNA Isolation

Plasmid and Cosmid DNA were isolated using the Wizard® Plus SV Miniprep kit (Promega) following the protocol (using a microcentrifuge) supplied. Chromosomal DNA was isolated using the DNAce Spin Cell Culture kit (Bioline) or the DNeasy Tissue kit (Qiagen). *R. leguminosarum* was freshly grown on a TY Slope and was resuspended into 10ml of TY broth, before being spun down and resuspended in 1ml of TY broth. When the DNAce Spin Cell Culture kit was used, protocol 4 as supplied with the kit, was performed. When the DNeasy Tissue kit was used, Isolation of Total DNA from Animal Tissue protocol (with Appendix D adaptations) as supplied with the kit, was performed. The method described by Chen & Kuo (1993) was also used to isolate chromosomal DNA.

2.6.2. Agarose Gel Electrophoresis, Staining and Extraction

DNA was separated on 1% agarose (Bioline) gels run at ~100V in a Tris Acetate (40mM) EDTA (1mM) (TAE) buffer. DNA samples were loaded onto each gel with a solution of 30% glycerol, 0.25% bromophenol blue used as a loading buffer (x6 concentration). A 1kb ladder, or 1kb plus ladder (both Invitrogen), was used as a size

marker. Gels were stained in ethidium bromide solution (0.8µg/ml) before the DNA was visualised under UV light. DNA was extracted from gels as required using the QIAquick Gel Extraction Kit (Qiagen) following the QIAquick Gel Extraction Kit microcentrifuge protocol as supplied with the kit.

2.6.3. DNA Digests

DNA samples were digested using restriction enzymes supplied by Invitrogen or New England Biolabs, following the guidelines supplied. If required, digested DNA was dephosporylated with Bacterial Alkaline Phosphatase (BAP) (Invitrogen), using the simplified protocol as supplied with the kit.

2.6.4. Ligation

DNA was ligated using T4 DNA Ligase (Invitrogen) following the guidelines supplied. Reactions were carried out overnight at ~15°C. TOPO ligations (pCR® 2.1-TOPO®, or pRU1097/D-TOPO® (Invitrogen)) were carried out following the guidelines supplied with each kit.

2.6.5. Transformation

Plasmid and cosmid DNA was routinely transformed into competent *E. coli* cells (TOP10, DH5α or DH5α T1). DNA was incubated with the competent cells on ice for 30 minutes before being heat shocked at 42°C for 30 seconds (90 seconds if using lab made cells). Cells were then incubated with rotation shaking in SOC media (as supplied with competent cells) for 1 hour at 37°C, before being plated onto LA that contained selective antibiotics.

Cells were either purchased from Invitrogen or made in-labs as follows. Cells were grown in LB to OD_{600}nm of 0.3 - 0.4 before being chilled on ice for 20 minutes. Cells were then washed, resuspended and concentrated (~x20) in cold 0.1M $CaCl_2$. Competent cells were stored at -80°C in 15% glycerol after snap freezing in liquid nitrogen.

2.6.6. Polymerase Chain Reaction (PCR)

Primers were designed using Vector NTI, version 6, 7 or 9 (InforMax), and were obtained from MWG-Biotech (see section 2.3).

DNA was amplified from *Rhizobium* by using the polymerases, BIOTAQ, BIO-X-ACT (long) (Bioline), Pfu turbo (Stratagene) or NEB Taq (New England Biolabs) as described in text. Buffers and additives (if needed), were used as guided in the protocol

supplied with each polymerase. 50pmol of each primer, 0.2mM of each dNTP and 2mM of additional $MgCl_2$ (if needed) was used in each final PCR mix (50µl).

Colony PCR was carried out in the same manner, except no DNA was added to the PCR mix. Instead, the 50µl final reaction mix was split into 10µl aliquots before a toothpick was stabbed into a colony and then stabbed (not mixed or shaken) into each aliquot.

PCR machines used were an Omn-E thermal cycler or a Px2 thermal cycler (both Thermo Hybaid)

PCR conditions were 95°C for 5 minutes (10 minutes for colony PCR) followed by 30 cycles of denaturing at 95°C for 1 minute, annealing at 55-60°C for 1 minute and extension at 72°C for 1 minute (plus 1 minute per kb of DNA product); followed by a final extension at 72°C for 10 minutes, unless otherwise stated in text.

PCR products were visualised on an agarose gel before purification or extraction, if required. If required, products were then cloned in suitable vectors as described in text.

2.6.7. Enzyme/Nucleotide Removal

Enzyme activity was removed from the samples when necessary by heat inactivation or by using the QIAquick Gel Extraction Kit (Qiagen), following the QIAquick Gel PCR Purification Kit microcentrifuge protocol supplied. The latter technique was also used to remove residual nucleotides from completed PCR reactions.

2.6.8. DNA Purification

DNA was purified using phenol extraction or ethanol precipitation.

Phenol extraction was carried out by adding 100% volume of phenol:chloroform:isoamyl alcohol to the DNA sample, which was vortexed before centrifuging at 13,000 rpm for 5 minutes. The aqueous layer was then removed and kept.

Ethanol precipitation was carried out by adding 10% volume of 3M sodium acetate, then 100% volume of 100% ethanol to the DNA sample. The sample was stored at -20˚C for at least 20 minutes. Sample was then centrifuged at 13,000 rpm for 20 minutes. The sample was then washed twice in 70% ethanol before the sample was left to air-dry. If required the sample was resuspended in the suitable amount of nH_2O.

2.6.9. DNA Sequencing

DNA sequencing was performed by either MWG-Biotech, as per their website (http://www.mwg-biotech.com), or by the AMSEQ Sequencing Unit, as per their website

(http://www.ams.rdg.ac.uk/SequencingService/index.html). DNA (~1µg) was provided for sequencing reactions, which was estimated from visualisation of DNA after gel electrophoresis with quantitative standards included.

Sequencing primers were designed using Vector NTI and were obtained from MWG-Biotech (see section 2.3).

DNA sequences were analysed using Vector NTI. Comparisons of DNA and/or amino acid sequences were made using the Artemis software (Genome Research Limited) in tandem with the preliminary genome sequence from the Sanger Institute website (http://www.sanger.ac.uk/Projects/R_leguminosarum). Comparisons were also made using the BLAST (Altschul *et al*., 1997) program hosted by the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/BLAST/) and Vector NTI. Homology searches were made using the BLAST programs from both the NCBI website and from the *Sinorhizobium meliloti* genome website (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/). Analysis of proteins, encoded for by obtained DNA sequences, were also carried out with the BLAST program and with the Pfam program too (http://www.sanger.ac.uk/Software/Pfam/).

2.7. Conjugation

Conjugations to transfer DNA into *Rhizobium* were accomplished via a tri-parental mating using the helper plasmid pRK2013 in *E. coli* 803 (Figurski & Helinski, 1979). Cultures of donor plasmid and helper plasmid were grown in LB plus antibiotics overnight and sub-cultured (100µl into 10ml fresh LB + antibiotics) the next morning. Sub-cultures were incubated shaking at 100rpm (to prevent pili from shearing) for a further 5 hours. 1 ml of each donor and helper plasmid was harvested and washed three times in TY broth to remove any antibiotics from the media, before resuspension in TY broth. *Rhizobium* that had been freshly grown on TY slopes was washed off with 3ml of TY broth. 400µl of donor plasmid, 400µl of *Rhizobium* and 200µl pRK2013 were mixed, span down and resuspended in a final volume of 30µl TY broth. This resuspension was transferred to a sterile filter placed on a TY plate and incubated overnight at 27°C. Bacteria were resuspended from the filter by vortexing with TY broth the next day before being streaked, or a serial dilution was plated, onto selective media.

2.8. Mutagenesis

2.8.1. Tn*5* Mutagenesis

Tn*5* mutagenesis was carried out on *Rhizobium* via conjugation (as for 2.7. as a bi-parental mating, i.e. without the need for pRK2013 helper plasmid), as described by Simon *et al*. (1983). Two Tn*5*::3841 libraries were made, a pooled one and an individual one.

For the pooled library, the conjugation mix was diluted and plated out onto TY (containing streptomycin to select for *R*. *leguminosarum* and kanamycin to select for the Tn*5*) to give ~500 colonies per plate. Once grown, 45 plates were then washed with TY broth, which was pooled, spun down and resuspended in 10ml TY broth + 15% glycerol. The wash was divided into 1ml aliquots and stored at -80°C after snap freezing in liquid nitrogen.

For the individual library, the conjugation mix was diluted and plated onto QTrays filled with TY, (containing streptomycin to select for *R*. *leguminosarum* and kanamycin to select for the Tn*5*) at a serial dilution so 1500 – 2000 cfu grew per tray. Colonies were then individually picked by a QPix colony picking robot (Genetix) into 96 well plates containing TY broth + 15% glycerol with streptomycin, kanamycin and nystatin. Each well therefore contained an individual 3841::Tn*5* clone. Microtitre plates were grown for 48 hours at 27°C shaking at 150rpm. The plates were then stored at -80°C.

2.8.2. pK19mob Mutagenesis

pK19mob (Schäfer *et al*, 1994) was used, using the methods previously described by Prell *et al*., 2002 to generate insertion mutants. PCR was used to amplify an internal region of the target gene. This product was then ligated into pK19mob via pCR® 2.1-TOPO®. The pK19mob-based plasmid was then conjugated into wild-type 3841 (see section 2.7). The conjugation mix was grown on TY with streptomycin and neomycin at 27°C. Any colonies isolated were colony PCR screened to confirm the presence of the pK19mob vector in the target gene.

2.9. Transduction

2.9.1. Phage Propagation

Bacteriophage RL38 (Beringer *et al*., 1978) was used in a method based on that of Buchanan-Wollaston, 1979.

Phage was propagated by mixing 0.1ml of phage (serial dilutions of 10^{-2} to 10^{-5}) with 0.1ml of *Rhizobium* (bacteria were freshly grown up on TY slope for 3 days and

resuspended in 3ml nH_2O). The mixture was added to 3ml of soft TY (50% molten TY: 50% TY broth) at 42°C before being poured on top of a TY agar plate. Plates were incubated for 3 days at 27°C. Plates with the dilution that contained confluent plaques and almost confluent plaques were then eluted with 10ml nH_2O. The solution was filtered through a 0.2µm micropore filter, had 5 drops of chloroform added (to kill any remaining bacteria) and was kept at 4°C until required.

2.9.2. Non-UV Transduction

Wild-type 3841 was grown up on a TY slope and was resuspended in TY broth, of which 200µl was mixed with 1µl of each phage propagation (at serial dilutions of 10^0 to 10^3). This mixture was left for 1 hour at 27°C before being spread onto TY plates containing 80µg/ml kanamycin. Plates were left to grow at 27°C and any colonies isolated were purified on TY plates containing 80µg/ml kanamycin, before being colony PCR screened to confirm the presence of the Tn*5* transposon. If required, antibiotic screening was also used to confirm that no other transposable elements (apart from the Tn*5*) were transduced from donor to recipient.

2.10. GFP-UV Quantification

The reporter gene *gfpuv* was utilized in promoter activity experiments and encoded for GFPuv. GFPuv, an improved version of GFP created by Crameri *et al*. (1995), has an 18 fold increase in fluorescence whilst keeping the excitation and emission maxima of wild-type GFP. GFPuv is also partially optimized for use in prokaryotes by replacing rarely used codons for those preferred by *E. coli*. Also, GFPuv is expressed as soluble, fluorescent, protein whereas the majority of wild-type GFP is expressed in non-fluorescent inclusion bodies.

GFP has an excitation maximum of 395nm, a minor peak at 470nm and an emission maximum of 509nm (Clontech). The protein was originally isolated from the bioluminescent jellyfish *Aequorea victoria* and is only fluorescent when located in the complete protein. When cloned into bacterial cells this fluorescence was visible in colonies and microscopically in single cells under UV light. No cofactors, substrates or other gene products are needed for fluorescent production.

From liquid cultures 200µl aliquots were pipetted into a 96-well microtitre plate and were read in a Genios plate reader (Tecan). Fluorescence was measured using a 490nm excitation filter and a 510nm emission filter, whilst optical density (OD) was measured using a 595nm absorbance filter. Blank readings were also made using 3841 containing no

GFP reporter fusion for the fluorescence value and uninoculated media for the OD value. Specific fluorescence was calculated by dividing the sample's fluorescence value (minus blank value) by its OD value (minus blank value).

Cultures grown on AMA were observed on a TL-33E transilluminator (UVP) with 365nm excitation bulbs and a long wavelength emission filter (420 nm). Fluorescence was scored (-, -/+, +, ++ or +++) by comparison with control strains RU1158 (+), RU1159 (+++) and RU1160 (-).

2.11. Plant Experiments

Pea seeds (*Pisum sativum* cv. Avola) were first surface sterilised to remove any micro-organisms present. This was accomplished by washing seeds in 95% ethanol for 30 seconds, before washing twice with sterile water. Seeds were then soaked in 2% sodium hypochlorite for 5 minutes, before washing at least 5 times in sterile water.

Two litre pots were filled with washed vermiculite (Vermiperl) and 800ml of N-free rooting solution (Allaway *et al*., 2000), before sterilisation. Three seeds were sown per pot and each was inoculated with 1ml of the suitable bacterial culture obtained from washing freshly grown *Rhizobium* from a TY slope with 3ml TY broth. Pots were then aseptically covered in cling-film to prevent contamination. On Germination, the cling-film was carefully pierced to allow seedlings through but still protect each pot from cross-contamination. At this stage seedlings were thinned to from 3 to 2, if all had germinated.

Plants were grown in a growth room under Sonti Agro grow lights, at 22°C with a 16 hour light cycle, for 6 weeks.

2.12. Transport Assays

Uptake of radiolabelled compounds by *R. leguminosarum* was determined by an adapted form of the rapid filtration method (Poole *et al*., 1985).

Cell cultures were grown overnight in AMS (or AMS modified as described in text). On the day of the uptake assay, cells were washed and resuspended in RMS to an OD_{600} of ~1. Cells were then left to starve for 1hour in RMS, shaking at 60rpm at 28°C. 200µl cells were used in an assay volume of 500µl in which amino acids where added to give a concentration of 25µM (0.125µCi $^{14}C/^{3}H$). For transport of ^{14}C proline betaine this concentration was increased to 40µM (0.06µCi) (Boscari *et al*., 2002) and for ^{14}C alanine transport via the monocarboxylate transport permease (MctP) it was increased to 500µM (0.5µCi) (Hosie *et al*., 2002b).

For assays carried out under hyper-osmotic conditions in Chapter 6, the above protocol was adapted in the following ways a) cells were grown up overnight in AMS (10mM glc, 10mM NH_4) + 100mM NaCl, b) whenever cells were washed and resuspended, it was with RMS + 100mM NaCl and c) there was no hour starvation period before assays were performed (LeRudulier & Dupont, personal communication).

For inhibition studies, the same cultures were used as in uninhibited assays, to reduce variance in data. Inhibitory solutes were added at the concentrations described in text.

2.13. Protein Assays

2.13.1. Periplasmic Fraction Isolation

Rhizobia cell cultures were grown overnight at 27°C. These were then spun down at 3800rpm in a Megafuge 1.0R centrifuge (Heraeus), washed once in RMS then resuspended in 10ml of Tris-HCL pH 8 with 20% sucrose and 1mg/ml of lysozyme. Cells were then incubated at room temperature for 15 minutes before addition of 10µl of 0.5M EDTA pH 8 20 minutes further incubation. Cells were then spun down at 3800rpm for 20 minutes and the supernatant containing the periplasmic fraction was removed.

2.13.2. SDS-PAGE

Samples were subjected to SDS-PAGE as previously described (Laemmli, 1970), using a 14% Separating gel (made up of 2.5ml 1.5M Tris-HCl pH 8.8, 0.1ml 10% SDS, 2.6ml nH_2O, 4.66ml 30% Acrylamide and 0.1ml 10% APS). This mix was degassed for 5 – 15 minutes before 2.5µl TEMED was added and the gel poured. Following polymerisation the gels were topped with a 5% stacking gel (made up of 1.25ml 0.5M Tris-Hcl pH 6.8, 50µl 10% SDS, 2.82ml nH_2O, 0.83ml 30% Acrylamide, 50µl 10% APS and 2.5µl TEMED). On completion of polymerisation, the gel was covered in wet tissue paper and stored overnight at 4°C.

The gel was set up in a tank with running buffer (made up of 5g Tris Base, 14.4g Glycine and 5g SDS in a litre of nH_2O). Prior to loading, samples were mixed with 4x loading buffer (made up of 12.5ml 0.5M Tris-HCl pH 6.8, 2g SDS, 10ml Glycerol, 0.402ml 1M DTT, 2.4mg 1% Bromophenol Blue and 25ml nH_2O) and boiled for 3 – 5 minutes. Samples were loaded alongside a SDS-PAGE standard, low range protein marker (Bioline) and run at 150V through the stacking gel and then at 200V through the separating gel. The gel was then stained for 30 – 60 minutes (in 2.5g Coomassie Blue R250, 450ml Methanol,

450ml nH_2O and 100ml acetic acid) and destained for 1 – 2 hours (in a 30% methanol, 10% acetic acid solution).

CHAPTER 3: IDENTIFICATION OF KEY STRESS CONDITIONS AND STRESS INDUCED FUSIONS

3.1. Introduction

In order to study stress responses, the conditions that elicit such a reaction must first be determined. Soil bacteria face a constantly changing environment, made up of many different elements – some interrelated, others not. The aim of this research was to identify genes that are induced under stress conditions. Some previous work has been conducted on *Rhizobium* and its response to stress (Keyser & Munns, 1979; Miller & Wood, 1996), especially with *S. meliloti*, which recently had its genome completely sequenced (Galibert *et al*., 2001). These previous reports provided a foundation upon which this research can build.

It is known that the sequenced rhizobia have no *rpoS* gene and therefore no product of this gene, sigma factor 38 (also called sigma S or σ^s) (Tanaka *et al*., 1993; Galibert *et al*., 2001; Wells & Long 2002). σ^s has been extensively characterised and is known to be the major sigma factor involved in the regulation of stationary phase and general stress response in a variety of bacteria (Tanaka *et al*., 1993; Zgurskaya *et al*.,1997). Consequently, the manner in which stress response is regulated in *Rhizobium* is of great interest. On identifying genes induced under stress conditions, the next goal of this project was to determine how they are regulated.

The pOT1 plasmid was specifically designed to detect promoter activity in the environment (Schofield, 1995) (Fig. 3.1). It contains the *gfpuv* reporter gene, an improved version of *gfp*, which has been optimised for use within bacteria and has increased fluorescence (Crameri *et al*., 1995). Other reporter genes were considered for pOT1, including *sacRB*, *gus*, *lacZ*, *luxAB* and *phoA*, but were rejected in favour of *gfpuv*. GFPuv is a superior reporter as it is both visible in colonies on an ultra-violet (UV) transilluminator and under the microscope (with UV) in single cells. No cofactors, substrates or other gene products are needed for production of GFP and its fluorescence. Most notably, GFPuv is not toxic and does not suffer from background expression through native genes intrinsic to *R. leguminosarum* (Schofield, 1995). GFP can therefore be used in a differential fluorescence induction (DFI) strategy to distinguish between genes that are induced or not.

A *R. leguminosarum* genomic library was created by taking DNA from wild-type 3841 and cloning it into the pOT1 vector at a unique *Sal*I site. The *Sal*I site was destroyed in the process of inserting the rhizobial DNA. The cloning strategy was specifically designed so that pOT1 would not self ligate and so that only a single DNA fragment would insert into the vector (Schofield, 1995). Before its destruction, the *Sal*I site was in the middle of a multiple cloning region (or polylinker) allowing for the manipulation and removal of the DNA inserted in its place (Fig. 3.2).

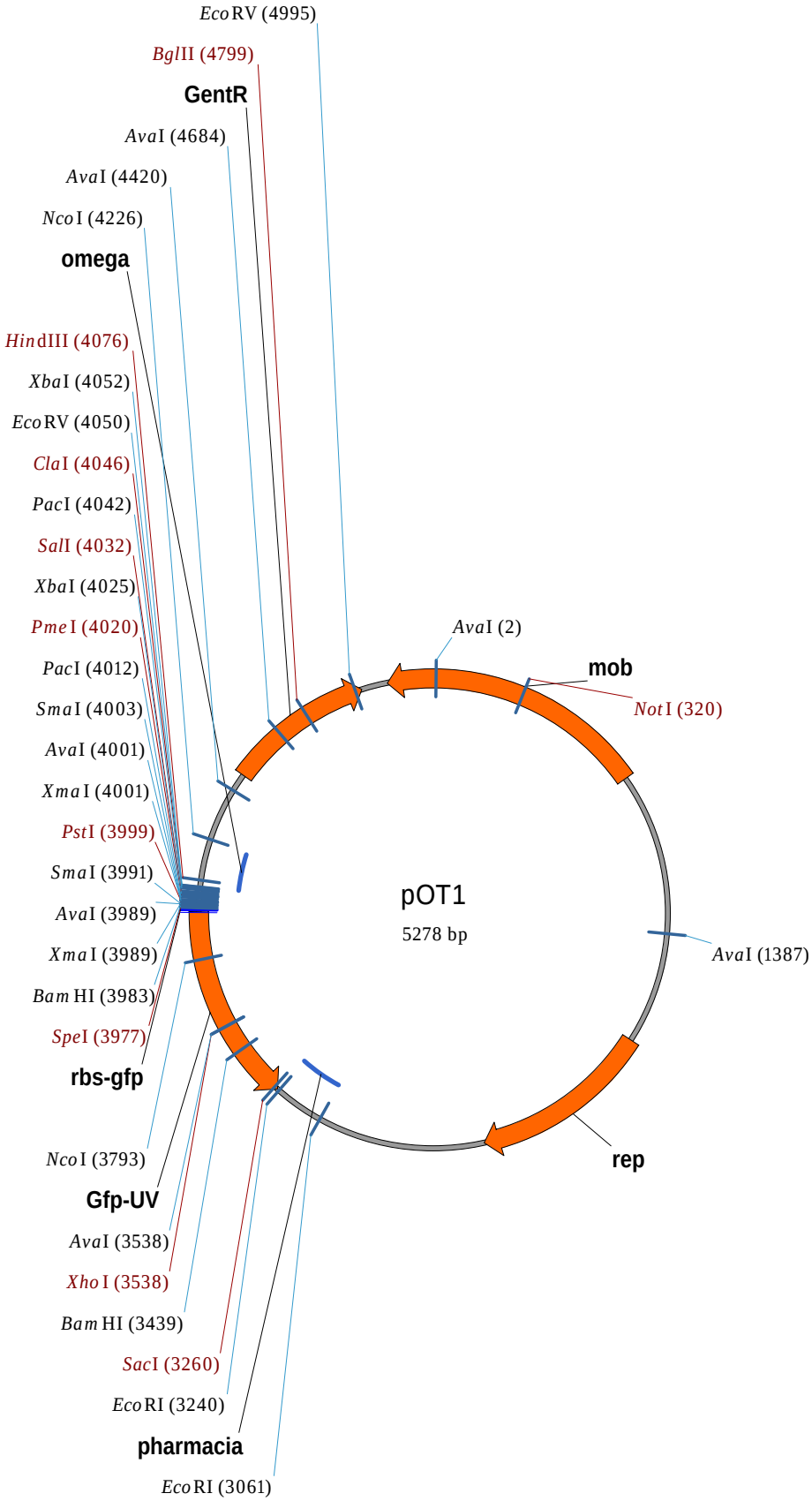


Figure 3.1. pOT1. The construct has a *gfpuv* (Gfp–UV) reporter gene flanked by omega and pharmacia transcriptional terminators. An artificial ribosome-binding site (rbs-gfp) was created in the 5' primer next to *gfpuv*. Unique restriction sites are shown in red and the others are shown in black. Also shown is the *Sal*I site, into which DNA inserts were cloned, as well as the genes for mobilisation (mob), replication (rep) and gentamycin resistance (GentR) (Allaway *et al*., 2001).

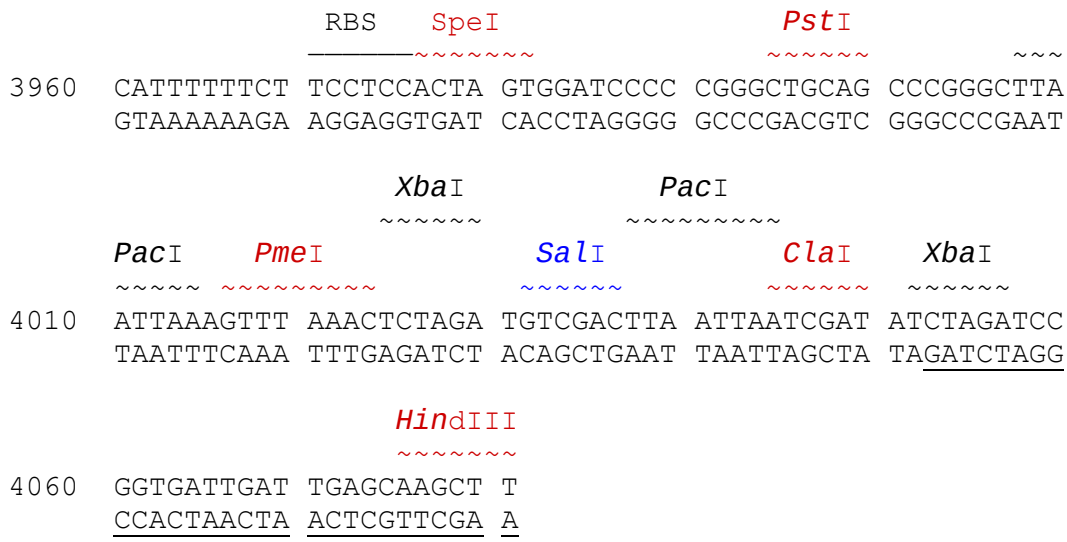


Figure 3.2. The pOT1 Polylinker. A simplified map of the pOT1 polylinker, showing unique restriction enzymes in red and the *Sal*I site, which was destroyed by the insertion of 3841 DNA fragments, in blue. Duplicated sites, shown in black, are only found within the polylinker and allow the removal of cloned inserts. The ribosome-binding site (RBS) is also shown, as well as the end of the omega transcription terminator (underlined) (Allaway *et al*., 2001).

This genomic library was conjugated into *R. leguminosarum* and divided into sub-libraries depending on the amount of GFP produced. LB3 contains clones that did not produce any GFP when screened on AMA (10mM glc, 10mM NH_4), therefore LB3 clones either lack a promoter or contain promoters that are inactive on standard minimal media (Schofield, 1995).

The aim of the work presented in this chapter was to test three or four different stresses in order to isolate stress-induced genes and also to investigate the possibility of cross-induction. To accomplish this, LB3 was used in an attempt to isolate promoters that are activated by stress. LB3 was screened on media designed to mimic stress found in *R. leguminosarum*'s natural environment.

3.2. Results

3.2.1. Minimal Induction Concentrations (MICs)

LB3 had been briefly investigated to establish if any fusions were inducible under various conditions (Schofield, 1995). The conditions tested were; phosphate limited (20μM instead of 500μM), added hesperitin (1μM) and the use of succinate (10mM) instead of glucose (also 10mM) as a carbon source. Thirteen colonies were identified; nine phosphate-limitation induced colonies, one succinate induced, two hesperitin induced and one induced on hesperitin and on succinate. This proved that LB3 library clones could be induced by altered growth conditions.

It had previously been described that 100mM sucrose added to AMA will bring about a stress response and can activate the GFP reporter gene in the pOT1 plasmid (Poole, personal communication). Therefore, this was the concentration that was used to induce hyper-osmosis and no MIC tests were conducted.

Previous studies have shown that, in general, the growth of *Rhizobium leguminosarum* is severely impeded around a pH of 5 (Richardson *et al*., 1988), but mutants do exist with a higher tolerance to acidity (Chen *et al*., 1991 & 1993). It was therefore decided to test the growth of 3841 at pH 5.5, 5.75, 6.0, 6.25, 6.5 and 6.75. During this MIC experiment it was discovered that once buffered to a pH lower than 6.0, AMA plates would not set and had a 'sloppy' consistency. In order to counter this, the amount of agar used in AMA was increased to 2% (w/v). After 3 days, colonies appeared on all AMA (10mM glc, 10mM NH_4) plates that were buffered to a pH greater then 6. After 5 days, colonies appeared on AMA (10mM glc, 10mM NH_4) plates at pH 5.75 and pH 6. There were no colonies on the pH 5.5 plates even after 7 days growth. A pH of 5.75 was therefore used as a screen for acidic growth.

Heavy metal toxicity can occur at relatively low concentrations (Keyser & Munns, 1979; Richardson *et al*., 1988; Tiwari *et al*., 1996a). Metals were therefore tested at the concentrations of 50μM, 100μM, 200μM, 500μM and 1mM. The metals chosen to be investigated were aluminium, copper and zinc (Keyser & Munns, 1979; Reeve *et al*., 1998). Their chlorides ($AlCl_3$, $CuCl_2$ and $ZnCl_2$ respectively) were mixed into AMA (10mM glc, 10mM NH_4) at each concentration. After 3 days, colonies appeared on the AMA (10mM glc, 10mM NH_4) plates containing 50μM aluminium and on the 50μM zinc plates. After 5 days, colonies grew on the AMA (10mM glc, 10mM NH_4) plates containing 100μM aluminium and on the 100μM zinc plates. After 7 days, no colonies had grown on any of the AMA (10mM glc, 10mM NH_4) plates containing copper; nor had any colonies grown on

AMA (10mM glc, 10mM NH_4) plates containing aluminium or zinc at 200μM and above. The copper tests were therefore repeated at concentrations of 5μM, 10μM, 20μM and 50μM. After 3 days, colonies had appeared on all AMA (10mM glc, 10mM NH_4) plates with a concentration lower then 20μM. After 5 days, colonies had appeared on the AMA (10mM glc, 10mM NH_4) plate containing 20μM copper. After 7 days, 1 colony had appeared on the AMA (10mM glc, 10mM NH_4) plates containing 50μM copper. Therefore, 3841 was tested in the on AMA (10mM glc, 10mM NH_4) plates containing 100μM aluminium or with 30μM copper added. (As the aluminium and zinc results were very similar, only aluminium was used in the initial screening process.)

3.2.2. Mass Screenings

Once the conditions that were to be investigated were established, LB3 was then diluted appropriately so that ~50 colony forming units (cfu) would grow on each plate. LB3 was spread onto 200 plates of AMA (10mM glc, 10mM NH_4) containing 100mM sucrose and 200 plates of AMA (10mM glc, 10mM NH_4) buffered to pH 5.75. This meant that approximately 10,000 cfu were screened on each of these stresses. LB3 was also spread onto 100 plates of AMA (10mM glc, 10mM NH_4) containing 30μM copper and 100 plates of AMA (10mM glc, 10mM NH_4) containing 100μM aluminium. This screened 5,000 cfu for copper and aluminium each, 10,000 cfu overall for metal toxicity. All AMA contained streptomycin and gentamycin. AMA containing 100mM sucrose also contained nystatin as contaminant fungi growth had previously been problematic on AMA with added sucrose.

Plates were incubated to allow colonies to grow. After 4 days, plates that were examined under UV light for any signs of GFP expression. Any colonies expressing GFP were isolated onto TY plates containing streptomycin and gentamycin. At this stage, the amount of GFP was not scored. Plates were examined daily under UV light, before being discarded after 10 days growth. Exposure to UV light was kept to a minimum to prevent any radiation-induced stress.

Microaerobic screens were also attempted, however, the reduced oxygen environment did not provide enough oxygen to activate the fluorophore of GFP, making its detection problematic and so a mass screen was not performed.

Table 3.1 shows the initial number of clones expressing GFP found under each screen, after 10 days of growth.

Table 3.1. Initial Screening Results. Number of clones initially found on each stress to have GFP present

Stress	No. of Clones expressing GFP
100mM Sucrose	33
Acidic (pH 5.75)	26
100μM Aluminium	1
30μM Copper	26

These clones all had to be confirmed as being stress-induced. Colonies had already been grown up on TY, after their initial isolation. A single colony from the TY plate was then streaked onto a standard AMA (10mM glc, 10mM NH_4) plate as a control and an AMA (10mM glc, 10mM NH_4) plate containing the appropriate stress, i.e. the stress from which they had originally been isolated. Again all AMA contained streptomycin and gentamycin. The plates were left to grow in tandem. Both sets of plates were then examined and compared under UV light for signs of GFP as before.

If no GFP was present in colonies on both the stress and the control plates, then the isolation of the original clone was erroneous. If GFP was present in colonies on both the stress and the control plates, then those clones must contain a fusion with an insert that is potentially constitutive and therefore not induced by the stress. Any colonies that matched either of these two criteria were discarded as they were not stress-induced. If GFP was present only in colonies on the stress plate, then those clones must contain a fusion with a promoter that was induced by the stress condition.

Table 3.2 shows the number of clones from each screen confirmed as expressing GFP under stressful conditions, after 10 days of growth.

Table 3.2. Confirmed Screening Results. Number of clones confirmed to have GFP present under stress

Stress	No. of Clones expressing GFP
100mM Sucrose	22
Acidic (pH 5.75)	8
100μM Aluminium	1
30μM Copper	1

Each one of these clones was stocked and given a unique strain number (Table 3.3). Additionally, each one of the plasmids was then transferred into *E. coli* DH5α via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7). Each plasmid was stocked and given a unique plasmid number (Table 3.3).

Table 3.3. Strain and Plasmid Numbers. Strain numbers given to each clone identified from LB3 as containing a stress induced pOT fusion and plasmid number given to said fusion.

Strain	Plasmid	Comment
RU1505	pRU860	Clone isolated under AMA (10mM glc, 10mM NH_4) containing 100mM sucrose
RU1506	pRU859	
RU1507	pRU843	
RU1508	pRU844	
RU1509	pRU845	
RU1510	pRU846	
RU1511	pRU847	
RU1512	pRU848	
RU1513	pRU849	
RU1514	pRU850	
RU1515	pRU851	
RU1516	pRU852	
RU1517	pRU853	
RU1518	pRU854	
RU1519	pRU855	
RU1520	pRU856	
RU1521	pRU857	
RU1522	pRU858	
RU1523	pRU861	
RU1524	pRU862	
RU1525	pRU863	
RU1526	pRU864	
RU1527	pRU865	Clone isolated under AMA (10mM glc, 10mM NH_4) buffered to pH 5.75
RU1528	pRU866	
RU1529	pRU867	
RU1530	pRU868	
RU1531	pRU869	
RU1532	pRU870	
RU1533	pRU871	
RU1534	pRU872	
RU1642	pRU1059	
RU1643	pRU1060	

3.2.3. Cross Induction of Stress-Induced Fusions in *R. leguminosarum*

To this point, the 32 isolated strains were known only to react to one stress. In order to determine how specific each fusion's induction was, each strain was tested on other stresses. To score the amount of GFP produced, the 30 strains that were isolated on hyper-osmotic or acidic stress were all screened on AMA, AMA + 100mM sucrose and AMA at pH 5.75. The 2 strains that were isolated under metal toxicity were screened on AMA, AMA + 100μM aluminium, AMA + 30μM copper and AMA + 100μM zinc. These strains were grown in tandem with the GFP control strains RU1158, RU1159 and RU1160, which were plated on standard AMA. All AMA contained 10mM glc, 10mM NH_4, streptomycin

and gentamycin. Comparisons between each strain and the GFP control strains allowed the GFP expression of each fusion to be as accurately scored as possible. This showed what fusions (if any) responded to another stress, other than the one on which they were originally isolated.

As before, plates were left to grow for 4 days before daily UV examinations took place. Strains were grown until 10 days old before being discarded. The results obtained from these sets of screenings are shown in tables 3.4 and 3.5.

Table 3.4. Metal Toxicity Scores. The GFP expression scored for each plasmid/strain tested on each stress. AMA contained 10mM glc and 10mM NH_4.

Plasmid	Strain	AMA + 100µM Aluminium	AMA + 30µM Copper	AMA + 100µM Zinc
pRU1059	RU1642	++	++	++
pRU1060	RU1643	-	++	-

In addition, each plasmid that was conjugated into *E. coli* was transferred back into 3841 to check they exhibited the same GFP profile as their original strain. This was done to verify that each plasmid had been isolated correctly and was not damaged or altered after undergoing the conjugation process into *E. coli*. As expected, plasmids had the same induction pattern.

Following these results, it was decided to concentrate on the fusions that responded to acidic and hyper-osmotic stress. This decision was made so that the research could be focused on the screens that yielded the most results, as the metal toxicity screens were not as successful.

3.2.4. Further Cross Induction

Although each fusion had been screened under acidic and hyper-osmotic stress generated by 100mM sucrose, it was decided that this was not enough data to establish whether each response was general or specific. Furthermore, organisms can respond to hyper-osmosis in different ways, depending on what generates the osmotic conditions (i.e. if the molecule is non-ionic or ionic) (Gloux & Le Rudulier, 1989; Breedveld *et al*., 1990). Therefore, in order to determine whether each plasmid responds to stress specifically or generally, further screens were performed. In all screens strains were grown in tandem on standard AMA (10mM glc, 10mM NH_4) as a control and alongside RU1158, RU1159 and RU1160 as GFP controls.

Each strain was tested on AMA (10mM glc, 10mM NH_4) + 100mM NaCl and AMA (10mM glc, 10mM NH_4) + 100mM mannitol, in order to test an ionic osmolyte and a

different non-ionic osmolyte. Another polar molecule was tested as strains could have been reacting to the sucrose specifically and not osmotic upshift. To prove that strains did not respond to sucrose, strains were also tested on AMA (10mM glc, 10mM NH_4) + 10mM sucrose and on AMA (5mM sucrose, 10mM NH_4), replacing glucose with sucrose as the carbon source.

From the data retrieved (Table 3.5), it could be seen that some strains fluoresced more strongly when a non-ionic molecule was used to induce an osmotic upshift (i.e., RU1519, RU1524 and RU1525); in fact, RU1519 only reacted strongly to hyper-osmosis generated by sucrose. This could have been due to an effect caused by a polar molecule and not an ionic one or it could be due because NaCl is more stressful than sucrose (Gloux & LeRudulier, 1989). These data show that none of the isolated fusions were induced solely on sucrose, as no GFP was observed in any of the strains when grown on AMA (5mM sucrose, 10mM NH_4), just as there was no GFP production on AMA (10mM glc, 10mM NH_4) (Table 3.5). However, the results for the strains on AMA + 100mM sucrose and AMA + 10mM sucrose were similar (Table 3.5). It was unexpected that 10mM sucrose would have brought about a hyper-osmotic stress response. However, 10mM glucose was also present, so that the only difference between standard plates and agar meant to elicit hyper-osmosis was the additional sucrose. Therefore a total of 20mM of osmolyte was present. These data indicates that a relatively small amount of osmolyte is required to generate a hyper-osmotic response in *R. leguminosarum*; sucrose present at 5mM or glucose present at 10mM was not enough to cause a stress response, but 10mM sucrose in addition to 10mM glucose was sufficient to induce an osmotic upshift.

To determine the minimum amount of osmolyte required to induce osmotic upshock in 3841, some of the strains (RU1507, RU1519, RU1524 and RU1525) were grown on AMA (10mM sucrose, 10mM NH_4) and AMA (10mM glc, 5mM sucrose, 10mM NH_4). These strains were also grown on AMA (100mM glc, 10mM NH_4) and AMA (100mM sucrose, 10mM NH_4) so that only one compound was present as both the carbon source and the osmolyte. No GFP production was observed with AMA (10mM sucrose, 10mM NH_4), but the 10mM glucose plus 5mM sucrose did activate the pOT fusions (Fig. 3.3); data only shown for RU1525 but the other strains had the same induction pattern as shown in Fig. 3.3. These data confirmed that it was not sucrose itself was not the cause of the GFP production, but was the osmotic upshift and also showed that at least 15mM of osmolyte must be present to generate hyper-osmotic conditions and induce pOT fusions in *R. leguminosarum*.

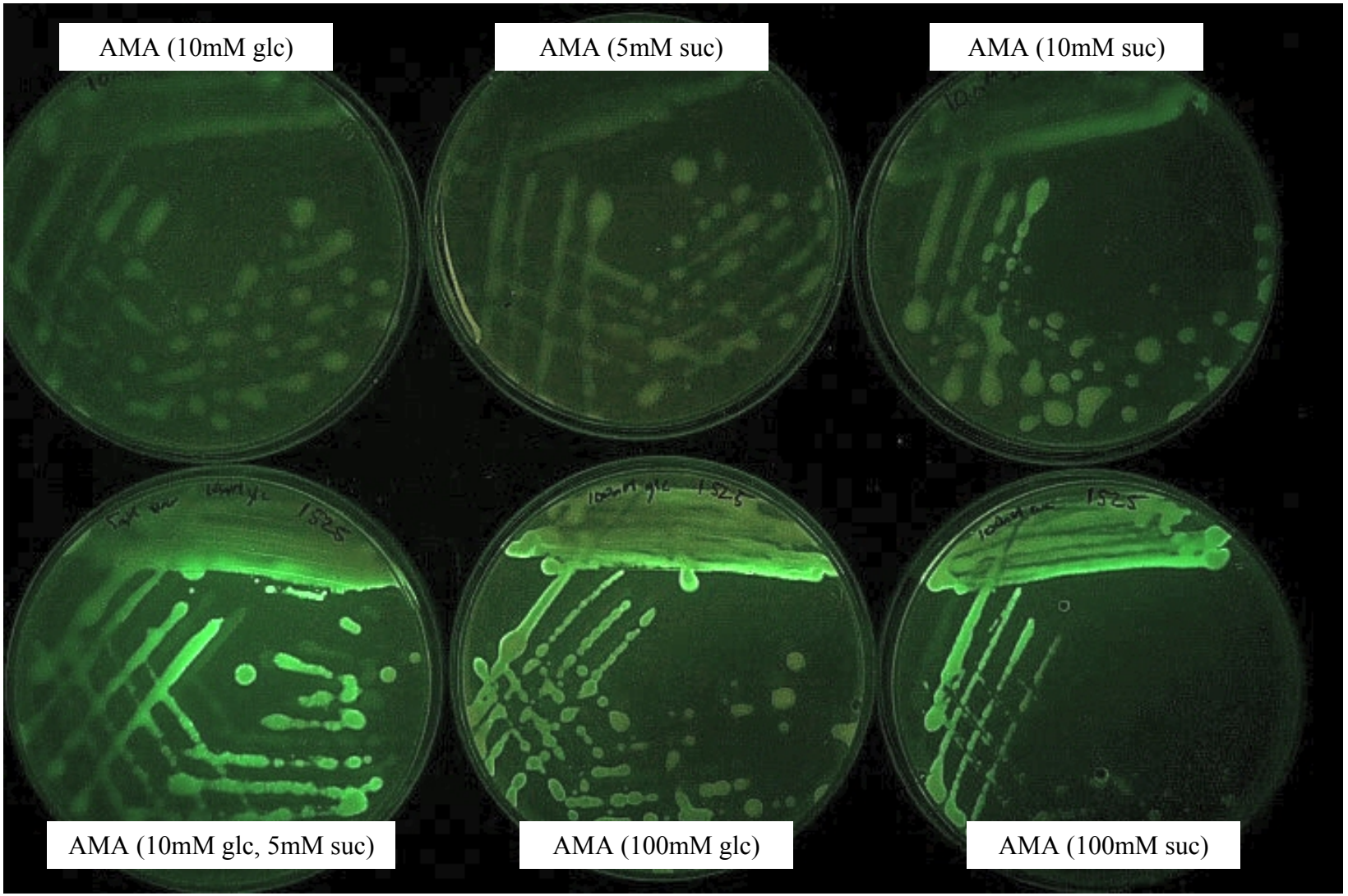


Figure 3.3. RU1525 on Various AMA (10mM NH_4). All plates also contained 10mM NH_4; suc = sucrose. GFP production is only seen in plates with 15mM or more osmolyte present.

As well as stress stimulated by hyper-osmosis, oxidative stress was also investigated in all strains. This was conducted in two ways; by a direct inducer of oxidative stress (hydrogen peroxide – H_2O_2) and by an oxidative stress generator (paraquat). H_2O_2 and paraquat were both chosen, as some oxidative stress responses respond to the stress itself, whilst some respond to generators of the stress (Sigaud *et al*., 1999; Santos *et al*., 2000). Concentrations used were taken from previous reports (Crockford *et al*., 1995; Sigaud *et al*., 1999; Santos *et al*., 2000).

Results for all these screens are shown in Table 3.5.

Table 3.5. Overall Screening Results. GFP expression scored for each plasmid/strain under each screen on AMA (10mM glc, 10mM sucrose) (except for 5mM sucrose, which had no glucose present) + either, sucrose (suc), sodium chloride (NaCl), mannitol (man), H_2O_2, paraquat (para) (at the concentrations indicated), or were buffered at pH 5.75.

Plasmid	Strain	5mM Suc	0.1M Suc	0.01M Suc	0.1M NaCl	0.1M Man	pH 5.75	1mM H_2O_2	0.25mM Para
pRU843	RU1507	-	+++	++	+++	++	+	-	-
pRU844	RU1508	-	++	++	+	++	+++	+	-
pRU845	RU1509	-	++	++	+	+++	+++	-	-
pRU846	RU1510	-	+++	+++	+++	+++	++	+	-
pRU847	RU1511	-	+	+	+	+	-	-	-
pRU848	RU1512	-	++	++	++	+	+	-	-
pRU849	RU1513	-	+	++	+	+	+	-	-
pRU850	RU1514	-	+	++	+	+	+++	-	-
pRU851	RU1515	-	+	+	+	+	-	-	-
pRU852	RU1516	-	+++	+++	+	+	+	-	-
pRU853	RU1517	-	++	++	+	+	+	-	-
pRU854	RU1518	-	+	++	+	+	-	-	-
pRU855	RU1519	-	+++	+++	+	+	+	-	-
pRU856	RU1520	-	+++	+++	+	+	+	-	-
pRU857	RU1521	-	++	++	+	+	+	-	++
pRU858	RU1522	-	+	++	+	+	+	-	-
pRU859	RU1506	-	+	+	+	+	-	-	-
pRU860	RU1505	-	+++	++	+++	++	+	-	-
pRU861	RU1523	-	++	++	+	+++	+++	+/-	+/-
pRU862	RU1524	-	+++	+++	-	+++	++	-	-
pRU863	RU1525	-	+++	+++	+	++	-	-	-
pRU864	RU1526	-	+++	+++	+	+	+	-	-
pRU865	RU1527	-	++	+	+	+	+	-	+
pRU866	RU1528	-	+	++	+	+	+++	+/-	++
pRU867	RU1529	-	++	++	+++	++	+++	-	+
pRU868	RU1530	-	-	+	-	-	+	-	-
pRU869	RU1531	-	-	-	+	-	+	++	+
pRU870	RU1532	-	++	++	-	+	+	-	+
pRU871	RU1533	-	-	+	-	-	+++	+	-
pRU872	RU1534	-	+++	+++	+++	++	+++	+	+/-

3.3. Discussion

3.3.1. Initial Screens

When the genomic libraries were created it was estimated that the main library (before it was sub-divided) covered 40% of 3841's genome (Schofield, 1995). This of course meant that 60% of the genome was not represented in any of the fusion libraries. It was therefore highly likely that some genes that could be activated under stress are not present in LB3. This could explain why some screens were more successful then others. Genomic coverage of LB3 will be discussed further in Chapter 4.

As expected, many fusions were found that were activated by osmotic stress (generated by 100mM sucrose) and at pH 5.75. Hyper-osmotic stress responses have been reported in many species of *Rhizobium*, as have acidic stress responses (Keyser & Munns, 1979; Miller & Wood, 1996). The DFI strategy worked even under acidic conditions. It has been documented, that GFP expression is hindered under low pH (Kneen *et al*., 1998; Llopis *et al*. 1998; Hansen *et al*., 2001). Fusions found from these two screening will prove to be very beneficial for this project, as later chapters will show.

Metal stress results were not as successful, with only two inducible reporter fusions isolated. Previous studies have shown that metal ions play an important role in eliciting a stress response in *Rhizobium*. Vital metals include calcium (O'Hara *et al*., 1989), cobalt (Watson *et al*., 2001), iron (Todd *et al*., 2001) and magnesium (Zahran, 1999; Kiss et al., 2004b). Toxicity studies conducted have shown that many metals can induce stress in *Rhizobium*, including aluminium (Keyser & Munns, 1979), cadmium and iron (Balestrasse *et al*., 2001), copper (Tiwari *et al*., 1996a), manganese (Slattery *et al*., 2001), nickel and zinc (Singh *et al*., 2001). Responses to some of these metals have been characterised; notably by production of thiols to counter heavy metal-induced oxidation (Balestrasse *et al*., 2001; Singh *et al*., 2001) or by the activation of a heavy metal-responsive regulator (Reeve *et al*., 2002). With so many different studies indicating metals in inducing a reaction in *Rhizobium*, it was surprising that only two fusions were isolated. However, some of these studies showed a link between acidity and metal toxicity. As pH decreases, the availability of metals also increases (Dilworth *et al*., 2001). As the AMA used in these screenings was buffered at pH 7.0, the metal ions present may not have been readily available to generate stress or could well have formed complexes with the buffer. Dilworth *et al*. (2001) showed that in a minimal media containing 100μmol/l total copper, the free Cu^{2+} concentration changed from 0.002 to 0.56 μmol/l when the pH was decreased from 7.0 to 5.5. This could explain the lack of results for the metal toxicity screens. Other research has shown that

when metal stress is present in the soil, it is sometimes the plants that are the limiting factor and not the symbiotic bacteria (Kidd et al., 2001; Hall, 2002). This shows that metal stress does play an important role in plant-microbe interactions, although this is not necessarily reflected in the results from these screens.

Overall, the work presented in this chapter shows a successful screening process, with 32 stress-induced fusions being isolated.

3.3.2. Cross Induction Screens

The data shows that 13 fusions were induced specifically under hyper-osmotic conditions. Of these, 2 had significantly more GFP produced under hyper-osmotic stress generated by polar molecules (pRU855 and pRU863); 1 of these being specifically due to sucrose (pRU855). Two fusions induced specifically under acidic conditions, 6 fusions were induced under both hyper-osmotic and acidic conditions and 9 fusions were induced under stressful conditions in general (Table 3.6).

Table 3.6. Stress Induced Fusions. Numbers of each plasmid and what stress/es under which they were induced.

Stress	Plasmid
Hyper-osmosis	pRU843, pRU847, pRU848, pRU851, pRU852, pRU854, pRU855, pRU856, pRU857, pRU859, pRU860, pRU863, pRU864
Acidic	pRU850, pRU871
Hyper-osmosis & Acidic	pRU844, pRU849, pRU853, pRU858, pRU862, pRU868
Hyper-osmosis, Acidic & Oxidative	pRU845, pRU846, pRU861, pRU865, pRU866, pRU867, pRU869, pRU870, pRU872

Few of the isolated strains were induced under oxidative stress (further to their induction under hyper-osmosis and/or acidic stress); 7 with hydrogen peroxide and 8 with paraquat. There are two possible reasons for this. The first is that the majority of the isolated fusions were specific to one stress and so little or no induction was seen under oxidative stress. The other possible reason is that these cross-induction screenings were all carried out on solid AMA plates, so as to correspond to the initial screenings. Previous work on oxidative stress in *Rhizobium* had been conducted in liquid cultures (Crockford *et al*., 1995; Sigaud *et al*., 1999; Santos *et al*., 2000). It is possible that the solid conditions in some way affected the nature of the oxidative stress.

Interestingly, the majority of fusions that displayed any signs of GFP production under oxidative stress were plasmids initially isolated under acidic conditions (Table 3.6). Whether this was mere coincidence or not, will be investigated more in the next chapter. However, both oxidative stress and a low pH can alter the electro-chemical gradients within a cell. Any genes involved in maintaining or restoring redox potentials in *Rhizobium* could therefore have been induced under both these conditions.

Another interesting discovery was the low amount of osmolyte required to induce hyper-osmosis (section 3.2.4). It is not believed that 15mM of disaccharide is potentially fatal to a cell, so why were some genes induced by such a relatively low concentration? It is possible that in order to survive the drastic changes to water availability in the environment, *Rhizobium* must respond quickly. Such a low concentration of osmolyte in the environment is not enough to harm cells, but could serve as an 'early warning system' allowing the bacteria to respond early in case hyper-osmotic conditions increased. To my knowledge, this is the lowest concentration recorded as inducing a hyper-osmotic stress response.

Overall, the work presented here was successful and showed that some of the isolated fusions were induced under specific conditions, whilst others were induced to stress in general. This work will be further investigated in the next chapter.

CHAPTER 4: CHARACTERISATION OF STRESS INDUCED FUSIONS

4.1. Introduction

In order to determine how *R. leguminosarum* responds to each stress, it is necessary to establish what genes are transcribed under such conditions. As reported in the previous chapter, 32 pOT fusions were isolated and identified as being stress-induced and from these 30 underwent further investigation. At this point it was only known what stresses induced GFP production in each fusion and not the genes and/or promoters contained within each plasmid. In this chapter the DNA of each fusion was sequenced and the stress-induced genes identified. Sequence data shows the size of each insert, which was previously indicated by PCR mapping to be between 1.5 and 2.5kb (Schofield, 1995). *R. leguminosarum* 3841 is currently being sequenced and fortunately this preliminary data has been made available and so greatly assisted in the identification of genes. It also allowed potential operons to be recognized. This is very important as a single promoter may govern the transcription of many genes and only some of which may be involved in a stress response.

This chapter also shows which fusions exhibit a specific stress response and which display a general stress response. Although the last chapter showed under what conditions each fusion was induced, there are many differences between a specific stress response and a general stress response. A general stress response tends to occur first, whilst a specific stress response is activated after prolonged exposure (Bremer & Krämer, 2000). However, some specific responses occur immediately (Stokes *et al*., 2003), whereas some bacteria enter stationary growth phase as part of a general stress response (the stringent response) (Hecker *et al*, 1996; Hecker & Völker, 1998). Therefore, ascertaining the timing of induction is a key part in determining the type of stress response.

It was expected that many of the genes associated with the stress-induced fusions to be hypothetical or unknown, as ~40% of the *Sinorhizobium meliloti* genome is made up of such genes (Galibert *et al*., 2001). It was also expected that some of the genes identified may show similarity to genes that have either previously been characterised within *Rhizobium*, or as part of a stress response in another bacteria. This would be of a great benefit to this research as it will provide a link between the 3841 genome and an earlier characterised gene.

4.2. Results

4.2.1. Sequencing Fusions

Once stress-induced fusions had been identified it was necessary to determine the sequence of the DNA insert each contained. Initially, the pOT fusions were sequenced using two primers, pOTfor and pOTrev. The pOTrev primer binds to pOT vector close to the 3' end of the insert (nearest the *gfpuv* gene) and sequences upstream. The pOTfor primer binds around the 5' end of the insert and sequences downstream towards *gfpuv* gene (Fig. 4.1).

Although both of these primers worked satisfactorily and provided good sequencing data, the start of the sequences obtained were not always ideal. Generally, the first twenty to fifty bases from where the primer binds and starts sequencing were incoherent, or missing. The pOTrev primer binds 51bp downstream from where the *Sal*I site was, whilst pOTfor binds 41bp upstream of the former site. This meant that for most of the fusions no sequence data was being acquired for the very start and/or the very end of each insert. The complete insert sequence was required in order to thoroughly analyse the DNA and identify potential promoters. In order to counter this problem two more primers were designed, namely pOTfor_far and pOTrev_gfp, which bind 131bp upstream and 132bp downstream from former *Sal*I site respectively (Fig. 4.1).

Not only did these two primers allow the very ends of each insert to be sequenced, but also some of the flanking polylinker region from pOT1 allowing the start and end of each insert to be determined precisely. Sequencing data acquired from all four primers for each insert was combined as described below.

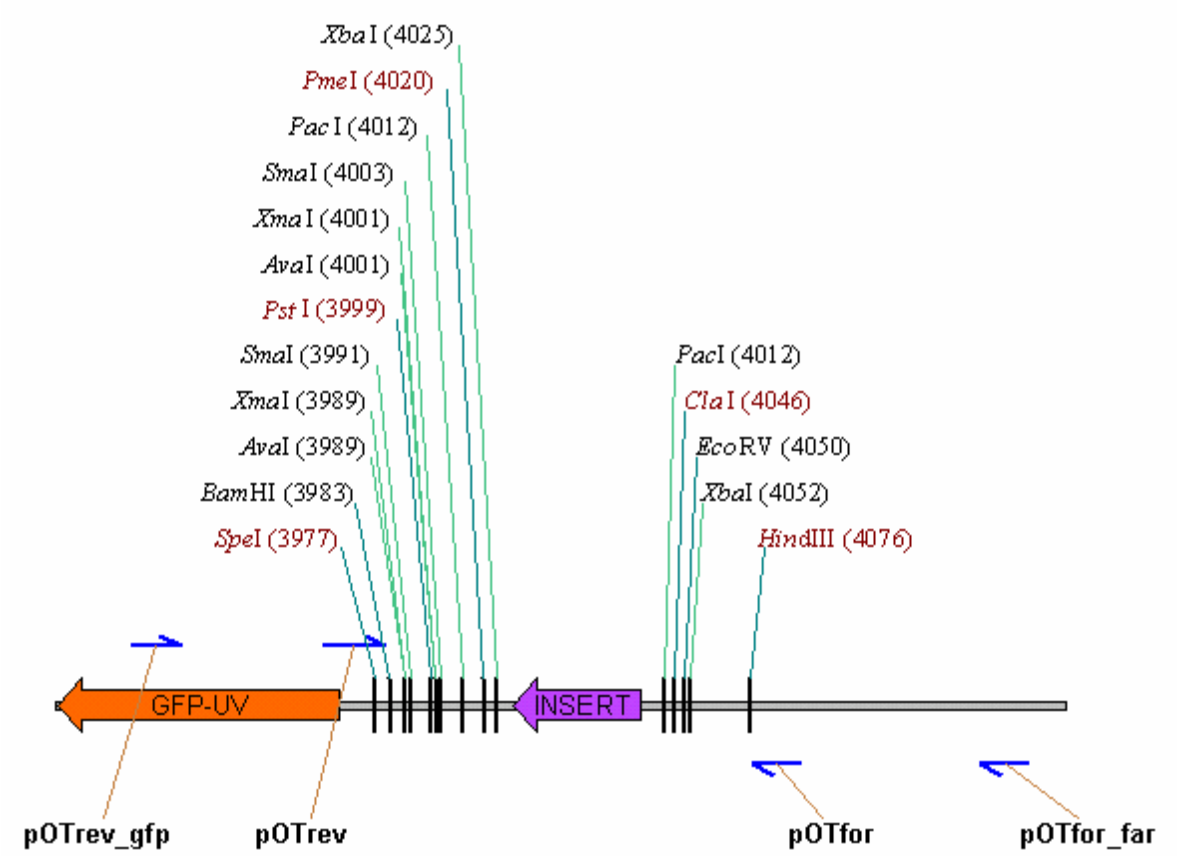


Figure 4.1. Binding Sites of the Four Sequencing Primers. The blue arrows indicate not only where each primer binds, but also the direction in which they sequence. Also shown is the start of the *gfpuv* gene and the polylinker region flanking the insert (unique restriction sites are shown in red, others in black).

4.2.2. Analysing Sequence Data

As mentioned above, the sequencing data obtained were aligned with the polylinker region of pOT1, so that both ends of each insert could be precisely identified. When the library was created it was predicted that inserts would be approximately 2kbp in length (Allaway *et al*., 2001). Most of the sequencing data obtained was between 800bp and 1000bp in length. This meant that there could potentially be missing sequence from the middle of some of the inserts. The complete sequence for each insert was therefore not determined from the two end sequences alone.

Fortunately, the Wellcome Trust Sanger Institute has sequenced the genome for *R. leguminosarum* 3841. At the time of writing, sequencing is in the final stages of annotation and the Sanger Institute have set up a BLAST server on their website (http://www.sanger.ac.uk/Projects/R_leguminosarum), allowing DNA sequences to be submitted and aligned with the preliminary genome. Sequencing data for each insert was submitted to the BLAST server, allowing the entire DNA sequence for each insert to be identified and any missing sequence was retrieved from the genome.

In addition to allowing the sequence of each insert to be completed, the availability of the genome meant that the DNA flanking each insert in its native genome could be identified. This meant that potential operons or the proximity of related genes could be recognized.

It was during the analysis of the sequencing data for each insert, that identical isolates amongst the fusions were discovered (Table 4.1).

Table 4.1. Siblings. Plasmid number (strain number in brackets) alongside its sibling(s)

Plasmid (Strain)	Sibling(s)
pRU843 (RU1507)	pRU860 (RU1505)
pRU855 (RU1519)	pRU852 (RU1519) pRU856 (RU1520) pRU864 (RU1526)
pRU859 (RU1506)	pRU847 (RU1511) pRU851 (RU1515)

Furthermore, it was discovered that the insert of pRU871 was itself contained within the insert of pRU868 (Fig. 4.2). The insert size of pRU868 is 4685bp, whereas pRU871 has an insert size of 2589bp (55% of pRU868). The insert of pRU871 starts 2014bp downstream of the start of pRU868's insert and ended 82bp before pRU868's insert ended (Fig. 4.2).

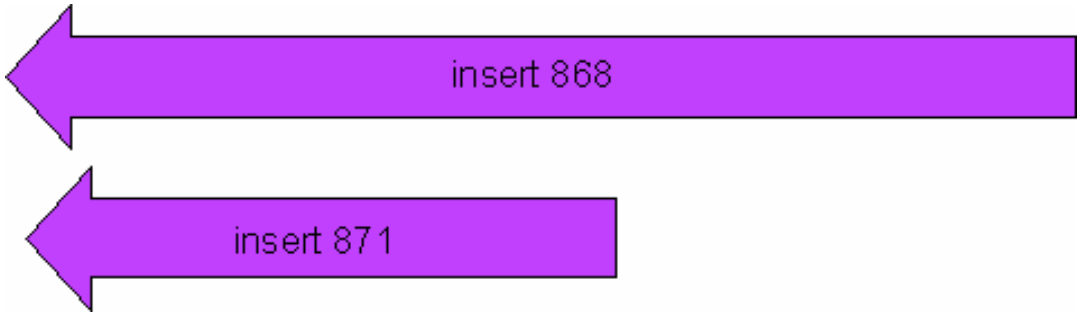


Figure 4.2. Inserts of pRU868 and pRU871. The insert of pRU871 is within the insert of pRU868.

Due to the recent release of the preliminary genome of 3841, the discovery of identical isolates amongst the fusions was made after most of the screenings had taken place. To accommodate the presence of siblings amongst the data, their results were pooled and an average taken where appropriate, however, as would be expected the GFP expression amongst siblings was practically identical. The pooled results for each sibling will subsequently be referred to by the plasmid or strain number in the first column of Table 4.1; i.e. pRU843/RU1507, pRU855/RU1519 and pRU859/RU1506.

pRU868 and pRU871 did not have their results pooled, as they are not identical isolates and so may not have the same results.

Sequencing data for each of the inserts are shown in full in the appendix.

Once the inserts and their native flanking regions had been identified, open reading frames (ORFs) and predicted genes were ascertained using Vector NTI and Artemis. Each ORF/predicted gene had its nucleotide sequence submitted to the BLAST servers supplied by NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) and was also compared to the *S. meliloti* genome (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/). The Sanger Institute has assigned each predicted gene with a stable number that will remain constant on release of the fully completed genome. These numbers are used in the data presented here.

From all of these data, the gene within each insert that is most likely to be activated by stress was deduced (section 4.2.4).

4.2.3. Quantifying GFP Induction in AMS Cultures

In order to quantify the level and timing of GFP produced by each fusion, each strain was grown in liquid media to allow fluorescence levels to be examined in a plate reader. Strains were freshly grown up on a TY slope and resuspended in AMS to an optical density (OD_{600}) of ~1.0 before 10ml cultures were inoculated with 100µl of each suspension. Strains were grown for 3 days in AMS, AMS + 100mM sucrose and AMS buffered at pH 5.75, shaking at 200rpm. All AMS contained 10mM glc, 10mM NH_4, streptomycin and gentamycin. A control of 3841 was also grown in identical media (but without gentamycin) in tandem with the strains. Samples were taken daily from each culture so their OD_{600} and relative fluorescence could be measured in a plate reader and the experiment was performed in triplicate.

The data produced from this is calculated by dividing the absolute fluorescence of each culture by its optical density, minus the appropriate blanks (see Chapter 2, section 2.10). These data are presented individually for each strain in section 4.2.4. Each figure shows an average of triplicate results for each strain grown in each media (outlined above) plus standard errors. Where appropriate, sibling strain data were combined. Figure 4.3 shows an example of the growth rates of 3841 under the three test media. All strains matched the pattern shown.

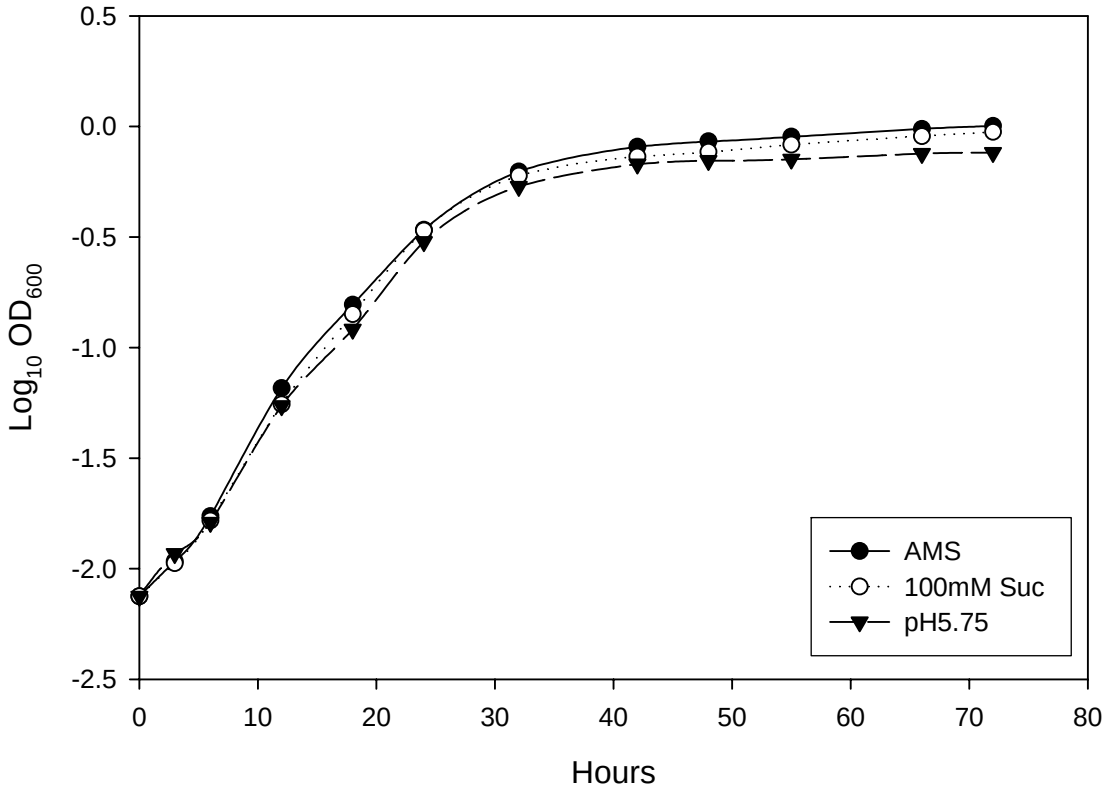


Figure 4.3. Growth Rates of 3841 on AMS, AMS + 100mM Sucrose and AMS Buffered at pH 5.75. Example of 3841 growth rates on the three test media (all supplemented with 10mM glc, 10mM NH_4). All strains showed similar rates.

4.2.4. Overall Results

The nature of this research has produced a large amount of data for each plasmid/strain. The data for each one, including results obtained from solid and liquid media, is presented individually below. The region of DNA from 3841 for each insert is indicated using the nucleotide numbers annotated by the Sanger Centre *R. leguminosarum* 3841 sequencing project.

Once the insert for each pOT fusion had been sequenced, the most probable activated promoter (and therefore gene) for each one was deduced. The best BLAST result that showed the most likely protein encoded by each gene is shown for each insert. The region of DNA present in each insert, and the neighbouring region in the genome, is also shown. All diagrams are proportional, with a 1kb scale shown. Genes have been labelled with their gene number (annotated by the Sanger Centre *R. leguminosarum* 3841 sequencing project). Where applicable, genes have also been labelled by the most likely function of the protein they encode for, as determined through BLAST analysis. If no homologous gene was found by BLAST analysis, then just the predicted gene number is given.

Some predicted gene starts have been trimmed from those calculated by the Sanger Centre, although the gene numbers have remained the same. This was because the data the Sanger Centre provided was only preliminary, so the exact translational start of each predicted protein may not be correct and trimming the gene made a more probable protein. In some BLAST results the low complexity filter was disabled in order to obtain a more precise match. These cases are indicated in text.

4.2.4.1. pRU843/RU1507

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+++	++	+++	++	+	-	-

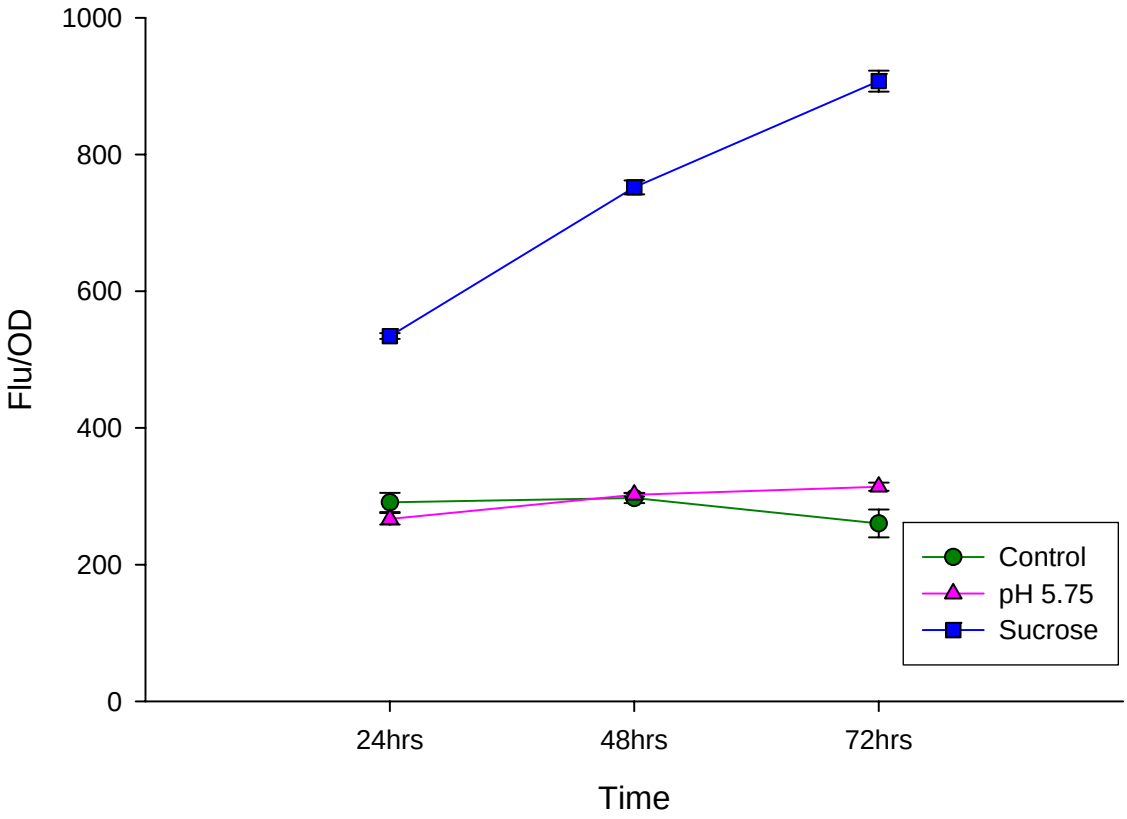


Figure 4.4. RU1507 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures of siblings RU1505 and RU1507 (as in explained in section 4.2.2).

The screening results on solid media for this fusion showed that it responded strongly to hyper-osmotic stress, whether caused by polar or ionic molecules. Liquid growth showed that this fusion was induced early in its growth cycle, with a 1.8-fold

induction after 24 hours, which led to a 3.5-fold induction after 72 hours (Fig. 4.4). *Rhizobium* DNA from this fusion mapped to the pRL10 plasmid, nucleotides 79468 – 81350 (Fig. 4.5).

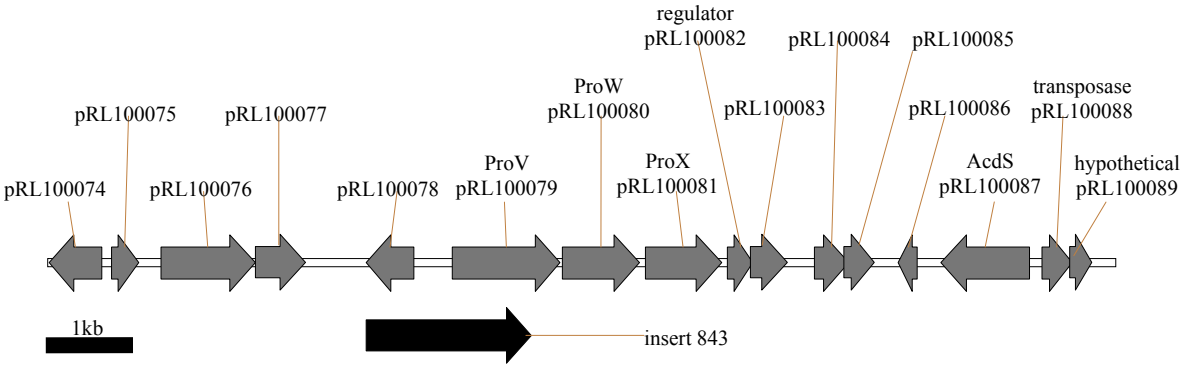


Figure 4.5. Genomic Region of pRU843's Insert. Black arrow shows the region of DNA within pRU843, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL100079 (pRL10 nucleotides 80451 to 81680, 410aa). The promoter within the insert could activate the transcription of pRL100080 to pRL100085 inclusive as an operon. However, as there is an intergenic region (that may contain a promoter) between most of these genes, this may not be the case so those genes were not investigated. As BLAST analysis of pRL100078 did not reveal a homologue it was not analysed further, besides which, it would be encoded in the opposite orientation to a promoter within the insert. The best BLAST result for pRL100079 is shown below.

Glycine Betaine/L–Proline transport ATP–binding protein ProV BMEII0548 (imported) [*Brucella melitensis* 16M] Length = 398aa

Score = 504 bits (1297), Expect = e-141
Identities = 265/396 (66%), Positives = 314/396 (79%)
Frame = +1

```
Query: 40   ATTKISLKNIYKVFGEHPKKAFALLRAGKTKSEIHAATGCSIGVNDASFDIRAGEIFVIM 219
            A TKISL  ++KVFG++P +A   L AGK+K++IH+  G +IGV++A+FDIR GE+FVIM
Sbjct: 15   AKTKISLNGVFKVFGDNPMRAMRELGAGKSKAQIHSDLGATIGVDNATFDIREGEVFVIM 74

Query: 220  GLSGSGKSTLLRLLNRLIEPSSGSIEIDGRDITGMSRSELIALRRRDISMVFQSVALLPN 399
            GLSGSGKSTLLRLLNRLIEP++GSIE++GRDI  MS+ ELI LRRRD+SMVFQS ALLPN
Sbjct: 75   GLSGSGKSTLLRLLNRLIEPTAGSIEVEGRDIVKMSKRELIDLRRRDMSMVFQSFALLPN 134

Query: 400  RTVLNNAAFGLEVAGVGEAGRKQKALAALKAVGLDGYADSRPDQLSGGMKQRVGLARALA 579
            R+VLNNAAFGLEVAG+GE  R QKAL AL AVGL+ YA S PDQLSGGMKQRVGLARALA
Sbjct: 135  RSVLNNAAFGLEVAGMGEVERHQKALKALAAVGLEPYAHSMPDQLSGGMKQRVGLARALA 194

Query: 580  SEPTILLMDEAFSALDPLIRTEMQDELVRLQSEHSRTIVFVSHDLDEAMRIGDRICIMQN 759
            SEPT+LLMDEAFSALDPLIRTEMQDEL RLQ++HSRTIVFVSHDLDEAMRIGDRICIMQ+
Sbjct: 195  SEPTVLLMDEAFSALDPLIRTEMQDELKRLQAQHSRTIVFVSHDLDEAMRIGDRICIMQH 254

Query: 760  GNVVQVGAPDEIVTQPANDYVRSFFRNVDVAHVFKAGDVARKSQVTIIEREGVSAAAALE 939
            G VVQVG P+EI++ PANDYVRSFFRNVDV+ VFKA DVAR  ++ + E E    A ALE
Sbjct: 255  GKVVQVGTPNEIISAPANDYVRSFFRNVDVSRVFKAADVARDDELIVFEAE--QLATALE 312

Query: 940  RMKNYDREYAIILGRDKTYHGMISQTSLIEKMRAKAADPYRGAFLTEIQAIPASEPLSNV 1119
            R     + Y +++  D+TY G++S+ +L    R           L +  AI A  PLS +
Sbjct: 313  RFDASGKAYGVLVDADRTYRGLVSRDALAAGQR-----------LNDFAAIEADAPLSGL 361

Query: 1120 LGKVAASPWPVPVVCDRNRYIGSISKSALLETLDRA 1227
            L  VA SPWPVPV   +NRY+G+IS+S LLETL RA
Sbjct: 362  LVHVAKSPWPVPVTDRQNRYVGAISRSVLLETLGRA 397
```

This result showed a good sequence identity to the ProV component of the ProU transport system in *B. melitensis* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). ProU (encoded by the *proVWX* operon) is a broad-specificity osmoprotectant uptake system, which belongs to the ABC transporter class. ProU is specifically activated when hyper-osmosis is imposed on bacteria grown in an environment of low-osmolality (Faatz *et al*., 1988). (This class of transporter is more thoroughly reviewed in Chapter 6.) Encouragingly, the product of pRL100080 shared sequence identity with ProW and the product of pRL100081 shared sequence identity with ProX. These data indicated that a region with sequence identity to the entire ProU operon was present in 3841 and as such was probably transcribed as an operon under the same promoter. In *E. coli*, ProU has been extremely well characterised as one of the systems responsible for rapid uptake of the compatible solutes proline and glycine betaine. This is a classic example of a bacterial response to hyper-osmotic stress. Bacteria accumulate compatible solutes in an attempt to raise internal osmolality to match that of the environment and so relieve the hyper-osmosis. These data shows that *R. leguminosarum*

may employ a similar system under hyper-osmosis. This set of BLAST results was well suited to the data obtained, as the induction of pRU843 was rapid and specific for hyper-osmotic stress.

This indicates that pRL100079, pRL100080 and pRL100081 were induced as an operon as part of a specific stress response to hyper-osmotic stress. This made pRU843 an ideal fusion for further investigation (see later chapters).

4.2.4.2. pRU844/RU1508

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	+	++	+++	+	-

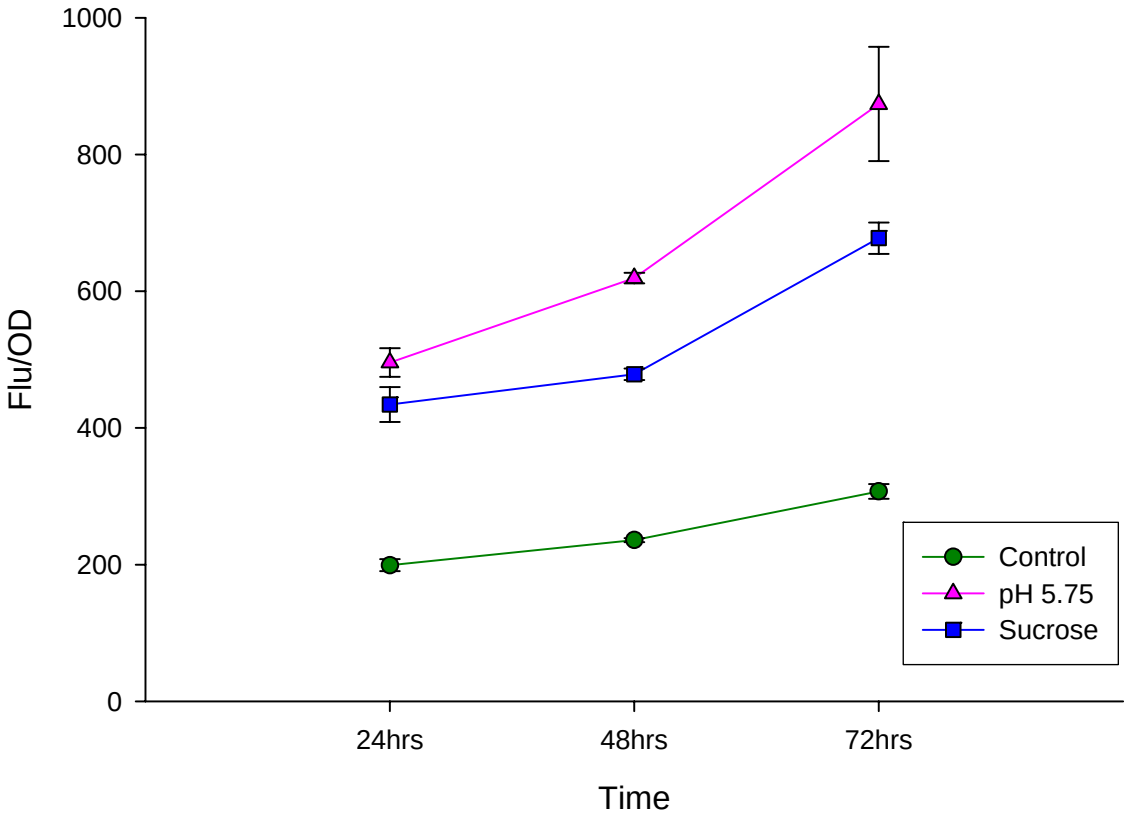


Figure 4.6. RU1508 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded to hyper-osmotic, acidic and slightly under oxidative stress. Liquid growth showed that this fusion induced early in its growth cycle, with a 2 to 2.5-fold induction after 24 hours, which

remained constant thereafter (Fig. 4.6). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 631647 – 634153 (Fig. 4.7).

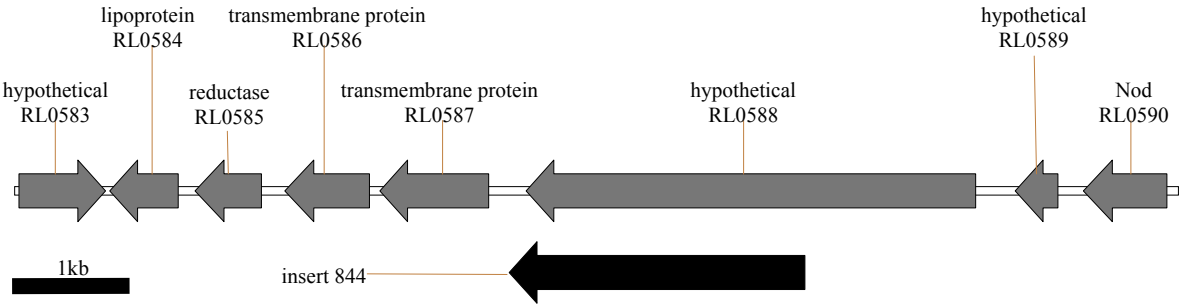


Figure 4.7. Genomic Region of pRU844's Insert. Black arrow shows the region of DNA within pRU844, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL0587 (compliment of chromosome nucleotides 630561 to 631481, 307aa). The insert starts within RL0588 and so it cannot contain its promoter. Although the insert ends before the predicted transcriptional start of RL0587, it may still contain its promoter, although it could be a cryptic promoter within the end of the RL0588 gene. As the annotation of the 3841 genome is only preliminary, the transcriptional start may have been predicted incorrectly and may be further upstream, which would put it within the DNA mapped from the insert. The best BLAST result for RL0587 is shown below.

Conserved hypothetical protein Atu0498 (imported) [*Agrobacterium tumefaciens* C58]
Length = 308aa **Low Complexity Filter disabled**

Score = 512 bits (1318), Expect = e-144
Identities = 259/304 (85%), Positives = 288/304 (94%)
Frame = +1

```
Query: 1   VTIYLPIAELSVNIFIILGMGAAVGFLSGMFGVGGGFLITPLLIFYNIPPVVAVATGANQ 180
           +T+YLPIAELSVNIFIILGMGAAVGFLSGMFGVGGGFLITPLLIFYNIPPVVAVATGANQ
Sbjct: 1   MTVYLPIAELSVNIFIILGMGAAVGFLSGMFGVGGGFLITPLLIFYNIPPVVAVATGANQ 60

Query: 181 VVASSISGAITHFRRGSLDVKLGTVLLVGGLAGATVGIWIFSLLRAIGQLDLIISLMYVI 360
           VVASS+SG+ITHFRRG+LD+KLGTVLLVGGL GATVG+WIFS LR+IGQLDLI+SL+YVI
Sbjct: 61  VVASSVSGSITHFRRGTLDIKLGTVLLVGGLVGATVGVWIFSFLRSIGQLDLIVSLLYVI 120

Query: 361 FLGTVGGLMLLESVNAMRRAARNEPPAPRKPGHQHWVHKLPLKVRFKKSKIFLSVIPIVA 540
            LGTVG LML ES++A+RRAARNE    R+PGH +WVH+LPLK+RFKKSKI+LS+IPIVA
Sbjct: 121 LLGTVGTLMLKESISALRRAARNETVTLRRPGHHNWVHRLPLKMRFKKSKIYLSIIPIVA 180

Query: 541 LGFAIGILTSIMGVGGGFIMVPAMIYLLRIPTNVVVGTSLFQIIFVTAYTTIVQAATNFS 720
           LGF IGILTSIMGVGGGFIMVPAMIYLLRIPT+VVVGTSLFQIIFVTAYTT+VQAATN+S
Sbjct: 181 LGFGIGILTSIMGVGGGFIMVPAMIYLLRIPTSVVVGTSLFQIIFVTAYTTVVQAATNYS 240

Query: 721 VDIVLAFILMVAGVIGAQYGVRVGQKLRGEQLRALLGLLVLAVGVRLAIALVVTPADVYS 900
           VD+VLAFILMVAGVIGAQYGVRVGQKLRGEQLRALL LLVLAV +RLA++LVV P D++S
Sbjct: 241 VDVVLAFILMVAGVIGAQYGVRVGQKLRGEQLRALLALLVLAVALRLAVSLVVRPEDLFS 300

Query: 901 VVMG 912
           V +G
Sbjct: 301 VAVG 304
```

This result showed an excellent sequence identity to a hypothetical protein in *A. tumefaciens* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). As the result was for a hypothetical protein, not much more could have been inferred from that data alone. Pfam analysis was carried out and showed the predicted protein contains a domain of unknown function (DUF81). This domain is known to include two duplicated modules of three transmembrane helices, which indicated that the molecule was found in the cell membrane. A protein with membrane-spanning regions could form a pore or channel in its host membrane. This channel could be used to maintain or restore balance to a cell's internal environment after encountering a general external stress. The pores may aid in the removal of protonated species (under acidic stress) or reactive oxygen species (under oxidative stress). Furthermore, aquaporins (pores that can selective transport water across cell membranes under stress) are known to have six transmembrane domains (Preston *et al.*, 1994). These channels have been characterised in both eukaryotes and prokaryotes (Calamita *et al.*, 1995). Under hyper-osmotic stress, cells passively lose water through diffusion and so cells must respond quickly to avoid this loss. As bacteria have active control on aquaporins, they could have helped maintain internal water concentrations by preventing water efflux through these channels. Mechanosensitive channels have been

observed to activate rapidly, just as pRU844 did, in cells responding to osmotic stress (Stokes *et al.*, 2003). Although these channels deal with hypo- and not hyper-osmotic stress, they illustrate how quickly cells can respond to stress.

These data, and the data recorded from tests with pRU844, support the theory that RL0587 is induced as part of a general stress response in *R. leguminosarum*.

4.2.4.3. pRU845/RU1509

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	+	+++	+++	-	-

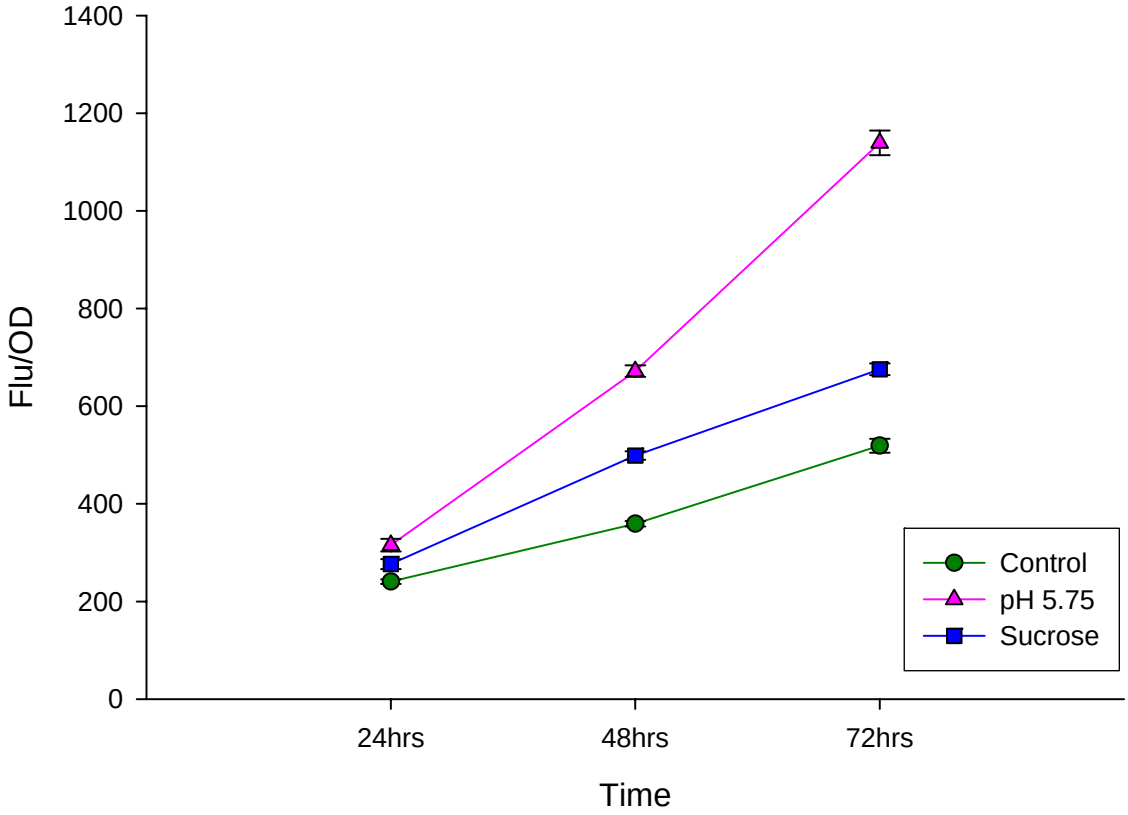


Figure 4.8. RU1509 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded to hyper-osmotic stress and acidic stress. Liquid growth showed that this fusion induced later in its growth cycle as it took 48 hours before any significant induction was observed (Fig. 4.8). *Rhizobium* DNA from this fusion mapped to the pRL9 plasmid, nucleotides 190856 – 193166 (Fig. 4.9).

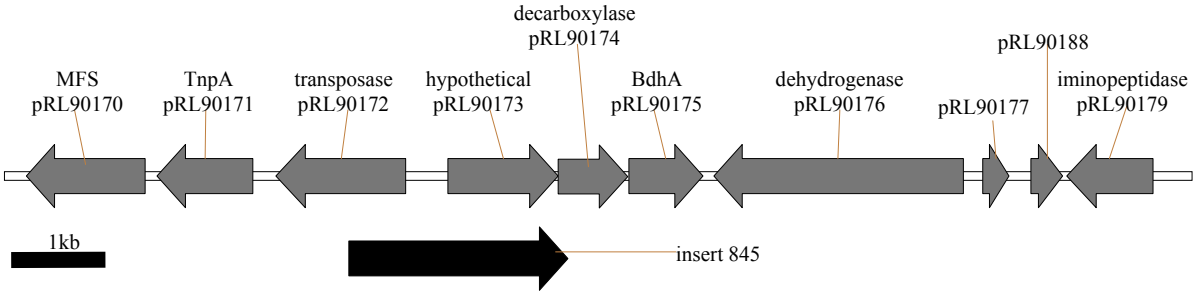


Figure 4.9. Genomic Region of pRU845's Insert. Black arrow shows the region of DNA within pRU845, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL90174 (pRL9 nucleotides 193066 to 193803, 246aa). Although the insert is also made up of DNA from pRL90172, it is encoded in the opposite orientation to the promoter. The promoter within the insert could also transcribe pRL90173 and pRL90175 as a part of an operon with pRL90174, as there was no intergenic region between these genes. As such, the products of these genes were taken into account during the analysis. The best BLAST result for pRL90174 is shown below.

Acetoacetate decarboxylase [*Burkholderia fungorum* LB400] Length = 252aa

Score = 366 bits (939), Expect = e-100
Identities = 174/246 (70%), Positives = 204/246 (82%)
Frame = +1

```
Query: 1   MKIEDVVRNAFAMPLTSPSYPPGPYRFVNREYMIITYRTDPEALRRVVPEPLQFDEPLVK 180
           M ++ V+ NAFAMP+TSP++P GPYRF+NRE++IITYRTDP+ LR VVPEPL+  EPLV
Sbjct: 1   MDVKSVLSNAFAMPITSPAFPMGPYRFINREFLIITYRTDPDKLRAVVPEPLEIGEPLVH 60

Query: 181 YEFIRMPDSTGFGDYTESGQVIPVTYQGVHGGYVHSMYLNDDAPIAGGREIWGFPKKLAE 360
           YEFIRMPDSTGFGDYTESGQVIPV+Y+GV GGY  +MYL+D  PIAGGRE+WGFPKKLA
Sbjct: 61  YEFIRMPDSTGFGDYTESGQVIPVSYKGVAGGYTLAMYLDDHPPIAGGRELWGFPKKLAN 120

Query: 361 PSLTSVKDALVGTLDYGGQRVATATMGFKHRSLDEAKILESLKQPNFMLKIIPHVDCTPR 540
           P L    D LVGTLDYG  R+AT TMG+KHR LD A+  + L+ PNF+LK+IPHVD TPR
Sbjct: 121 PVLAVHTDTLVGTLDYGPVRIATGTMGYKHRQLDLAQQKKRLETPNFLLKVIPHVDGTPR 180

Query: 541 ICELVRYYLEDLTVKGAWEGPGALALFPHALAPVADLPVLEVKSAVHILSDLTLGLGEVV 720
           ICELVRYYL+D+ +KGAW GP AL L PHALAPVA LPVLEV  A H+++DLTLGLGEVV
Sbjct: 181 ICELVRYYLQDIDLKGAWTGPAALELAPHALAPVAALPVLEVVEARHLIADLTLGLGEVV 240

Query: 721 HDYLAK 738
            DYL +
Sbjct: 241 FDYLGQ 246
```

This result showed a good sequence identity to an acetoacetate decarboxylase in *B. fungorum* (a member of the β-proteobacteria, whereas *R. leguminosarum* is a member of the α-proteobacteria). An acetoacetate decarboxylase is a transferase which could have many possible functions. It is likely that it converts a non-metabolisable compound into one that the 3841 can use. pRL90175 appears to encode for a protein with sequence identity to BdhA. BdhA is a β-hydroxybutyrate-dehydrogenase, responsible for breaking down the storage compound polyhydroxybutyrate (PHB). pRL90175, like pRL90174, may also have a metabolic role within the cell and be used to generate an energy source (Hofmann *et al*., 2000). The PHB degradation pathway has been well characterised in *S. meliloti* and involves PHB being broken down to β-hydroxybutyrate, which is then converted to acetoacetate by BdhA (Aneja & Charles, 1999). Acetoacetate is then converted into acetoacetyl-coA, which can then re-enter the PHB cycle or go through to the TCA cycle (Fig. 4.10) (Lodwig & Poole, 2003). pRL90175 has a 40% identity and 57% similarity with the *bdhA* of *S. meliloti*, whereas another gene from *R. leguminosarum*, RL3569, has a 69% identity and 84% similarity with the same gene. (pRL90175 and RL3569 share a 62% identity and a 62% similarity with each other.) This indicates that RL3569 is the equivalent *bdhA* gene of *R. leguminosarum* and pRL90175 is either a homologue or paralogue. Interestingly, the product of pRL90174 is a decarboxylase of acetoacetate, the product of β-hydroxybutyrate degeneration by BdhA, and converts acetoacetate to acetone releasing CO_2. This suggest that in *R. leguminosarum* PHB degradation may also follow another pathway, utilising the products of pRL90174 and pRL90175, perhaps as a method of removing acid from the cells. PHB is converted to β-hydroxybutyrate by a dehydrogenase, and BdhA converts this to acetoacetate (Fig. 4.10). Both β-hydroxybutyrate and acetoacetate have a carboxylic acid group as part of there structure, making them acidic in nature. During acidic stress, cells need to remove acidic compounds in order to counter the stress. As mentioned above, the product of pRL90174 converts acetoacetate to acetone, a neutral compound. This combination of the products of pRL90175 and pRL90174 seems to be a novel mechanism of removing protons from acid stressed cells and appears to be linked to PHB, one of the major storage compounds of the bacteria. The presence of these two genes in other α-proteobacteria was investigated by sequence comparisons and homology studies, and no gene similar to pRL90174 could be found indicating this mechanism may be unique to *R. leguminosarum*. Whilst novel to rhizobia, previous work has shown the use of other decarboxylases in response to a low pH (as described here) in *E. coli*; e.g. the *gad* genes (glutamic acid decarboxylation) involvement with the conversion of

glutamic acid to GABA under acidic conditions (Castanie-Cornet & Foster, 2001; Hommais *et al*., 2004).

pRL90173 appears to encode a hypothetical protein. As the result was a hypothetical protein, not much more could have been inferred from that data alone. Pfam analysis was carried out and showed the predicted protein contains a patatin-like phospholipase. Patatin can be a storage protein but it also has the enzymatic activity of a lipid acyl hydrolase, which catalyses the cleavage of fatty acids from membrane lipids. Changes in fatty acid composition have been observed in *Bradyrhizobium japonicum* under osmotic stress (Boumadhi *et al*., 2001). A similar response may be seen in *R. leguminosarum* and the product of pRL90173 could be responsible for catalysing changes in fatty acid composition in the membrane. Membrane composition also undergoes a change in *Mesorhizobium loti* during acidic stress (Correa *et al*., 1999), further supporting this hypothesis for the role of pRL90173's product, as pRU845 induced under hyper-osmosis and acidic stress. Whether pRL90173's product is involved in the same pathway as the products of pRL90174 and pRL90175 is currently unknown.

These data indicates that pRL90174 and pRL90175 were induced as an operon as part of an acidic stress response.

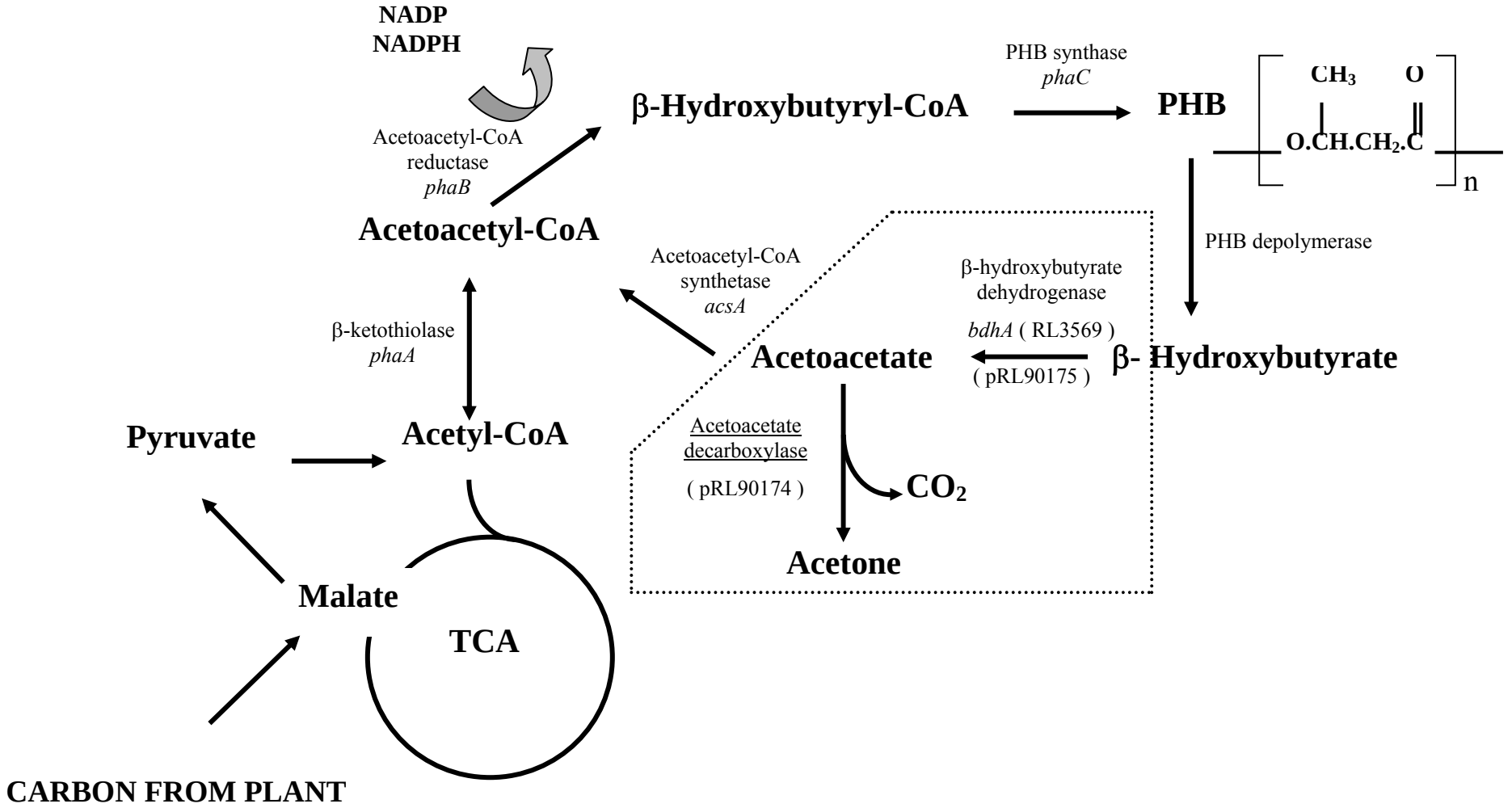


Figure 4.10. Pathways leading to the Synthesis and Degradation of PHB in rhizobia (adapted from Lodwig & Poole, 2003). Additional pathways suggested by this work are indicated by dotted outline.

4.2.4.4. pRU846/RU1510

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+++	+++	+++	+++	++	+	-

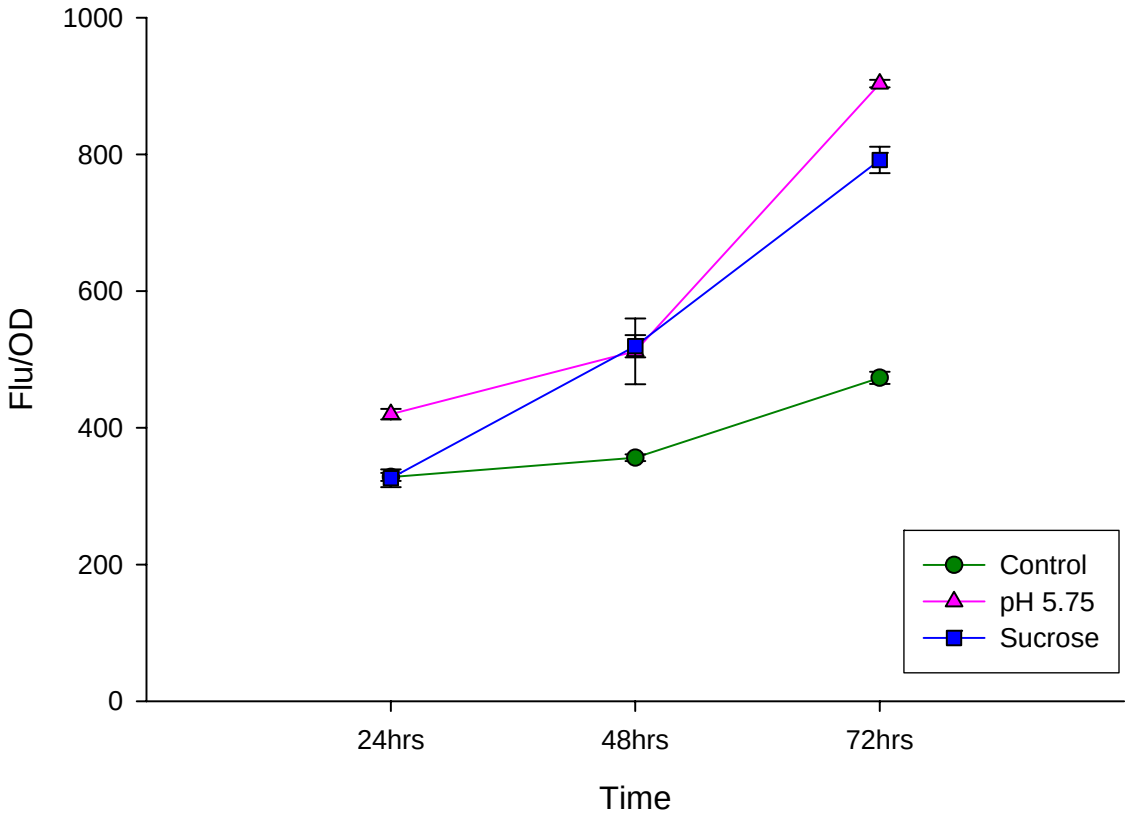


Figure 4.11. RU1510 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded to hyper-osmotic, acidic and slightly under oxidative stress. Liquid growth showed that this fusion induced later in its growth cycle as it took 48 hours before any significant induction was observed (Fig. 4.11). *Rhizobium* DNA from this fusion mapped to the chromosome, nucleotides 387562 – 389867 (Fig. 4.12).

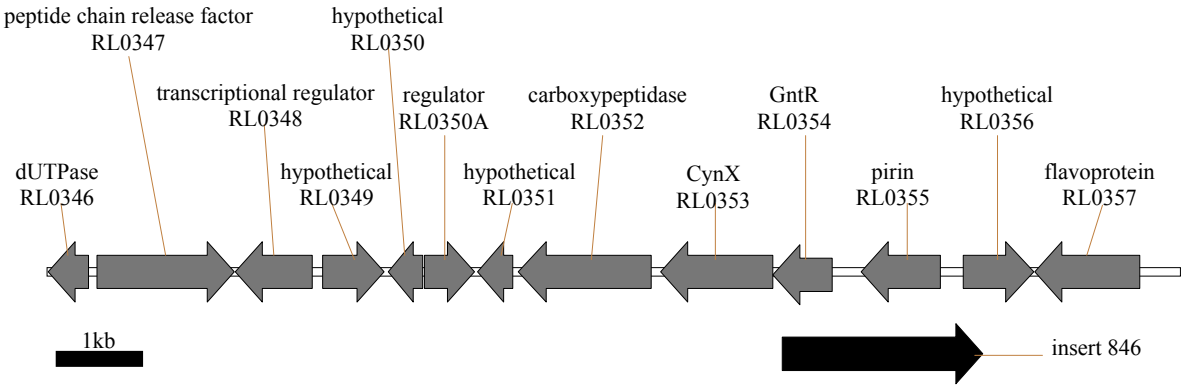


Figure 4.12. Genomic Region of pRU846's Insert. Black arrow shows the region of DNA within pRU846, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre except RL0350A, which had not been assigned a number at time of writing. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL0356 (chromosome nucleotides 389637 to 390446, 270aa). Although most of the insert is made up of DNA from RL0354 and RL0355, they are both encoded in the opposite orientation to the promoter. The best BLAST result for RL0356 is shown below.

Further analysis showed that the insert for pRU855 mapped in close proximity to where pRU846 mapped (Fig. 4.24).

Hypothetical protein SMc01162 [*Sinorhizobium meliloti* 1021] Length = 271aa

Score = 417 bits (1071), Expect = e-115
Identities = 199/255 (78%), Positives = 221/255 (86%), Gaps = 2/255 (0%)
Frame = +1

```
Query: 1   MCRWAAYRGDPLYLEELVSSPAHSLIEQSHCATRAKTATNGDGFGIAWYGDRPEPGRYRD 180
           MCRWAAYRG+PLYLEELV+SP HSLIEQSHCA RAKTATNGDGFGIAWYGDRPEPGRYRD
Sbjct: 1   MCRWAAYRGEPLYLEELVTSPKHSLIEQSHCAVRAKTATNGDGFGIAWYGDRPEPGRYRD 60

Query: 181 ILPAWSDCNLKSLARQIRSPLFLAHVRAATGGGTRRDNCHPFTQGTWSFMHNGQISGFER 360
           ILPAWSDCNLKS+ARQIRSPLFLAHVRAATGGGTRRDNCHPF  G WSFMHNGQI  FE
Sbjct: 61  ILPAWSDCNLKSIARQIRSPLFLAHVRAATGGGTRRDNCHPFVFGRWSFMHNGQIGDFEH 120

Query: 361 LRRPMEAMLDDELFNARGGTTDSELMFLLALQFGLREAPVAAMAEMIGVIEDLAESVIGS 540
           LRRPME MLD+EL++AR GTTDSEL+FLLALQFGL   P+ A+AE +  +E LAE +
Sbjct: 121 LRRPMETMLDNELYSARSGTTDSELLFLLALQFGLDRDPLGAVAEALAFVERLAERLGRP 180

Query: 541 ILLRFTAAFSDGKALYAIRYATDRKAPTLYASPVGA--GYCLVSEPLNDDVDAWAEIPDG 714
            L+RFTAAFSDG+ LYA+RYATD KAPTLYA P+G+  GYCLVSEPLNDD  AW E+PDG
Sbjct: 181 ALVRFTAAFSDGRDLYAVRYATDWKAPTLYAGPMGSSGGYCLVSEPLNDDDSAWVEVPDG 240

Query: 715 SAVTVGKDGIDVADF 759
           SA+ VG++G+DV  F
Sbjct: 241 SAIIVGENGVDVRLF 255
```

This result showed an excellent sequence identity to a hypothetical protein in *S. meliloti* (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). As the result was a hypothetical protein, not much more could have been inferred from that data alone. Pfam analysis was carried out and showed the predicted protein contains a glutamine amidotransferases class-II domain. These are a large group of biosynthetic enzymes that can catalyse the removal of the ammonia group from glutamine and then to transfer this group to a substrate to form a new carbon-nitrogen group. Perhaps this gene was involved in biosynthesis of carbon or nitrogen sources, providing the *Rhizobium* with a much needed energy supply (Nelson & Cox, 2000). Furthermore, glutamine amidotransferases are induced as part of the stringent response in *E. coli* (Smulski *et al.*, 2001). A stringent response would fit the data collected, as pRU846 was induced later in its cell cycle and under many stresses.

These data indicates that RL0356 was induced as part of a general stress response. Further analysis of pRU846 is given in section 4.2.4.10.

4.2.4.5. pRU848/RU1512

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	++	+	+	-	-

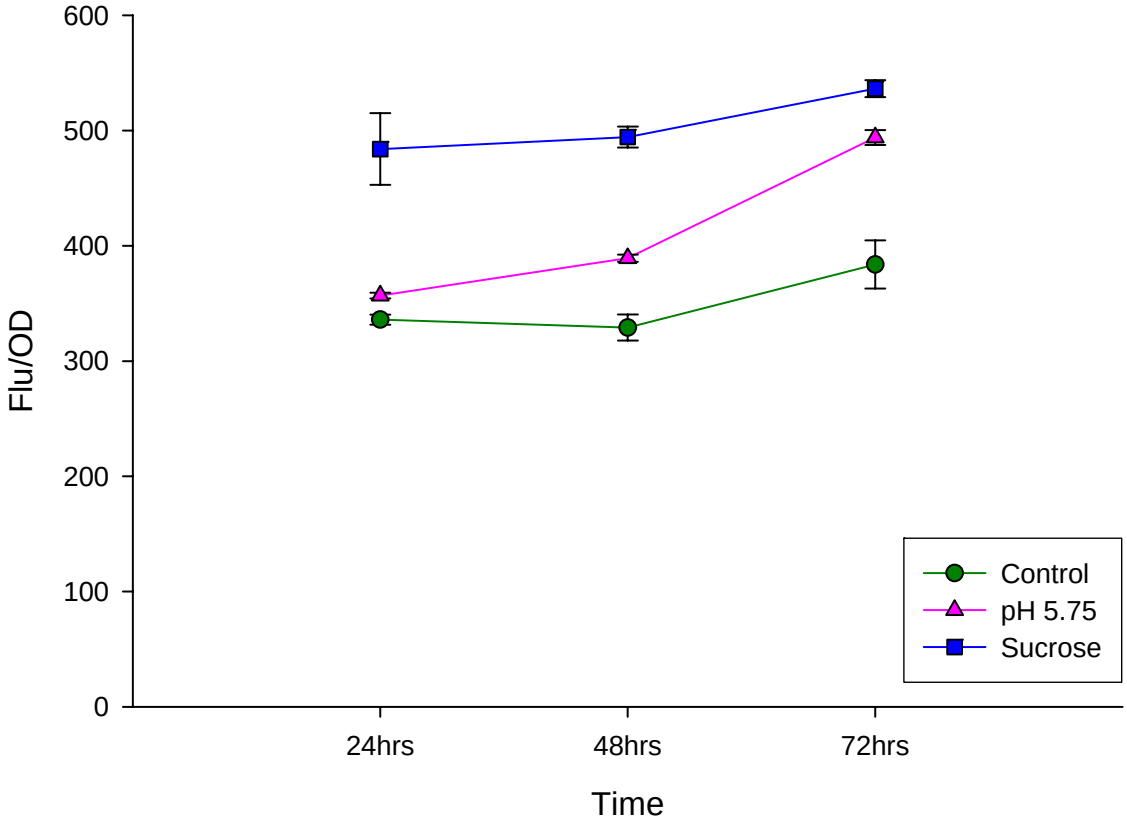


Figure 4.13. RU1512 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded mainly to hyper-osmotic with some induction under acidic conditions. Liquid growth showed that this fusion induced later in its growth cycle as it took 48 hours before any significant induction was observed (Fig. 4.13). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 2420820 – 2423356 (Fig. 4.14).

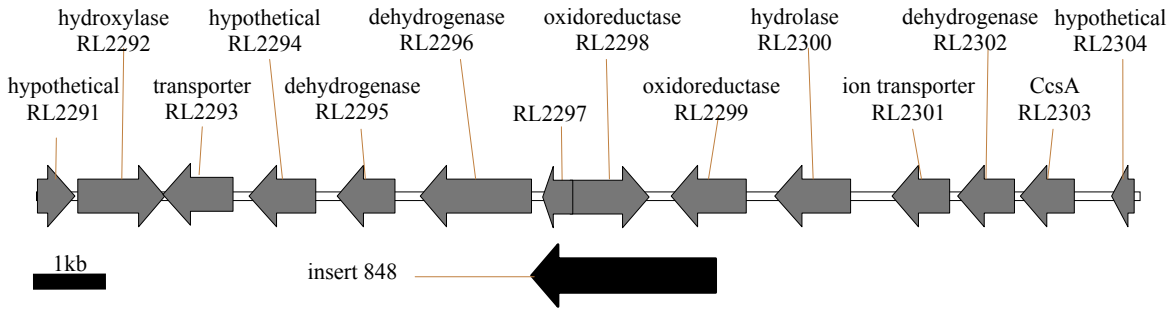


Figure 4.14. Genomic Region of pRU848's Insert. Black arrow shows the region of DNA within pRU848, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL2296 (compliment of chromosome nucleotides 2419291 to 2420817, 509aa). Although most of the insert is made up of DNA from RL2298 it is encoded in the opposite orientation to the promoter. The insert started within RL2299 and so it cannot contain its promoter. As BLAST analysis of RL2297 did not reveal a homologue it was not analysed further. The promoter in the within the insert could activate the transcription of RL2296 to RL2293 inclusive as an operon. However, as there is an intergenic region (that may contain a promoter) between each of these genes, this may not be the case and so only RL2296 was investigated. The best BLAST result for RL2296 is shown below.

Fatty aldehyde dehydrogenase (imported) [*Caulobacter crescentus* CB15] Length = 528aa
Low Complexity Filter disabled

Score = 614 bits (1583), Expect = e-174
Identities = 318/507 (62%), Positives = 379/507 (74%), Gaps = 2/507 (0%)
Frame = +1

```
Query: 7    LSGDLIIGGANVRGAAAAFSAINPANGNPMEPSFAGATKEQVEEATSLAWDAFPVYKETS 186
            L+G+L+IGG    G       +NPA G  +EP+F GAT   VE A +LA +AF  Y+
Sbjct: 4    LTGELLIGGERRFGIHGEIKGVNPATGETLEPAFGGATTADVEAACALAAEAFGPYRSLP 63

Query: 187  LDDRARFLEAIAEGIVALGDDLVMRAIDETGLPRGRIEGERARTVGQLRLFAKEVRDGRF 366
             + RA+FLE+IAE I A+GDDL++R + ETGLPR R+EGER RTVGQLRLFA  +RDG F
Sbjct: 64   YETRAQFLESIAEHIEAIGDDLIVRTMAETGLPRPRLEGERGRTVGQLRLFAGVLRDGGF 123

Query: 367  QELRFDPADTERRPVAKPDLRLRNVALGPVAVFGASNFPLAFSVAGGDTASALAAGCPVV 546
             E R DPA  +R+P+ +PDLRLRNV LGPVAVFGASNFPLAFSVAGGDTASALAAGCPV+
Sbjct: 124  LEARIDPAMPDRKPLPRPDLRLRNVPLGPVAVFGASNFPLAFSVAGGDTASALAAGCPVI 183

Query: 547  VKAHSAHPGTSELVGRAVADAVAACGLPRGTFGLLFDAGFEVGQTLVADHRIRAVGFTGS 726
            VKAH AHPG SELVGRA+  AVAACGLP G F L+ D+G+EV Q LVAD R++A GFTGS
Sbjct: 184  VKAHPAHPGASELVGRAIQAAVAACGLPPGVFALIHDSGYEVSQALVADPRVKAAGFTGS 243

Query: 727  RRGGTALMKIASERKQPIPVYAEMSSINPVILYPNALRSRGAEIGKSFASSLILGAGQFC 906
            RRGG AL+ IA  R +PIP YAEMSSINPVIL P AL++R  +I   F ++L LGAGQFC
Sbjct: 244  RRGGLALLAIAQGRPEPIPFYAEMSSINPVILLPAALKARADKIAPDFVAALTLGAGQFC 303

Query: 907  TNPGLIIAVEGAGLDAFIGSAAAALSESQAQTMLTGGIFDAYCNGVARLSSSPSVSQVAA 1086
            TNPGLI+A++G  LDAFI +A  A+  + A  MLT GI  A+ +GVA L+ +  V+ VA
Sbjct: 304  TNPGLILAIDGPELDAFIEAAGKAVEAAPASVMLTPGICQAFAHGVAALTDAAEVTTVAR 363

Query: 1087 GKDGT--PNQAAAALFETTAAAFLANPELQEEVFGASGLIVRCRDDEELRTVVGSLEGQL 1260
            G  G    +   AALF  TAA FLANP L EEVFGA+ L+VRC    EL  V+ +LEGQL
Sbjct: 364  GVPGPDGSHTGRAALFSVTAADFLANPHLHEEVFGAASLVVRCAGQAELEAVIAALEGQL 423

Query: 1261 TIALHVDAGDIGAASPMISQLELLAGRLLVNGFGTGVEVSPAMVHGGPYPATSDGRSTSV 1440
            TIALH+D  D G A  ++  LEL AGR+LVNGFGTGVEV+PAMVHGGP+P+TSDGR+TSV
Sbjct: 424  TIALHMDEADHGIAGALLPALELKAGRILVNGFGTGVEVAPAMVHGGPFPSTSDGRTTSV 483

Query: 1441 GTLAIYRFLRPVSYQDFPIDLLPEPLK 1521
            GTLAI RFLRPVSYQ+ P  LLP  LK
Sbjct: 484  GTLAIARFLRPVSYQNLPEALLPAELK 510
```

This result showed a good sequence identity to a fatty aldehyde dehydrogenase in *C. crescentus* (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). Dehydrogenases are enzymes involved in cell metabolism, specifically in generating energy and/or in electron transfer systems. It could be that this gene, like RL0356 above, is involved in energy generation.

However, cells that have undergone membrane stress, like that caused by hyper-osmosis or acidity, may also have been affected by a change in electrochemical-gradient. Aldehyde dehydrogenases have been shown in *Bacillus subtilis* to maintain the redox balance of bacteria as part of a general stress response (Petersohn *et al*., 2001). It could be that RL2296 performs a similar function in *R. leguminosarum*.

Additionally, fatty aldehyde dehydrogenases have also been key components in the detoxification of aldehydes formed under oxidative stress (Demozay *et al*., 2004). Although pRU848 did not react to oxidative stress, this information shows that aldehyde dehydrogenases can be involved in a stress response.

These data indicates that RL2296 was induced as part of a general stress response.

4.2.4.6. pRU849/RU1513

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+	++	+	+	+	-	-

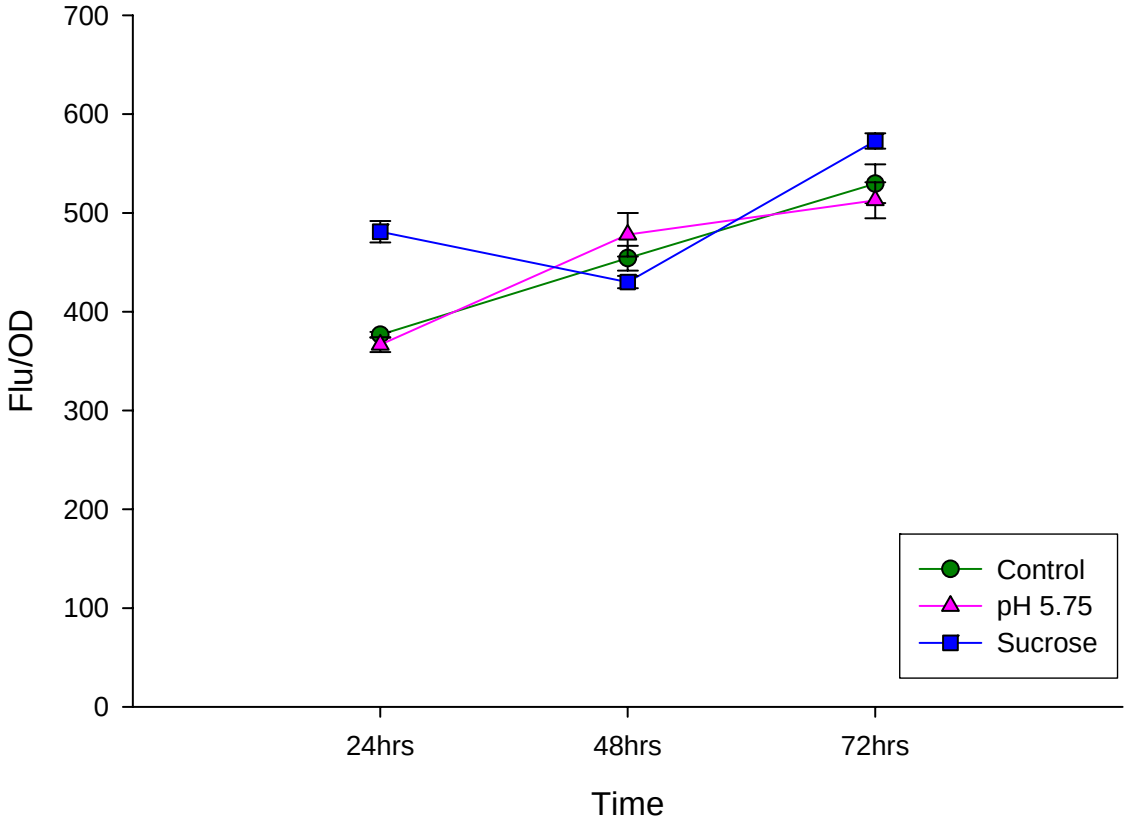


Figure 4.15. RU1513 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded weakly to hyper-osmotic and acidic stress, however, liquid growth showed little induction at any point (Fig. 4.15). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 4471789 – 4475121 (Fig. 4.16).

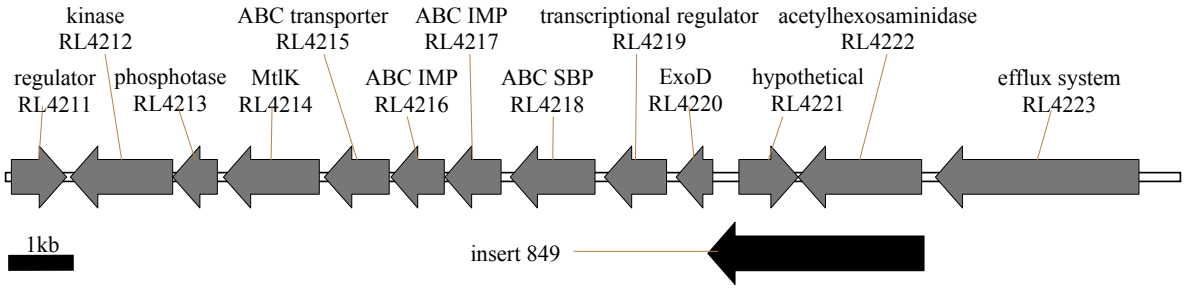


Figure 4.16. Genomic Region of pRU849's Insert. Black arrow shows the region of DNA within pRU849, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL4220 (compliment of chromosome nucleotides 4471294 to 4471860, 189aa). Although most of the insert is made up of DNA from RL4221 it is encoded in the opposite orientation to the promoter. The insert starts within RL4222 and so it cannot contain its promoter. The promoter within the insert could activate the transcription of RL4220 to RL4212 inclusive as an operon. However, as there is an intergenic region (that may contain a promoter) between most of these genes, this may not be the case and so only RL4220 was investigated. The best BLAST result for RL4220 is shown below.

ExoD [EDTA-degrading bacterium BNC1] Length = 217aa **Low Complexity Filter disabled**

Score = 214 bits (544), Expect = 1e-54
Identities = 104/169 (61%), Positives = 127/169 (75%)
Frame = +1

```
Query: 1   MAADQSRERISIGDIFDTMGDRAISALMLIFALPNAFPTPPGTSAVLGAPLVFLAAQLTF 180
           MA D +RERIS+ D+   MGDRA  ALML+FALPN  PTPPGTS +LG PLV+LAAQL
Sbjct: 1   MAEDTARERISVNDLVVAMGDRAFGALMLVFALPNVLPTPPGTSGLLGLPLVYLAAQLML 60

Query: 181 GLKPWLPKVIANRSMRREDFETIVGRIHRWLAWAERMLKPRLAIFAEPPAEYLAGAACLL 360
           G +PWLPK I +RS+RREDF   + R   WL  AER+L+PRL     P AE L GA CL+
Sbjct: 61  GQRPWLPKFIGSRSIRREDFAAFIERAAPWLRRAERLLRPRLVFLVSPAAERLVGAVCLV 120

Query: 361 LSIVLLLPVPLGNILPAVTISVFAFGILGRDGLFALIGFVMTAVSLVVA 507
           L+IVL LPVPLGN+LPA++ISVF+FGILGRDGL+ + G ++  +S VVA
Sbjct: 121 LAIVLFLPVPLGNMLPALSISVFSFGILGRDGLWVICGVLLAVISAVVA 169
```

This result showed a good sequence identity to ExoD in EDTA-degrading bacterium BNC1 (a member of the α-proteobacteria, the same group as *R. leguminosarum*). ExoD is involved in exopolysaccharide (EPS) synthesis and is one of the genes required for nodule invasion (Reed & Walker, 1991b). It was suggested that EPS may have had a protective role, enabling *Rhizobium* producing greater amounts of EPS to survive in acidic conditions better then *Rhizobium* that produce smaller amounts (Cunningham & Munns, 1984). Furthermore, ExoD mutants are alkaline sensitive, but can nodulate in slightly acidic conditions (Reed & Walker, 1991a). This data indicates that ExoD, and its transcription, is significantly linked to changes in pH. The fact that ExoD can only deal with slightly acidic conditions is corroborated by the result of pRU849 and its low GFP induction on acidic stress. Perhaps in order to survive a greater decrease in pH, more EPS is required, but the pRU849 data clearly showed that RL4220 induction is not great enough to meet these demands.

Regarding the induction under hyper-osmotic conditions, acidic stress could induce a hyper-osmotic stress response (and visa versa) by altering electrochemical gradients; an effect that has been previously reported (Leyer & Johnson, 1993). However, the data strongly suggested low pH as the main cause of induction. RL4220 showed 25% identity and 40% similarity to the *S. meliloti* version of the gene that encodes ExoD and analysis indicated it to be the only copy of an ExoD-like gene in 3841.

These data indicates that RL4220 was induced as part of a stress response specific to acidic stress.

4.2.4.7. pRU850/RU1514

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+	++	+	+	+++	-	-

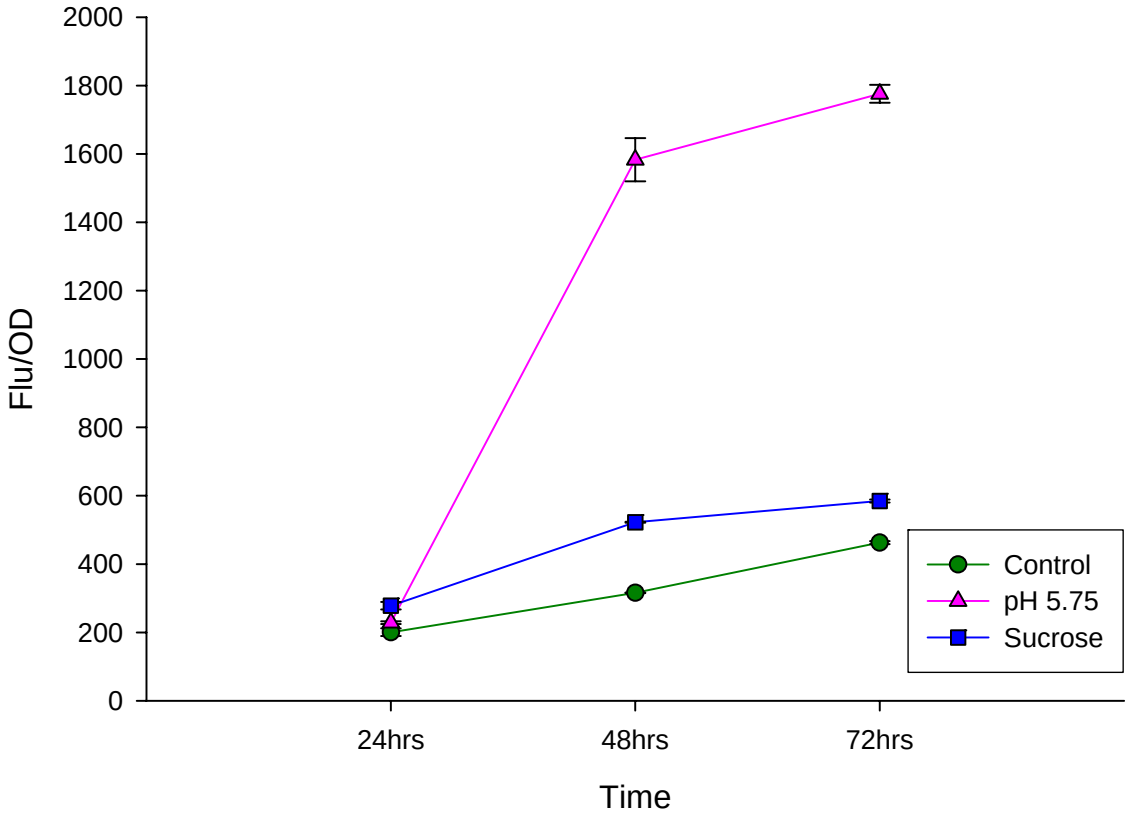


Figure 4.17. RU1514 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded strongly to acidic stress, with some hyper-osmotic activation. Liquid growth showed that

this fusion induced later in its growth cycle as it took 48 hours to get a 5-fold induction (Fig. 4.17). *Rhizobium* DNA from this fusion mapped to the chromosome, nucleotides 1360288 – 1361851 (Fig. 4.18).

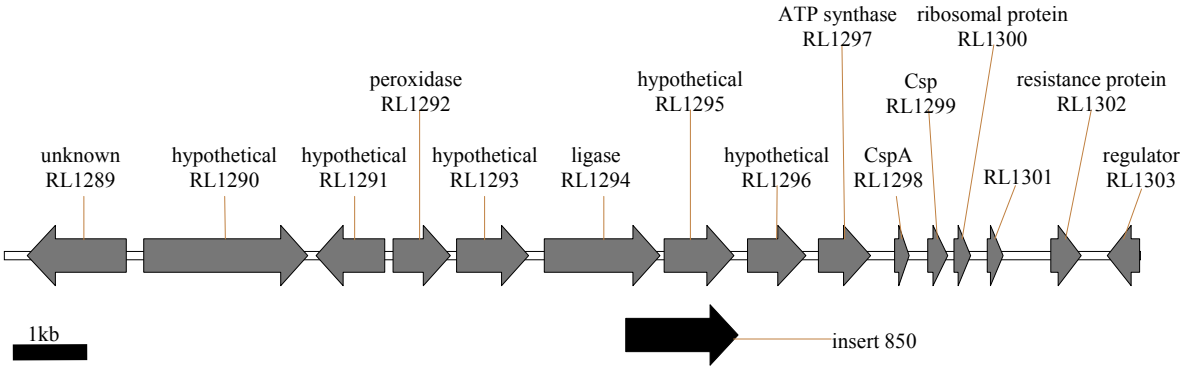


Figure 4.18. Genomic Region of pRU850's Insert. Black arrow shows the region of DNA within pRU850, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was either RL1295 or RL1296 (chromosome nucleotides 1360817 to 1361800, 328aa and 1361993 to 1362796, 268aa respectively). The insert started within RL1294 and so it cannot contain its promoter. Although most of the insert is made up of DNA from RL1295, its transcription should end before the insert does and so another promoter (one for RL1296) could be downstream of it. However, the insert stops before the predicted transcriptional start of RL1296 and so it may not contain its promoter. Even if this is not the case RL1296 may be transcribed by the same promoter as RL1295 as an operon. As such both RL1295 and RL1296 were analysed.

The promoter in the within the insert could also transcribe RL1297 to RL1302 inclusive as an operon. However, as there is an intergenic region (that may contain a promoter) between most of these genes, this may not be the case so those genes were not investigated. The best BLAST results for RL1295 and RL1296 are shown below.

RL1295 – Conserved hypothetical protein Atu0841 [*Agrobacterium tumefaciens* str. C58] Length = 321aa **Low Complexity Filter disabled**

Score = 281 bits (718), Expect = 3e-74
Identities = 144/297 (48%), Positives = 195/297 (65%), Gaps = 5/297 (1%)
Frame = +1

```
Query: 97  LQRRAVLAGLAG----ALILPRMAAAFDVPDEPRLAKRDYAKVRHQFRTKLLQKGPAPDK 264
           L RR++LAG  G    A + PR  A    P+   L ++DYA+ R +F T LL+K  AP+K
Sbjct: 23  LTRRSLLAGAMGLSAAAFLPPRAKAGIVAPEVLPLERQDYAEARKRFHTHLLRKMAAPEK 82

Query: 265 YEPLNAPADADKIFYRSGY-GELELAAWVSKYKRERAARPAVLFLHGGNAMGIGHWQLMK 441
              L  P  A+++ Y  G  G +EL AW+S Y+  +  +PAVLFLHGGNA G GHW LMK
Sbjct: 83  SSVLGTPPGAERVTYPGGPDGSIELVAWLSHYQPSKTLKPAVLFLHGGNATGDGHWALMK 142

Query: 442 PYMDAGYVVMMPSLRGENGQMGNFSGFYDEVDDVLAATERLAHLPGVDPERLFIAGHSIG 621
           PY +AGYVV++PS RGENGQ G++SGFY+E  D LAA   L +LPG+D  R FIAGHS G
Sbjct: 143 PYWEAGYVVLLPSFRGENGQSGHYSGFYNETADALAAATYLENLPGIDRNRFFIAGHSNG 202

Query: 622 GTLTMLTAMTTHKFRAAAPISGNPDAFRFFNRYPQDIRFDDSNAHEFEVRSALCYAHSFK 801
           GTLT+L AM + KFRAAAPIS   +++R+FNRY  +I FD+++  EF +RS++C+  S K
Sbjct: 203 GTLTLLAAM-SRKFRAAAPISAGVNSWRYFNRYSDEICFDETDEREFIMRSSVCFGPSLK 261

Query: 802 CPVRVVHGTEEPHFNDRADLLARRARGAGVHIETETVAGNHTSALPAEIEQSIRFFH 972
           CP  ++ GTEE  F+    L   RA  +G  I+ + + G H   +P  +E+SIR F+
Sbjct: 262 CPALLLRGTEERPFDADHQLFVDRALTSGFKIDKKLLPGTHNGVVPHAVEESIRLFN 318
```

RL1296 – Conserved hypothetical protein Atu0844 (imported) [*Agrobacterium tumefaciens* C58] Length = 302aa

Score = 328 bits (840), Expect = 1e-88
Identities = 170/296 (57%), Positives = 200/296 (67%), Gaps = 32/296 (10%)
Frame = +1

```
Query: 13  RRVFPFAGLVIAVASLSGCNILIPDVAADSPARFVQETSPVFYQPPGVDPRRVRPIPDQP 192
           RR     GL + +  +SGC + + D     P  FV+ET+PVF     V   R   +P QP
Sbjct: 10  RRAASLLGLAL-LPLMSGC-LFVTDTTRMDPDVFVRETAPVF-NFNSVSSNRQPELPPQP 66

Query: 193 VP-QTR--ELYKTQFHQ-----------------------------TYGLPVTNPVHMAM 276
               TR  +L++T+FHQ                             YGLPV+NP+H  M
Sbjct: 67  GQLSTRPPDLFRTRFHQEYGPPVRGQGLQAPQVQGYNVPQAQGQTMAYGLPVSNPLHRVM 126

Query: 277 YGQQRDEDFTLPAIPVSRVQPQFLRQEVDYQTTERPGTVVIDTKTHFLYFVEGNGKAMRY 456
           YG  RDED +LPAIP  R+ P++LRQEV YQT E PGT+V+DT+ HFLY V+  GKA+RY
Sbjct: 127 YGPIRDEDRSLPAIPYGRIDPRYLRQEVSYQTAEAPGTIVVDTRQHFLYLVQSGGKAIRY 186

Query: 457 GVGLGRDGYAWSGRGVIQWKQKWPRWTPSVEMVSRQPEVRPFGAENGGMNPGLMNPLGAR 636
           GVGLGRDGYAWSGRG IQWK KWPRWTP  EMV RQPE+    A NGGM PGL NPLGAR
Sbjct: 187 GVGLGRDGYAWSGRGKIQWKAKWPRWTPPDEMVKRQPELTSISAANGGMTPGLNNPLGAR 246

Query: 637 AMYIFKDGQDTLYRVHGTPDWQSIGKATSSGCVRMLNQDVIDLYDRVPARAEIVVM 804
           A+YIFKDG+DTLYRVHGTPDWQS+GKATSSGCVRMLNQDVIDLY+RVP  A+IVV+
Sbjct: 247 ALYIFKDGKDTLYRVHGTPDWQSVGKATSSGCVRMLNQDVIDLYERVPQGAQIVVI 302
```

RL1295 showed a reasonable sequence identity to a hypothetical protein in *A. tumefaciens* (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). As the result was a hypothetical protein, not much more could have been inferred from that data alone, however, Pfam analysis was carried out and revealed no domains for within the protein. Therefore, not much more can be predicted about RL1295 with the data retrieved.

RL1296 showed a reasonable sequence identity to a hypothetical protein in *A. tumefaciens* (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). As the result was a hypothetical protein, not much more could have been inferred from that data alone. Pfam analysis was carried out and showed the predicted protein contains an ErfK/YbiS/YcfS/YnhG domain. This function of this domain remains largely unknown, but it is found in a wide range of bacteria. The conserved region contains a conserved histidine and cysteine, suggesting that these proteins have an enzymatic activity. Several members of this family contain peptidoglycan binding domains, indicating they may use peptidoglycan or a precursor as a substrate. Therefore, RL1296 could have encoded an enzyme that in some way helps alleviate cells from stress. Unfortunately not much else can be predicted from these data.

Interestingly, the BLAST results for predicted proteins encoded by the genes immediately downstream of this insert showed that they share sequence identity with previously characterised stress related proteins. RL1297 resembles an ATP-synthase component, an enzyme reasonable for energy production in the form of ATP (general stress response), RL1298 and RL1299 both resemble cold shock proteins (cold shock response) and RL1302 resembles a hydroperoxide resistance protein (oxidative stress response). Although none of these genes would have been necessarily activated by low pH, their close proximity to an acid inducible promoter indicates that this region of *R. leguminosarum*'s DNA may be important to the stress response of 3841.

These data indicates that RL1295 and RL1296 were induced as an operon as part of a stress response specific to acidic stress.

4.2.4.8. pRU853/RU1517

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	+	+	+	-	-

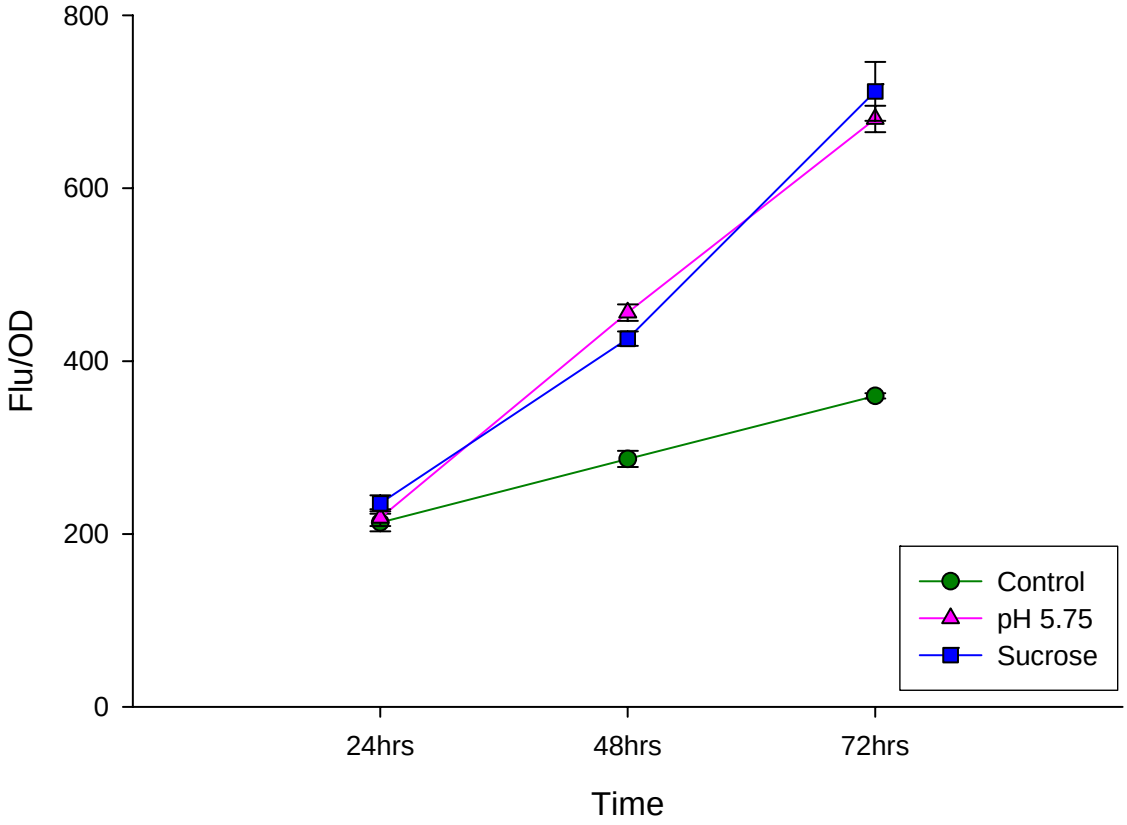


Figure 4.19. RU1517 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded equally to acidic and hyper-osmotic stress. Liquid growth showed that this fusion induced later in its growth cycle as it took 48 hours before any significant induction was observed (Fig. 4.19). *Rhizobium* DNA from this fusion mapped to the pRL9 plasmid, compliment of nucleotides 14407 – 15441 (Fig. 4.20).

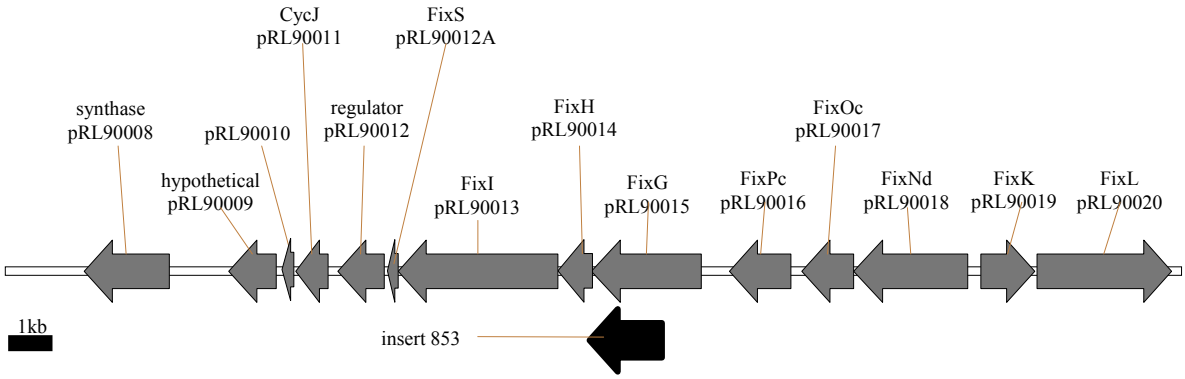


Figure 4.20. Genomic Region of pRU853's Insert. Black arrow shows the region of DNA within pRU853, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre except pRL90012A, which had not been assigned a number at time of writing. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL90014 (compliment of pRL9 nucleotides 13840 to 14322, 161aa). The entire insert starts and ends within pRL90015 and so it cannot contain its promoter. This also means that there may be a cryptic promoter for pRL90014 within pRL90015, which was induced by the stress and caused the GFP production observed with pRU853. If such a promoter is within the insert, it could activate the transcription of pRL90014 to pRL90008 inclusive as an operon. However, as there is an intergenic region (that may contain a promoter) between most of these genes, this transcription may be limited to pRL90012, pRL90012A, pRL90013 and pRL90014. The best BLAST result for pRL90014 is shown below.

FixH protein [*Rhizobium leguminosarum*] Length = 158aa **Low Complexity Filter disabled**

Score = 274 bits (700), Expect = 7e-73
Identities = 136/142 (95%), Positives = 141/142 (99%)
Frame = +1

```
Query: 1   MKTSAQGFTGLHMLLATSAFFGVVIAVNVTMAFYASSSWSGLVVKNTYVASQEFNRKAAA 180
           MKTSAQGFTGLHMLL+TSAFFGVVIAVNVTMAFYASSSWSGLVV+NTYVASQEFNRKAAA
Sbjct: 1   MKTSAQGFTGLHMLLSTSAFFGVVIAVNVTMAFYASSSWSGLVVENTYVASQEFNRKAAA 60

Query: 181 MKAMAASGIEGNLSIKGHEIRYDIHDKSGSPAIVDDVVLNFKRPVGDHEDFLLTLKKAAA 360
           MKAMAASGIEGNLSIKGHEIRYDIHDKSGSPAIVDDV+LNFKRPVGDHEDFLLTL+K AA
Sbjct: 61  MKAMAASGIEGNLSIKGHEIRYDIHDKSGSPAIVDDVILNFKRPVGDHEDFLLTLRKTAA 120

Query: 361 GRFEAEHDLAEGDWIVEAISRN 426
           GRFEAEHDLA+GDWIVEAISRN
Sbjct: 121 GRFEAEHDLADGDWIVEAISRN 142
```

Score = 33.9 bits (76), Expect = 1.7
Identities = 14/15 (93%), Positives = 15/15 (100%)
Frame = +2

```
Query: 431 AWSSCMRQDASIPRS 475
           AWSSCMRQ+ASIPRS
Sbjct: 144 AWSSCMRQNASIPRS 158
```

This result showed an excellent sequence identity to FixH found within *R. leguminosarum* itself. The whole *fixGHIS* operon is present in the pRL10 plasmid and so pRL90014 appears to be a homologue to FixH. This explains why the sequence of pRL90014 did not get 100% identical to *fixH* during BLAST analysis, as it was compared to the sequence of the *fixH* in pRL10 (pRL100209). FixH is known to be transcribed as part of the *fixGHIS* operon (Kahn *et al*., 1989). FixG has five transmembrane helices and a central region resembling bacterial-type ferredoxins (Neidle & Kaplan, 1992), indicating a potential role with metal stress/acquisition or in redox reactions. The *fixGHIS* operon encodes proteins that form a membrane-bound complex involved in symbiotic nitrogen fixation, combining the FixI cation pump with a redox process catalyzed by FixG (Kahn *et al*., 1989). Acidic stress could cause protonated species to form and a cation pump would aid in their removal; whilst a redox reaction would restore altered electrochemical gradients that could also be caused by the lowered pH.

However, it has been postulated that FixI is involved in copper uptake in *Bradyrhizobium japonicum* (Preisig, *et al*., 1996) and as mentioned in Chapter 3, copper availability increases as pH decreases (Tiwari *et al*., 1996a; Dilworth *et al*., 2001). This could mean that low pH caused the concentration of free copper to increase, which led to the induction of pRL90015 and its operon. If this is the case, then pRL90015 may be

induced by copper but this was directly due to the acidic conditions and it would remove copper and other reactive metals from the cell.

As the *fixGHIS* is known to be induced as an operon it seems strange that a promoter exists that would only activate three of the four required genes, given that the promoter within pRU853 cannot be for pRL90015. One explanation is that *fixG* is activated by the stress and contains a cryptic promoter, which in turn transcribes the *fixHIS* operon.

These data indicates that pRL90015 to pRL90012A were induced as an operon as part of a stress response specific to acidic stress. However, due to the unknown nature of the transcription of this operon, this fusion was not investigated further.

4.2.4.9. pRU854/RU1518

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+	++	+	+	-	-	-

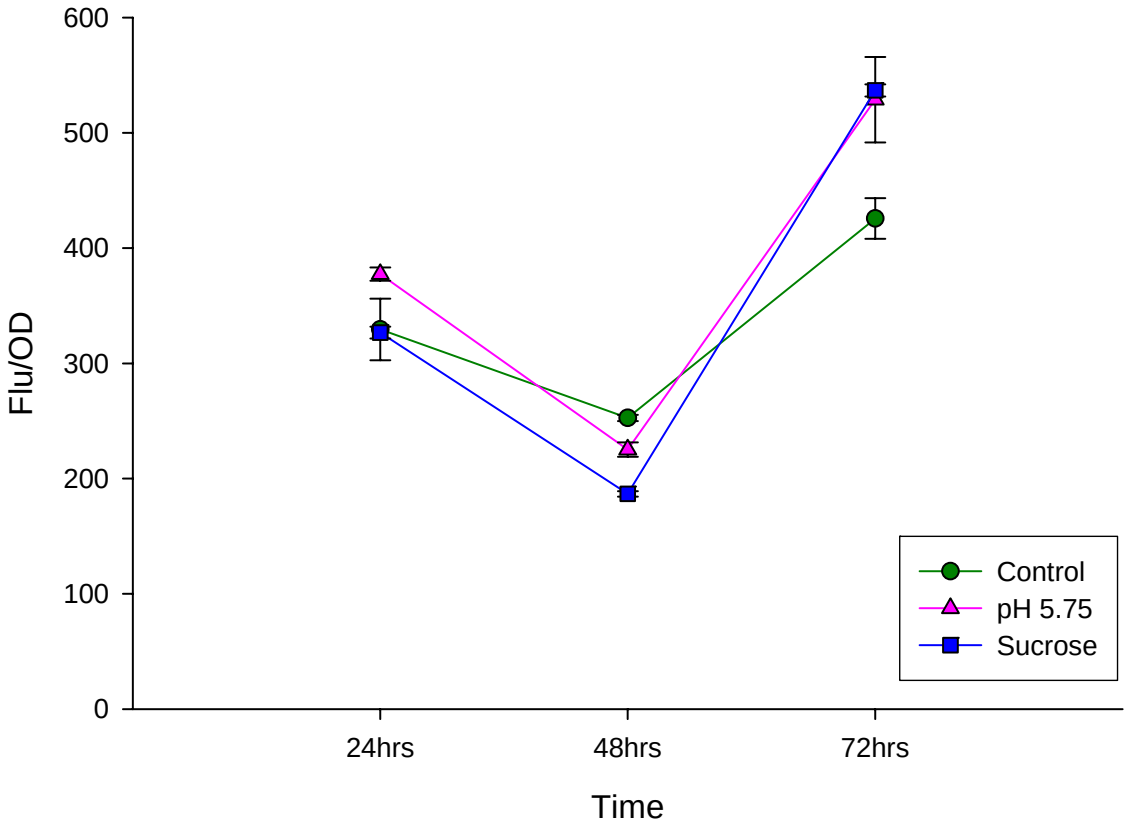


Figure 4.21. RU1518 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded specifically to hyper-osmotic stress, however, liquid growth showed little induction at any point (Fig. 4.21). *Rhizobium* DNA from this fusion mapped to the pRL12 plasmid, compliment of nucleotides 615295 – 617202 (Fig. 4.22).

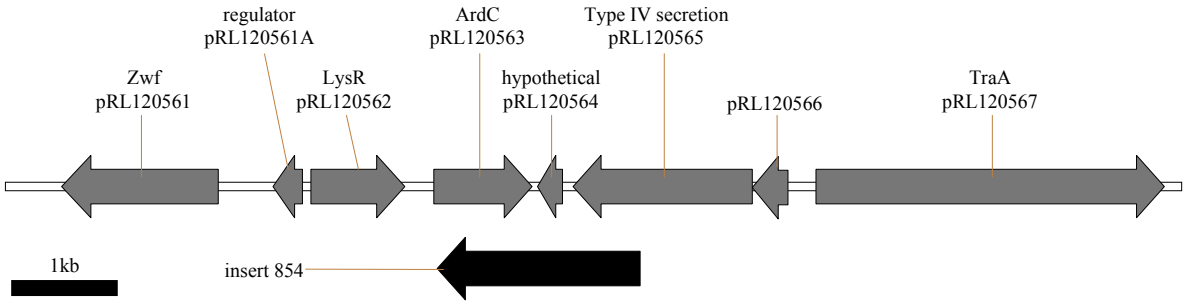


Figure 4.22. Genomic Region of pRU854's Insert. Black arrow shows the region of DNA within pRU854, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre except pRL120561A, which had not been assigned a number at time of writing. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL120564 (compliment of pRL12 nucleotides 616236 to 616478, 81aa). The insert starts within RL120565 and so it cannot contain its promoter. Although the insert contains all of DNA from pRL120564, its transcription should end before that of the insert and so another promoter (one for pRL120563) could be downstream of it. However, pRL120563 was transcribed in the opposite orientation to that of the insert and so that it is impossible that the promoter for pRL120563 is present in the insert. This made it highly likely that the promoter is for pRL120564 but transcription reads through pRL120563, which led to the production of GFP observed in pRU854. The best BLAST result for pRL120564 is shown below.

Hypothetical 22.9 kD protein Y4dW [*Rhizobium* sp. NGR234] Length = 204aa **Low Complexity Filter disabled**

Score = 112 bits (279), Expect = 4e-24
Identities = 51/78 (65%), Positives = 61/78 (78%)
Frame = +1

```
Query: 1   MTREPYRLYIERIDPSKNMARYYALSIEPNLFGGTSLVRSWGRIGSRGQQKIHVFDSEAK 180
           M  +PYRLY+ER+DPS+NMARYYA+SIEPNLFG   L+R WGRIG++GQ  +H F  E
Sbjct: 84  MISQPYRLYVERLDPSRNMARYYAMSIEPNLFGDICLLRKWGRIGTKGQMMVHHFGQEED 143

Query: 181 AVDLLLTLLRKKRSRGYR 234
           AV L L LLR+KR RGYR
Sbjct: 144 AVRLFLDLLRQKRKRGYR 161
```

This result showed a good sequence identity to a hypothetical protein in *Rhizobium*. As the result was a hypothetical protein, not much more could be inferred from that data alone. Pfam analysis was carried out and showed the predicted protein is made up of a WGR domain. This domain is found in a variety of polyA polymerases, as well as in the *E. coli* molybdate metabolism regulator and in other proteins of unknown function. PolyA polymerases are known to add a polyA tail to mRNA post-transcriptionally, which regulates the stability of the mRNA and the initiation of its translation (Gallie, 1991). As mentioned above, when bacteria undergo stress they need to produce novel compounds to counter the shock. Production of protein obviously requires mRNA, and therefore the role for a polyA polymerase could be pivotal to stress response.

These data indicated that pRL120564 was induced as part of a stress response specific to hyper-osmotic stress.

4.2.4.10. pRU855/RU1519

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+++	+++	+	+	+	-	-

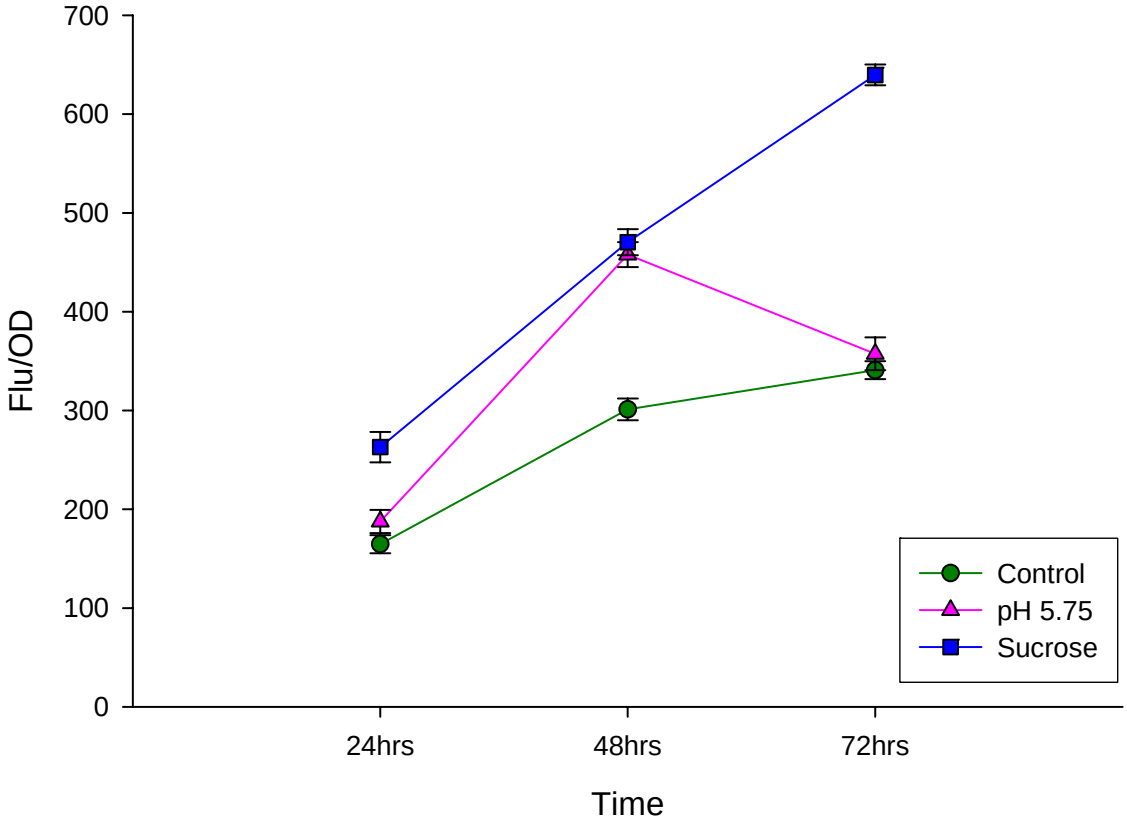


Figure 4.23. RU1519 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures of siblings RU1516, RU1519, RU1520 and RU1526 (as in explained in section 4.2.2).

The screening results on solid media for this fusion showed that it responded to hyper-osmotic stress and a little under acidic stress. Liquid growth showed significant induction after 24 hours under hyper-osmotic stress (Fig. 4.23), although as mentioned in Chapter 3 this was specific to an osmotic upshift generated by sucrose. *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 385144 – 387536 (Fig. 4.24).

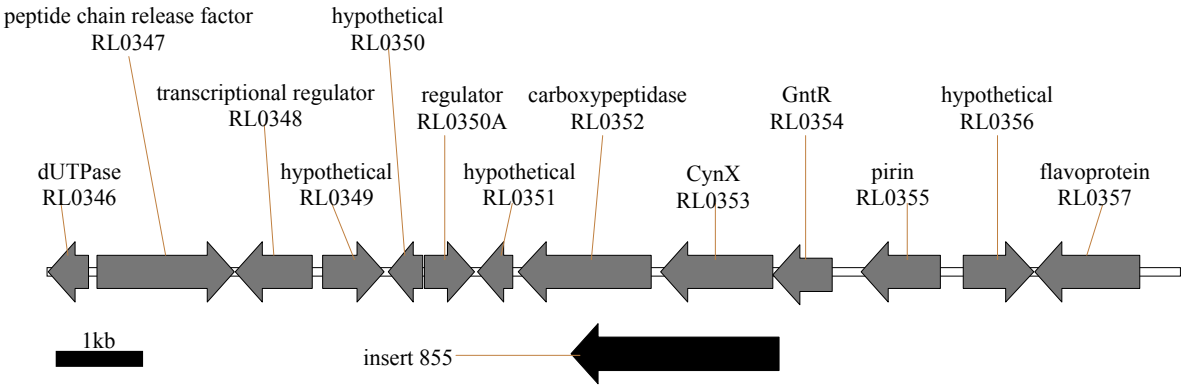


Figure 4.24. Genomic Region of pRU855's Insert. Black arrow shows the region of DNA within pRU855, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre except RL0350A which had not been assigned a number at time of writing. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL0352 (compliment of chromosome nucleotides 384542 to 386068, 309aa). Although most of the insert is made up of DNA from RL0353, its transcription should end before that of the insert and so another promoter (one for RL0352) could be downstream of it. Even if this is not the case it would suggest that RL0352 was activated by the same promoter as RL0353. In both scenarios RL0352 would be transcribed from the promoter within the insert of pRU855 (RL0353 was also investigated during analysis).

The promoter within the insert could also transcribe RL0351 as an operon with RL0352. However, as there was an intergenic region (that may contain a promoter) between RL0351 and RL0352, this may not be the case and so RL0351 was not investigated. The best BLAST result for RL0352 is shown below.

Further analysis showed that the insert for pRU846 mapped in close proximity to where pRU855 mapped (Fig 4.12).

Carboxypeptidase C (cathepsin A) [*Mesorhizobium* sp. BNC1] Length = 511aa **Low Complexity Filter disabled**

Score = 298 bits (762), Expect = 3e-79
Identities = 185/511 (36%), Positives = 257/511 (50%), Gaps = 14/511 (2%)
Frame = +1

```
Query: 10   RTLFLLATFMASLAPALSHAQESNQPRANVQS--------GARDGVLKLLPPDSVTEHAL 165
            + L    A  +  ++    HAQE  +P +   S         + +     LP    T H L
Sbjct: 6    KRLLSAAALVLLISTQSLHAQEEAEPSSQQTSEQEERPAEASSENKQSRLPEARRTLHRL 65

Query: 166  TIGDRKFAYTATAGTLDLFGQDGAQTGAIFYTAYVARDSG-ANRPLTFAFNGGPGAASAY 342
             I   +  + ATAG + L G +      I + +Y    +  ANRP+TFA NGGPGAASAY
Sbjct: 66   QIDGEELPFQATAGAITLTGTNDRPEAEIAFVSYTKEGADKANRPVTFAINGGPGAASAY 125

Query: 343  LHLGLVGPKVLDF-GPDGRDGANAKLVDNPQSWLDFTDLVLIDPIGTGWSRTTKADDA-- 513
            LHLG +GP +L   G       +  LV+NP +WL FTDLV IDP GTG+SR     D
Sbjct: 126  LHLGAIGPWLLPMSGERIVPSQSIALVENPDTWLSFTDLVFIDPAGTGFSRLIDPTDRLR 185

Query: 514  -NYYNVDADAQSIAKAIALYVAHNNRSTSPKYLLGESYGGFRAAKVASVLQQSQGIIVAG 690
              YY+V  D  ++A+ I  ++  N+R  SPKY +GESYGGFR  +VA  LQ   G+ + G
Sbjct: 186  ERYYSVRGDVDALARFIRQWLVENDRLVSPKYFVGESYGGFRGPRVAEALQTRYGVALKG 245

Query: 691  AVMLSPLIEGQLMFNADQFALGAALELPSLAAAELDRHKAFDEEKQKEAETFALGDYLTT 870
              ++SP+++       D   L  A  LPSL  AA L+R  AF E  Q+EAE +A G ++T
Sbjct: 246  LTLVSPVLDFGWFDQPDYSPLAKASYLPSLVAAALERRGAFSEAAQREAEAYAAGPFVTD 305

Query: 871  LAGPPPTGADAAAFYGRIARLTGIPEDIVSRNRGFLGSS-FVKHSDAGSGEVMSSYDASF 1047
            L       A  A   GR+A LTG+  D +    G +      +      G + S YD +
Sbjct: 306  LLRGTQDEAAKARIVGRVAELTGVSPDALRDFEGRIDMEVLTRELLRDGGRIASFYDTTI 365

Query: 1048 AAPDPYPESDYDRGDDAILDGFTRAYGGAFADYARNELGFRTEMTYSLLDGDISRRWEWG 1227
             A  P   S  + G + +LD        A  ++ R  L +  +  Y LL   ++   WEWG
Sbjct: 366  VADAPDHGSILENGPEPVLDAMLAPITSAMLNHYRQNLQWLPQRPYHLLARNLD--WEWG 423

Query: 1228 GGRGGGSRFQASATDDIRQLLAANPAFHLLIAHGYSDVVTPYGVSRYVVDHLPPSLAGGR 1407
             G+      Q  A   +R +LA +P F +L+ HGY+D+VTPY  +  ++  L P    GR
Sbjct: 424  EGKE-----QPEAVSALRNVLALDPEFRVLVVHGYTDLVTPYFGTELILRQLRPFEPLGR 478

Query: 1408 VGLKLYRGGHMFYTRAEQRAAFTADAKAFYA 1500
            V  K Y GGHMFYTRA+ R A   DA   Y+
Sbjct: 479  VLRKNYEGGHMFYTRADSRHALRQDAFQLYS 509
```

This result showed a sequence identity to carboxypeptidase C in *Mesorhizobium* species BNC1 (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). Carboxypeptidases catalyze the hydrolysis of the terminal amino acid of a polypeptide from the end that contains a free carboxyl group and are usually employed as digestive enzymes within eukaryotic organisms. Carboxypeptidase C has a broad specificity and is at its optimum between pH 4.5 and 6.0 (interesting to note, given there was some induction observed with pRU855 at pH 5.75). Despite extensive research, no previously recorded link between bacterial stress response and carboxypeptidases could be found. However, Nelson and Cox (2000) did report that carboxypeptidases target precursors of signalling molecules and/or enzymes. The cleavage of the carboxyl end of such a protein could activate any

signalling or enzymatic properties it has. Signalling sequences are usually at the amino terminus of a protein (Nelson and Cox, 2000). If this activated protein is a signal molecule it may bring about other stress responses in the cell as part of a signalling cascade. This would mean the carboxypeptidase has an indirect role in stress responses.

Alternatively, as carboxypeptidases can as a as digestive enzyme involved in breaking down molecules, it may function in a role similar to proteases. These enzymes are used to degrade misfolded proteins that have been generated by stress, although they are usually associated with an acid (or heat) stress response.

These data indicated that RL0353 may be induced as part of a stress response specific to hyper-osmotic stress.

Interestingly, RL0354 appears to encode for a protein with sequence identity to a GntR regulator. This class of regulator repress the transcription of genes until such time when they are required (Haydon & Guest, 1991). RL0354 could encode for a similar type of regulator involved in regulating RL0353 and stress response in *R. leguminosarum*. The proximity of the insert of pRU846 and pRU855 to RL0354 made them useful fusions and so RL0354 was investigated further (see later chapters).

4.2.4.11. pRU857/RU1521

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	+	+	+	-	++

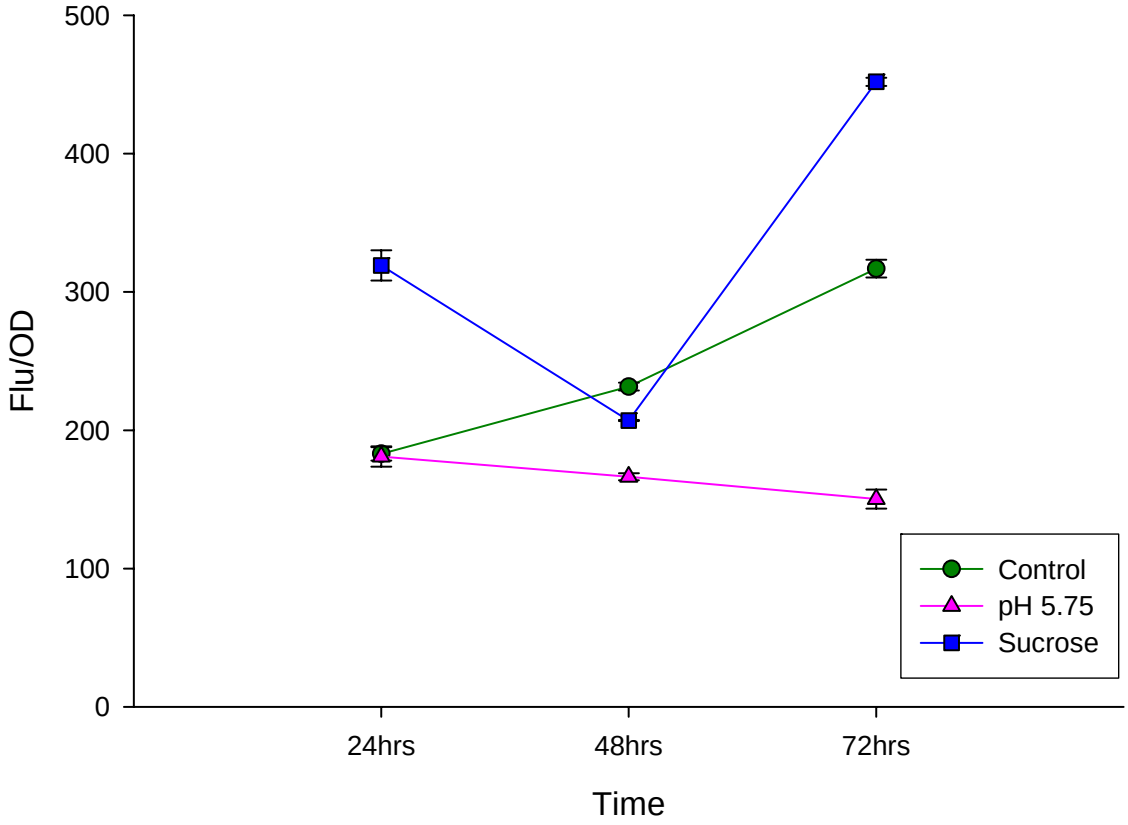


Figure 4.25. RU1521 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded to hyper-osmotic and oxidative stress and a little under acidic stress. Liquid growth showed significant induction after 24 hours under hyper-osmotic stress, but little induction for acidic stress (Fig. 4.25). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 1529872 – 1532220 (Fig. 4.26).

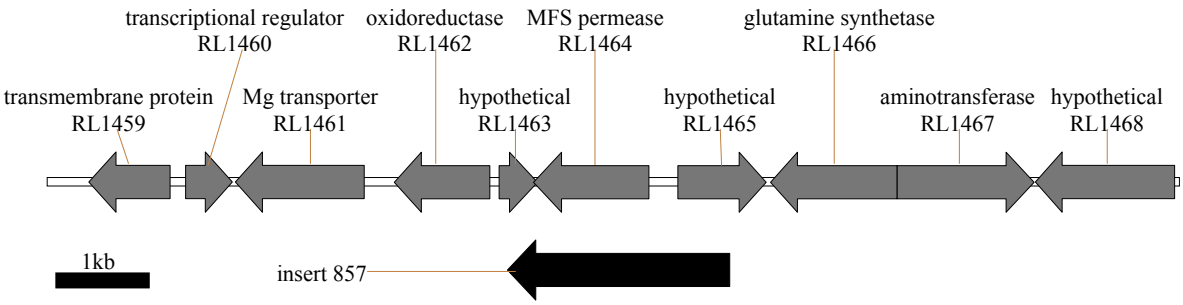


Figure 4.26. Genomic Region of pRU857's Insert. Black arrow shows the region of DNA within pRU857, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL1464 (compliment of chromosome nucleotides 1530146 to 1531369, 308aa). Although the insert contains the DNA from RL1465, it is encoded in the opposite direction to the promoter. RL1464's transcription should end before the insert but RL1463 is encoded in the opposite direction to the promoter within pRU857, a similar situation to that encountered with pRU854 (section 4.2.4.9). It was therefore theorised that RL1464 is induced by the stress and the transcription reads through RL1463, which led to GFP production in pRU857. However, unlike pRU854 where the final 894bp (47% of insert) are encoded in the 'wrong' orientation, only the last 310bp (13% of insert) are encoded in the 'wrong' orientation in pRU857 and so this theory could be tested. To do so it was necessary to get rid of the last ~300bp of the insert (thus removing the hypothetical gene region) and fortunately a *Pst*I site was present upstream of the end of RL1464 (Fig. 4.27).

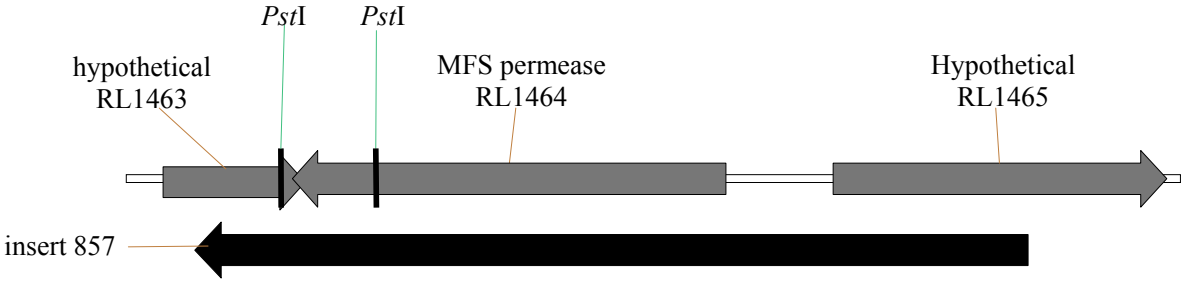


Figure 4.27. Close-up of Figure 4.26. Section of Figure 4.26, black arrow shows the region of DNA within pRU857, grey arrows show RL1463, RL1464 and RL1465, Where *Pst*I cuts this region of DNA is also shown.

The same site is present in pOT downstream of the *Sal*I insertion site but upstream of the *gfpuv* gene, which allowed this enzyme to be used to remove 543bp from pRU857 and therefore all of the RL1463 DNA that was present. This removal did not affect the potential promoter containing region upstream of RL1464's DNA. This trimmed fusion (pRU1216) was tested in *R. leguminosarum* (strain RU1848) on AMA (10mM glc, 10mM NH_4) + 100mM sucrose and AMA (10mM glc, 10mM NH_4) buffered at pH 5.75. Results were identical as those recorded with pRU857. This result proved that the promoter within the insert of pRU857 is nothing to do with RL1463 and therefore transcribes RL1464. These data supported the theory that pRL120564 is induced by stress (section 4.2.4.9), which led to GFP production in pRU854.

The best BLAST result for RL1464 is shown below.

MFS permease (imported) [*Agrobacterium tumefaciens* C58] Length = 409aa **Low Complexity Filter disabled**

Score = 550 bits (1416), Expect = e-155
Identities = 275/386 (71%), Positives = 325/386 (84%)
Frame = +1

```
Query: 55   SAMTVALVQLALACGGFGIGTGEFAIMGLLPNVAQTFSVTTPQAGYVISAYALGVVVGAP 234
            S + +AL++LALA GGFGIGTGEFAIMGLLP+VA T+ VT PQAGYVI+AYALGVV+GAP
Sbjct: 23   SPLAIALIELALAAGGFGIGTGEFAIMGLLPDVATTYGVTVPQAGYVITAYALGVVIGAP 82

Query: 235  VIAVLAAKMARRTLLLTLMLIFAAGNISSAMAPTFESFTLLRFVSGLPHGAYFGVAALVA 414
            +IAVLAA++ RRTLLL LM +FAAGNI SA+AP F SFTLLRFV+GLPHGAYFGVAALVA
Sbjct: 83   IIAVLAARITRRTLLLGLMGLFAAGNILSALAPDFLSFTLLRFVTGLPHGAYFGVAALVA 142

Query: 415  ASMVPAHRRARAVGRVMLGLTVATLLGTPFTTFFGQSLDWQVAFFSVGVLGLLTVVLIWF 594
            ASM P H+RARAVGRVMLGLT+ATLLGTP  TFFGQ L W+ AF  VG +GLLTV L+W
Sbjct: 143  ASMAPIHKRARAVGRVMLGLTIATLLGTPLATFFGQLLSWRAAFMLVGGIGLLTVALLWL 202

Query: 595  YVPKDRVSAEAGFLRELGAFRRPQVWLTLGIAAVGYGGMFAMFSYIASTTTEVAMLPETA 774
            + P+D+V   A   RELGAFRR QVWLTL IAAVG+GGMF++FSYIA TTT+VAM+P +
Sbjct: 203  FQPRDKVEEGASVWRELGAFRRVQVWLTLAIAAVGFGGMFSVFSYIAKTTTDVAMMPVST 262

Query: 775  VPIMLVLFGVGMNAGNFIGSWLADKSLLGTIGGSLIYNIVVLTTFSLTAANPYLLGLSVF 954
            V ++L LFG+GMN GN +GS LAD SL GTIGG L +N++ +T F +TA NP++L + VF
Sbjct: 263  VSMVLALFGIGMNVGNVVGSRLADISLNGTIGGMLAFNVLAMTVFGMTADNPFMLCICVF 322

Query: 955  LVGCGFAAGPALQTRLMDVAADAQTLAAASNHSAFNIANAIGAWLGGLVIAGGYGFAATG 1134
            L+GCGFAA PA+QTRLMDVA DAQTLAAASNHSAFNIANA+GAWLGGLVIA G+G+A+TG
Sbjct: 323  LIGCGFAACPAVQTRLMDVAQDAQTLAAASNHSAFNIANALGAWLGGLVIAMGFGYASTG 382

Query: 1135 YVGAALSFLGLFVFAASLRLERRDRS 1212
            YVGA LS LGL VF  S+ +ERR ++
Sbjct: 383  YVGAVLSLLGLGVFLVSVTVERRAKA 408
```

This result showed a good sequence identity to a major facilitator superfamily (MFS) permease in *A. tumefaciens* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). The MFS class of permeases is the second largest family of membrane transporters found, after the ABC transporters (Pao *et al*., 1998). MFS permeases belong to the secondary, shock-insensitive transporters that are energized by the electrochemical gradient (Saier, 2000). Many MFS transporters have been shown to be involved in compatible solute uptake in cells that have undergone hyper-osmotic stress and ProP from *E. coli* is one of the best characterised (Culham *et al*., 1993). Uptake via ProP is mainly activated post-translationally, although transcription of ProP is induced two- to five-fold under hyper-osmosis (Csonka & Epstein, 1996). This MFS could therefore act in a similar way to ProP.

Regarding the acid stress induction, many bacteria use inducible systems to raise the internal pH of the bacterium, in order to counter any intruding acidic molecules. These systems employ ATPases and other transport mechanisms to either move acidic molecules

out of the cell, or take in basic compounds (Foster, 2000; Priefer *et al*, 2001). This process is only usually successful if the difference between internal and external pH is of approximately 1 pH unit (Foster, 2000), which was the case with the pRU857 data. This MFS could therefore act in as a pump to remove acidic molecules and relieve stress in the cell. However, whilst this response is possible the data strongly suggested hyper-osmosis as the main cause of induction.

These data indicated that RL1464 was induced as part of a stress response specific to hyper-osmotic stress.

4.2.4.12. pRU858/RU1522

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+	++	+	+	+	-	-

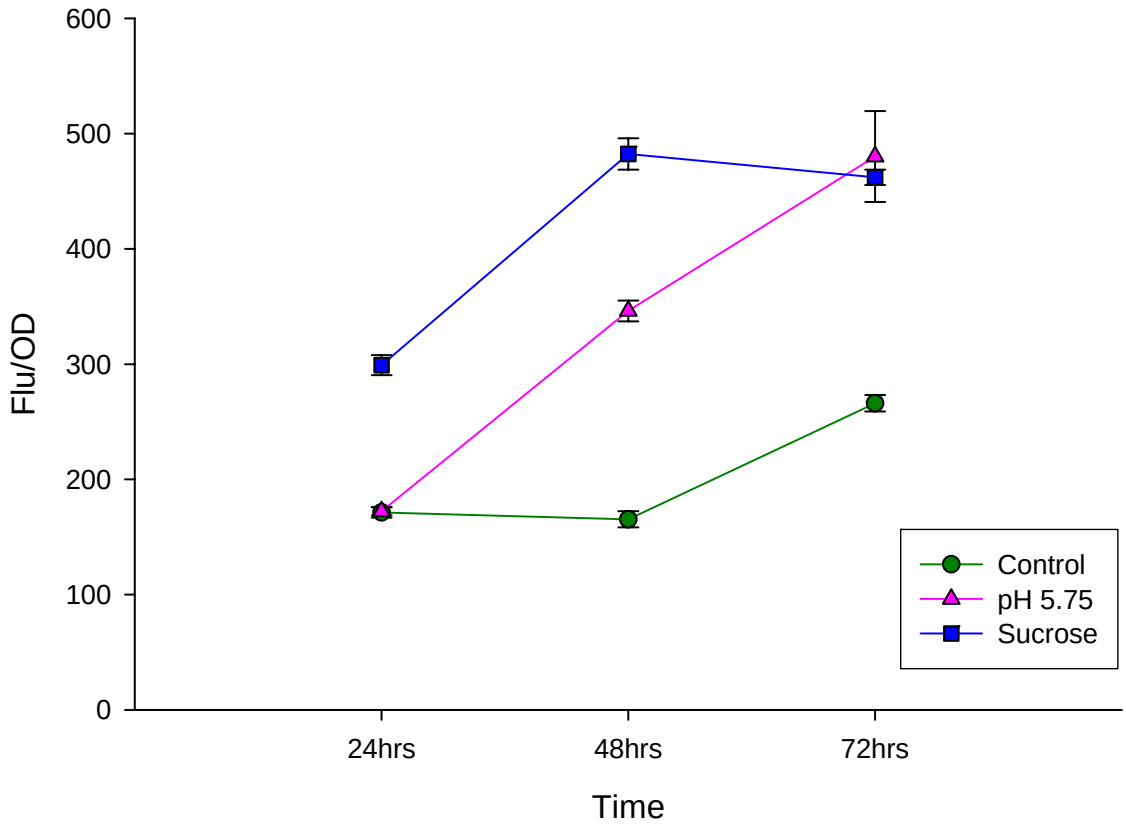


Figure 4.28. RU1522 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded equally to hyper-osmotic stress and acidic stress. Liquid growth showed significant induction after 24 hours under hyper-osmotic stress, but it took 48 hours under acidic stress before significant induction was observed (Fig. 4.28). *Rhizobium* DNA from this fusion mapped to the pRL12 plasmid, nucleotides 82068 – 83577 (Fig. 4.29).

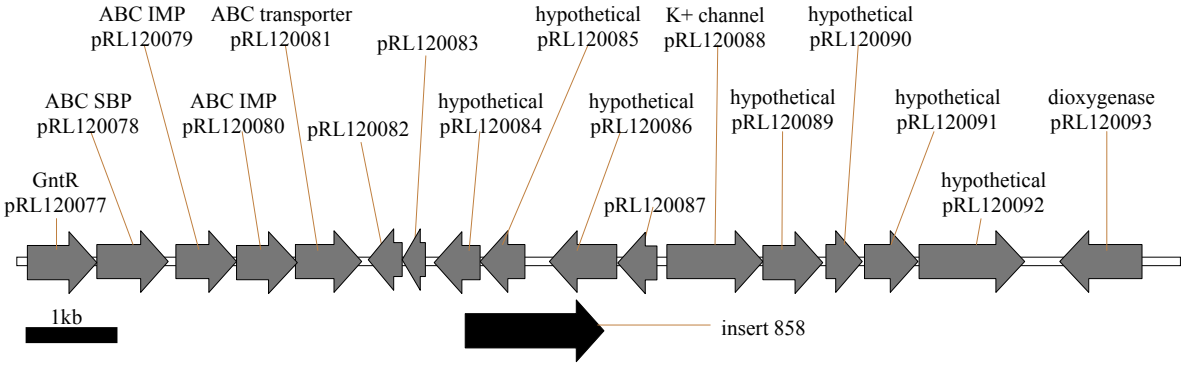


Figure 4.29. Genomic Region of pRU858's Insert. Black arrow shows the region of DNA within pRU858, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that all of the potential genes in and adjacent to the insert are transcribed in the opposite orientation to that of the insert. This made it highly improbable that a promoter for any of these genes was within pRU858. As annotation of the 3841 genome is preliminary, it could be that a gene is present in this region of DNA in the 'correct' orientation and has yet to be identified. However, BLAST analysis of this region did not reveal any such gene.

Given the dubious nature of the transcription of this fusion/gene and the unlikelihood of transcription of nearby genes, no gene could be associated with pRU858 and so it was not used in any further research.

4.2.4.13. pRU859/RU1506

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+	+	+	+	-	-	-

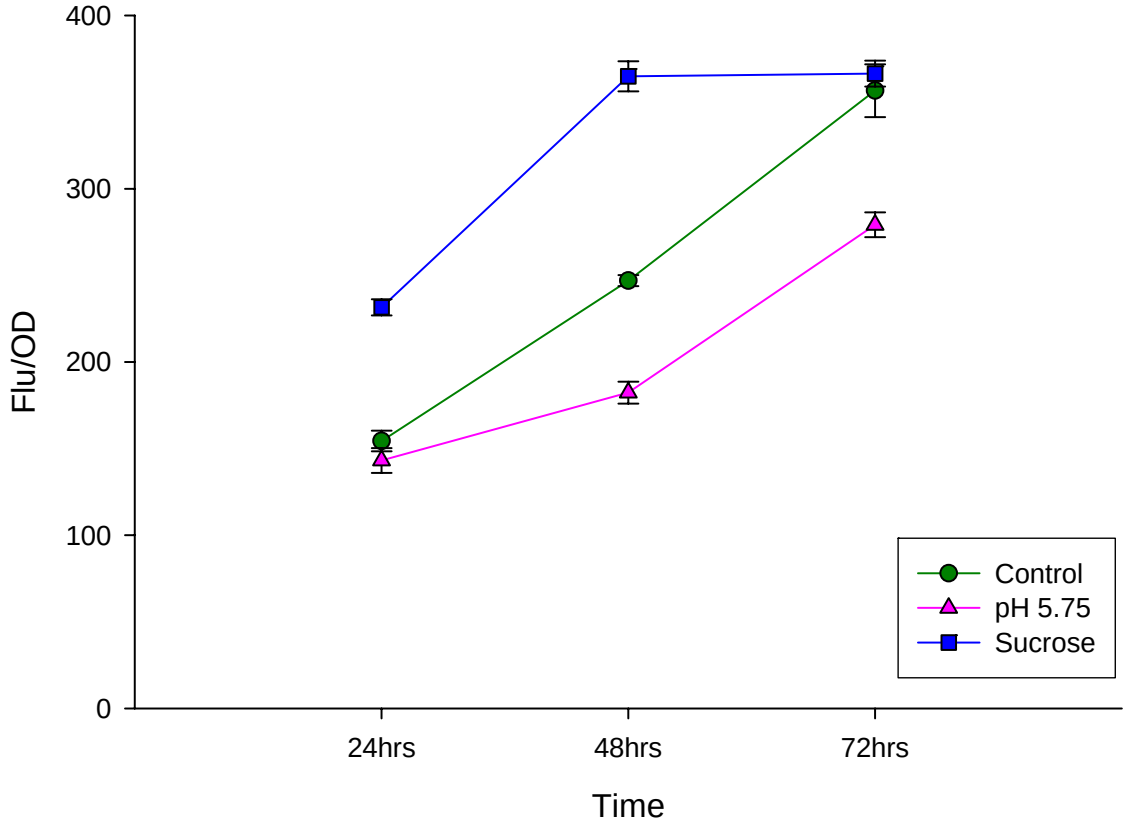


Figure 4.30. RU1506 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures of siblings RU1506, RU1511 and RU1515 (as in explained in section 4.2.2).

The screening results on solid media for this fusion showed that it responded to hyper-osmotic stress alone. Liquid growth showed that this fusion induced early in its growth cycle with significant induction observed after 24 hours, although it also appeared to induce on entry into stationary phase (Fig. 4.30). *Rhizobium* DNA from this fusion mapped to the DNA from pRL10 plasmid, compliment of nucleotides 148627 – 150043 (Fig. 4.31).

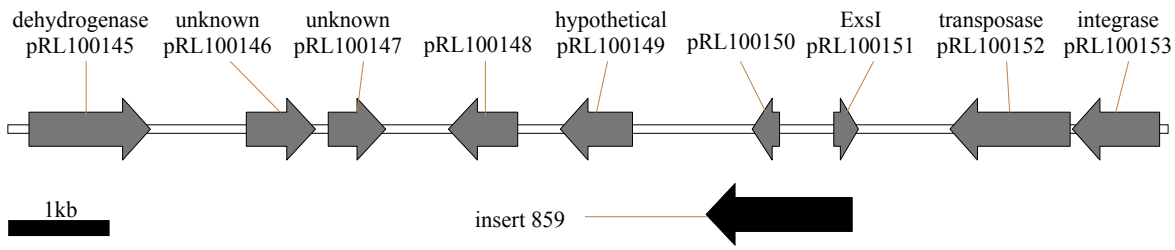


Figure 4.31. Genomic Region of pRU859's Insert. Black arrow shows the region of DNA within pRU859, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL100149 (compliment of pRL10 nucleotides 147202 to 147909, 236aa). The insert contains the DNA from pRL100151, but it is encoded in the opposite direction to the promoter. Although the insert contains the entire DNA from pRL100150, BLAST analysis did not reveal a homologue and so it was not analysed further. The promoter within the insert could activate the transcription of pRL100148, however, BLAST analysis did not reveal a homologue for that gene either and so it was not analysed further. Also, if the insert did contain the promoter for pRL100149, it would be quite far upstream from the start of the gene, which is unusual. As the annotation of the 3841 genome is only preliminary, maybe the transcriptional start has been predicted incorrectly and it is actually further upstream within the DNA mapped from the insert.

The best BLAST result for pRL100149 is shown below.

Conserved hypothetical protein Atu6162 (imported) [*Agrobacterium tumefaciens* C58]
Length = 236aa

Score = 407 bits (1046), Expect = e-112
Identities = 206/236 (87%), Positives = 216/236 (91%)
Frame = +1

```
Query: 1   MSLCSAKCIRIAALAATAIVLGTQIAIAQEPSIGSQWLNTPASRVEALAVLQTLNANLLS 180
           MSLC +KC RIAALAA +IVLGTQIAIAQEPSIGSQWLNTPASRVEALAVLQTLNANLLS
Sbjct: 1   MSLCPSKCFRIAALAAASIVLGTQIAIAQEPSIGSQWLNTPASRVEALAVLQTLNANLLS 60

Query: 181 NASATLTLDRWCAAHKLAPKGSKIVAQRVDRQGKPADEHIHELLTVGPGELIAYRRVRLV 360
           NASATLTLDRWCA HKLAP+GSKI+AQRV  Q KPAD+HI ELLTVGP E IAYRRVRLV
Sbjct: 61  NASATLTLDRWCAGHKLAPEGSKILAQRVRGQAKPADDHIRELLTVGPDEPIAYRRVRLV 120

Query: 361 CGDRVLSEADNWYVPAKLTAEMNRALNTSDIAFGRAVQALNFTRTNLSAKLLWSPLSEGW 540
           CGDRVLSEADNWYVPA+LTAEMN+ALNTSDIAFGRAVQALNFTRTNLSAKLLWSPLSEGW
Sbjct: 121 CGDRVLSEADNWYVPARLTAEMNQALNTSDIAFGRAVQALNFTRTNLSAKLLWSPLSEGW 180

Query: 541 DMDGLITHKTSSLSLPPFLLEHRAILKLQDGTPFSALVESYTDKVLDFPVPRLLSQ 708
           DMDGLI     S +LPPFLLEHRA+LKL DGTPFS LVESYT KVLDFPVPR L+Q
Sbjct: 181 DMDGLIASGRGSPTLPPFLLEHRAVLKLPDGTPFSTLVESYTSKVLDFPVPRSLAQ 236
```

This result showed an excellent sequence identity to a hypothetical protein in *A. tumefaciens* (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). As the result was a hypothetical protein, not much more could be inferred from that data alone. However, Pfam analysis was carried out and did not identify any domains within the gene. With no further insight into the make up of the protein that pRL100149 encoded for, it made it very difficult to assign any potential function to this gene.

These data indicated that pRL100149 was induced as part of a stress response to hyper-osmotic stress, although it also appeared to be induced on entry into stationary phase,

so the specificity of this response is unknown. However, as no function could be attributed to its predicted protein and as such pRU859 was not investigated further.

4.2.4.14. pRU861/RU1523

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	+	+++	+++	+/-	+/-

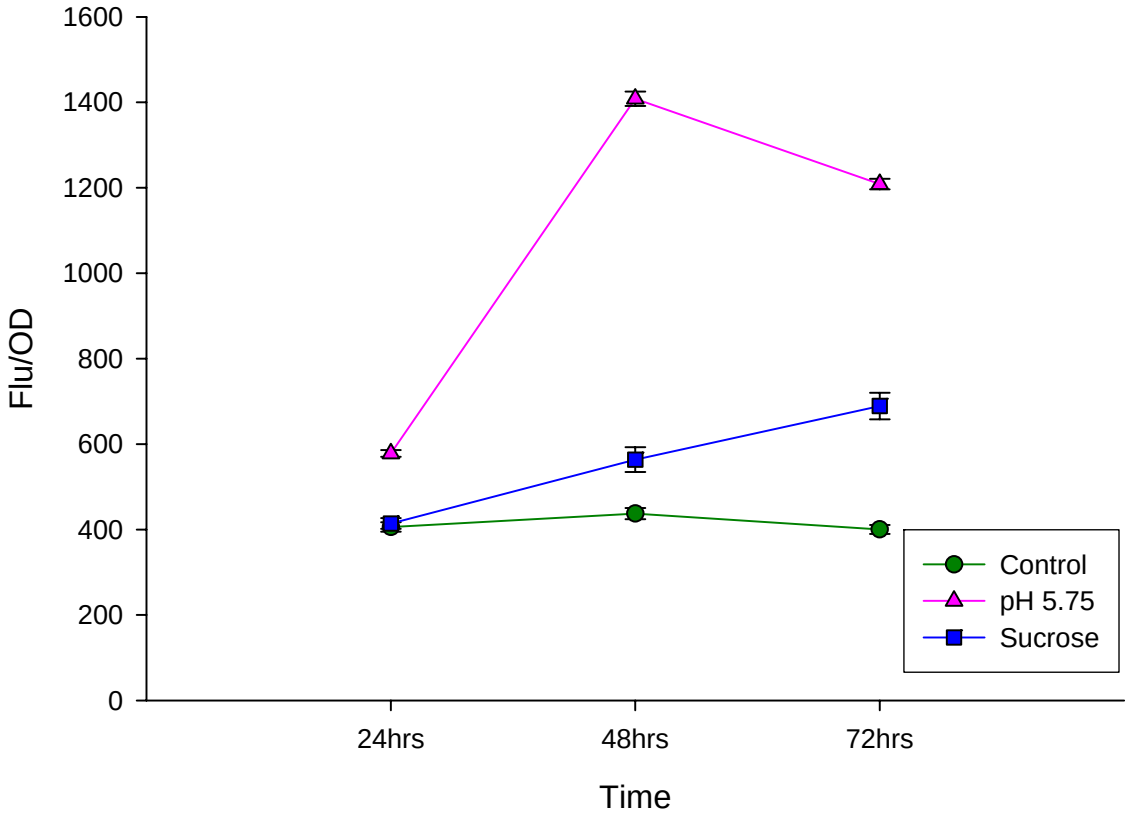


Figure 4.32. RU1523 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard errors produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded equally to hyper-osmotic and acidic stress, with fractional induction under oxidative stress. Liquid growth showed that this fusion induced early in its growth cycle under acidic stress, with a 3-fold induction after 24 hours, but only after 72 hours was significant induction observed for hyper-osmotic stress (Fig. 4.32). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 1234091 – 1239833 (Fig. 4.33).

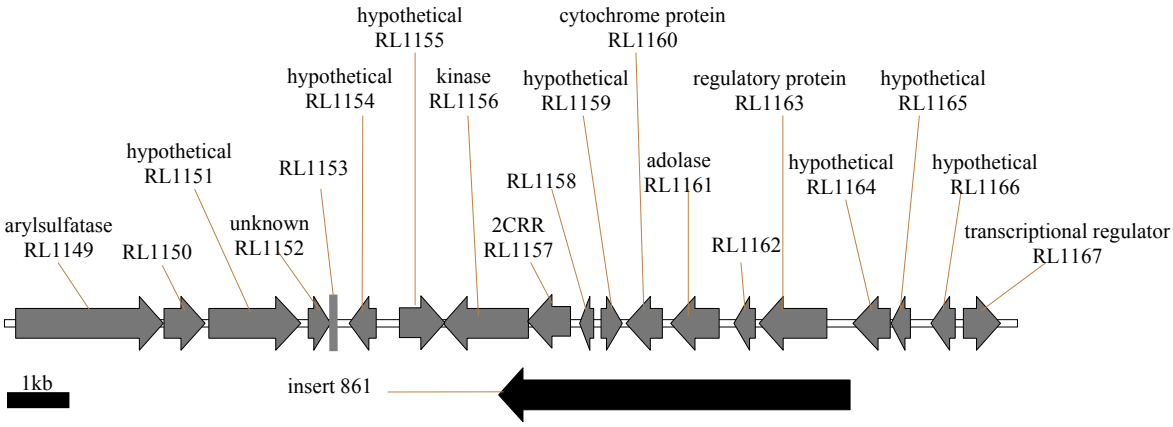


Figure 4.33. Genomic Region of pRU861's Insert. Black arrow shows the region of DNA within pRU861, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL1157 (compliment of chromosome nucleotides 1234570 to 1235256, 232aa). The insert contains the entire DNA for RL1163 to RL1158, but their transcription should end before that of the insert and so another promoter (one for RL1157) could be downstream of them. Also, RL1159 is transcribed in the opposite orientation to that of the insert. This means that it is unlikely that a promoter for RL1163 to RL1160 would read through and transcribe RL1159 and subsequent genes as transcription from that promoter may terminate before RL1159. BLAST analysis revealed no homologue for RL1158 and so it was not further analysed. RL1157 overlaps RL1156, and so RL1157's promoter probably also transcribes RL1156 as an operon and as such RL1156 was also investigated during analysis. The best BLAST result for RL1157 is shown below.

Further analysis showed that the insert for pRU862 mapped in close proximity to where pRU861 mapped (Fig 4.35).

Two-component response regulator [*Mesorhizobium loti*] Length = 227aa

Score = 358 bits (920), Expect = 5e-98
Identities = 173/224 (77%), Positives = 200/224 (89%)
Frame = +1

```
Query: 22  MRILLIEDDTKTSDYIAKGFSEAGHVCDVVGDGRDGLFQAQREAYDVIVVDRMLPGLDGL 201
           MR+LL+EDD KT+DYI +G +EAGHVCD++ +G D LF A   +YDVIV DRM+PGLDGL
Sbjct: 1   MRLLLVEDDQKTADYIVRGLTEAGHVCDLLRNGHDALFAATSGSYDVIVADRMIPGLDGL 60

Query: 202 AIVRSLRAAKVGTSALFLTSIGGVDDRVEGLEAGGDDYLVKPFAFSELMARVNALGRRPP 381
           ++V++ RAA V T A+FLTSIGG+DDRVEGLEAGGDDYLVKPFAFSEL+AR+NALGRRP
Sbjct: 61  SMVKAARAAGVRTPAIFLTSIGGIDDRVEGLEAGGDDYLVKPFAFSELLARINALGRRPA 120

Query: 382 VQEQRTVLKVADLELDLIRREARRAGQVIELQPREFTLLEVLMRGEGRVITKTMLLERVW 561
            QEQ+TVL+VADLE+DLI R   R GQ I+LQPREF+LLEVLMRGEGRVIT+TMLLERVW
Sbjct: 121 AQEQKTVLRVADLEMDLIMRRVTRQGQPIDLQPREFSLLEVLMRGEGRVITRTMLLERVW 180

Query: 562 DFHFDPKTSVVETHISRLRAKVDKPFQIQLLHTVRNTGYSLHAP 693
           DFHFDPKTSVVETHISRLRAKVDKPF+ QL+HT+RNTGYSLHAP
Sbjct: 181 DFHFDPKTSVVETHISRLRAKVDKPFEAQLIHTIRNTGYSLHAP 224
```

This result showed an excellent sequence identity to a two-component response regulator in *M. loti* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). Bacteria use two-component signal transduction systems to detect and respond to changes in the environment. This system consists of a sensor histidine kinase and a response regulator (Albright *et al.*, 1989). On detection of an external stimulus, such as environmental stress, the kinase component autophosphorylates in a histidine residue. The phosphate is then transferred to a highly conserved receiver domain of the response regulator (Forst *et al.*, 1989; Bearson *et al.*, 1998). Phosphorylation activates a variable effector domain of the response regulator, which triggers the cellular response (Albright *et al.*, 1989).

There are many examples of two-component response regulators being involved in stress response. The EnvZ/OmpR sensor/regulator pair in *E. coli* is probably the best characterised two component system involved with osmoregulation (Forst *et al.*, 1989). This pair is directly responsible for the production of the porins OmpC and OmpF, two pores involved in maintaining osmotic gradients in *E. coli*. The PhoP/Q regulator/sensor pair is transcribed under acid shock in *Salmonella typhimurium* and is responsible for the induction of several genes involved in acid tolerance response (Bearson *et al.*, 1998). As RL1156 encoded for protein with sequence identity to a sensory histidine protein kinase, it was likely that the products of RL1156 and RL1157 formed a two-component sensor/regulator pair similar to the examples outlined above. This predicted regulator could be responsible for the regulation of stress response in *R. leguminosarum* and as such made pRU861 a very usual fusion (see later chapters).

These data indicated that RL1157 and RL1156 was induced as an operon as part of a general stress response.

4.2.4.15. pRU862/RU1524

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+++	+++	-	+++	++	-	-

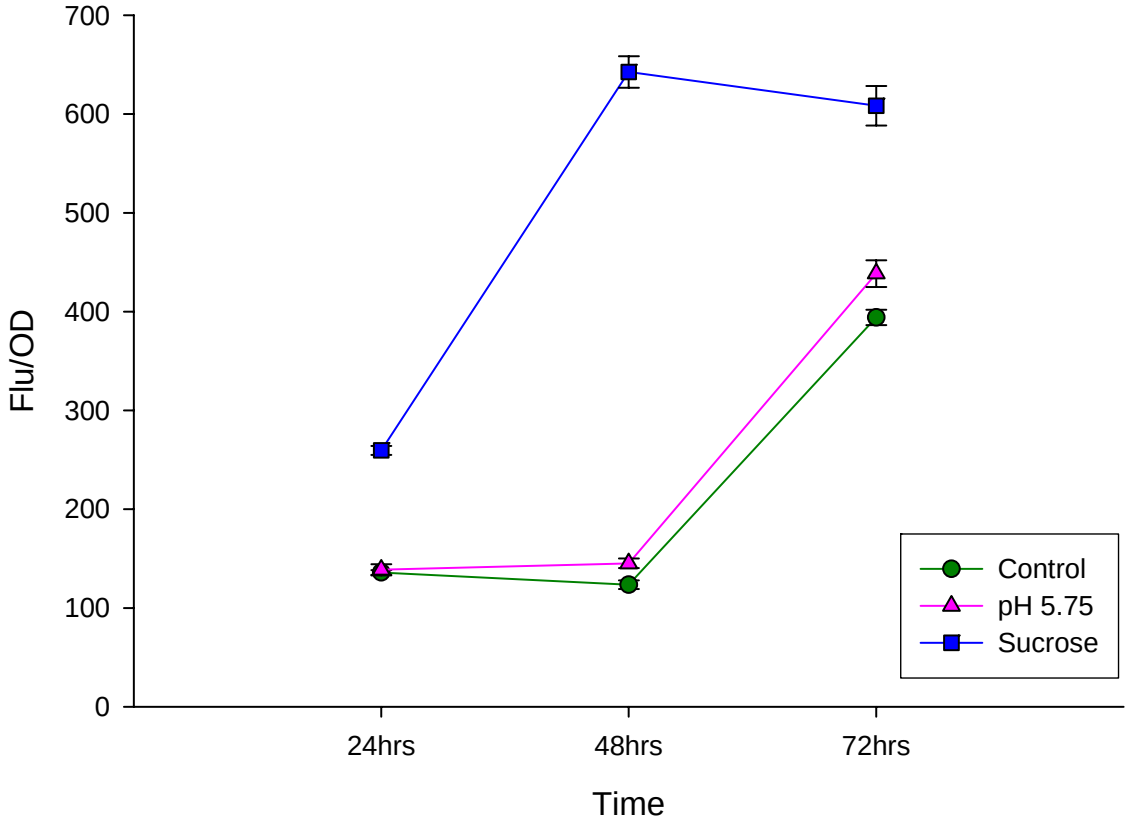


Figure 4.34. RU1524 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded equally to hyper-osmotic and acidic stress. Liquid growth showed a significant induction after 24 hours under hyper-osmotic stress and it also appeared to induce on entry into stationary phase (Fig. 4.34). *Rhizobium* DNA from this fusion mapped to the chromosome, nucleotides 1232278 – 1232671 (Fig. 4.35).

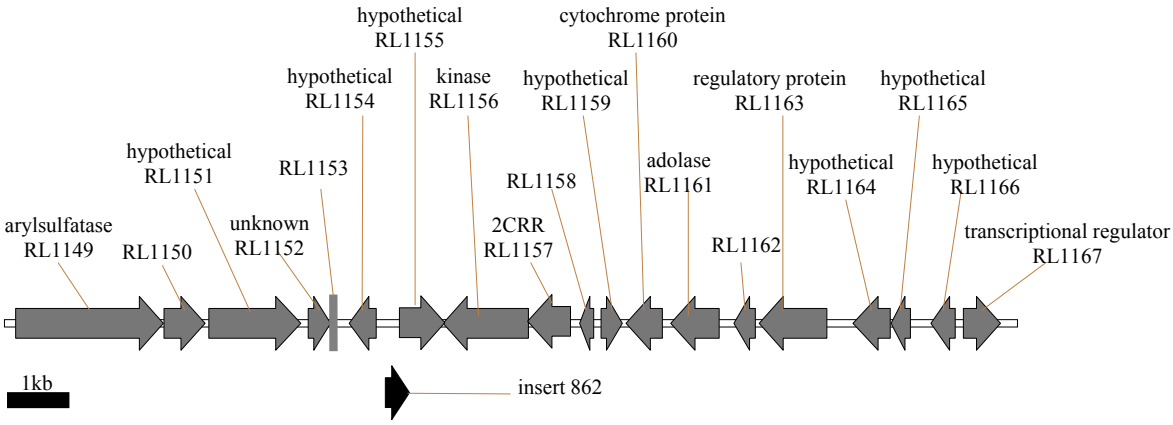


Figure 4.35. Genomic Region of pRU862's Insert. Black arrow shows the region of DNA within pRU862, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL1155 (chromosome nucleotides 1232474 to 1233214, 247aa). The best BLAST result for RL1155 is shown below.

Further analysis showed that the insert for pRU861 mapped in close proximity to where pRU862 mapped (Fig 4.33).

Conserved hypothetical protein Atu2058 (imported) [*Agrobacterium tumefaciens* C58]
Length = 300aa **Low Complexity Filter disabled**

Score = 230 bits (587), Expect = 2e-59
Identities = 124/233 (53%), Positives = 157/233 (67%)
Frame = +1

```
Query: 10  YLINLDRAPLRRFRMERLLASFGLAFERVAAVDGAGLSLPHPGFDDAAYLSRHGRRPNPF 189
           +LIN+D A  R   M   L + GL  ERVA V+G  L  P P F + +Y+  HGRR +P
Sbjct: 44  FLINMDSATKRLTDMNARLDAMGLKAERVAGVNGRELQYPIPEFSEISYMLMHGRRTSPP 103

Query: 190 EIGCYLSHVECAKRFLGSPAEFALILEDDLDFDDDLAELLDAALDHQARWDILRLSTVNS 369
           EIGCYLSHV CA +F+   A+ ALILEDD+ F+DD  + +D A+ +   WDILRL+TV++
Sbjct: 104 EIGCYLSHVACANKFMTGDADIALILEDDVVFEDDFLDAIDEAVLNGNDWDILRLTTVSN 163

Query: 370 GKKHKVEPLTASRSLAIALTREKGSGAYLINRKAAGWIAGVLVPMRLPYDLAFDLEFDDG 549
           G+K     L+  RSLA+ALTREKGSGAYL+NR+A  WI+  L+PMRL YD+AFDLE+  G
Sbjct: 164 GRKFAFRALSNGRSLAVALTREKGSGAYLVNRRAGKWIS-KLIPMRLAYDIAFDLEYLSG 222

Query: 550 LSACFVDPLPVSQRADPCSQIQAGLSAYRLGRRRPWSVLPYRAAAELRRFAAR 708
           L A F+ PL  +Q AD  SQIQ  L  YRL R R ++VLPYRA  E  RF  R
Sbjct: 223 LKAAFIYPLCATQDADGESQIQNNLRIYRLPRWRYFTVLPYRAYLETSRFLLR 275
```

This result showed a reasonable sequence identity to a hypothetical protein in *A. tumefaciens* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). As the result was a hypothetical protein, not much more could be inferred from that data alone.

Pfam analysis was carried out and showed the predicted protein is made up of a glycosyltransferase family 25 domain. This is a family of glycosyltransferases is involved in lipopolysaccharide (LPS) biosynthesis. These enzymes catalyse the transfer of various sugars onto the growing LPS chain during its biosynthesis (Jennings *et al***.**, 1995).

LPSs are major components of the outer membrane of Gram-negative bacteria, including *Rhizobium* species. Changes in the structure of LPS have been related to adaptation to different environmental situations in *R. leguminosarum*, such as a change pH and/or osmotic pressure (Kannenberg & Brewin, 1989; Bhat & Carlson, 1992; Tao *et al***.**, 1992; Zahran *et al***.**, 1994). Therefore, it was highly likely that RL1155 was involved in the LPS adaptation under environmental stress.

Data gathered from liquid growth of RU1524 showed an initial strong induction under hyper-osmosis (5-fold) but also induction on entering stationary phase. These data indicated that RL1155 was induced as part of a general stress response.

The extremely close proximity of RL1155 to RL1156 and RL1157 was very interesting, especially as data obtained for pRU861 and pRU862 were very similar. These data suggest that the putative two-component sensor/regulator could regulate RL1155. This made both pRU861 and pRU862 useful fusions and so the RL1155, RL1156 and RL1157 were investigated further (see later chapters).

4.2.4.16. pRU863/RU1525

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+++	+++	+	++	-	-	-

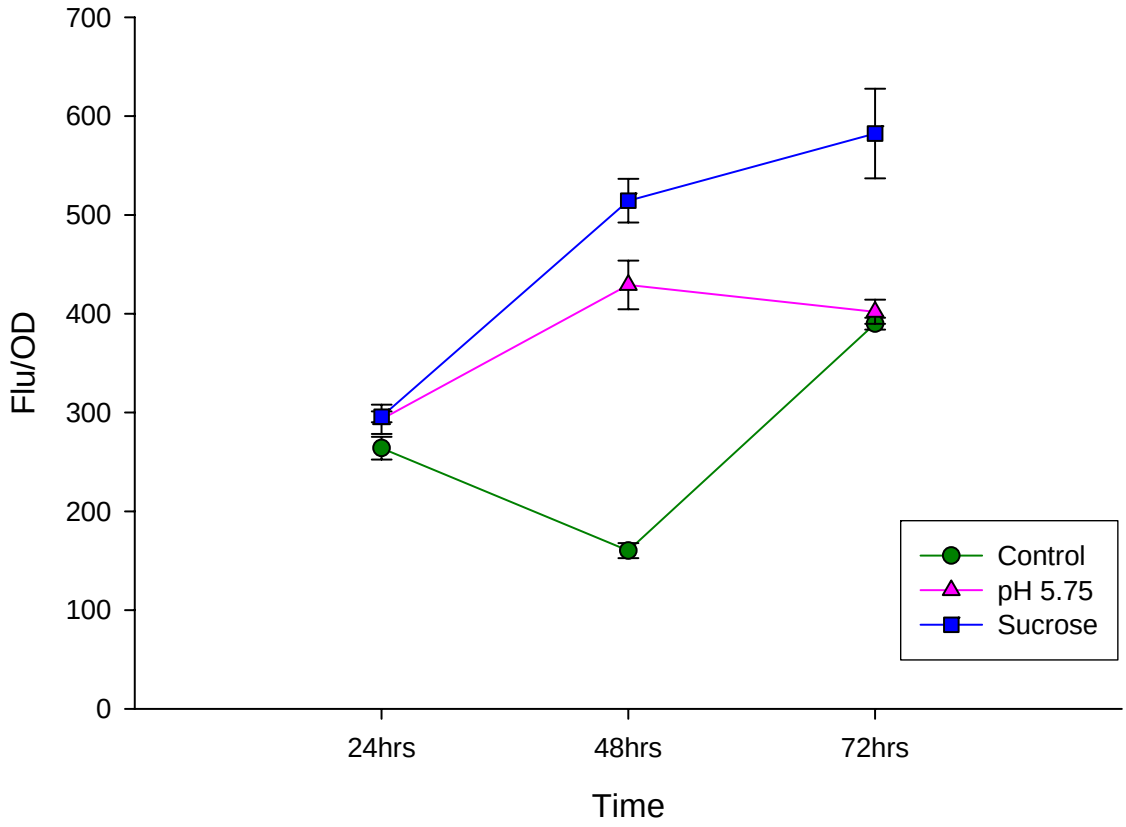


Figure 4.36. RU1525 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded specifically to hyper-osmotic stress. Liquid growth showed that this fusion induced later in its growth cycle as it took 48 hours before significant induction was observed (Fig. 4.36). *Rhizobium* DNA from this fusion mapped to the pRL12 plasmid, compliment of nucleotides 723528 – 725390 (Fig. 4.37).

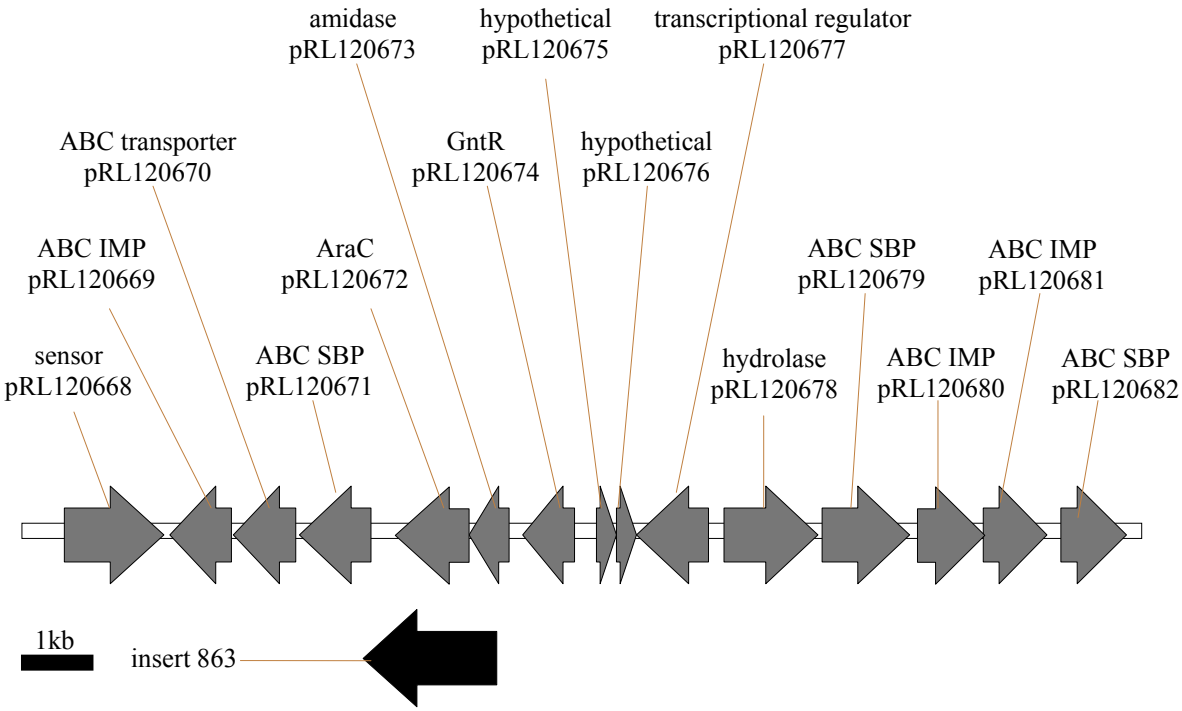


Figure 4.37. Genomic Region of pRU863's Insert. Black arrow shows the region of DNA within pRU863, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL120671 (compliment of pRL12 nucleotides 722616 to 723626, 337aa). The insert starts within pRL120673 and so it cannot contain its promoter. Also, the insert contains the entire DNA for pRL120672, but its transcription should end before that of the insert and so another promoter (one for pRL120671) could be downstream of it. The promoter within the insert could activate the transcription of pRL120670 and pRL120669, as well as pRL120671, as an operon, however, as there is an intergenic region (that may contain a promoter) between each of these genes, this may not be the case and so only pRL120671 was investigated. The best BLAST result for pRL120671 is shown below.

Further analysis showed that the insert for pRU863 mapped in close proximity to where pRU870 mapped (Fig 4.49).

ABC transporter, substrate-binding protein [*Mesorhizobium loti*] Length = 337aa **Low Complexity Filter disabled**

Score = 521 bits (1343), Expect = e-147
Identities = 253/337 (75%), Positives = 297/337 (88%)
Frame = +1

```
Query: 1    MRSSRSLFHTVAFSALLAAASFATSAAHAADKITIMVGGYEKQIYLPAKLAESLGYFKDE 180
            M  +R L  + A +AL+A  + +  + HAADK++IMVGGYEKQIYLPAKL E+LGYFKDE
Sbjct: 1    MSLARILLDSAATTALVATMALSAPSVHAADKVSIMVGGYEKQIYLPAKLTEALGYFKDE 60

Query: 181  GLDVELLNEAAGVDAENQLLAGAVQGVVGFYDHCVDLQAKGKFVESIVQFSQAPGEVEMV 360
            GLDVELLNE AGVDAEN++LAGAVQGVVGFYDHC+DLQAKGKFVES+VQFSQAPGEVE+V
Sbjct: 61   GLDVELLNEPAGVDAENEMLAGAVQGVVGFYDHCIDLQAKGKFVESVVQFSQAPGEVELV 120

Query: 361  SSKYPDIKSPADFKGKTLGVTGLGSSTNFLTLFMASKAGLKPGDVVTVPVGAGGTFIAAM 540
            S+K+P+IKSPADFKG +LGVTGLGSSTNFLT ++A K GLK G+  +VPVGAG TFIAAM
Sbjct: 121  STKHPEIKSPADFKGMSLGVTGLGSSTNFLTEYLAVKNGLKLGEFTSVPVGAGNTFIAAM 180

Query: 541  QQDQIQAGMTTEPTISRMIKTGEASVLVDMRTVESTRQALGGTYPAASLYMEASWVDAHK 720
            QQD+IQAGMTTEPTI+R++KTGEA VL+DMRT+E T+ ALGGTYPAASLYM+  WV+AHK
Sbjct: 181  QQDKIQAGMTTEPTITRLLKTGEAKVLIDMRTMEGTKAALGGTYPAASLYMQTDWVEAHK 240

Query: 721  EEAQKLANAFVKTLRYINTHSAAEIADKMPKDFYVGDKDGYIKALDDGKGMFTPDGVMPE 900
            +  QKLANAFVKT ++INTHS AEIADKMPKD+YVGDK+GY+KALD GK MFTPDG+MPE
Sbjct: 241  DIVQKLANAFVKTQKFINTHSGAEIADKMPKDYYVGDKEGYVKALDAGKAMFTPDGIMPE 300

Query: 901  DGPKTVLAVLSEFSKNVKGKQIDLSKTYTTEFVKNVK 1011
             GP+TVL VLS F K ++GKQIDL+KTYT+EFVKN K
Sbjct: 301  GGPETVLTVLSAFKKELQGKQIDLAKTYTSEFVKNAK 337
```

This result showed an excellent sequence identity to a substrate-binding component of an ABC transporter (mlr7949 – the NitT family) in *M. loti* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). Encouragingly, the product of pRL120670 shared sequence identity with an ABC domain and the product of pRL120669 shared sequence identity with an IMP domain. These data indicated that a region with sequence identity to an entire ABC operon is present and is probably transcribed as an operon. As mentioned in section 4.2.4.1, ABC transporters have been shown to be involved in the uptake of compatible solutes (see section 4.2.4.1 for review). This set of BLAST results is well suited to the data obtained, as the induction of pRU863 was specific for hyper-osmotic stress. The ABC transporter encoded by pRL120669, pRL120670 and pRL120671 could very well play a role identical to that theorised for the ProU-like system in section 4.2.4.1.

These data indicated that pRL120669, pRL120670 and pRL120671 were induced as an operon as part of a stress response specific to hyper-osmotic stress.

Interestingly, pRL120674 appears to encode for a protein with a sequence identity of a GntR regulator. This may be involved in the regulation of pRL120669, pRL120670 and

pRL120671 and the stress response in *R. leguminosarum*, as mentioned above (section 4.2.4.10).

4.2.4.17. pRU865/RU1527

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	+	+	+	+	-	+

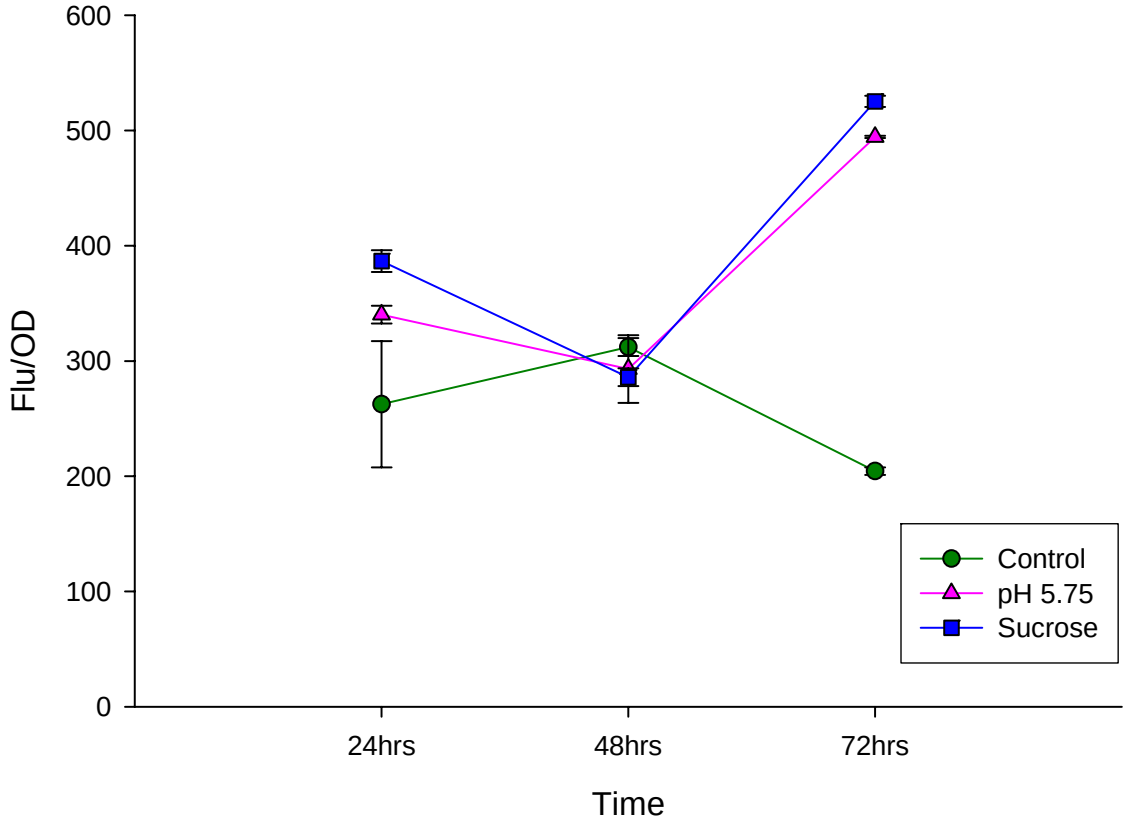


Figure 4.38. RU1527 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded equally to hyper-osmotic and acidic stress, with some induction under oxidative stress. Liquid growth showed that this fusion induced early in its growth cycle with a significant induction after 24 hours that led to a ~2.5-fold induction after 72 hours stress under both stresses (Fig. 4.38). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 3771546 – 3773151 (Fig. 4.39).

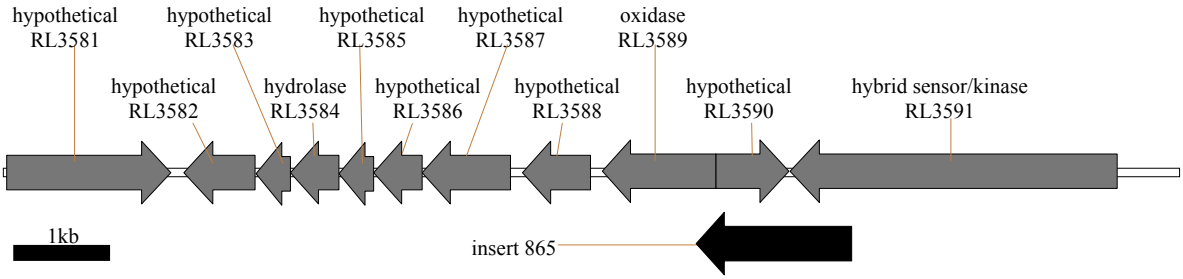


Figure 4.39. Genomic Region of pRU865's Insert. Black arrow shows the region of DNA within pRU865, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL3589 (compliment of chromosome nucleotides 3770573to 3771748, 392aa). The insert starts within RL3591 and so it cannot contain its promoter. Although most of the insert is made up of DNA from RL3590 it is encoded in the opposite orientation to the promoter. The promoter within the insert could activate the transcription of RL3589 to RL3582 inclusive as an operon, however, as there is an intergenic region (that may contain a promoter) between most of these genes, this may not be the case and so only RL3589 was investigated. The best BLAST result for RL3589 is shown below.

Probable D-amino acid oxidase (PA4548) (imported) [*Agrobacterium tumefaciens* C58]
Length = 410aa **Low Complexity Filter disabled**

Score = 466 bits (1199), Expect = e-130
Identities = 245/402 (60%), Positives = 297/402 (73%), Gaps = 12/402 (2%)
Frame = +1

```
Query: 1    MEKADSPAPVSGQS-------SCELLIVGGGIMGLWVAVHAERRGIRTFVADAGSLGGGA 159
            M KAD+PA   GQS       S  LLIVGGGIMGLW AV AER GI T + +AG LGGGA
Sbjct: 7    MAKADNPAFEQGQSPRMRLPDSVPLLIVGGGIMGLWAAVKAERLGIGTLLVEAGRLGGGA 66

Query: 160  SGGLLGALMPHMPDRWSEKKQFQFDALVSLEAEIAGLEAETGLSACYNRSGRLIPLPKPH 339
            SGGLLGALMPHMPDRWS+KKQFQFDALV+LEAEIAGLEAETGLS  Y R GR+IPLPKPH
Sbjct: 67   SGGLLGALMPHMPDRWSDKKQFQFDALVALEAEIAGLEAETGLSGGYRRCGRIIPLPKPH 126

Query: 340  LNKIALGHSKDAEHHWRSSDRRFHWHVLDRPPVDGWIEASAGESGFVHDTLAARVAPRAL 519
            L  IA  H +DA  +WRS +RRF WHV +RP V GW++ +AGE+GFV DTLAARV+PRAL
Sbjct: 127  LRGIAERHERDAAENWRSGERRFQWHVGERPSVAGWVDDAAGEAGFVFDTLAARVSPRAL 186

Query: 520  IAVLIAFLRRARHVRIMEHAGVTGLDPERGIAEV-GGETVAFGRCIIAAGHQSFPLLE-- 690
            IA+L AFLR+ARHV++ E   V  LD + G A +  GE ++FG  ++A GH+SFPL+
Sbjct: 187  IALLSAFLRKARHVQVAEGCRVVSLDADAGRAALSSGEEISFGHVVVANGHESFPLIRDA 246

Query: 691  -GLTPGLKQPLGQAVKGQAALLKADIDPALPTIFLDGLYVVAHEGGHAAIGSTSENRFED 867
             GL  G+   LGQAVKGQAALL A  DPA+P +FL+GLY+V HE G  AIGSTSE+ F +
Sbjct: 247  LGLEAGV--ALGQAVKGQAALLDASADPAMPVVFLNGLYIVPHEDGTVAIGSTSEDCFSE 304

Query: 868  PTSTDAQLDALIDAARAIVPALRSAPVVERWAGLRPKAIDRDPMVGCHPDHPRLIALTGG 1047
            P STD +L+ L+  A  +VP+L  APV+ERWAGLRPKA+ RDPMVG      +L+AL+GG
Sbjct: 305  PFSTDEKLEKLLVDACTVVPSLAGAPVLERWAGLRPKAVGRDPMVGAMAGTAKLVALSGG 364

Query: 1048 FKVSFGLAHRLAEAAI-CIAGDEPHEFSLPQSFAISSHIAVA 1170
            FKVSFGLAH LA+AA+  + G  P    +P  F +  H+++A
Sbjct: 365  FKVSFGLAHFLADAALETVCGHTP---VIPSGFRLQEHVSIA 403
```

This result showed a good sequence identity to a D-amino acid oxidase in *A. tumefaciens* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). An amino acid oxidase is an enzyme that catalyzes the oxidative deamination of an amino acid to a keto acid. This enzyme has been reported as being involved in biosynthesis and cell metabolism in *Rhizobium* (Miranda-Rios *et al*., 1997). Like some of the genes identified above, RL3589 could have played a role in generating a metabolisable source, or some form of energy, which would have been of great benefit to the cell and its survival (Nelson & Cox, 2000). In addition, as this gene probably encoded for an oxidase it may generate a redox reaction, which could restore altered electro-chemical gradients that may have also been caused from the stress. Maintaining cell viability by either producing a metabolite or restoring electro-chemical gradients is a well characterised general stress response, which usually occurs after bacteria have acclimatised to the initial shock of the stress and they begin a specific stress-response (Hengge-Aronis, 2000). This was reflected in the results obtained with pRU865, as not only was it induced by more then one stress, but also this induction increased over time and after the initial shock of stress.

These data indicated that RL3589 was induced as part of a general stress response.

Interestingly, RL3591 appears to encode for a protein with sequence similarity to a sensor/kinase hybrid which could regulate RL3589 and stress response in *R. leguminosarum* (see section 4.2.4.21).

4.2.4.18. pRU866/RU1528

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+	++	+	+	+++	+/-	++

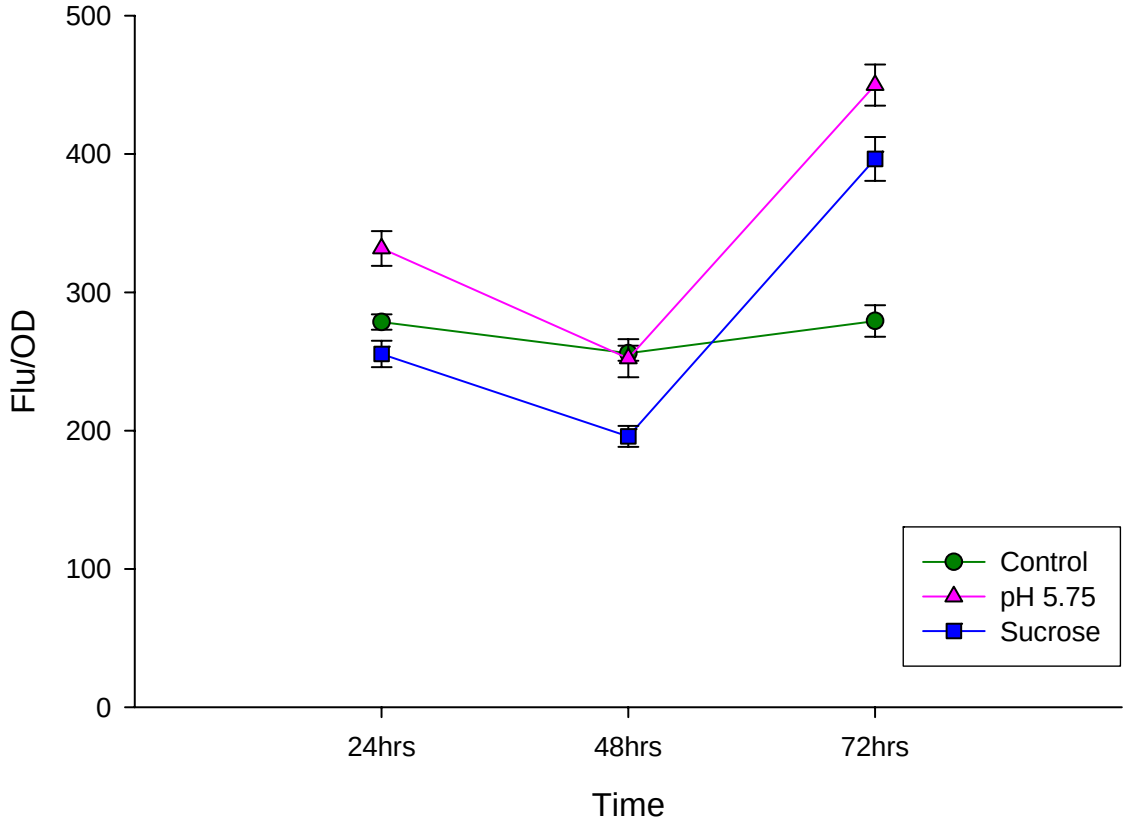


Figure 4.40. RU1528 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded mainly to acidic stress, with some induction under hyper-osmotic and oxidative stress. Liquid growth showed that this fusion induced later in its growth cycle as it took 72 hours before any significant induction was observed under stress (Fig. 4.40). *Rhizobium* DNA from this fusion mapped to the chromosome, compliment of nucleotides 4876419 -878126 (Fig. 4.41).

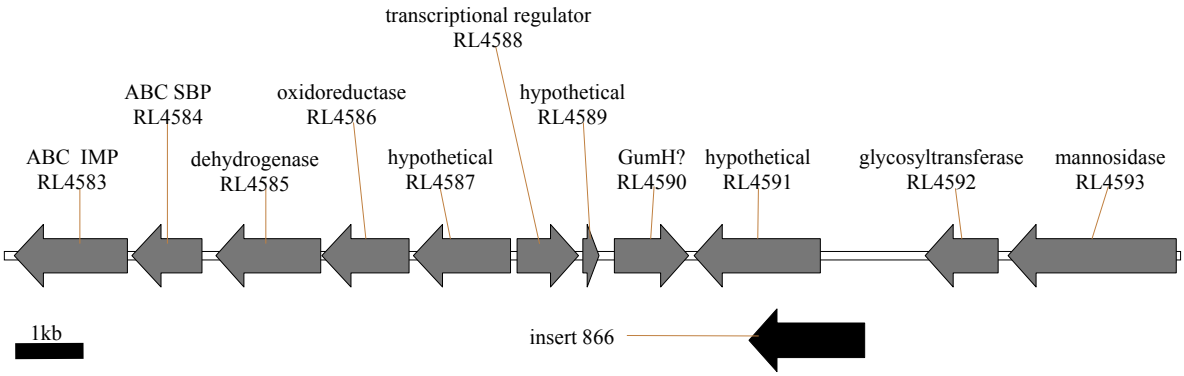


Figure 4.41. Genomic Region of pRU866's Insert. Black arrow shows the region of DNA within pRU866, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL4591 (compliment of chromosome nucleotides 4875611 to 4877470, 620aa). The best BLAST result for RL4591 is shown below.

Hypothetical protein Atu2576 (imported) [*Agrobacterium tumefaciens* C58] Length = 619aa **Low Complexity Filter disabled**

Score = 446 bits (1146), Expect = e-123
Identities = 259/582 (44%), Positives = 341/582 (58%), Gaps = 4/582 (0%)
Frame = +1

```
Query: 121  GAGRSAALPVVFDGTIGLFMPENPLAKKRSAAVLFVSPWGFEEMCSRKFFRVAAEHFSDI 300
            G   + A PV FDG  G+F P     ++ + AVLFVSPWG EE+CSRKF RV AE  +
Sbjct: 37   GLDGAIAHPVSFDGLAGIFHPARRDVRQ-AHAVLFVSPWGMEELCSRKFQRVLAERLAAR 95

Query: 301  GVPSLRFDYRGTGDALDFDALPARLETWEDSIRAATDKLKSLSGCDRIILIAQGLGATLA 480
            GV SLRFDY G GDA D +    R+  W    RAA   LK LSGC  ++++AQGLG  +A
Sbjct: 96   GVASLRFDYLGAGDAFDPEDA-GRVADWLSDTRAAFAYLKRLSGCAEVMVVAQGLGCLIA 154

Query: 481  HRVGSSIEGVDSLVMLAPVLSGRAYLRELNMWSKIIDADLGLGKEHIQTAKVQIAGLVMP 660
             +  +   G  S+  LAPV+SGRAYLREL MWS +ID  LGL           IAG+ MP
Sbjct: 155  AQALTDAVGAGSMAFLAPVVSGRAYLRELAMWSSMIDDGLGLRPGQRLAEAGAIAGMTMP 214

Query: 661  EEIAAELGKLNITSPQGLATSRYLILERPAKAEDTGFADALKALGADVEQKAFEGYDELA 840
            E +A  + K N+ +         L+L RP +  D  FA  L A+G +VE+ AF GYD+L
Sbjct: 215  EGVADAVKKTNLANLAAAPAHTILVLSRPGRVTDADFAKHLAAIGCEVEEAAFSGYDDLV 274

Query: 841  TNPLFAKTPMTVVALLTAWLETRTTETSAAHSSAAID-NPPLAGDGFAETPVRFGSHNHL 1017
            ++P  +K    VV  L  W+ ++T   S A++   I  N P  G GF E  V+FG    L
Sbjct: 275  SSPTLSKISGDVVNRLVDWVLSQTHAESPANTGEDIILNAPQRGRGFIEQSVQFGDGGRL 334

Query: 1018 VGVVSRPLGEIKGNAVLFLSTAYDRHAGWGRTTVDMARELARQGVVSLRFDSANVGDSPP 1197
             GV  RP      ++VL L  AYDRHAGWGR +V MAR LAR+GV SLRFD+AN+ DSPP
Sbjct: 335  FGVFCRPHDREAVSSVLLLGAAYDRHAGWGRLSVQMARTLAREGVASLRFDAANIADSPP 394

Query: 1198 RPDAPEQVLYSDTQTADAVAALDLLESVVAGPVMVAGRCSGGYVAFRAGVADERLKAVVS 1377
              +AP+QVLY   Q  D  AALD L +   GP + AGRCSG Y+AF   +AD+R+ AV++
Sbjct: 395  VKNAPDQVLYDAAQNDDVAAALDFLGTRGKGPFIAAGRCSGAYLAFNGALADDRIGAVIA 454

Query: 1378 INPFVYYWDPDMPVRREHVVSVPRSLDDYSQRLARLDTLKRLLRGQVDVVSALQNIVIAA 1557
            +NP V++W   + V  E +   PRS  +YSQR  +  T KRL+ G VDV SA  NI+ A
Sbjct: 455  VNPVVFHWRKGLSV-DEALHKRPRSFGEYSQRFRQGATFKRLISGDVDVASAGLNILKAT 513

Query: 1558 GRRLSPWIAPLLELLPDRRHIAREVRHSFALFGKRKVPLT---LIYSEGDVGLDHVYFHF 1728
             +RLS   A L       R  + E R  F  F + K   T   L+YS+ D GL+H  ++F
Sbjct: 514  MKRLSTKTARLF------RRGSEEGRAVFGAFDRLKAKNTAVHLLYSDNDDGLEHFQYYF 567

Query: 1729 GPHGAKLSRYPNVRLLMLPDADHNLTPPQSRKFVLDEIIRLA 1854
               G  L+ Y NV L ++PDADHNL+ P+++   ++ + RLA
Sbjct: 568  DADGDGLAAYRNVSLTIIPDADHNLSTPEAKTIYIETVKRLA 609
```

This result showed a reasonable sequence identity to a hypothetical protein in *A. tumefaciens* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). As the result was a hypothetical protein, not much more could be inferred from that data alone. However, Pfam analysis was carried out and did not identify any domains within the gene. With no further insight into the make up of the protein that RL4591 encoded for, it made it very difficult to assign any potential function to this gene.

These data indicated that RL4591 was induced as part of a general stress response, although no function could be attributed to its protein and as such pRU866 was not investigated further.

4.2.4.19. pRU867/RU1529

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	+++	++	+++	-	+

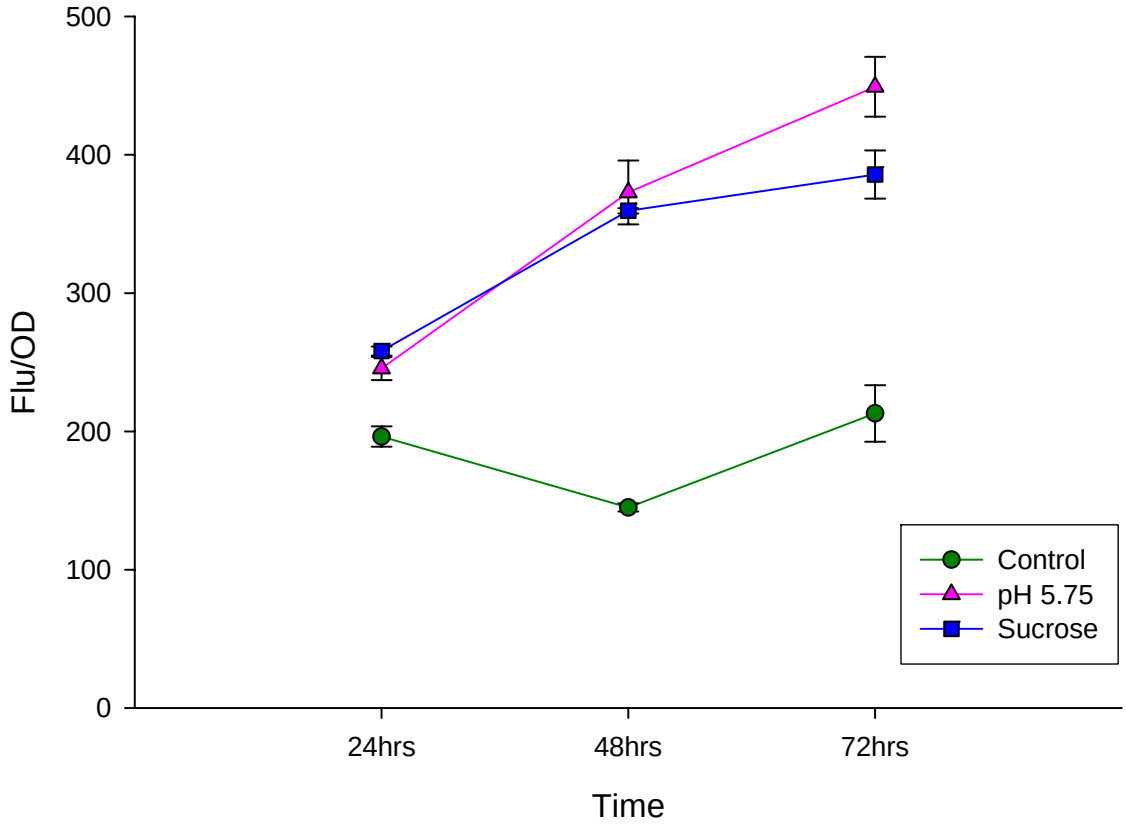


Figure 4.42. RU1529 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded equally to hyper-osmotic and acidic stress, with some induction under oxidative conditions. Liquid growth showed that this fusion induced later in its growth cycle, with a 2.5-fold induction after 48 hours for both stresses (Fig. 4.42). *Rhizobium* DNA from this fusion mapped to the DNA from the pRL10 plasmid, nucleotides 300050 – 304524 (Fig. 4.43).

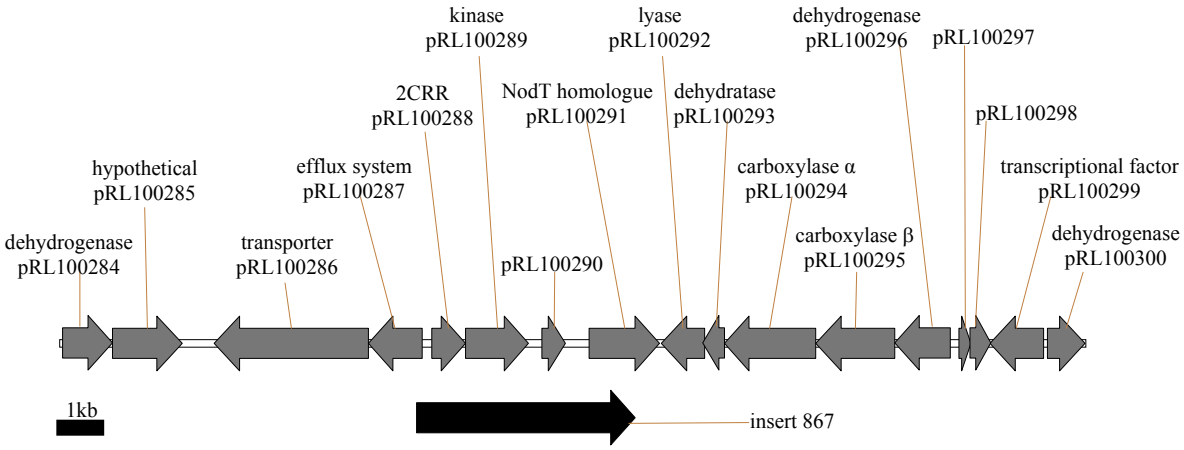


Figure 4.43. Genomic Region of pRU867's Insert. Black arrow shows the region of DNA within pRU867, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL100291 (pRL10 nucleotides 303571 to 305022, 484aa). The insert contains the entire DNA for pRL100288 to pRL100290, but their transcription should end before that of the insert and so another promoter (one for pRL100291) could be downstream of them. Also, as BLAST analysis for pRL100290 did not reveal a homologue it was not analysed further. The best BLAST result for pRL100291 is shown below.

Hypothetical protein 471 [*Rhizobium leguminosarum*] Length = 471aa **Low Complexity Filter disabled**

Score = 605 bits (1561), Expect = e-172
Identities = 316/468 (67%), Positives = 375/468 (80%)
Frame = +1

```
Query: 31   SLRYAASAITLLLAGCVSGPDHAPPQMPLPAKFQEGGSKSNGDVVAAQWWTAYRDKQLDG 210
            SLR+A  A+ LLL+GCV GPD  PP+MPLPAKF EGG+KS+GDV  A WWTA++D +L+G
Sbjct: 3    SLRFATPALLLLLSGCVVGPDLVPPEMPLPAKFGEGGTKSDGDVATAAWWTAFKDSRLNG 62

Query: 211  LVAHGLSENLDVLQALERINSASANVTVAGAGGLPSLDVGASHTVYGEKGSQRTTIGTKN 390
             V  GL +NL V QA+ERIN+ASANVT AGAGGLPSL VGAS TV G+K   RT + T+N
Sbjct: 63   YVKAGLDQNLTVQQAIERINAASANVTTAGAGGLPSLTVGASQTVSGQKAELRTQLDTRN 122

Query: 391  TTGGEASLSWLLDFFGQYRRSKESAIASLGAAYATADDAKLTFLKDLVSSYVDARYYQQR 570
            T+ G+ SLSWLLD FG Y+R+ ESA+ASL +AYA+AD A+LT ++DLVSSY+D R+YQQR
Sbjct: 123  TSAGDVSLSWLLDLFGLYKRNTESALASLDSAYASADVARLTLIQDLVSSYIDVRFYQQR 182

Query: 571  IALSQANLKSRQETYELTQLQLKAGAASRLDVVQAEGLVQSTKADIPGLEQSFTVSAHHI 750
            +A+S+ANLKSRQETYELT+ QL+AGA   LDVVQAEGLVQST A+IPGLE +  +SAHHI
Sbjct: 183  LAVSKANLKSRQETYELTKFQLEAGARP-LDVVQAEGLVQSTLAEIPGLETNIRISAHHI 241

Query: 751  ALLLGMPAASLMNELQRSTGQPVFRGDIRAGIPADLIRNRPDIRKAERDLAAAVADIGAA 930
            A LLG+PA+ L++EL + +GQPVFRG I +GIPADLIRNRPDIR  ERDLAAA A+IG A
Sbjct: 242  ATLLGLPASRLVDELLKGSGQPVFRGGITSGIPADLIRNRPDIRSRERDLAAATANIGVA 301

Query: 931  EAQLYPSISLSGSISPSWVKSSGASGGTMTSWSFGPTLNLPIFDGGKLRANVDIEKSDAK 1110
            +AQLYPSISLSGSISP  V   G  GG +  WSFGPTLNLPIFDGG+LRANV   +SDA
Sbjct: 302  QAQLYPSISLSGSISPLHVNQRGIHGG-LDRWSFGPTLNLPIFDGGRLRANVKSAQSDAA 360

Query: 1111 AQYLAWKEAVLNGVEEVENALSAVRRDTQTLAPLRRQVQTAEESLALSTTSYKDGASSLL 1290
              YL WK  VL  VE+VENALSAVRRD QT+A L+ QV+T  E+L LST SYKDGASSLL
Sbjct: 361  TAYLNWKSTVLTAVEQVENALSAVRRDAQTVAALQAQVKTTTETLELSTASYKDGASSLL 420

Query: 1291 DVLDAQRSVSDAQASLAATVQEVAKDYVDLYVAIGAGYLTEQGQNAPE 1434
            DVLDAQ+ VS AQASLAA+VQ++AKDYV L +AIG G+   Q   A E
Sbjct: 421  DVLDAQQ-VSLAQASLAASVQQMAKDYVSLNIAIGGGFAPAQNHRASE 467
```

This result showed a good sequence identity to a hypothetical protein in *R. leguminosarum* itself. In fact this hypothetical gene had previously been identified as homologous to NodT, a gene essential in nodulation in *R. leguminosarum* (Rivilla & Downie, 1994). However, as pRL100291 did not share 100% identity to the NodT homologue, it was believed this was not the same gene as characterised by Rivilla & Downie. (Genome searches revealed *nodT* to be pRL100178 and the *nodT* homologue to be RL3856.) This is therefore the third copy of a *nodT*-like gene.

It has previously been suggested that the NodT homologue plays a role in nodulation redundancy, as strains with a nod mutation could still formed nodules. Pfam analysis was carried out and showed the predicted protein is made up of two outer membrane efflux protein (OEP) domains. OEPs form trimeric channels that allow the export of a variety of substrates in Gram-negative bacteria. Examples of such channels are TolC in *E. coli* and obviously NodT in *R. leguminosarum* (Rivilla *et al*., 1995; Koronakis *et al*., 2000). These

channels have all been identified as having a role in the secretion of compounds from cells. As such the product of pRL100291, like some of the genes identified above, could play a role in the removal of harmful species acquired under environmental stress.

These data indicated that pRL100291 was induced as part of a general stress response.

Interestingly, pRL100288 and pRL100289 appear to encode for proteins that have sequence identity to a two-component sensor/regulator system that could regulate pRL100291 and stress response in *R. leguminosarum* (see section 4.2.4.14).

4.2.4.20. pRU868/RU1530

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
-	+	-	-	+	-	-

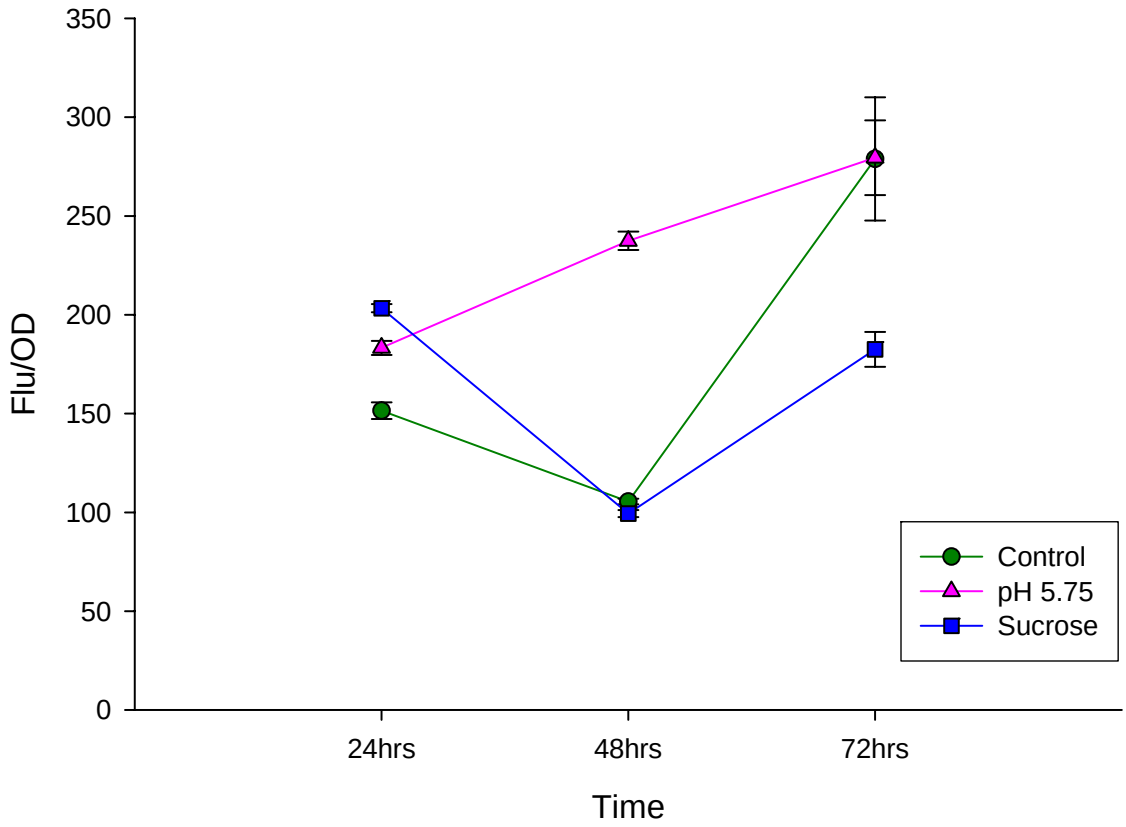


Figure 4.44. RU1530 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded weekly under hyper-osmotic and acidic stress. Liquid growth showed that this fusion induced with

a 2.3-fold induction after 48 hours for acidic stress, but had no significant induction under hyper-osmosis at any time (Fig. 4.44). *Rhizobium* DNA from this fusion mapped to the DNA from the chromosome, nucleotides 1388691 – 1393375 (Fig. 4.45).

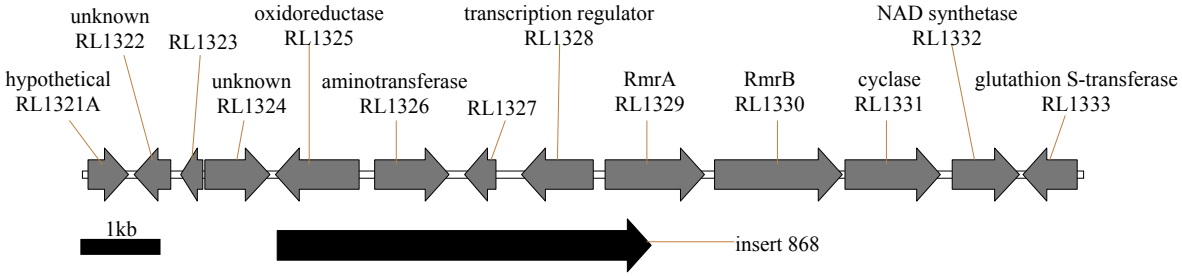


Figure 4.45. Genomic Region of pRU868's Insert. Black arrow shows the region of DNA within pRU868, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre except RL1322A, which had not been assigned a number at time of writing. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL1329 (chromosome nucleotides 1392794 to 1394038, 415aa). The insert contains the entire DNA for RL1325 to RL1328, but their transcription should end before that of the insert and so another promoter (one for RL1329) could be downstream of them. Also, RL1325, RL1327 and RL1328 are transcribed in the opposite orientation to that of the insert, although BLAST analysis for RL1327 did not reveal a homologue and so it was not analysed further. This also means that it is unlikely that a promoter for RL1326 would transcribe RL1329 as the subsequent genes are in opposite orientation and so transcription from that promoter should terminate before RL1329. The promoter within the insert could activate the transcription of RL1329 to RL1332 inclusive as an operon, however, as there is an intergenic region (that may contain a promoter) between all of these genes, this may not be the case and so only RL11329 was investigated. The best BLAST result for RL1329 is shown below.

As mentioned earlier (section 4.2.2) pRU868 shares some of its insert with the insert of pRU871. As such this homology between inserts is further reviewed with the pRU871 data in section 4.2.4.23.

Probable secretion protein [*Mesorhizobium loti*] Length = 417aa **Low Complexity Filter disabled**

Score = 445 bits (1145), Expect = e-123
Identities = 231/364 (63%), Positives = 284/364 (78%)
Frame = +1

```
Query: 91   PMPVSEAPVADAPVADAPKKTGRRIVKRAVIAAVLLAGVAFAGDFGYRYWTVGRFIESTD 270
            P+P+   PV    +   P  TGRR  KR + A V+LAG A A  +G+ YWT GR++ESTD
Sbjct: 27   PLPLL-TPVTSNEIVPVPP-TGRRNFKRVLAAVVVLAGTAAAAYYGHDYWTTGRYLESTD 84

Query: 271  DAYVKADYTTVAPKVAGYIKAVLVNDNDAVKAGQVLARIDDRDFQAALSQAKADVKAAEA 450
            DAYVKAD TT+APKV+GYI  VLV DN  V  GQVLARIDDRDF+AAL QA+AD++AAEA
Sbjct: 85   DAYVKADSTTIAPKVSGYIAEVLVRDNQKVTVGQVLARIDDRDFRAALDQAQADMRAAEA 144

Query: 451  AITNIDAQISLQQSVIEQAKATVDASQASFDFAVSDAARSARLITNGAGTQSRAEQTQSA 630
             + N+DAQI LQ+++IEQA+ATV A+QAS  FA  DA R A L  +G GT  +AE +++
Sbjct: 145  TVRNLDAQIVLQRALIEQARATVAATQASLRFAAVDADRYATLAKSGTGTTQKAEASRAG 204

Query: 631  RDQAAAAVERDRAALVTAQNKVPVLQTEREQTVAQRDRAAAAAQQAELNLSYTDIVAAVD 810
             DQ AA + RD+AA+V A+ ++ VL TER++ +AQ DRA AA +QA LNLSY  I A VD
Sbjct: 205  ADQLAAGLARDQAAVVAAEVRIDVLATERDKALAQVDRAQAAGEQARLNLSYATITAPVD 264

Query: 811  GTVGARSIRVGQYVTSGTQLMAVVPLHAVYVVANFKETQLTYISPGQSVEIKIDSFPDIS 990
            GTVGAR++R+GQYV +GTQLMAVVP +AVYVVANFKETQLTY+  GQ V + ID FP +
Sbjct: 265  GTVGARTLRIGQYVGAGTQLMAVVPQNAVYVVANFKETQLTYVRGGQPVRVAIDGFPGVE 324

Query: 991  IKGHVDSVSPASGLEFSLLPPDNATGNFTKIVQRIPVKIVIDDEALSGLLRSGMSVEPEI 1170
            ++GHVDS+SPASGLEF+LLPPDNATGNFTKIVQRIPVKI+I+D+ L GLLR+GMSVEP I
Sbjct: 325  LEGHVDSLSPASGLEFALLPPDNATGNFTKIVQRIPVKIMIEDQELGGLLRAGMSVEPTI 384

Query: 1171 DTKA 1182
            DTKA
Sbjct: 385  DTKA 388
```

This result showed a good sequence identity to a secretion protein in *M. loti* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). This secretion protein belongs to the HlyD family, a group of proteins that is only found in Gram-negative bacteria and have been reported to have many functions. A HlyD-like protein has previously been identified in *Rhizobium*, RmrA in *R. etli* (Gonzalez-Pasayo & Martinez-Romero, 2000). RmrA (*Rhizobium* multidrug resistance) is part of a multidrug efflux pump involved in the resistance to various antibiotics and toxic compounds. (Incidently, although it was not the best hit, BLAST analysis showed that the product of RL1329 had a 55% identity and 67% similarity with RmrA, further supporting this theory.) RmrA requires the translocase RmrB in order to cross the cell envelope and the two proteins are transcribed as part of an operon (Gonzalez-Pasayo & Martinez-Romero, 2000). This system shares homology to a well characterised multidrug efflux system in *E. coli*, EmrAB (*Escherichia* multidrug resistance) (Lomovskaya & Lewis, 1992). Interestingly, the product of RL1330 shared a good sequence identity to RmrB, which indicates that RL1329 and RL1330 could work in a similar manner to the RmrAB system in *R. etli*. The *R. etli* system was proposed

as an inducible export system that prevents the accumulation of toxic compounds within the bacterial cell (Gonzalez-Pasayo & Martinez-Romero, 2000). It is therefore highly likely that the products of RL1329 and RL1330 form a similar efflux system responsible for removal of toxic compounds acquired or formed in the cell under stress.

These data indicated that RL1329 and RL1330 were induced as an operon as part of a stress response specific to acidic stress. However the levels of induction are quite low and so this fusion may not be as reliable as others (see section 4.2.4.23).

Interestingly, RL1328 appears to encode for a protein that has sequence identity to a transcriptional regulator that could regulate RL1329 and RL1330 and stress response in *R. leguminosarum*.

4.2.4.21. pRU869/RU1531

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
-	-	+	-	+	++	+

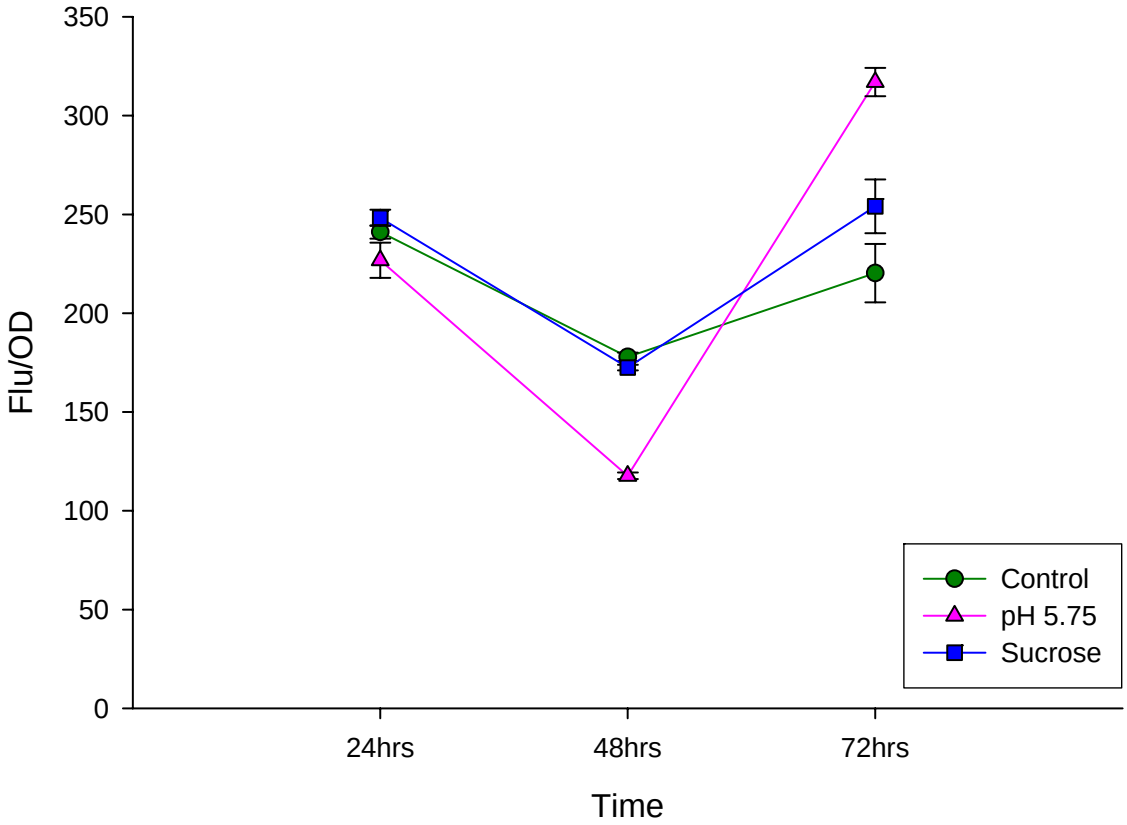


Figure 4.46. RU1531 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded to hyper-osmotic, acidic and oxidative stress, however, liquid growth showed no significant induction under stress at any time (Fig. 4.46). *Rhizobium* DNA from this fusion mapped to the DNA from the pRL10 plasmid, compliment of nucleotides 448931 – 451019 (Fig. 4.47).

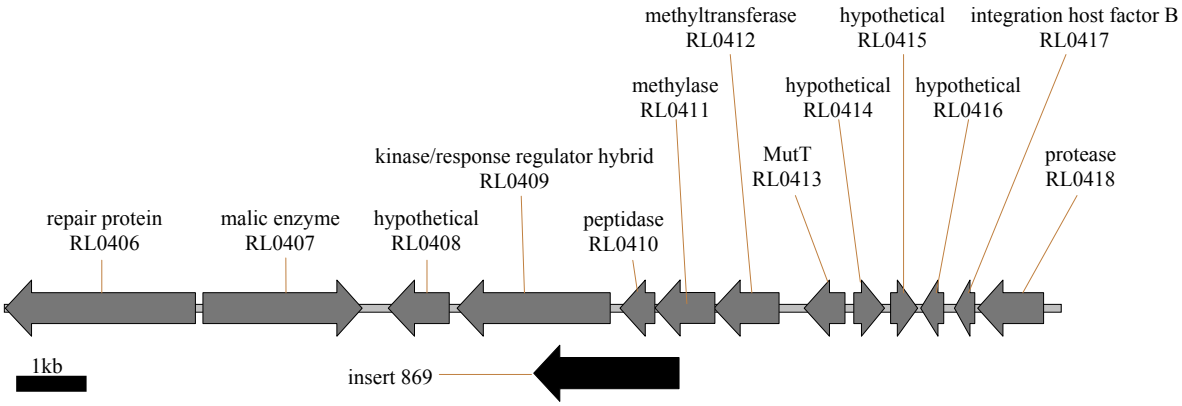


Figure 4.47. Genomic Region of pRU869's Insert. Black arrow shows the region of DNA within pRU869, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL0409 (compliment of nucleotides 447830 to 450031, 734aa). The insert starts within RL0411 and so it cannot contain its promoter. Although the insert contains the entire of DNA from RL0410, its transcription should end before the insert and so another promoter (one for RL0409) could be downstream of it. Even if this is not the case it would mean that RL0409 is transcribed by the same promoter as RL0410. In both scenarios RL0409 is transcribed from the promoter within the insert of pRU869. However, RL0412, RL0411 and RL0410 overlap each other, which indicates they may be transcribed as an operon. If this is so, the operon would have its promoter before RL0412, a region of DNA that is not present in the insert of pRU869. This suggests that it is more likely that the promoter within the insert of pRU869 was for RL0409 and not RL0410.

The promoter within the insert could also transcribe RL0408 with RL0409 as an operon, however, as there is an intergenic region (that may contain a promoter) before this gene, this may not be the case and so only RL0409 was investigated. The best BLAST result for RL0409 is shown below.

Putative sensor histidine kinase transmembrane protein [*Sinorhizobium meliloti*]
Length = 730aa **Low Complexity Filter disabled**

Score = 600 bits (1546), Expect = e-170
Identities = 332/664 (50%), Positives = 430/664 (64%), Gaps = 12/664 (1%)
Frame = +1

```
Query: 61   GGPAPTRRLLFYWLGGAGLLAATILMLLAHAGDPLLLSGGLVVLGLAVIASYALLMVRSR 240
            G P+     + Y L G   + A +L+L   AG  L L   ++  G  ++ ++ LL  R
Sbjct: 15   GDPSAGGARVAYALAG---IVAAMLLLAVGAGAGLHLLPAIIAAG-GLVGAFLLLSGRDE 70

Query: 241  RAGQR---PGQTMPD----GNGGAKLFADVHDVLGDITVSRTMDRRIISANDTFRRLTGR 399
             +G+R    G+T  D    G   A L A +HD +GD+ + R +D +I+ AN  F  L G
Sbjct: 71   ASGKRVAGAGETAQDPARHGIETAALLATIHDAMGDLAIVRDLDGKIMQANGAFHELCGC 130

Query: 400  LRPEGKTCEEIGLAFRPGPIPHCYDVEISTPEGQRIFLWRDVVTRDPANGRLLLQSVARD 579
                G TC E+GL F P   P  Y V I TP G R++ W DV+ RDPA GR +  S+ARD
Sbjct: 131  ADARGLTCAELGLRFEPKTGPDRYFVHIYTPSGTRLYDWHDVLVRDPARGRPMRHSIARD 190

Query: 580  VTDERLIAQGREEARQKAEYNSAAKSRLLATVSHEVRTPLSGILGMTHLIAETRLTQEQQ 759
            VT+E L A  REEAR++AE  S AKSRLLATVSHE+RTPLSGILGM+ L+AETRL++EQ+
Sbjct: 191  VTEEMLAASQREEARRRAEEASRAKSRLLATVSHEIRTPLSGILGMSRLLAETRLSEEQK 250

Query: 760  NYLASIRQSGHALTQLVEDLLDFSTIEVGRFALHPRSESLRKLLESVVEMLAHRAHEKGI 939
            NYLA ++QSGH L QLVEDL+DFS++ VGRF L P  E LR+ +E+VVEML+ RAHEK I
Sbjct: 251  NYLAGMQQSGHTLVQLVEDLIDFSSLAVGRFQLRPSQEDLRQTVENVVEMLSPRAHEKNI 310

Query: 940  EIGATVSSDVPENMSFDPARLRQVLFNVIGNAVKFTQVGGVFIRVSLDGDDLSITVTDSG 1119
            EIGATV+ +VPE M FD ARLRQVLFNV+GNAVKFT+ GGVF+ V ++   + I + DSG
Sbjct: 311  EIGATVAIEVPERMLFDAARLRQVLFNVVGNAVKFTEKGGVFVSVDIENGSVRIRIDDSG 370

Query: 1120 PGMTAEEQARVFGEFEQGGSVTDKSSGTGLGLAISARIMREFNGALTVASEKGRGSEFTI 1299
            PGM+A+E ARVF EFEQ G    ++ GTGLGLAIS RIM  F G+LT  S  G+GS F I
Sbjct: 371  PGMSADELARVFEEFEQAGDDAQRAKGTGLGLAISRRIMEAFGGSLTATSMSGKGSRFEI 430

Query: 1300 RFPV-DIGSERPDRRNTLLAGNSVVLLAPAGAARTAIAETITALGGLCHLVGDGETARAT 1476
            RFP+   G      R  +LAG  V+++AP G +  A+A TI  LGG CH       A
Sbjct: 431  RFPMAGAGLSGVPVRRGILAGAQVLVMAPEGPSSAALAATIETLGGTCHRASTLAVA-GR 489

Query: 1477 LLELAKGGRRP-TDIIIDHRMSAEFSAHLADRADIAALGLRKVLLVNPEERSAHP---LD 1644
            ++  A G R P TD+I+DHR +A+F   LA    IA L LR+  L++PEER++HP   L
Sbjct: 490  VVAGALGNRLPLTDVIVDHRHAAQFRELLALEPAIAGLRLRRTYLISPEERTSHPVSRLG 549

Query: 1645 LFDAWLIRPLREQSLIDVLRGRMRGMEKRDALNDNQPGFGLSVTETMVATRGLSILLGED 1824
             ++AWLIRPLRE+SL++VL GR+RGMEKRDA+NDN+P      T T++ +    ILL ED
Sbjct: 550  GYEAWLIRPLRERSLVEVLLGRLRGMEKRDAINDNRPVLREEPTATVMTSDVRGILLAED 609

Query: 1825 DPINAMLVRVVLEKGGHKVRHVEDFETLLDYALCEANDRPDIIISDLSMPGGNGIDMLGR 2004
            DP+NA+++R +L + G  V HV DF+ L        +  P +I++DL+MPGG+G+D+L R
Sbjct: 610  DPVNALVLRSLLSRAGRAVDHVGDFKALEAALRSAGSAPPPLIVTDLNMPGGDGLDVLRR 669
```

This result showed a reasonable sequence identity to a sensor histidine kinase transmembrane protein in *S. meliloti* (a member of the α-proteobacteria, the same group as ***R. leguminosarum***). RL0409 is similar to the genes of interest in section 4.2.4.14 (RL1156 and RL1157). The products of those genes shared sequence identities with a histidine kinase and a sensor, and together probably formed a two-component response regulator. RL0409 however, appears to encode for a protein that is a hybrid of these two proteins,

something that is not uncommon in rhizobia. *R. leguminosarum* bv. *viciae* VF39 contains a *fixL* gene that encodes for a sensor/regulator hybrid that combines the functions of FixL and FixJ and controls the induction of genes under low oxygen concentrations (Lopez *et al*, 2001).

As mentioned in section 4.2.4.14, bacteria use two-component signal transduction systems to detect and respond to changes in the environment. The product of RL0409 would have the function of both components in one protein and so act and respond in the same way.

The data indicated that RL0409 was induced as part of a general stress response.

4.2.4.22. pRU870/RU1532

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
++	++	-	+	+	-	+

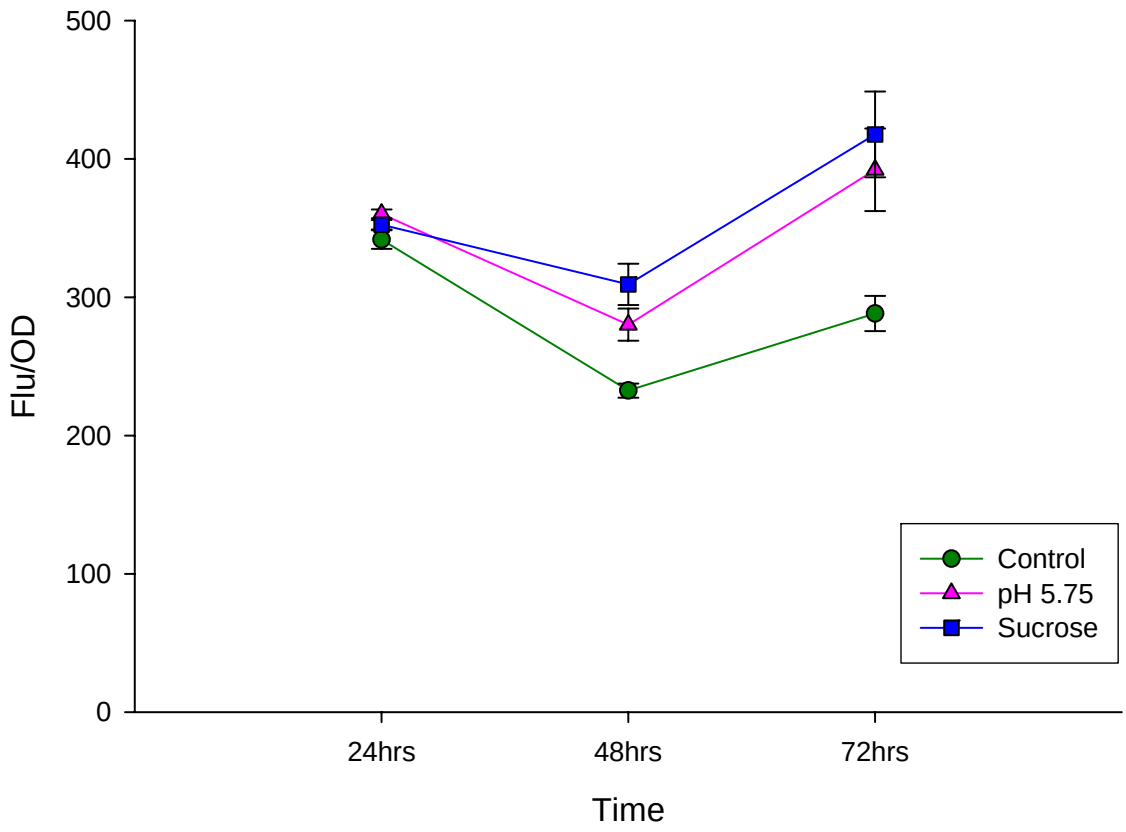


Figure 4.48. RU1532 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded to hyper-osmotic, acidic and oxidative stress, however, liquid growth showed no significant induction under stress at any time (Fig. 4.48). *Rhizobium* DNA from this fusion mapped to the DNA from the pRL12 plasmid, compliment of nucleotides 736574 – 738649 (Fig. 4.49).

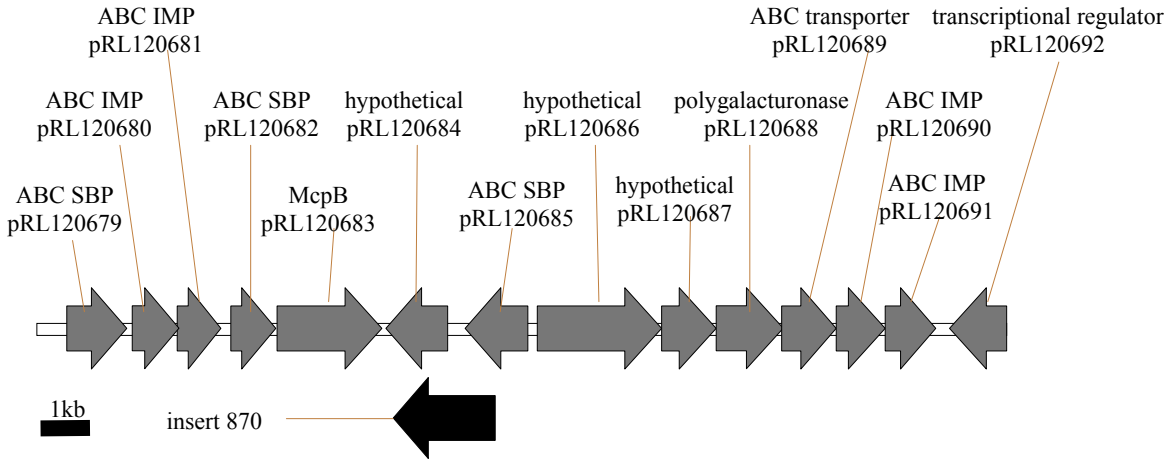


Figure 4.49. Genomic Region of pRU870's Insert. Black arrow shows the region of DNA within pRU870, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was pRL120684 (compliment of pRL12 nucleotides 736415 to 737674, 420aa). The insert starts within pRL120685 and so it cannot contain its promoter. The best BLAST result for pRL120684 is shown below.

Further analysis showed that the insert for pRU870 mapped in close proximity to where pRU863 mapped (Fig 4.37).

Hypothetical protein Rrub02002832 [*Rhodospirillum rubrum*] Length = 430aa **Low Complexity Filter disabled**

Score = 223 bits (568), Expect = 8e-57
Identities = 151/433 (34%), Positives = 233/433 (53%), Gaps = 25/433 (5%)
Frame = +1

```
Query: 1    MPMGSEHPLRRELHNELHARPSLYFDGDTDVWHVAIV-GENAPPQIPNS---LPGLEDVS 168
            +P   EHPLR  L NELHARPS   +    + H+A++ GE   P + +       + +
Sbjct: 2    IPPYREHPLREVLINELHARPSETIEAPVRLSHLAVLTGEITDPSLDHLGLLCARMGETV 61

Query: 169  TTREGNHGIGRIGDGRLKWEAHTEFLTLTFV-------VPASADPGSNPPEAFQACCRQI 327
              +        +G   L++E HTEF T TF        +P          P  + A    +
Sbjct: 62   PAKGATRFNANLGGLGLQFERHTEFCTYTFQRRGRTGDLPFDQPALDMVPPDWLAT---L 118

Query: 328  DGKVIAAVRVLVRDEKDGQRPEKPKFDY--------VASQVGGGDAEVHSNFRLTDSGFV 483
             G+V++AV ++V + KD       +           V S VGGG A  +S+ RL D   +
Sbjct: 119  PGQVLSAVHLVV-EPKDTPEYSIEELSIRHFAGNPVVGSAVGGGAAFAYSDLRLHDDRCL 177

Query: 484  EFLFFNRNLNAYRTGRMVRRFLEIETYRMMALLALPMARETVSKLSVFDRRLDLLIAHMQ 663
              L  + +LN    GR+V+R LE+ TYR +ALLALPMARE+   L    RR+D+++AH+
Sbjct: 178  RMLIRDVDLNPRHAGRLVQRLLELNTYRALALLALPMARESSPGL----RRIDMVLAHVA 233

Query: 664  S------AVKVDKALLSEVTKLSSDVLNFSALARHRFGATKAYAEIVASRSSELREERVE 825
            +       V  D  LLSE++ L+++V + +A   +R  AT+AY  +V  R  ELRE R++
Sbjct: 234  ARMADPNGVDSDAELLSELSNLTAEVESLAAANSYRIAATRAYHALVQRRLEELREVRLD 293

Query: 826  QRQRIGTFIDRRFQPAVRAVHAAERRLDELAERVSLAGDLLRTTVQVQLEDQNASLLTSM 1005
              G F+DRR  PA+  V +   R++ L+ER +    LLR  V+V L+ QN  LL SM
Sbjct: 294  GVVTFGAFMDRRLTPAMATVDSVSERIESLSERGARVASLLRARVEVDLQAQNKRLLESM 353

Query: 1006 EERARIQVHIQQAVEGFSVIAITYYTVGLAKICLESISELGVDPHVTKLAVLGAIPLVLF 1185
            + RARIQ+ +Q+AVEG SV+AI+YY VGL     + +S  GV+   + +  + A+P+++
Sbjct: 354  NRRARIQLRLQEAVEGLSVVAISYYLVGLVGYMAKGVSGAGVEVKESVVTAI-AVPVIVA 412

Query: 1186 AVWAAVRHVRKSI 1224
            +VW  +R  ++++
Sbjct: 413  SVWLVLRRAKRAM 425
```

This result showed a sequence identity to a hypothetical protein in *R. rubrum* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). As the result was a hypothetical protein, not much more could have been inferred from that data alone. However, Pfam analysis was carried out and did not identify any domains within the gene. With no further insight into the make up of the protein that pRL120684 encoded for, it made it very difficult to assign any potential function to this gene.

These data indicated that pRL120684 was induced as part of a general stress response, although no function could be attributed to its protein and as such pRU870 was not investigated further.

4.2.4.23. pRU871/RU1533

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
-	+	-	-	+++	+	-

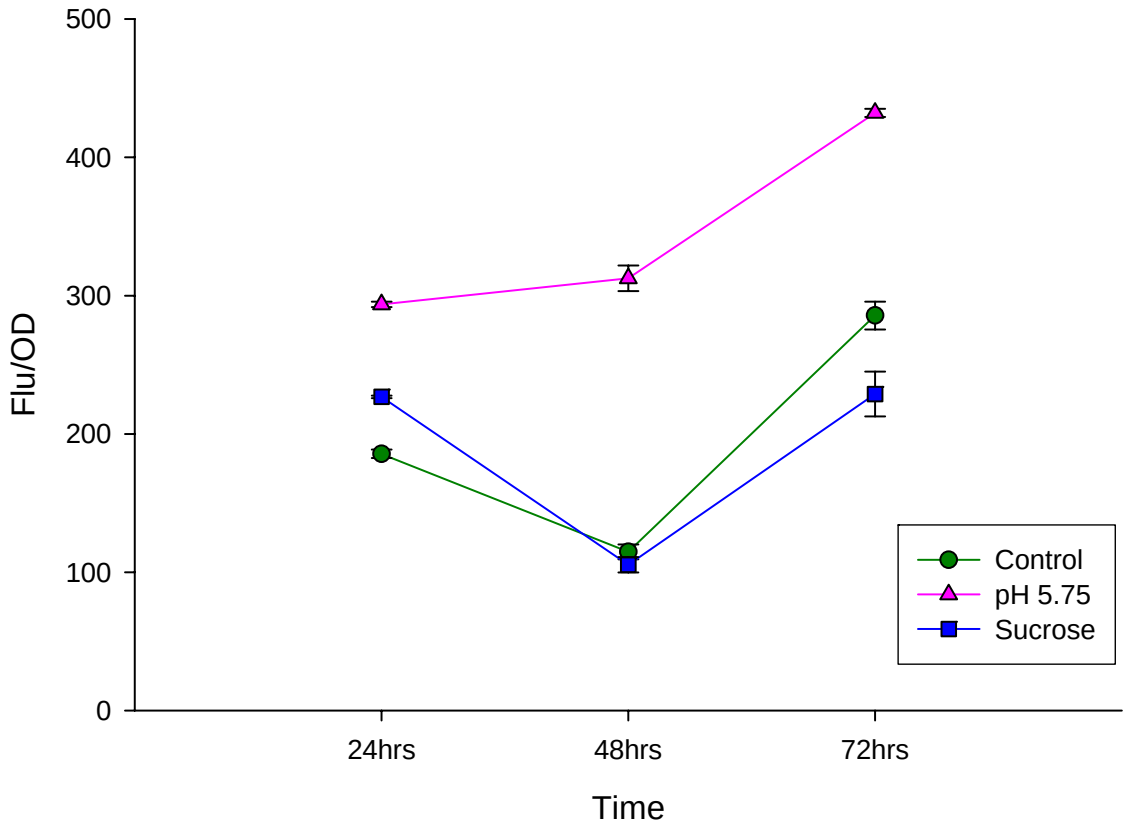


Figure 4.50. RU1533 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded mainly to acidic stress and also slightly under hyper-osmotic and oxidative stress. Liquid growth showed that this fusion induced significantly after 24 hours under acidic stress, which led to a 2.7-fold induction after 48 hours, but had no significant induction under hyper-osmosis at any time (Fig. 4.50). *Rhizobium* DNA from this fusion mapped to the DNA from the chromosome, nucleotides 1390705 – 1393293 (Fig. 4.51).

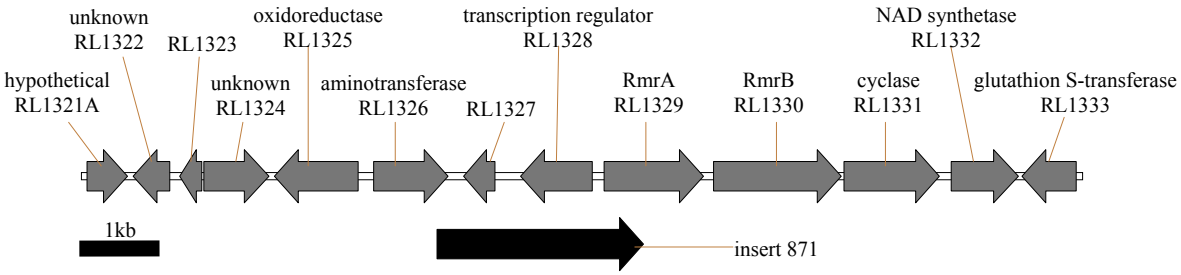


Figure 4.51. Genomic Region of pRU871's Insert. Black arrow shows the region of DNA within pRU871, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre except RL1322A, which had not been assigned a number at time of writing. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL1329 (chromosome nucleotides 1392794 to 1394038, 415aa). The insert starts within RL1326 and so it cannot contain its promoter. Also, the insert contains the entire DNA for RL1327 and RL1328, but they are transcribed in the opposite orientation to that of the insert. Furthermore, BLAST analysis for RL1327 did not reveal a homologue and so was not analysed further. The promoter within the insert could activate the transcription of RL1329 to RL1332 as an operon, however, as there is an intergenic region (that may contain a promoter) between all of these genes, this may not be the case and so only RL1329 was investigated. The best BLAST result for RL1329 is shown below.

As mentioned earlier (section 4.2.2) pRU871 shares of its insert with some of pRU868. As such this homology between inserts is further reviewed below.

Probable secretion protein [*Mesorhizobium loti*] Length = 417aa **Low Complexity Filter disabled**

Score = 445 bits (1145), Expect = e-123
Identities = 231/364 (63%), Positives = 284/364 (78%)
Frame = +1

```
Query: 91   PMPVSEAPVADAPVADAPKKTGRRIVKRAVIAAVLLAGVAFAGDFGYRYWTVGRFIESTD 270
            P+P+   PV    +   P  TGRR  KR + A V+LAG A A  +G+ YWT GR++ESTD
Sbjct: 27   PLPLL-TPVTSNEIVPVPP-TGRRNFKRVLAAVVVLAGTAAAAYYGHDYWTTGRYLESTD 84

Query: 271  DAYVKADYTTVAPKVAGYIKAVLVNDNDAVKAGQVLARIDDRDFQAALSQAKADVKAAEA 450
            DAYVKAD TT+APKV+GYI  VLV DN  V  GQVLARIDDRDF+AAL QA+AD++AAEA
Sbjct: 85   DAYVKADSTTIAPKVSGYIAEVLVRDNQKVTVGQVLARIDDRDFRAALDQAQADMRAAEA 144

Query: 451  AITNIDAQISLQQSVIEQAKATVDASQASFDFAVSDAARSARLITNGAGTQSRAEQTQSA 630
             + N+DAQI LQ+++IEQA+ATV A+QAS  FA  DA R A L  +G GT  +AE +++
Sbjct: 145  TVRNLDAQIVLQRALIEQARATVAATQASLRFAAVDADRYATLAKSGTGTTQKAEASRAG 204

Query: 631  RDQAAAAVERDRAALVTAQNKVPVLQTEREQTVAQRDRAAAAAQQAELNLSYTDIVAAVD 810
             DQ AA + RD+AA+V A+ ++ VL TER++ +AQ DRA AA +QA LNLSY  I A VD
Sbjct: 205  ADQLAAGLARDQAAVVAAEVRIDVLATERDKALAQVDRAQAAGEQARLNLSYATITAPVD 264

Query: 811  GTVGARSIRVGQYVTSGTQLMAVVPLHAVYVVANFKETQLTYISPGQSVEIKIDSFPDIS 990
            GTVGAR++R+GQYV +GTQLMAVVP +AVYVVANFKETQLTY+  GQ V + ID FP +
Sbjct: 265  GTVGARTLRIGQYVGAGTQLMAVVPQNAVYVVANFKETQLTYVRGGQPVRVAIDGFPGVE 324

Query: 991  IKGHVDSVSPASGLEFSLLPPDNATGNFTKIVQRIPVKIVIDDEALSGLLRSGMSVEPEI 1170
            ++GHVDS+SPASGLEF+LLPPDNATGNFTKIVQRIPVKI+I+D+ L GLLR+GMSVEP I
Sbjct: 325  LEGHVDSLSPASGLEFALLPPDNATGNFTKIVQRIPVKIMIEDQELGGLLRAGMSVEPTI 384

Query: 1171 DTKA 1182
            DTKA
Sbjct: 385  DTKA 388
```

This result showed a good sequence identity to a secretion protein in *M. loti* (a member of the α-proteobacteria, the same group as *R. leguminosarum*), the same result as for the gene of interest from insert of pRU868 in section 4.2.4.20.

The data implied that the inserts of both pRU868 and pRU871 were induced under stress because they contain the promoter for RL1329. If this was true, it would have been expected that both fusions would have had identical GFP profiles for their screening results. Although results obtained in liquid cultures are very similar for both strains, the data collected from growth of AMA (10mM glc, 10mM NH_4) at pH 5.75 is quite different. As for all strains, this screening was repeated several times and consistent results were obtained. This difference was undoubtedly due to the difference in insert size (section 4.2.2). The insert from pRU868 was approximately twice the size of that from pRU871 and because of this, DNA from pRU868's insert mapped further upstream of RL1329 than that from pRU871. It may be that a promoter for another gene (i.e. not RL1329) was present in the pRU868 insert and that this promoter may have in some way disrupted the GFP expression of this plasmid, when compared against pRU871.

This data confirms the likely induction of RL1329 and RL1330 under acidic stress.

4.2.4.24. pRU872/RU1534

0.1M Sucrose	0.01M Sucrose	0.1M NaCl	0.1M Mannitol	pH 5.75	1mM H_2O_2	0.25mM Paraquat
+++	+++	+++	++	+++	+	+/-

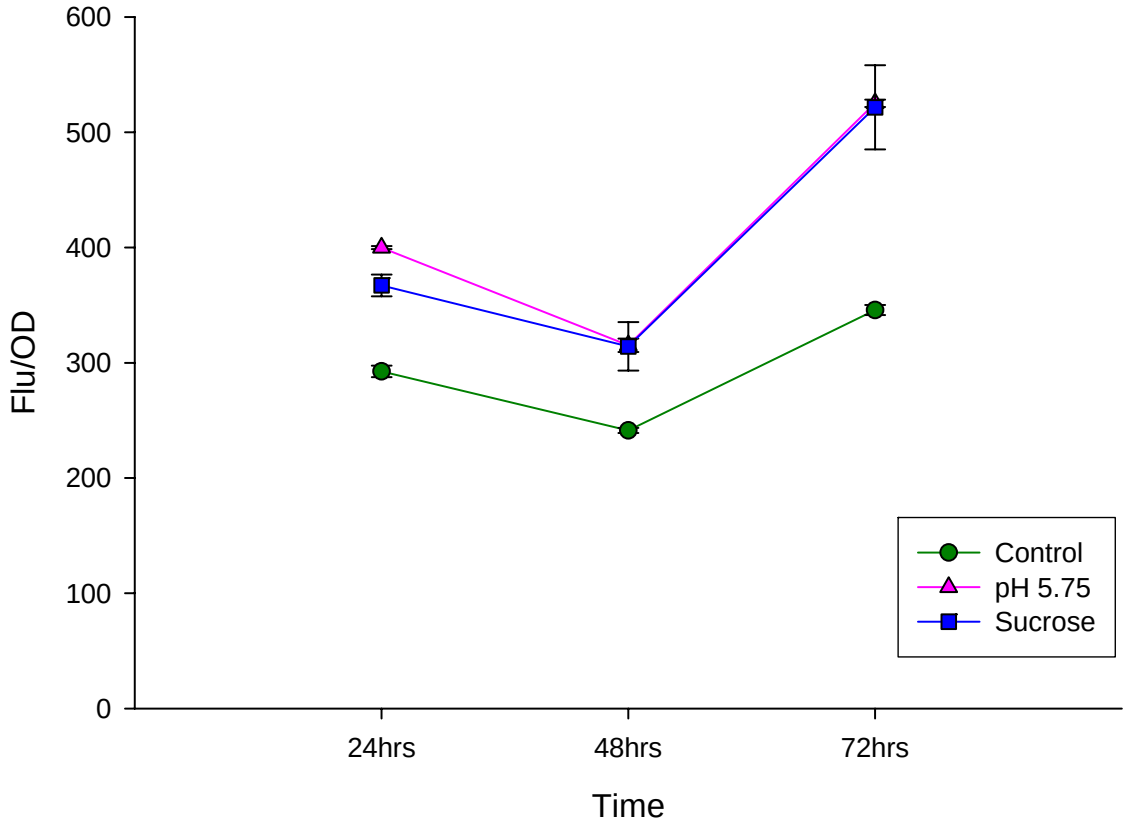


Figure 4.52. RU1534 GFP Values. Readings were taken daily from strains grown over 3 days. Values given are the units of relative fluorescence measured per unit of OD_{600}. Each line is the average and standard error produced from three independent cultures.

The screening results on solid media for this fusion showed that it responded mainly to hyper-osmotic and acidic stress, and slightly under oxidative stress. Liquid growth showed that this fusion induced later in its growth cycle as it took 72 hours stress before any significant induction was observed (Fig. 4.52). *Rhizobium* DNA from this fusion mapped to the DNA from the chromosome, nucleotides 4241206 – 4243351 (Fig. 4.53).

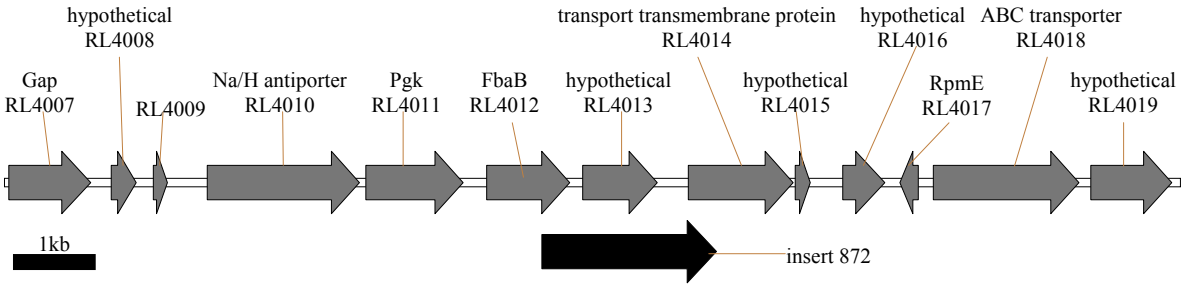


Figure 4.53. Genomic Region of pRU872's Insert. Black arrow shows the region of DNA within pRU872, grey arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown.

Analysis of this DNA indicated that the gene most likely activated by stress was RL4014 (chromosome nucleotides 4243010 to 4244305, 432aa). The insert starts within RL4012 and so it can not contain its promoter. Also, the insert contains the entire DNA for RL4013, but its transcription should end before that of the insert and so another promoter (one for RL4014) could be downstream of it. The promoter within the insert could also transcribe RL4015 and RL4016 with RL4014 as an operon, however, as there is an intergenic region (that may contain a promoter) between all of these genes, this may not be the case and so only RL4014 was investigated. The best BLAST result for RL4014 is shown below.

Putative transport transmembrane-protein [*Sinorhizobium meliloti*] Length = 433aa **Low Complexity Filter disabled**

Score = 592 bits (1525), Expect = e-168
Identities = 294/427 (68%), Positives = 351/427 (82%)
Frame = +1

```
Query: 1    MKRNLLSVAALLFGTLFLFMGNGLQGILLPVRGNLEGYATTTLGLLGTSWAGGFVIGCLI 180
            M+ NLL VAALL GTLFLF+GNGLQG+LLPVRG  EGY TT LGL GT WA GFV+GC
Sbjct: 1    MRNNLLPVAALLLGTLFLFLGNGLQGLLLPVRGTAEGYPTTILGLFGTLWATGFVLGCFF 60

Query: 181  APKIVRRVGHVRAFSGFISIIAIIALVSGIIIDPVWWVVLRAVTGFSTAGTSMIIESWLN 360
            AP +V+R+GHVRAFS F ++IAI++L++GI+IDP+WW+ LRAVTGFSTAGTSMIIESWLN
Sbjct: 61   APNVVKRIGHVRAFSVFTALIAIVSLLTGILIDPIWWLALRAVTGFSTAGTSMIIESWLN 120

Query: 361  ERASNESRGMIFSLYIGITLIGVVGGQMMIPLEDVRTPVLFMICGIFYCIAMLPTTLSTA 540
            ERA+NESRG+IFSLYI ITL GVVGGQMMIP  +  T   FMICGI YC+AMLPT LS A
Sbjct: 121  ERATNESRGVIFSLYIAITLFGVVGGQMMIPFGETSTTFFFMICGILYCVAMLPTLLSRA 180

Query: 541  ASPQPLKAVRLDLPALYRNSPVSCLGILLVGIANGAYGTLGAVFGAGAGLSDTSIAVMMS 720
            ASPQPLK VRLDL  LYRNSPVS LGILL+GIANGA+GTLGAVFG  AGLSD+++A MMS
Sbjct: 181  ASPQPLKQVRLDLRGLYRNSPVSFLGILLIGIANGAFGTLGAVFGRQAGLSDSTVAAMMS 240

Query: 721  ATIFAGAVMQLPAGRLSDRIDRRYVLAAMSGIAALAGLLIFLLHPTSPALLIGLVVLYGA 900
              IF+GAVMQLPAGR+SDRIDRRYVLAA++G+ ALAGLLIFL+ P    +++ L+ +YGA
Sbjct: 241  VAIFSGAVMQLPAGRISDRIDRRYVLAALAGVGALAGLLIFLVEPGQVWIVLTLIAIYGA 300

Query: 901  VANTLYPIAVAHANDFAASEDFVKVSGGLLLLYGIGTVIGPTLSGPVMSAITPHALFLVT 1080
             AN LYPIAV+HANDFA  EDFVKVSGGLLLLYGIGT+IGPT+ GP+M+A  P+ LF++T
Sbjct: 301  AANALYPIAVSHANDFATPEDFVKVSGGLLLLYGIGTIIGPTIGGPIMTASGPYGLFMIT 360

Query: 1081 AIAHVLITVYAIIRSRIRAAVPASDRDAYTTIPTGTSAILTPQSMSLADRGAGKPPETGK 1260
            A AH+LIT YAI+RSR RA VPA++R+ ++ +  GT+   TP+S+ L+ R A  P E
Sbjct: 361  ACAHMLITAYAIVRSRRRAPVPAAERENFSPVNAGTAT--TPESLQLSPRAA--PLEELP 416

Query: 1261 SPESDDP 1281
              +DDP
Sbjct: 417  DEGADDP 423
```

This result showed a good sequence identity to a transport transmembrane protein in *S. meliloti* (a member of the α-proteobacteria, the same group as *R. leguminosarum*). Pfam analysis was carried out and showed the predicted transport transmembrane protein belongs to the MFS family, just like RL1464 in section 4.2.4.11 (see section 4.2.4.11 for MFS review).

The data indicated RL4014 was induced as part of a general stress response.

4.2.4.25. Summary

Table 4.2 shows a summary of these data for each plasmid/strain.

Table 4.2. Most Probable Gene. Table shows the insert size for each plasmid and the predicted specificity of each stress response. Also shown is the Sanger assigned number for the most probable gene activated by stress and the gene with which its shares sequence identity (plus the identity (ID) and similarity (Sim) percentage).

Plasmid	Strain	Insert	Sanger No.	Sequence Identity with	ID%	Sim%	Stress
pRU843	RU1507	1883bp	pRL100079	Glycine Betaine/L-Proline transport ATP-binding protein ProV	66	79	Hyper-osmotic
pRU854	RU1518	1908bp	pRL120564	Hypothetical protein (mRNA regulation)	65	78	Hyper-osmotic
pRU855	RU1519	2393bp	RL0352	Carboxypeptidase C (cathepsin A)	36	50	Hyper-osmotic
pRU857	RU1521	2349bp	RL1464	MFS permease	71	84	Hyper-osmotic
pRU863	RU1525	1863bp	pRL120671	ABC transporter, substrate-binding protein	75	88	Hyper-osmotic
pRU845	RU1509	2311bp	pRL90174	Acetoacetate decarboxylase	70	82	Acidic
pRU849	RU1513	3333bp	RL4220	ExoD (exopolysaccharide synthesis protein)	61	75	Acidic
pRU850	RU1514	1564bp	RL1296	Conserved hypothetical protein (enzyme activity)	57	67	Acidic
pRU853	RU1517	1035bp	pRL90014	FixH	95	99	Acidic
pRU868	RU1530	4685bp	RL1329	Probable secretion protein, homologous to RmrA	63	78	Acidic
pRU871	RU1533	2589bp	RL1329	Probable secretion protein, homologous to RmrA	63	78	Acidic
pRU844	RU1508	2507bp	RL0587	Hypothetical protein (transmembrane-related)	85	94	General
pRU846	RU1510	2306bp	RL0356	Hypothetical protein (energy generation)	78	86	General
pRU848	RU1512	2537bp	RL2296	Fatty aldehyde dehydrogenase	62	74	General
pRU859	RU1506	1417bp	pRL100149	Hypothetical protein (unknown)	87	91	General
pRU861	RU1523	5743bp	RL1157	Two-component response regulator	77	89	General
pRU862	RU1524	394bp	RL1155	Conserved hypothetical protein (membrane-related)	53	67	General
pRU865	RU1527	1606bp	RL3589	Probable D-amino acid oxidase	60	73	General
pRU866	RU1528	1708bp	RL4591	Hypothetical protein (unknown)	44	58	General
pRU867	RU1529	4475bp	pRL100291	Hypothetical outer membrane protein, homologous to NodT	67	80	General
pRU869	RU1531	2089bp	RL0409	Putative sensor histidine kinase transmembrane hybrid	50	64	General
pRU870	RU1532	2076bp	pRL120684	Hypothetical protein (unknown)	34	53	General
pRU872	RU1534	2146bp	RL4014	Putative transport transmembrane protein	68	82	General
pRU858	RU1522	1510bp	?	?	?	?	Unknown

4.3. Discussion

The work presented here was very successful with the insert ends of each fusion being sequenced precisely. The majority of inserts proved to be between 1.5 and 2.5kb, as was previously estimated (Schofield, 1995). The average insert size was 2,351bp.

Of the 24 fusions, DNA for 15 originated from the chromosome, 2 from plasmid pRL9, 3 from plasmid pRL10 and 4 from plasmid pRL12. No isolated fusions contained DNA from plasmids pRL7, pRL8 or pRL11 (Table 4.3).

Table 4.3. DNA Coverage of 3841 Representing in Fusions. Numbers of each plasmid and where the DNA of their insert originates.

Genome	Plasmids	Total	Percent
Chromosome	pRU844, pRU846, pRU848, pRU849, pRU850, pRU855, pRU857, pRU861, pRU862, pRU865, pRU866, pRU868, pRU869, pRU871, pRU872	15	63%
pRL7	None	0	0%
pRL8	None	0	0%
pRL9	pRU845, pRU853	2	8%
pRL10	pRU843, pRU859, pRU867	3	13%
pRL11	None	0	0%
pRL12	pRU854, pRU858, pRU863, pRU870	4	17%

As mentioned in Chapter 3, when the original genomic library was made it was estimated that it contained only 40% of 3841's entire DNA. It was expected that the library would have been representative of the entire genome and although it was possible, it was highly unlikely that the 40% would not include any DNA from three plasmids. It could be that there are no stress-induced genes in pRL7, pRL8 or pRL11 indicating stress response in 3841 is cistronic; however, the sample size of fusions used in this research is not large enough to conclude if this finding is significant.

If stress-induced genes were not cistronic than the coverage of such genes would be spread evenly throughout the genome of 3841; e.g. if ~80% of the genome was made up of the chromosome, then ~80% of stress-induced genes would be found in the chromosome. On comparison of the percentages in tables 4.3 and 4.4, it can seen that 65% of 3841 DNA is found in the chromosome (the rest in plasmids) and ~65% of the genes identified in this research were from the chromosome (the rest in plasmids). This indicates that stress-induced genes are evenly distributed amongst the genome of *R. leguminosarum*.

Table 4.4. DNA Coverage in 3841.

Genome	Size (bp)	Percent
Chromosome	5,047,142	65%
pRL7	161,564	2%
pRL8	147,463	2%
pRL9	352,782	5%
pRL10	488,135	6%
pRL11	684,202	9%
pRL12	870,021	11%

However, sequence data from some fusions supports a cistronic system within *R. leguminosarum*. Plasmids pRU846 and pRU855 contained stress-induced promoters that mapped to genes in close proximity (4 genes apart), as did pRU861 and pRU862 (2 genes apart) and pRU863 and pRU870 (13 genes apart). Furthermore, genes near to the region of DNA where the insert from pRU850 mapped (RL1297, RL1298, RL1299 and RL1302) all appeared to be stress-related. These data indicated that stress-related genes are in close proximity in 3841, and so stress-response maybe a cistronic system in *R. leguminosarum*. Clearly more data is required before a definite conclusion can be made.

It was expected that many of the genes identified would be hypothetical, given that 40% of the *S. meliloti* genome is made up of hypothetical genes. This proved to be true, as inserts from 9 out of the 24 fusions were predicted to contain a promoter that transcribes a hypothetical protein. Although these genes are predicted to encode hypothetical proteins, an attempt was made to predict what role they could play in stress response and fortunately Pfam analysis greatly aided this process. Characterisation of these proteins should greatly contribute to the understanding of stress response in *R. leguminosarum* and should also help assign a definite function to each of these hypothetical proteins.

It was hoped that some genes identified would have shared sequence identity with those genes already characterised as part of a stress response. The predicted gene of pRU843 shows 79% sequence identity with part of an osmoprotectant uptake system in *B. melitensis*. This was an excellent result as pRU843 was only significantly induced under hyper-osmotic conditions (generated by all tested osmolytes). This made pRU843 a very useful fusion and it was used in the further investigation of stress response of *R. leguminosarum* (see later chapters).

It was also anticipated that some genes may have matched genes already found in *R. leguminosarum*. This too was observed, as the products of genes associated with pRU853 and pRU867 have a high sequence identity to genes already characterised from 3841 (pRL90015, from pRU853, is 97% similar to FixH and pRL100291, from pRU867, is 80% similar to a NodT homologue). Interestingly, pRL90015 and pRL100293 are not actually

the genes with which they shared sequence identity, both are homologues. pRL100291 is in fact a second homologue of pRL100178; pRL100291 is 80% similar to RL3856 (NodT homologue) which in turn is 79% similar to pRL100178 (NodT) (pRL100291 is itself 77% similar to pRL100178). This research also showed that 3841 contains two copies of the *bdhA* (pRL90175 and RL3569). Having at least three similar copies of one gene says something as to the origins of 3841 and indicates that these copies may in fact be paralogues with different functions, shown here to have a possible connection with stress response. It also suggests that *R. leguminosarum* may have a high degree of redundancy in some of its more vital systems.

One of the aims for the work presented in this chapter was to determine what kind of stress response is associated with each fusion. This was accomplished as best as possible, as shown for each gene in section 4.2.4. Some fusions in particular stood out; pRU843 (RU1507) is an excellent example of a hyper-osmotic specific response, pRU850 (RU1514) is an excellent example of an acidic specific response and pRU861 is an excellent example of a general response. pRU862 (RU1524) is also interesting as it showed strong induction initially under hyper-osmosis but also induced on entry into stationary phase. This fusion proved the care that must be taken when classifying stress response, as if experiments had been stopped after 48 hours pRU862 would have been classed as inducing specifically under hyper-osmosis, which was not the case. pRU845 led to the discovery of a possible novel system used by 3841 to remove acid when countering an environment with a low pH. The putative mechanism removes acid from the cell by coupling a decarboxylation pathway to PHB, by converting PHB to acetone and removing the carboxylic acid group. BLAST analysis indicated that this system is not present in other sequenced α-proteobacteria and so may be unique to 3841, but the use of decarboxylases to remove carboxylic acid groups from molecules (e.g. GABA) in response to a low pH has been reported in other bacteria (Castanie-Cornet & Foster, 2001; Hommais *et al*., 2004).

In order to further categorise each stress-induced gene/operon, Table 4.5 shows them grouped depending on the predicted function of each gene. As some fusions led to the identification of an operon and not a single gene, the table shows the original fusion and not the gene/s themselves. In the case of hypothetical genes, they have been characterised according to their Pfam analysis where possible but are shown in bold.

Table 4.5. Predicted Functions of Fusions. Numbers of each plasmid and the predicted function of the gene/s associated with each. Hypothetical genes are shown in bold.

Function	Plasmid
Transporters	pRU843, **pRU844**, pRU853, pRU857, pRU863, **pRU867**, pRU868, pRU871, pRU872
Cell Metabolism	pRU845, **pRU846**, pRU848, pRU865
Structural Elements	pRU849, **pRU862**
Cell Signalling	pRU855
Regulation	**pRU854**, pRU861, pRU869
Unknown	**pRU850**, pRU858, **pRU859**, **pRU866**, **pRU870**

With the preliminary genome sequence available, genes (which are near to those shown to induce under stress) that encode putative regulators can be identified. Whilst the data has only been made available and so whilst potential regulator could not be specifically targeted in this research, it is planned for future work. Fusions that are associated with genes near potential regulators include pRU846 & pRU855 (RL0354), pRU862 (RL1156 & RL1157), pRU863 (pRL120674), pRU865 (RL3591), pRU867 (pRL100288 & pRL100289) and pRU868 (RL1328).

The overall aim of the work presented in these last two chapters was to isolate and identify stress-inducible pOT fusions from the LB3 library. This was successfully accomplished and allowed experiments investigating the regulation of stress response in *R. leguminosarum* 3841 to be carried out. This further research would use some of the fusions identified here, as the next chapter describes.

CHAPTER 5: ISOLATION AND CHARACTERISATION OF MUTATIONS IN STRESS-INDUCED GENES

5.1. Introduction

Work has been presented showing the isolation (Chapter 3) and characterisation of stress-induced genes (or operons) in *R. leguminosarum* 3841 (Chapter 4), all of which was accomplished through the use of the pOT vector and the LB3 library. Although these genes had been identified, it was unknown how vital each gene was to the stress response of *R. leguminosarum* and its survival. One way of determining the necessity of a gene is to create a mutation in that gene and then examine how the mutant grows. As the gene was stress-induced, a mutant may not grow (or have restricted growth) in stressed conditions. Also, the way the mutant undergoes symbiosis with a legume may change, as some genes that have been shown to be involved in stress response are also involved in nodulation (Djordjevic, 2003). The aim of the work presented here was to determine if any of the stress-induced genes identified, are important for survival of 3841 under stressed conditions.

In a second strategy, a Tn*5* library was used to isolate mutants unable to grow under severe hyper-osmosis (or that could only grow in under hyper-osmosis) in an attempt to isolate any other genes vital to the stress response of *R. leguminosarum*

5.2. Results

5.2.1. Identification of Stress Regulation Pathways

Eight key fusions (see section 5.2.2) were chosen and used to investigate the regulation of stress-induced genes. Fusions were transferred into a Tn*5* mutant library, which was screened on AMA (10mM glc, 10mM NH_4) + 100mM sucrose as before with the LB3 library. However, for this screen colonies that no longer fluoresced were isolated. Approximately 80,000 colonies were screened and 32 mutants that no longer possessed the fluorescent phenotype associated with the plasmid were isolated and confirmed. Plasmids from these isolated mutants were shown to be functioning correctly, as they could produce GFP, as before, when transferred back into wild-type 3841 and screened on AMA (10mM glc, 10mM NH_4) + 100mM sucrose. However, transduction showed that the all thirty-two of the isolated mutants were spontaneous and not due to the presence of the Tn*5* transposon. Furthermore, these mutants had a high reversion rate, indicating that they were very unstable and making complementation impossible. Inserts from two of the fusions (pRU843 and pRU857) were cloned from their pOT plasmid into a new reporter vector, pRU1064 (Karunakaran *et al*., 2005) that used *gusA* as a reporter gene, as well as the *gfpuv* marker, in order to determine whether the pOT vector was at fault. The pRU1064 based plasmids were transferred into the Tn*5* library and ~20,000 colonies were screened on AMA (10mM glc, 10mM NH_4) + 100mM sucrose for those no longer expressing the reporter gene; only one was found. As before, the plasmid isolated from mutant functioned properly when returned to the wild-type 3841 and again transduction showed that the isolated mutant was spontaneous and not due to the presence of the Tn*5* transposon. As no Tn*5* mutant could be isolated, no further investigation into the regulation of stress-induced genes was carried out with this technique.

5.2.2. Generation of Specific Mutants

After a highly successful screening process (Chapter 3), it was decided to focus the work on specific fusions. As 24 separate plasmids were isolated, it would not have been possible to fully investigate every single one so the number of fusions needed to be reduced. However, at this stage there was no preliminary *R. leguminosarum* genome and although some sequencing data relating for each insert had been obtained, it was incomplete (Chapter 4). As such, the decision on what fusions to screen was made primarily on the conditions under which the fusions were induced and the strength of this induction, measured by the amount of GFP produced. As part of this choice, it was decided to concentrate research on

the fusions that had been induced under hyper-osmosis. In the end, eight fusions were chosen, each of which had shown a medium (++) to high (+++) GFP expression when screened on AMA (10mM glc, 10mM NH_4) + 100mM sucrose. These plasmids were pRU843, pRU845, pRU846, pRU848, pRU855, pRU857, pRU862 and pRU867. The regulatory pathways of these eight fusions were first tested (see above), before the genes associated with each plasmid were investigated. The genes associated with these fusions are a *proV* homologue (pRL100079), a hypothetical (pRL90173), a hypothetical (RL0356), a fatty aldehyde dehydrogenase (RL2296), a carboxypeptidase (RL0352), an MFS permease (RL1464), a hypothetical (RL1155) and a *nodT* homologue (RL1529) respectively. As well as these eight, the two-component response regulator associated with pRU861 (RL1157) and the GntR regulator (RL0354) found between the inserts of pRU846 and pRU855 were also selected for mutation. This was because they could have been involved in the regulation of the genes associated with other pOT fusions: RL0354 with pRU846 and pRU855 (Chapter 4, section 4.2.4.10) and RL1157 with pRU862 (Chapter 4, section 4.2.4.15).

In order to make a mutant in each of these genes the pK19mob vector (Schäfer *et al*., 1994) was utilised, using the methods previously described by Prell *et al*. in 2002. This method involved amplifying an internal region of the target gene by PCR (Fig. 5.1) and cloning the product into the pK19mob vector. Suitable primers were designed for each gene (Table 5.1)

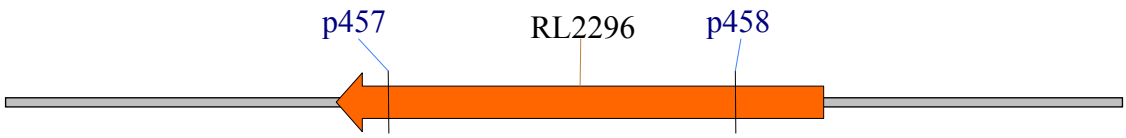


Figure 5.1. RL2296 (Fatty Aldehyde Dehydrogenase). Figure shows an example of the initial phase of the method used to generate a mutant, the RL2296 gene and where PCR primers (p457 & p458) will bind in order to amplify an internal region.

Table 5.1. Target Genes and Primers. Table shows the target gene and the plasmid it is associated with, the primers used to amplify an internal region and the size of the product.

Gene	Plasmid	Primers	Product Size (bp)
RL0354	pRU846/855	p430 & p431	424
RL1157	pRU861	p453 & p454	483
RL0352	pRU855	p455 & p456	1022
RL2296	pRU848	p457 & p458	1106
pRL90173	pRU845	p459 & p460	973
RL0356	pRU846	p461 & p462	566
RL1157	pRU862	p463 & p464	568
RL1464	pRU857	p465 & p466	792
RL1529	pRU867	p467 & p468	933
pRL100079	pRU843	p469 & p470	871

The internal regions were amplified by PCR using BIO-X-ACT (Chapter 2, section 2.6.6), the products of which were then cloned into the pCR® 2.1-TOPO® vector (Chapter 2, section 2.6.4), transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and stocked. Products were inserted into the pCR® 2.1-TOPO® vector before pK19mob, in order to use the polylinker within pCR® 2.1-TOPO®, as the primers used did not contain any sites for restriction enzymes needed to clone into pK19mob and so sites from the pCR® 2.1-TOPO® vector were used (Fig. 5.2). Restriction mapping was used to determine in which orientation the products had gone into pCR® 2.1-TOPO®. The internal regions were then digested out of their pCR® 2.1-TOPO® host by *Eco*RI (except for the plasmid that housed the RL1157 region, which contained an internal *Eco*RI site so a *Hin*dIII/*Xba*I double digest was used), isolated by gel extraction (Chapter 2, section 2.6.2) and ligated in pK19mob that had been digested by the same enzyme(s) (Chapter 2, section 2.6.4) (Fig. 5.3). Restriction mapping was used to determine in which orientation the products had gone into the pK19mob vector. Each pK19mob vector containing an internal region was transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and stocked. All of this data is summarised in Table 5.2.

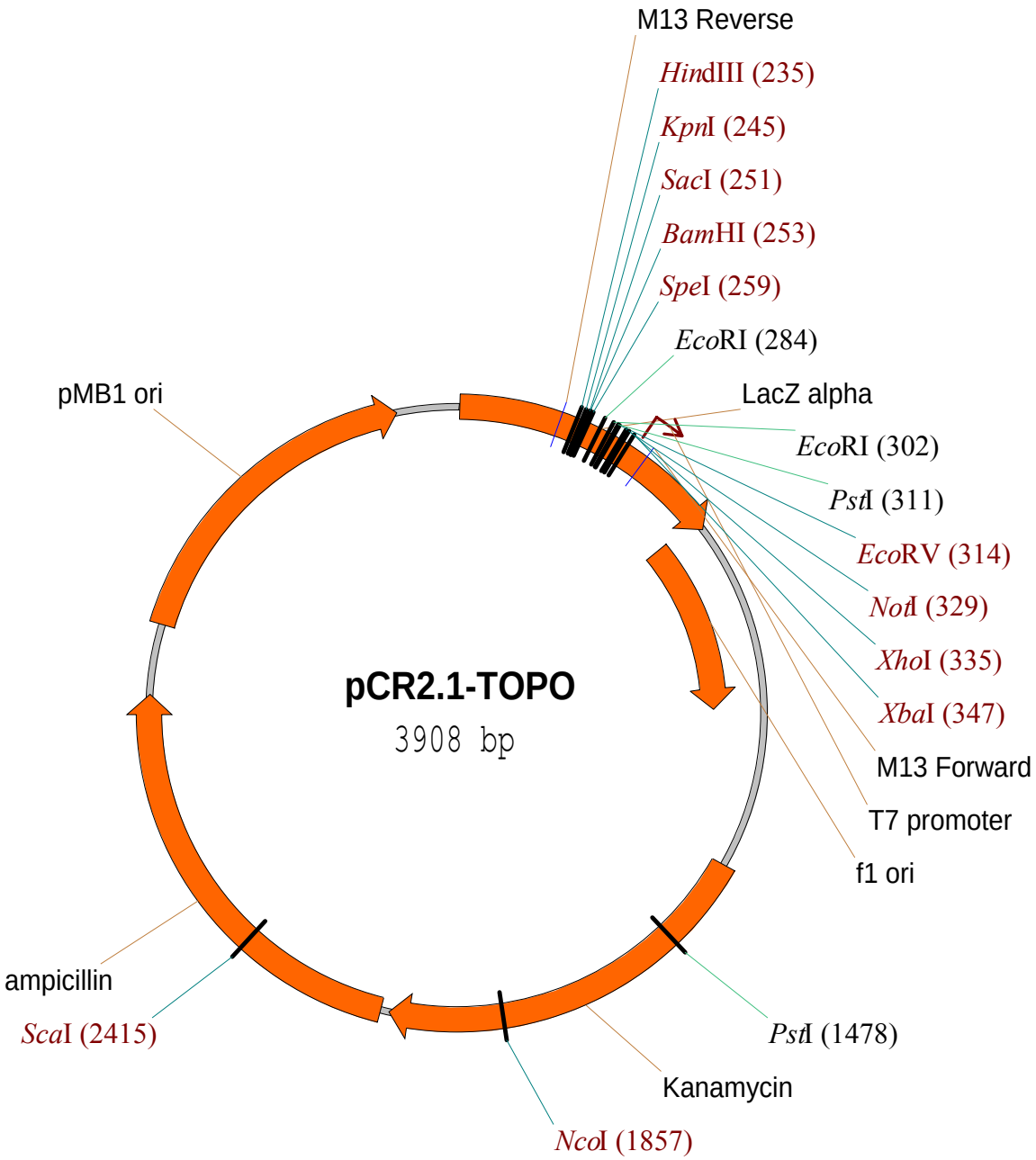


Figure 5.2. The pCR® 2.1-TOPO® Vector. Highlighted are the kanamycin resistance genes, the *lacZ* gene and some restriction sites. Unique restriction sites are shown in red and the others are shown in black. PCR products were inserted in between the two *Eco*RI sites.

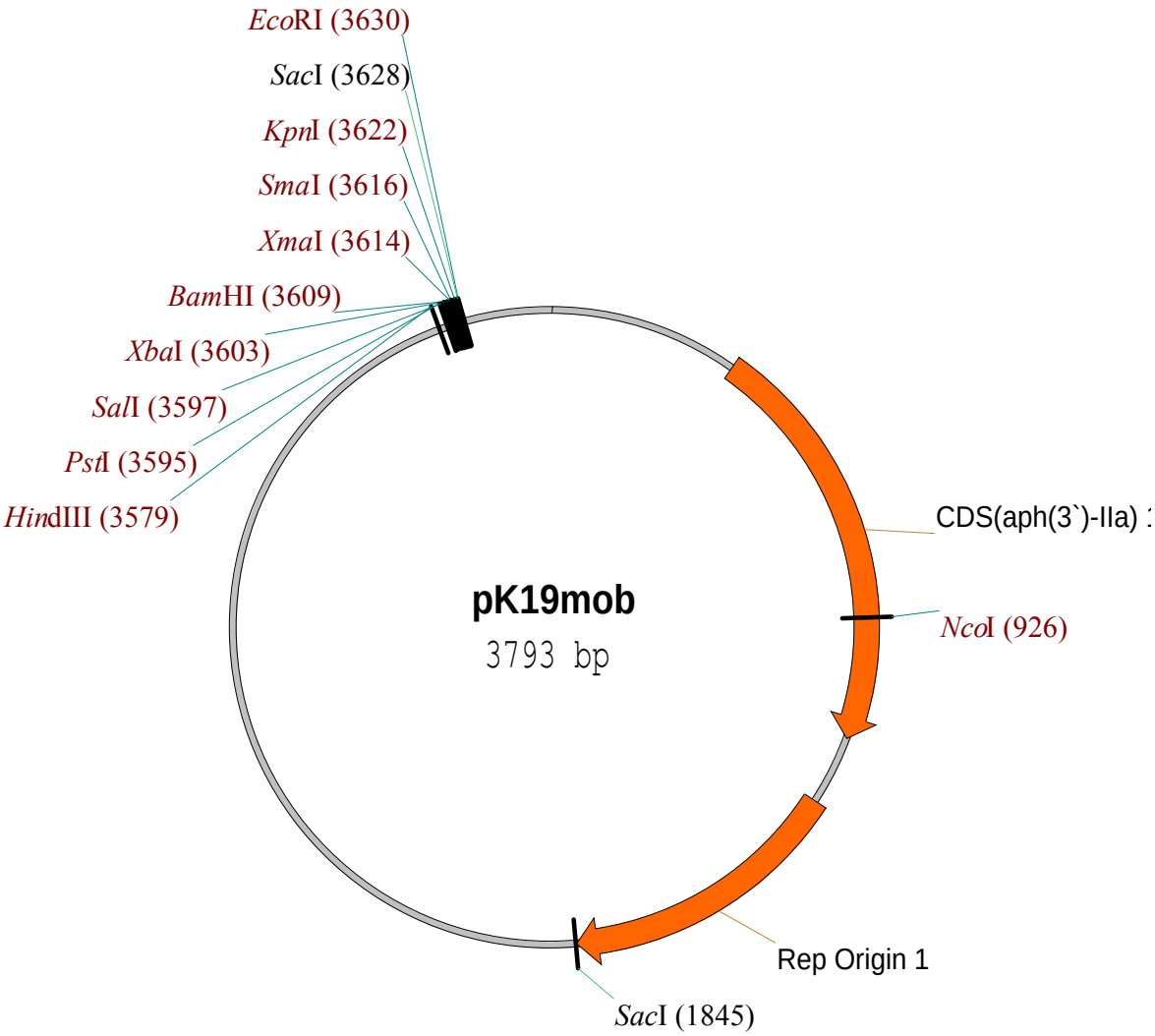


Figure 5.3. pK19mob Plasmid. Unique restriction sites are shown in red and the others are shown in black. pK19mob is kanamycin/neomycin resistant.

Table 5.2. Plasmid Numbers for the Internal Regions Cloned into the pCR® 2.1-TOPO® Vector and the pK19mob Vector.

Primers	pCR® 2.1-TOPO®	pK19mob
p430 & p431	pRU1181	pRU1189
p453 & p454	pRU1195	pRU1451
p455 & p456	pRU1196	pRU1336
p457 & p458	pRU1197	pRU1337
p459 & p460	pRU1198	pRU1338
p461 & p462	pRU1199	pRU1339
p463 & p464	pRU1200	pRU1340
p465 & p466	pRU1201	pRU1341
p467 & p468	pRU1202	pRU1342
p469 & p470	pRU1203	pRU1343

The pK19mob plasmids were then transferred into wild-type *R. leguminosarum* via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7) and the conjugation mixes were spread as serial dilutions (from 10^{-1} to 10^{-3}) on to TY plates containing streptomycin and neomycin. pK19mob does not have the necessary genes required to replicate in 3841 as a plasmid and so the only way that the *Rhizobium* can grow in the presence of neomycin is for the plasmid to integrate into the genome so that it can acquire the necessary resistance gene (Schäfer *et al*, 1994). This integration occurs by homologous recombination, which inserts the pK19mob vector into the middle of the target gene and prevents it from working properly (Fig. 5.4).

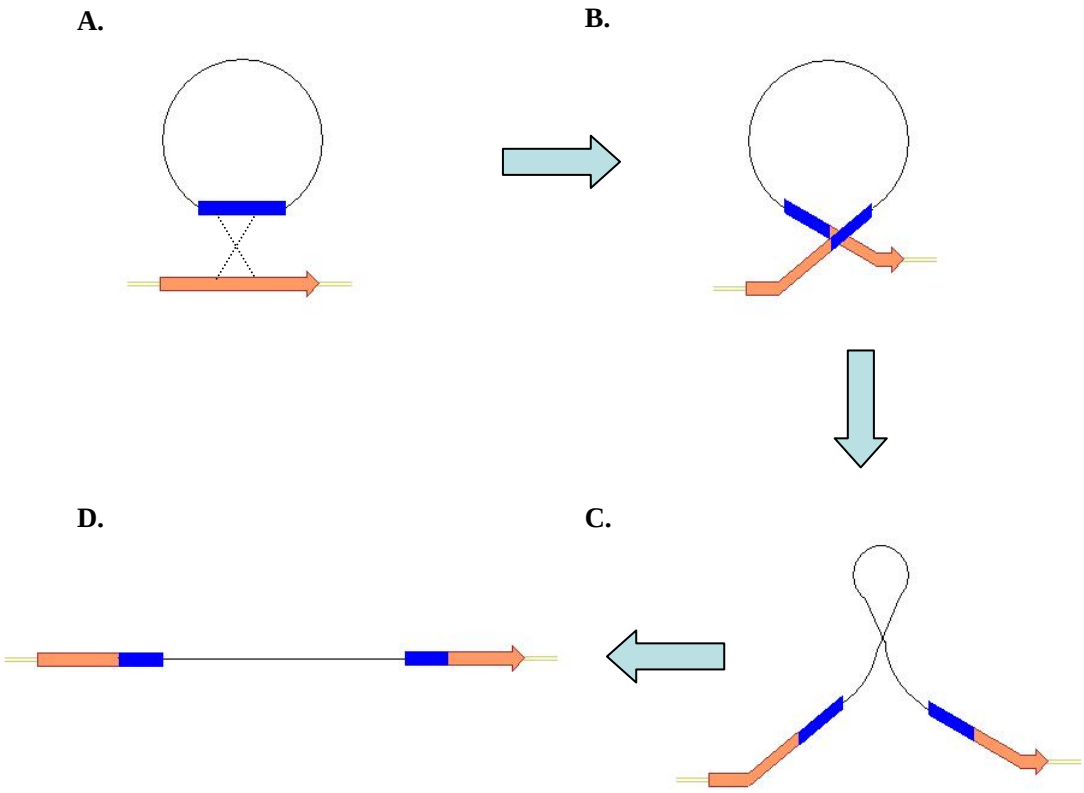


Figure 5.4. Integration of pK19mob Plasmid. A. pK19mob vector containing internal region of target gene lines up against target gene. **B. & C.** Homologous recombination occurs leading to the pK19mob vectors insertion. **D.** Vector inserted, mutated gene generated.

In order to prove the pK19mob vector had inserted correctly and the mutant had been made, colony PCR screening (Chapter 2, section 2.6.6) was used on any colonies that grew on the TY plates. As the orientation of the internal region within the pK19mob vector was known, the way it inserted into the target gene could be determined. This meant that a primer specific to pK19mob could be used with a primer specific to a region near the target

gene (but not in the plasmid clone itself) to carry out PCR and a product would only be obtained if the plasmid had inserted correctly. Two primers were designed specific to pK19mob (named pK19/18A and pK19/18B) and two primers were designed specific to the flanking regions of each target gene (Table 5.3). This method tested both ends of the target gene (Fig. 5.5). Once a mutant had been confirmed as being correct it was given a strain number and stocked (Table 5.3).

Table 5.3. Primers Used to Test Each Insertion Mutant. Table shows each pK19mob plasmid and the primers used with either pK19/18A or pK19/18B (along with the product sizes) to prove each conjugation was successful. The strain number given to the confirmed mutants is also given.

Plasmid	pK19/18A	Product(bp)	pK19/18B	Product(bp)	Strain
pRU1189	p525	1546	p535	1485	RU2193
pRU1336	p529	1955	p519	2306	RU2185
pRU1337	p520	2208	p530	2102	RU2186
pRU1338	p531	1824	p521	2169	RU2187
pRU1339	p532	1429	p522	1762	RU2188
pRU1340	p538	1431	p528	1455	RU2189
pRU1341	p546	1783	p545	1870	RU2190
pRU1342	p533	1864	p523	2226	RU2191
pRU1343	p534	1875	p524	2154	RU2192
pRU1451	p537	1469	p527	1439	RU2184

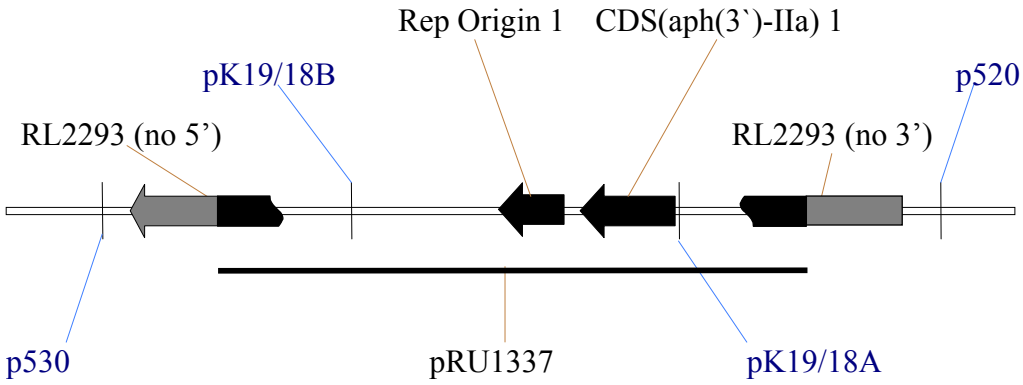


Figure 5.5. PCR Screen of pK19mob Generated Mutant. Figure shows an example of the PCR screening method on an insertion mutant. The RL2293 gene has had pRU1337 successfully integrated (native genes in grey, genes from pRU1337 in black) and so a PCR with primers pK19/18A & p520 and with pK19/18B & p530 will both produce a product. (N.B. although it appears as though there are now two copies of RL2293, they are incomplete and so would not produce a functional protein.)

5.2.3. Hyper-Osmotic MICs

It was already known that 100mM sucrose (or NaCl) was a high enough concentration to induce a stress response and induce GFP expression (Chapter 3), but it was unknown if this concentration would actually inhibit the growth on *R. leguminosarum*. In order to establish this, a MIC of the growth rate *R. leguminosarum* was carried out with various concentrations of sucrose (Fig. 5.6) and NaCl (Fig. 5.7) added to AMS (10mM glc, 10mM NH_4). A 3841 culture was grown up on a TY slope, washed in TY broth and resuspended to an OD_{600} = ~1. The AMS (with various levels of osmolyte) was then inoculated with the same volume of this suspension and cultures were left to grow. OD_{600} readings of each culture were taken at regular intervals over ~3 days and from these the mean generation times (MGT) of 3841 under each condition were calculated (Tables 5.4 & 5.5) with the following formula:

$$MGT = \frac{\text{Time of Growth}}{\left(\frac{\text{LogOD}\mathbf{b} - \text{LogOD}\mathbf{a}}{\text{Log2}}\right)}$$

Where **a** is the reading from the first time point, **b** is the reading from the second time point and the time of growth is the differencew between the times that **a** and **b** were taken. Reading were always taken during exponential growth phase, e.g. for growth on 100mM sucrose **a** = 3 hours and **b** = 24 hours.

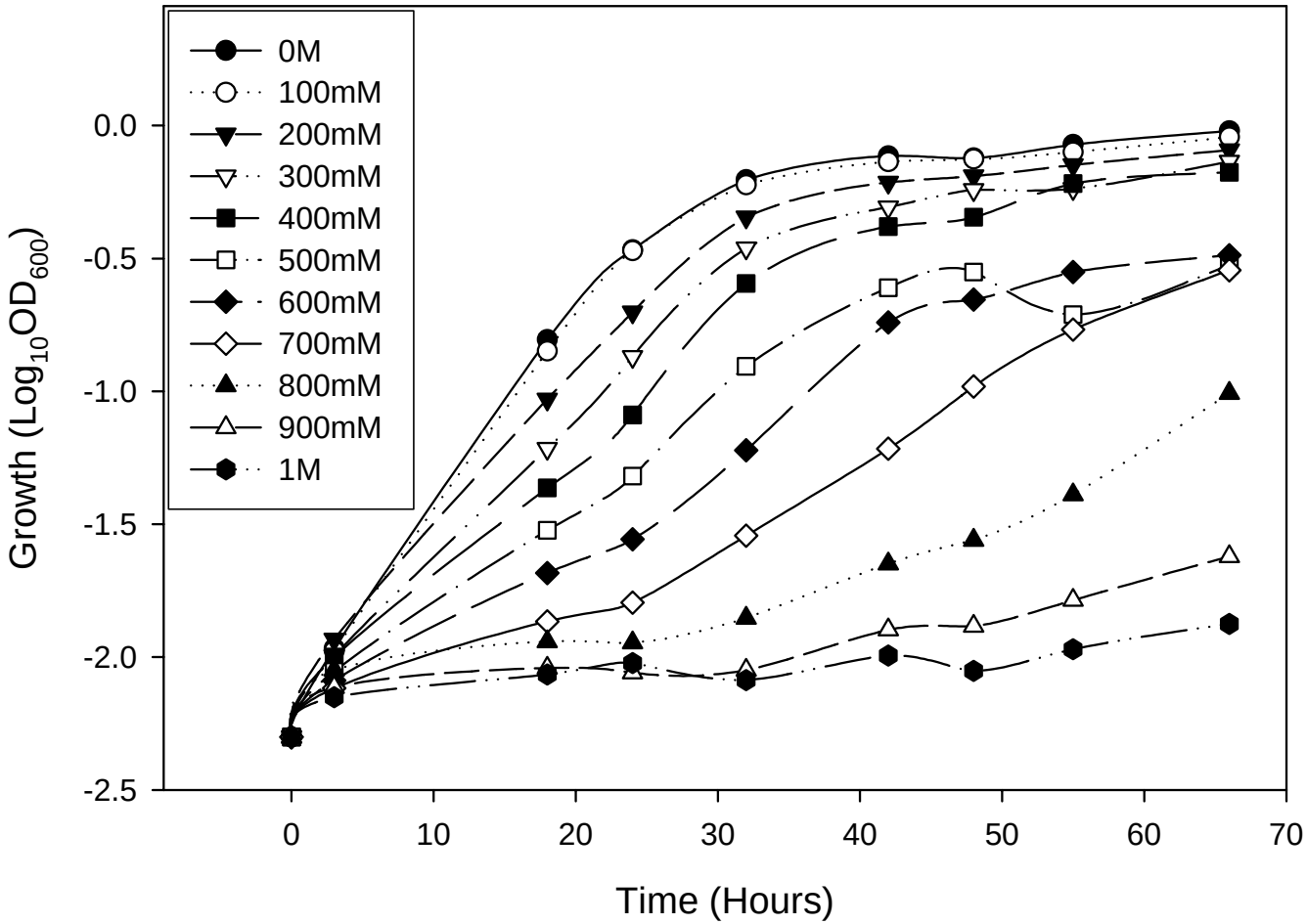


Figure 5.6. Sucrose Concentration MIC. Cells were grown in AMS (10mM glc, 10mM NH_4) containing the concentrations of sucrose indicated. OD_{600} readings are expressed as logarithms (base 10). Readings are the average of three experiments.

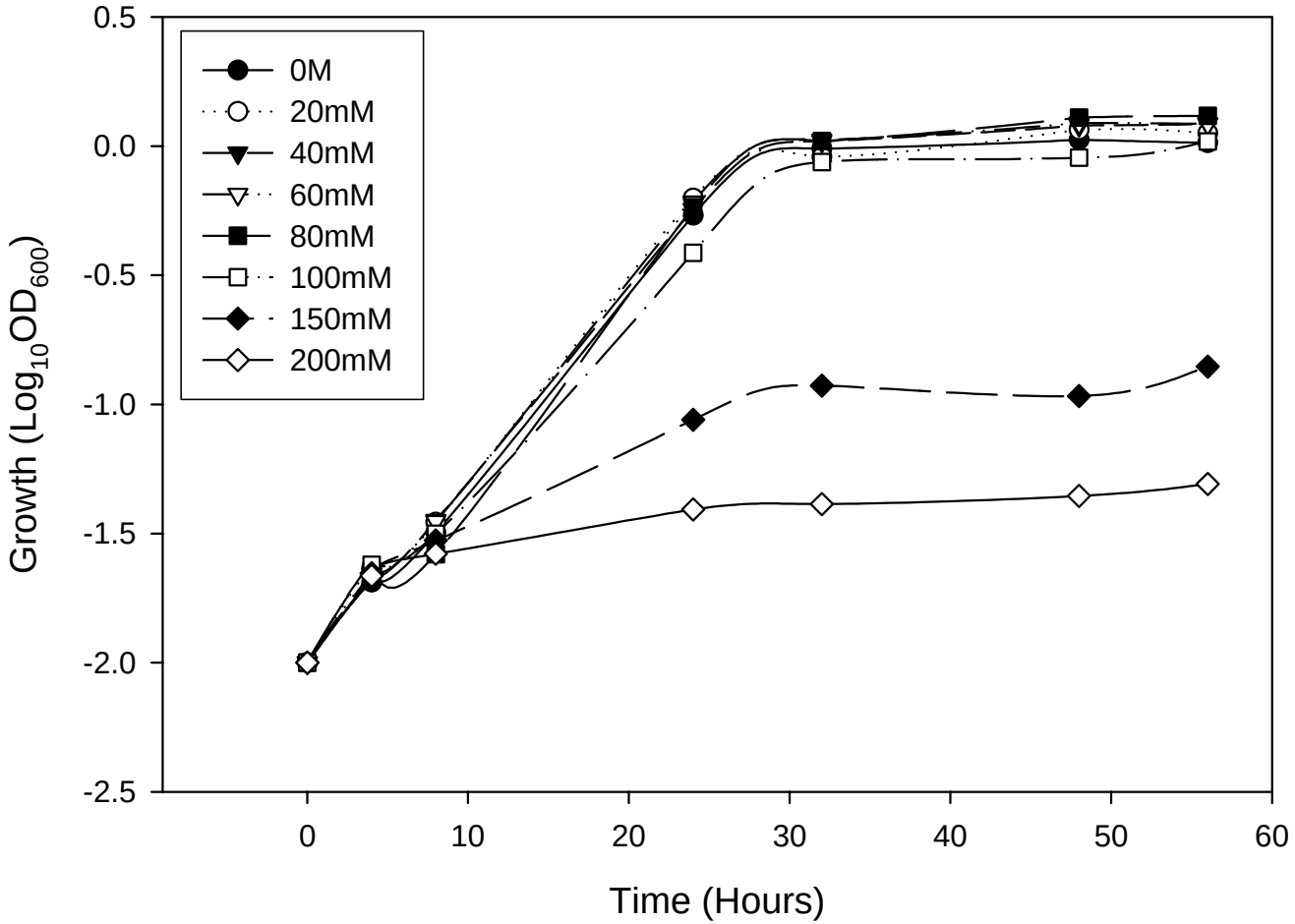


Figure 5.7. NaCl Concentration MIC. Cells were grown in AMS (10mM glc, 10mM NH_4) containing the concentrations of NaCl indicated. OD_{600} readings are expressed as logarithms (base 10). Readings are the average of three experiments.

Table 5.4. MGTs of 3841 Grown in the Presence of Various Concentrations of Sucrose. MGT mean and standard errors for 3841 grown in increasing amounts of sucrose from three separate experiments.

Sucrose Conc. (mM)	MGT (hours)
0	4.2 ± 0.7
100	4.2 ± 0.5
200	5.1 ± 0.4
300	5.6 ± 0.6
400	6.9 ± 0.5
500	8.1 ± 0.5
600	8.7 ± 0.5
700	11.8 ± 1.3
800	17.9 ± 0.9
900	37.5 ± 0.5
1000	67.2 ± 0.4

Table 5.5. MGTs of 3841 Grown in the Presence of Various Concentrations of NaCl. MGT mean and standard errors for 3841 grown in increasing amounts of NaCl from three separate experiments.

NaCl Conc. (mM)	MGT (hours)
0	4.2 ± 0.1
20	4.2 ± 0.1
40	4.2 ± 0.4
60	4.2 ± 0.2
80	4.3 ± 0.1
100	5.0 ± 0.2
150	10.1 ± 0.2
200	23.5 ± 0.1

From these data, it was decided to test the mutants in AMS (10mM glc, 10mM NH_4) with 300mM sucrose, 500mM sucrose and 125mM NaCl. As growth of 3841 was limited under these conditions, a mutant in a gene vital to stress response would cause further limitations and may entirely stop growth.

5.2.4. Mutant Growth Rates and in Planta Phenotypes

As with the MIC experiments above, each mutant was grown up on a TY slope, washed in TY broth and resuspended to an OD_{600} = ~1. The AMS (10mM glc, 10mM NH_4) with the appropriate concentration of osmolyte was then inoculated with the same volume of this suspension (100µl in 10ml) and cultures were left to grow. A 3841 control strain

was also grown up alongside the mutants. OD_{600} readings of each culture were taken at regular intervals over 32 hours, in order to monitor the initial growth rates. From these readings the MGT of each mutant under each condition were calculated (Table 5.6).

As can be seen, all the mutants were able to grow under all the stressed conditions tested. However, some of the mutants took significantly longer to grow under some of the stressed conditions then the others. The only mutant to have an increased MGT under all of the tested conditions was RU2185 (mutant of RL0352, the carboxypeptidase gene associated with pRU855). However, this difference in growth rate was not shown to be significant by T test (at the 95% confidence level or $p<0.05$), on any of the tested media (Table 5.6).

Although no changes in growth rates were observed with the mutants, the way they interacted with pea plants could have been altered. To investigate this, three lots of three pea seeds (*Pisum sativum* cv. Avola) were surface sterilised and sown into pots. Seeds were inoculated with a suspension of each mutant, thinned to two plants per pot on germination and allowed to grow for 6 weeks (Chapter 2, section 2.11). A 3841 control inoculation was also performed alongside the mutants, as well as plants that had only been inoculated with nH_2O. After the 6 weeks, the growth of the pea plants inoculated with wild-type 3841 and with water was compared to the plants inoculated with each mutant. In every case, plants inoculated with mutant strains showed no significant difference when compared to those inoculated with wild-type (Table 5.7 & Fig. 5.8). The nodules produced from plants inoculated with mutant strains were also identical to those seen on plants inoculated with wild-type (Fig.5.9).

Table 5.6. MGTs of Mutants Grown with and without Sucrose or NaCl plus Errors. Table shows the calculated MGT mean and standard errors for each strain grown in AMS, AMS + 300mM sucrose, AMS + 500mM sucrose and AMS + 125mM NaCl from three separate experiments. The percentage value shows the difference in growth rate between strains grown unstressed and those grown in the presence of sucrose; e.g. the MGT for RU2184 was 107% more when 300mM sucrose was present. A T test was carried out on the mutants with a MGT greater than that observed with the wild-type to show the significance of the difference and the p value for each is shown.

Strain	AMS (hours)	300mM Sucrose (hours)	%	T Test p value	500mM Sucrose (hours)	%	T Test p value	125mM NaCl (hours)	%	T Test p value
RU2184	3.9 ± 0.5	4.2 ± 0.7	107	0.72	7.3 ± 0.4	186	-	4.7 ± 0.5	119	-
RU2185	3.8 ± 0.5	5.3 ± 1.1	137	0.27	10.3 ± 2.3	269	0.68	5.6 ± 0.6	145	0.04
RU2186	3.8 ± 0.6	4.5 ± 0.7	119	0.98	9.9 ± 1.3	261	0.99	5.5 ± 1.0	144	0.11
RU2187	3.8 ± 0.6	4. 7 ± 0.6	124	0.17	9.2 ± 0.9	245	-	4.7 ± 0.4	124	-
RU2188	3.7 ± 0.2	4.9 ± 0.5	131	0.23	10.7 ± 2.6	289	0.60	4.9 ± 0.7	132	0.66
RU2189	4.1 ± 0.6	4.1 ± 0.5	100	-	10.7 ± 4.2	259	0.95	4.7 ± 0.5	114	-
RU2190	3.7 ± 0.2	4.7 ± 0.5	125	0.34	8.5 ± 2.0	228	-	4.5 ± 0.4	121	-
RU2191	4.1 ± 0.8	4.0 ± 0.2	96	-	7.7 ± 0.6	186	-	5.0 ± 0.7	120	-
RU2192	3.6 ± 0.2	4.6 ± 0.1	127	0.39	6.9 ± 0.7	192	-	5.7 ± 1.1	159	0.03
RU2193	3.6 ± 0.3	5.0 ± 0.3	138	0.09	6.7 ± 1.3	184	-	4.6 ± 0.5	128	-
3841	3.6 ± 0.5	4.2 ± 0.6	117		9.2 ± 3.3	254		4.7 ± 0.4	129	

Table 5.7. Dry Weights of Pea Plants Inoculated with Wild-Type, Water or a PK19mob Mutant. Plants were isolated after 6 weeks growth and dried in an oven for two days at 80°C before being weighed. Mean weight with standard error is shown along with the p value obtained when a T test was performed compared to the plants inoculated with 3841. All mutant stress had a p value greater then 0.05 indicating no significant difference st the 95% confidence level. (n/g = no plant growth)

Strain	RU2184	RU2185	RU2186	RU2187	RU2188	RU2189	RU2190	RU2191	RU2192	RU2193	3841	H_2O
Plant 1	0.787	0.387	1.534	1.185	0.867	1.022	1.028	0.565	0.182	0.644	1.085	0.418
Plant 2	0.261	0.393	0.721	1.080	0.526	0.613	0.522	0.836	0.702	1.069	0.134	0.147
Plant 3	0.825	0.845	1.521	1.030	0.797	0.803	0.121	1.115	0.894	0.216	0.951	0.340
Plant 4	0.764	0.586	1.039	0.675	0.405	0.503	0.338	0.867	0.839	0.671	1.349	0.433
Plant 5	1.057	0.399	1.089	n/g	1.077	1.126	n/g	1.166	0.602	0.042	1.662	0.133
Plant 6	1.323	1.284	0.662	n/g	0.618	0.633	n/g	n/g	0.856	0.758	n/g	0.144
Mean	0.836	0.649	1.094	0.993	0.715	0.783	0.502	0.910	0.679	0.567	1.036	0.269
Error	0.1443	0.1463	0.1533	0.1107	0.1003	0.1008	0.1935	0.1081	0.1090	0.1532	0.2562	0.0586
P Value	0.52	0.23	0.85	0.88	0.29	0.40	0.14	0.67	0.25	0.16	-	0.04

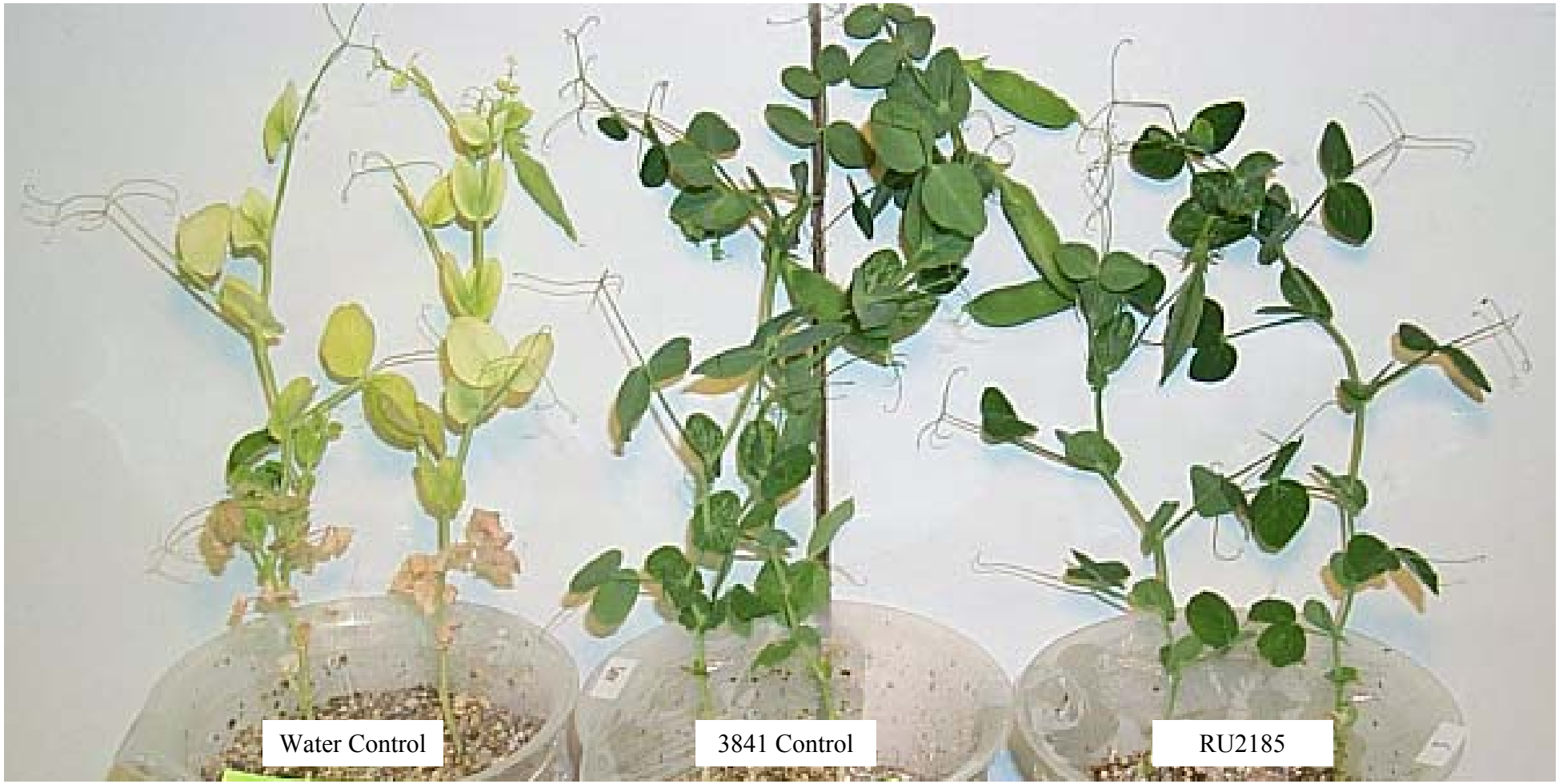


Figure 5.8. Plant Phenotypes. As can be seen, there are no differences between wild-type and mutant inoculated phenotypes. All plants inoculated with each mutant had an identical phenotype to RU2185 and 3841.

Figure 5.9. Nodule Phenotypes. Example of root sections from plants inoculated with water (left) and wild-type 3841 (right). Healthy pink round nodules can be seen on the plant inoculated with 3841 whereas none are seen on the water control. Nodules derived from all mutant strains appeared identical to those obtained with wild-type.

It was clear from this data, that none of the mutated genes were vital for successful symbiosis with pea plants. This was not necessarily a surprise result with RU2191 (the mutated NodT homologue, RL1529, associated with pRU867) as there are 2 other copies of this gene present in 3841 that could have taken any role RL1529 may have had in symbiosis/nodulation.

5.2.5. Plasmid Phenotypes in Regulator Mutants

Two mutations were made in genes with a predicted regulatory function, RU2184 (the mutated two-component response regulator, RL1157) and RU2193 (the mutated GntR repressor, RL0354), and although tested above with the other mutants, they were primarily generated to see if they regulated nearby genes, represented by pRU862 and pRU846 & pRU855 respectively. pRU862 was transferred into RU2184 and pRU846 & pRU855 were independently transferred into RU2193 from their *E. coli* hosts via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7); each plasmid was also conjugated into wild-type 3841 and were grown alongside, as controls. Strains were then grown on AMA (10mM glc, 10mM NH_4) and

AMA (10mM glc, 10mM NH_4) + 100mM sucrose for 10 days and were examined daily under UV light from 4 days. No change in GFP expression was observed with pRU846 or pRU855 in the RU2193 background, compared to their expression in wild-type 3841, i.e. both sets of colonies still fluoresced under UV light on AMA + 100mM sucrose plates. However, pRU862 showed no signs of GFP production in RU2184 (Fig. 5.10) under hyper-osmotic condition, unlike its 3841 counterpart. This result was also observed when RU2184 containing pRU862 was grown on AMA (10mM glc, 10mM NH_4) buffered to pH 5.75 (Fig. 5.10).

This result indicates that expression of pRU862 (and therefore RL1155) is controlled by the two-component response regulator predicted to be encoded by RL1157. In order to prove the mutated gene RL1157 was responsible for this phenotype, the mutant was complemented. As RL1156 (a sensory histidine protein kinase) and RL1157 may be encoded by the same promoter as an operon (Chapter 4, section 4.2.4.14), the mutation of RL1157 may have had a polar effect on RL1156. For this reason, RU2184 was complemented with RL1157 and with RL1156 and RL1157.

Primers were designed to amplify the region of DNA containing RL1157 and its predicted promoter (termed p2CRR) (p527 & 637) and the region of RL1157, RL1156 and p2CRR (p527 & p638). The regions were amplified by PCR using BIO-X-ACT (Chapter 2, section 2.6.6), the products of which were then cloned into the pCR® 2.1-TOPO® vector (Chapter 2, section 2.6.4), transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and stocked (pRU1611 and pRU1613 respectively). Products were inserted into the pCR® 2.1-TOPO® vector first in order to use the polylinker within pCR® 2.1-TOPO®, as with the pK19mob plasmids above. The pRK415 plasmid (Keen *et al*., 1988) was chosen as the vector to be used to house the genes for complementation, mainly because it is tetracycline resistant and therefore compatible with RU2184 containing pRU862, which was already resistant to streptomycin, gentamycin and kanamycin. pRK415 also had the advantage of containing a constitutive promoter (plac), next to the polylinker into which the PCR products were cloned. This meant that if the promoter for RL1157 had been incorrectly predicted, the genes within pRK415 could still be activated under plac control. The products were digested out of the pCR® 2.1-TOPO® vector with *Kpn*I, isolated by gel extraction (Chapter 2, section 2.6.2) and were ligated into pRK415 (Chapter 2, section 2.6.4), which had also been digested by the same enzyme. Products were cloned into pRK415 in both possible orientations (checked by restriction mapping) so that they were under the control of plac or the promoter for RL1157. The pRK415 plasmids were then transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and stocked (Table 5.8).

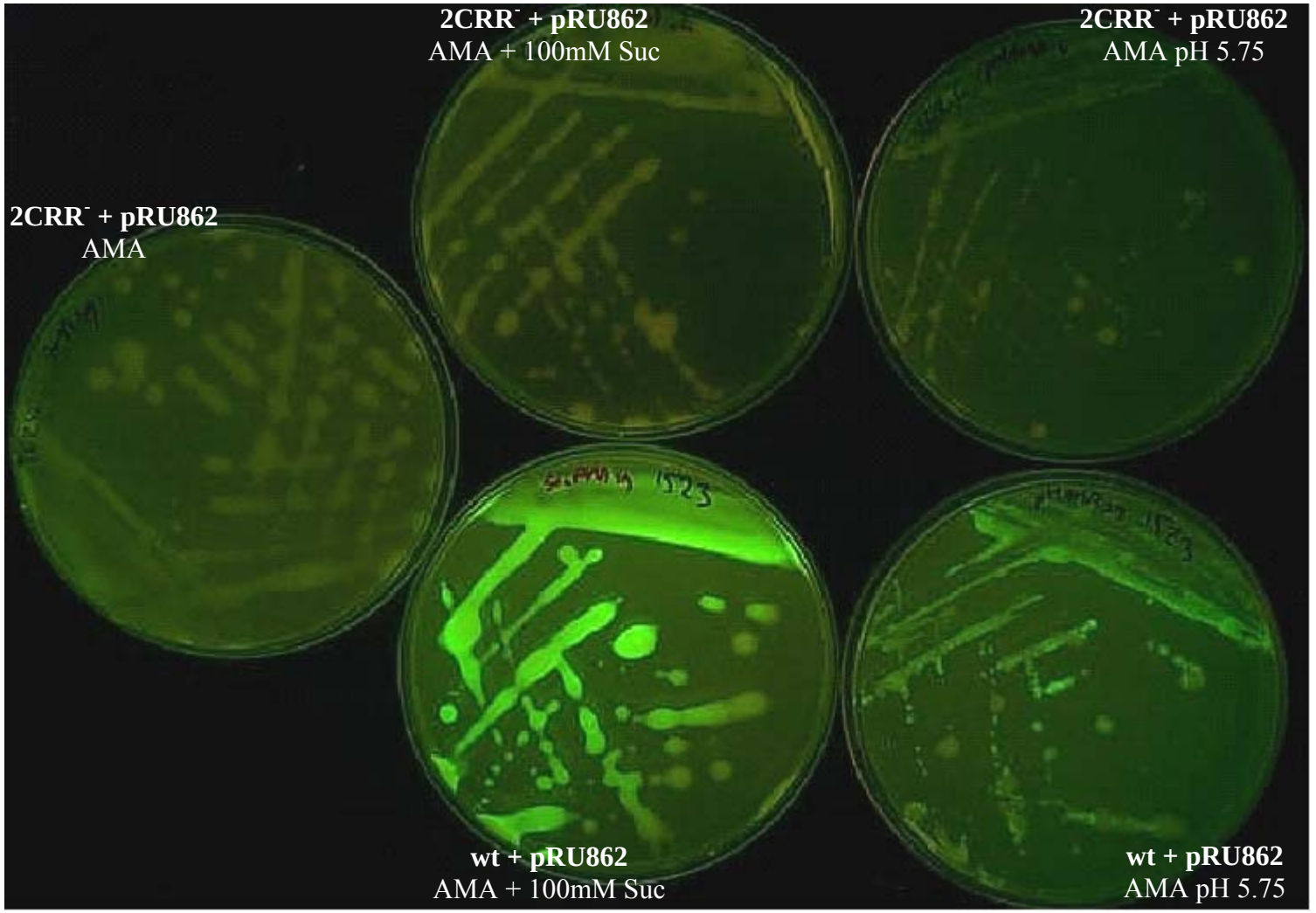


Figure 5.10. pRU862 Expression in 3841 and RU2184. pRU862 in RU2184 (top 3 plates) and pRU862 in 3841 (bottom 2 plates) grown in conditions as shown. No GFP expression of pRU862 is seen in RU2184. (2CRR$^-$ = Two component response regulator mutant, Suc = sucrose)

Table 5.8. Plasmids used for Complementation of RU2184. Shown are the gene(s) present in each plasmid, and the promoter controlling their transcription.

Primers	Product	Promoter	Plasmid
p527 & p637	Regulator	p2CRR	pRU1645
		plac	pRU1646
p527 & p638	Regulator & Kinase	plac	pRU1647
		p2CRR	pRU1683

Each complementing plasmid was then transferred into RU2184 containing pRU862 via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7) and were screened on AMA (10mM glc, 10mM NH_4) and AMA (10mM glc, 10mM NH_4) + 100mM sucrose. pRK415 was also transferred into RU2184 containing pRU862 and screened on the same media to check that any effect observed was not due to the pRK415 vector itself. All four complementing plasmids restored GFP production to RU2184 containing pRU862 when grown under hyper-osmotic conditions, whilst pRK415 had no effect (Fig. 5.11).

pRU1645	**pRU1646**	**pRU1647**	**pRU1683**	**pRK415**
2CRR (p2CRR)	2CRR (plac)	2CRR + Kin (plac)	2CRR + Kin (p2CRR)	Negative control

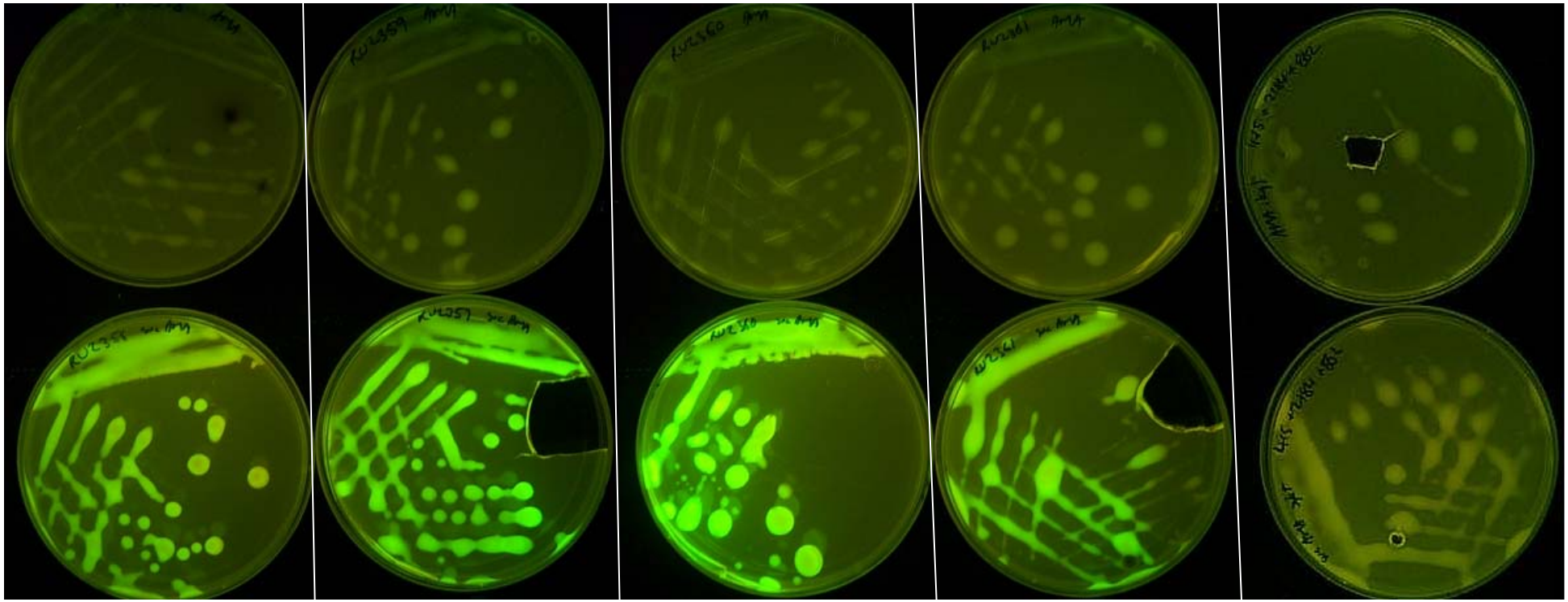


Figure 5.11. RU2184 containing pRU862 Complemented with Various Plasmids. RU2184 with pRU862 and the plasmid as indicated above on AMA (10mM glc, 10mM NH_4) (top row) and AMA (10mM glc, 10mM NH_4) + 100mM sucrose (bottom row). All plasmids restored GFP production under stressed conditions except for pRK415 on its own.

This result proved that pRU682 (and therefore RL1155) was regulated by the two-component response regulator predicted to be encoded by RL1157. It also showed that the promoter for RL1157 (p2CRR) had been correctly identified, as genes were transcribed from the pRK415 vector under p2CRR control and the control of plac. The results also showed that the complementation of RL1156 was not required, as GFP expression was identical between all the complemented mutants. This meant that either, RL1156 is not involved in the regulation of RL1155 or that its transcription was not effected by the insertion of pRU1451 into RL1157 and so it was still functioning correctly within RU2184. As a response regulator/histidine kinase pair is well known to act together as a regulator, it was assumed that the latter reason was true. Perhaps there is a cryptic promoter at the end of RL1157 that was unaffected by the pRU1451 insertion, or a promoter is present within the pK19mob vector that allowed the transcription of RL1156, even though RL1157 was mutated. Either way, a gene involved in the regulation of stress response in *R. leguminosarum* was discovered.

In order to determine how widespread the effect of this regulator was, all of the other pOT fusions were transferred individually into RU2184 (and wild-type 3841, as a control) from their *E. coli* hosts via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7). These were then grown, alongside their original strains (Chapter 3), on AMA (10mM glc, 10mM NH_4) and AMA (10mM glc, 10mM NH_4) + 100mM sucrose for 10 days and were examined daily under UV light from 4 days. No change in GFP expression was observed with any other pOT fusions in the RU2184 background and indicated that its regulatory control was limited to RL1155 (and perhaps other neighbouring genes).

5.2.6. Screening Tn5 Mutant Library for Growth Phenotypes

As none of the mutants reported above showed an inability to grow under highly stressed conditions (section 5.2.4), it was decided to screen an individual Tn*5* mutant library for such mutants. This work was initially conducted in conjunction with another experiment within the laboratory and so was carried out in a RU1736::Tn*5* mutant library (RU1736 is mutant of 3841 with an *aap*$^-$ *bra*$^-$ *gsp*$^+$ genotype, White, unpublished). Mutants were individually spotted onto AMA (10mM glc, 10mM NH_4) and AMA (10mM glc, 10mM NH_4) + 300mM sucrose and allowed to grow (Fig. 5.12).

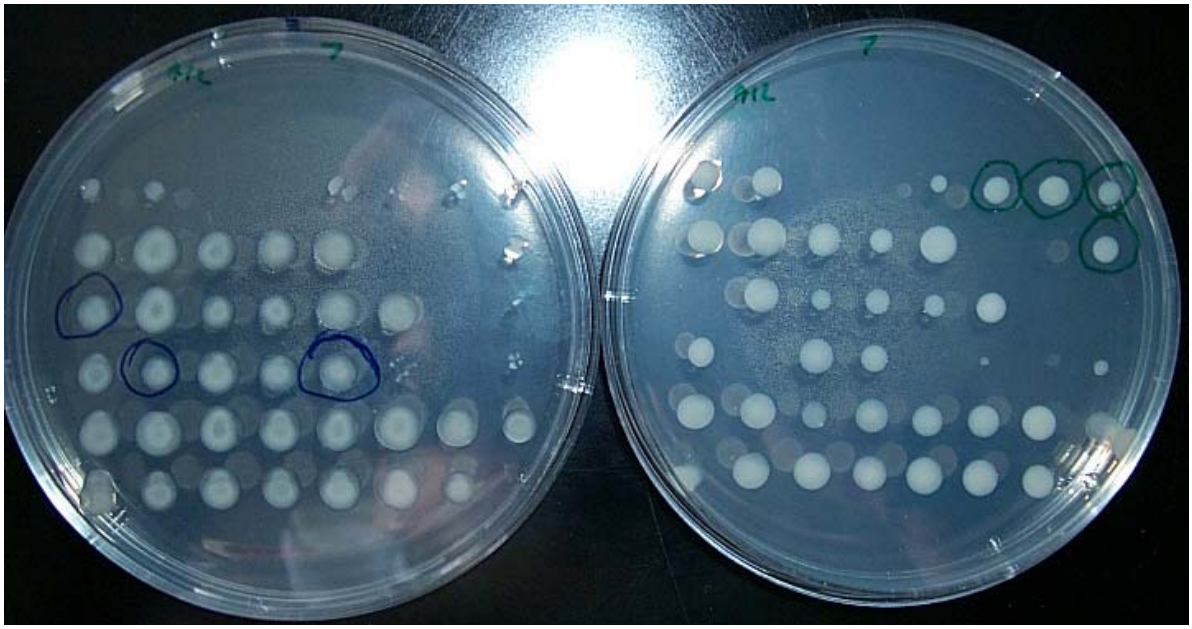


Figure 5.12. Tn*5* Mutant Library Screening. Some colonies could grow on AMA (10mM glc, 10mM NH_4) + 300mM sucrose (left) but not AMA (10mM glc, 10mM NH_4) (right), and visa versa.

Any mutants that could grow under one condition but not the other were isolated and then streaked onto the two conditions again, to confirm the non-growth phenotype was genuine and not due to an error made in inoculation. From this screen, only one mutant was found that was unable to grow on AMA with 300mM present, which was stocked (RU2248). RU2248 then underwent general transduction using the bacteriophage RL38 (Beringer *et al*. 1978; Buchanan-Wollaston, 1979) to confirm that the Tn*5* transposon was responsible for the growth phenotype. Recipient *R*. *leguminosarum*, which underwent transduction, were plated on to TY containing high amounts of kanamycin (80μg/ml) so that only successfully transduced *Rhizobium*; i.e. only bacteria now containing the Tn*5* transposon, would grow (Chapter 2, section 2.9). Colonies that grew on the TY plates were checked to prove that they had the Tn*5* transposon present by colony PCR (Chapter 2, section 2.6.6). Primers (p473 and p474) were designed to amplify a region, containing the kanamycin resistance gene, specific to Tn*5* and so is not found in wild-type 3841 (Figs. 5.13 & 5.14). (RU2248 underwent transduction with RU1736 – its original host, and 3841 – wild-type *R*. *leguminosarum*.)

A transductant in a 3841 background (RU2415) and a transductant in a RU1736 background (RU2421) were screened on AMA (10mM glc, 10mM NH_4) and AMA (10mM glc, 10mM NH_4) + 300mM sucrose alongside RU2248. None of the strains could grow on AMA containing 300mM sucrose, confirming that the Tn*5* was responsible for the phenotype and also that the phenotype was present in 3841 and not just in the RU1736

mutant. (Following this confirmation, research was switched from RU2248 to RU2415, in keeping with the investigation of 3841 and eliminating any effect RU1736 may have had during further study.)

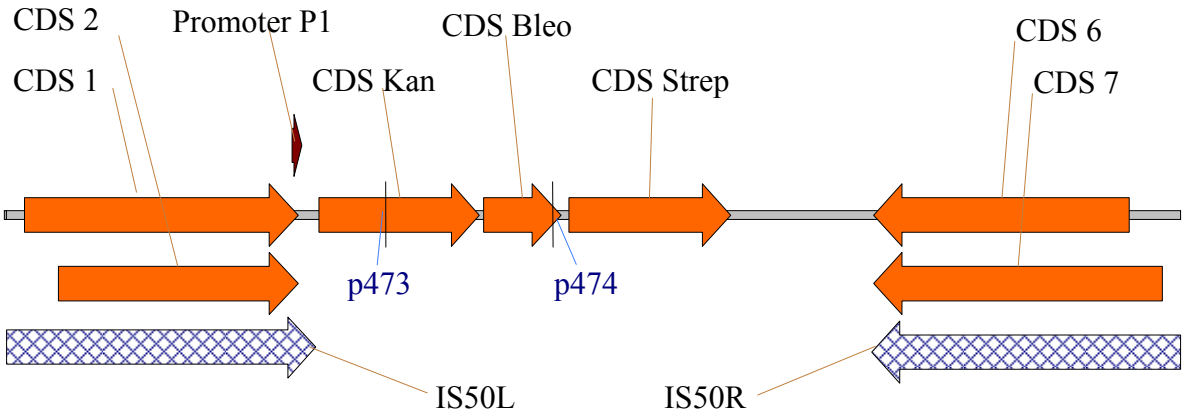


Figure 5.13. Tn*5* Transposon. The Tn*5* transposon flanked by the insertion sequence regions (IS50), its coding regions and where primers p473 and p474 bind. p473 and p474 amplify a region containing the kanamycin resistance gene to make an ~800bp PCR product, which is not present in the *R. leguminosarum* 3841 genome and so proves the presence of the Tn*5* transposon (Fig. 5.14).

This experiment was repeated but with an individual 3841::Tn*5* mutant library, eliminating any phenotype that may be associated with RU1736. From this, a mutant was isolated that would only grow on AMA containing 300mM sucrose, which was stocked (RU2283). RU2283 also underwent general transduction using the bacteriophage RL38 to confirm that the Tn*5* transposon was responsible for the growth phenotype, the Tn*5* transposon from the mutant into wild-type 3841 in the corresponding genomic region. Recipient *R. leguminosarum*, which underwent transduction, were plated on to TY containing high amounts of kanamycin (80µg/ml) so that only successfully transduced 3841; i.e. only bacteria now containing the Tn*5* transposon, would grow (Chapter 2, section 2.9). 300mM sucrose was also added to the TY as transductants of RU2283 may not have been able grown on TY alone. A successful 3841 transductant (RU2422) was screened on AMA (10mM glc, 10mM NH_4) and AMA (10mM glc, 10mM NH_4) + 300mM sucrose alongside RU2283. Neither strain would grow on AMA unless 300mM sucrose was present, confirming that the Tn*5* was responsible for the phenotype.

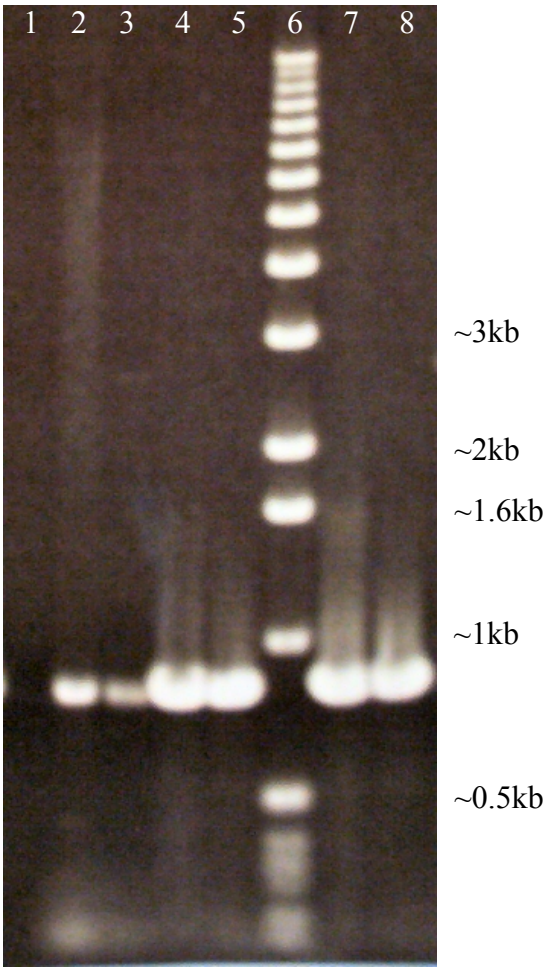


Figure 5.14. Colony PCR Screen. Lanes: 1) 3841 control; 2-5) Colonies from transduction; 6) 1kb Ladder (key sizes shown); 7-8) Colonies from transduction. As can be seen from ~800bp band, all transduced colonies contained the Tn*5* transposon, whilst no band is produced from a wild-type colony.

5.2.7. Characterisation of Tn*5* Mutants

In order to fully characterise the mutants, each (alongside a 3841 control) was screened on a range of media that covered a wide variety of different carbon and/or nitrogen sources and different osmolytes were used to generate hyper-osmosis (Fig. 5.15 & Table 5.9).

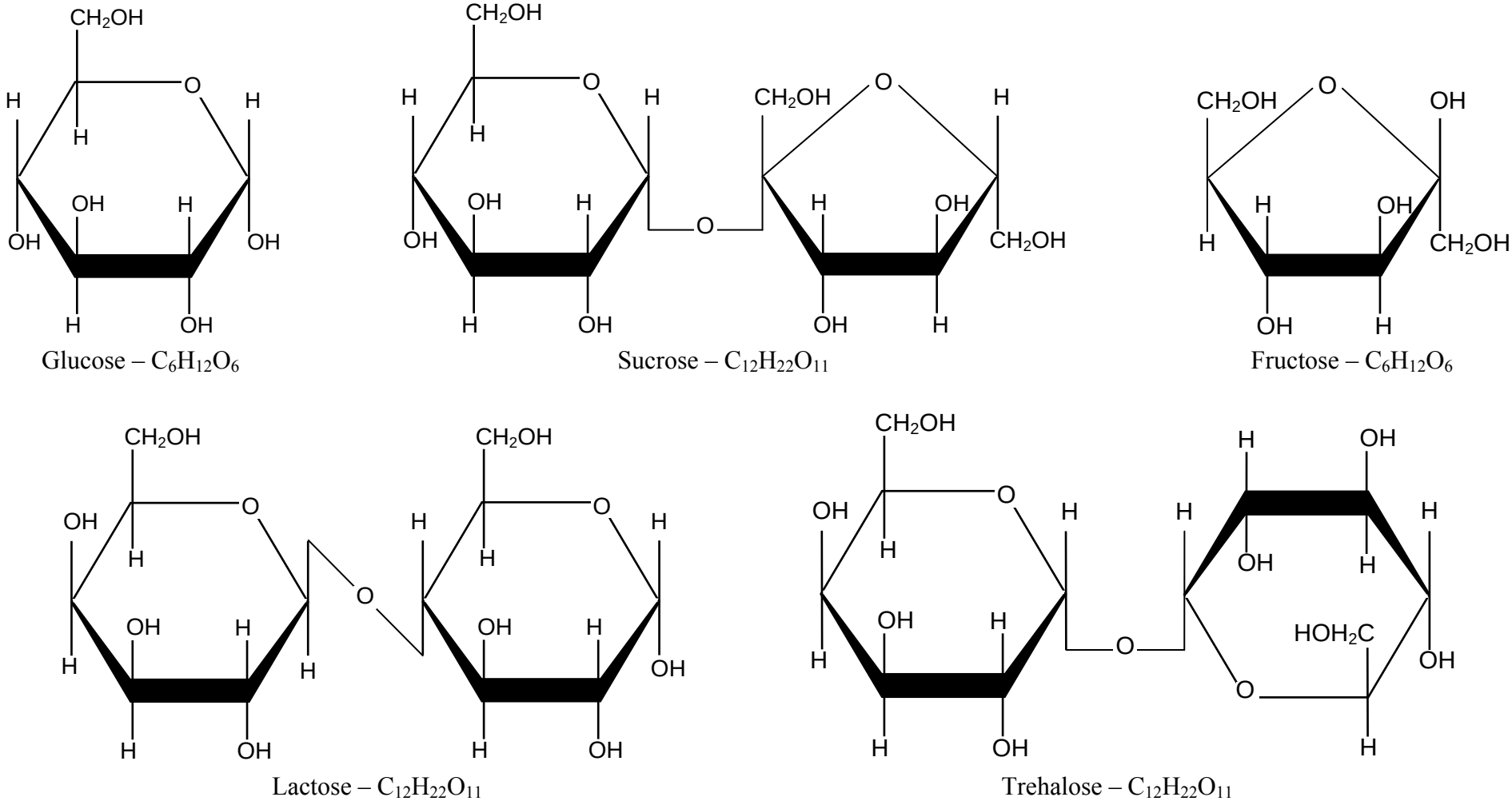
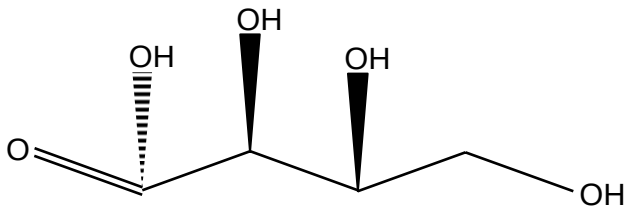
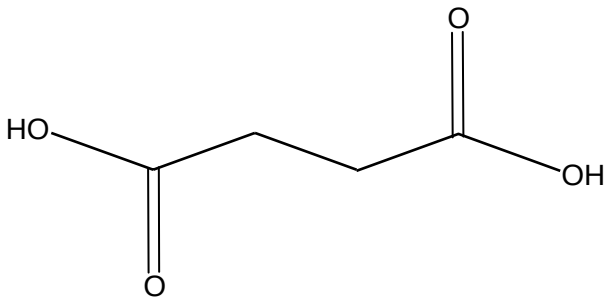


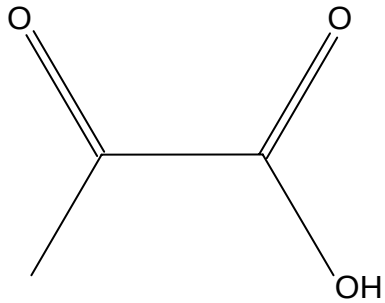
Figure 5.15. Compounds with which Mutants were Tested. Structure of NaCl not included, as it is a simple ionic compound. (Continued next page.)



Arabinose – $C_5H_{10}O_5$

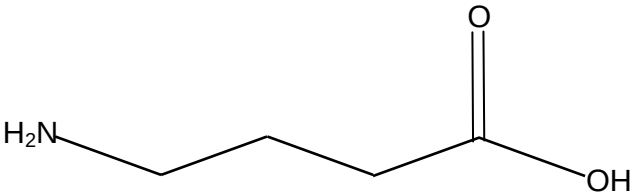


Succinate – $C_4H_6O_4$



Pyruvate – $C_3H_4O_3$

Mannitol – $C_6H_{14}O_6$



GABA – $C_4H_9NO_2$

Figure 5.15. (cont.) Compounds with which Mutants were Tested. Structure of NaCl not included, as it is a simple ionic compound.

Table 5.9. Growth of 3841, RU2415 and RU2283 on Various Media. Symbols indicate the amount of growth on plates after strains were left to grow for 7 days.

AMA	3841	RU2415	RU2283
10mM glc 10mM NH_4	+++	+++	-
100mM glc 10mM NH_4	+++	+++	-
300mM glc 10mM NH_4	+++	+++	-
5mM sucrose 10mM NH_4	+++	+++	+++
10mM sucrose 10mM NH_4	+++	++	+++
100mM sucrose 10mM NH_4	+++	-	+++
300mM sucrose 10mM NH_4	+++	-	+++
10mM glc 5mM sucrose 10mM NH_4	+++	+++	+
10mM glc 10mM NH_4 100mM NaCl	+++	+++	-
10mM fructose 10mM NH_4	+++	++	+++
300mM fructose 10mM NH_4	+++	-	+++
5mM lactose 10mM NH_4	+++	+++	+++
300mM lactose 10mM NH_4	+++	+++	+++
5mM trehalose 10mM NH_4	+++	+++	+++
300mM trehalose 10mM NH_4	+++	+++	+++
10mM arabinose 10mM NH_4	+++	+++	+++
300mM arabinose 10mM NH_4	+++	+++	+++
10mM succinate 10mM NH_4	+++	+++	+++
10mM mannitol 10mM NH_4	+++	+++	+++
10mM GABA	++	++	++
20mM pyruvate 10mM NH_4	++	++	++

From this set of screenings, it was observed that RU2283 could not grow on AMA containing glucose as the carbon source (Table 5.9 & Fig. 5.17), indicating that a mutation was present in a metabolic gene. RU2283 could grow on AMA (5mM sucrose, 10mM NH_4) but showed limited growth on AMA (10mM glc, 5mM sucrose, 10mM NH_4), which indicates that not only is the strain incapable of using glucose as a carbon source, but the presence of it in the media above a certain concentration inhibits growth. Growth of RU2283 was not significantly different to that of 3841 on all other media (Table 5.9). As this mutant is involved in glucose metabolism and not hyper-osmotic stress, no further research was conducted with it.

Also from this set of screening, it was observed that RU2415 could not grow on AMA plates containing 100mM (or more) sucrose or fructose, but the mutant showed no such inhibition when these sugars were present at 10mM or lower. Growth of RU2415 was not significantly different to that of 3841 on all other media, even those with high concentrations of other osmolytes used to induce an osmotic upshift (Table 5.9). This indicated that RU2415 was not sensitive to hyper-osmotic stress, but to fructose; sucrose probably inhibited growth because sucrose is a disaccharide made up of glucose and fructose. Interestingly, RU2415 could grow with sucrose or fructose when present at low levels (<10mM) as the carbon source. The result indicated that RU2415 was capable of using fructose as a carbon source for growth but that the presence of the sugar in large amounts prevented growth.

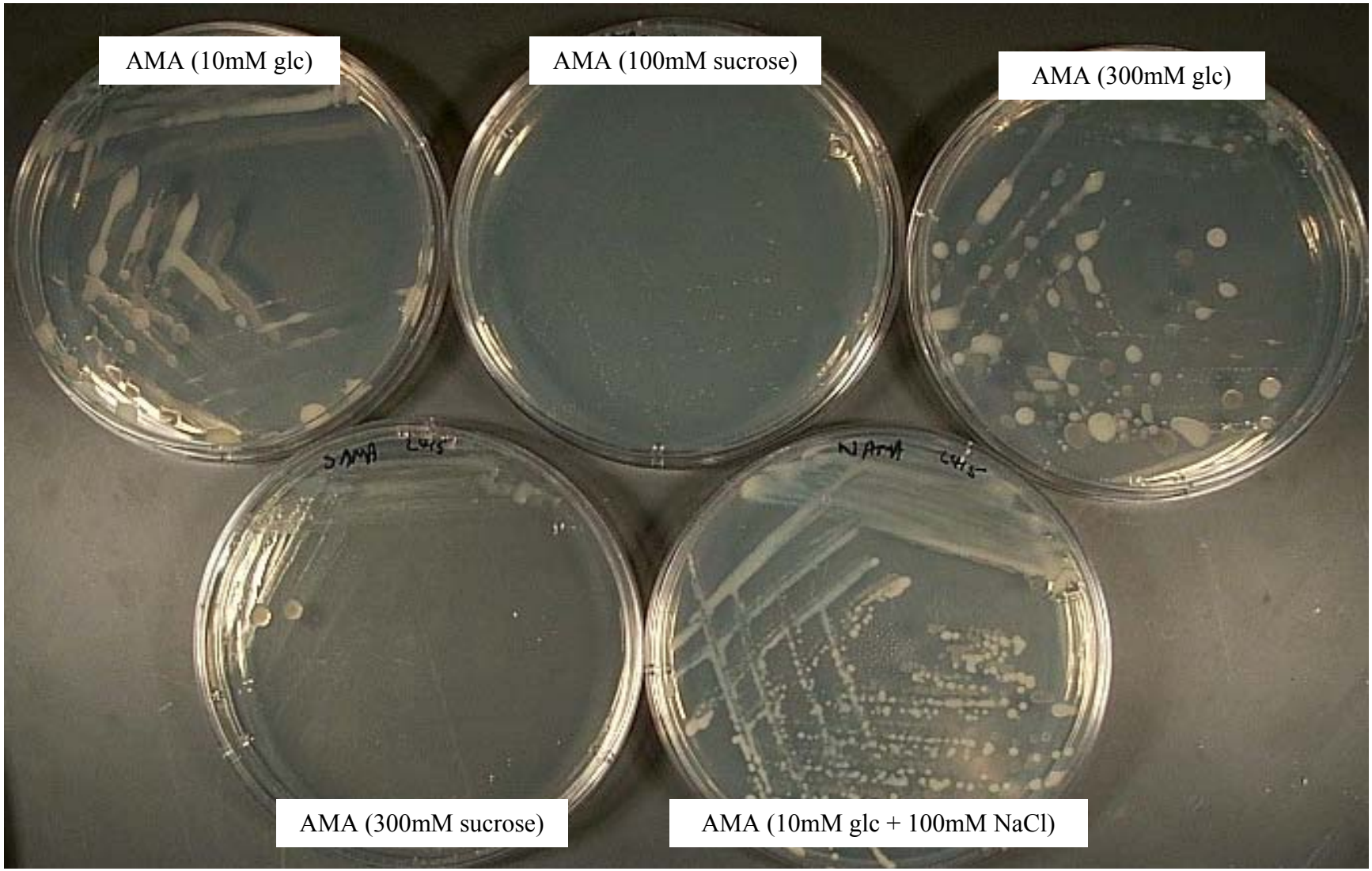


Figure 5.16. Growth of RU2415. All plates also contained 10mM NH_4.

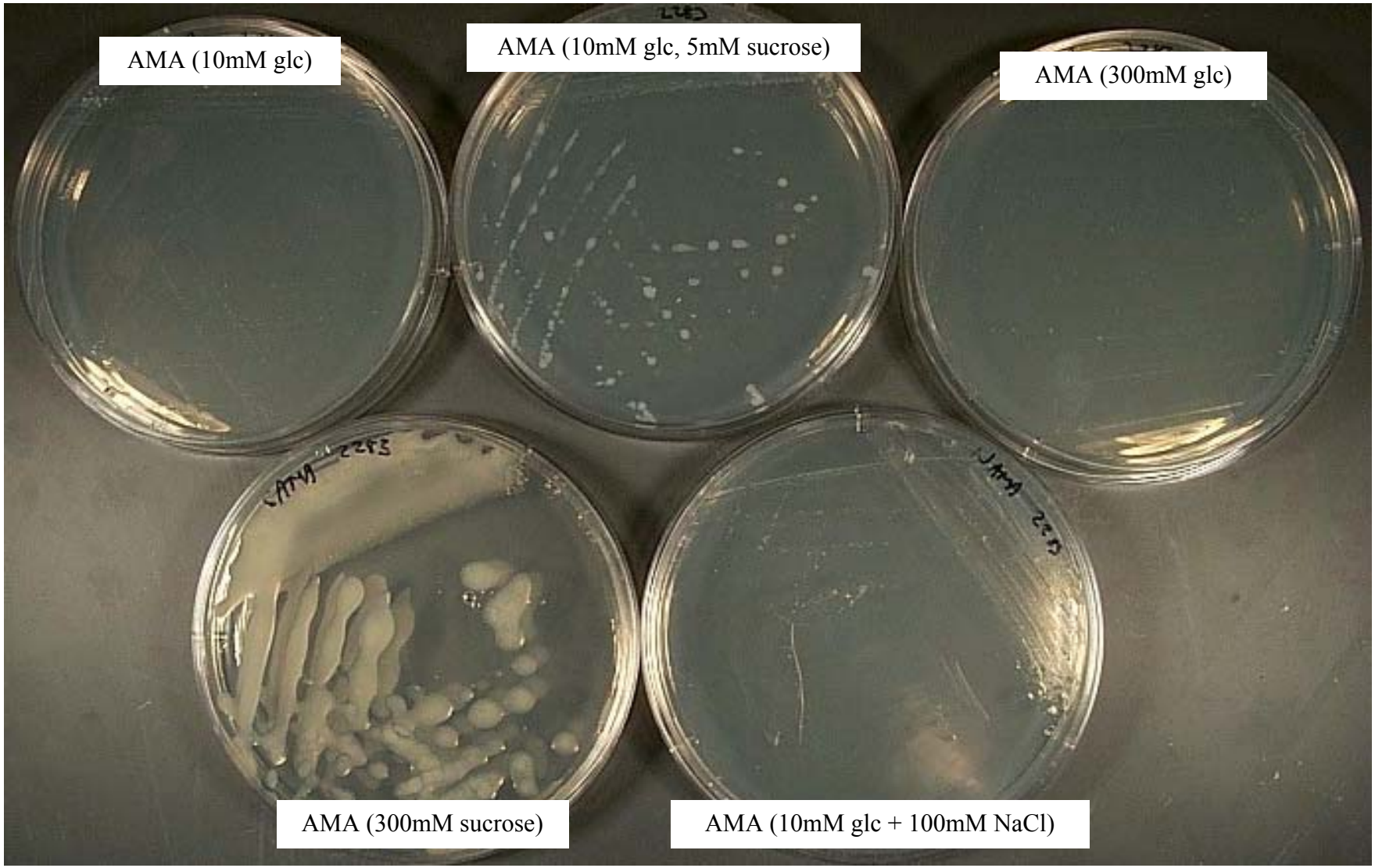


Figure 5.17. Growth of RU2283. All plates also contained 10mM NH4.

As it was known that the Tn*5* transposon was responsible for the phenotype of RU2415, it was possible to obtain sequence data from it using a primer specific to the end of the transposon that sequenced outwards and into the flanking genome. However, obtaining sequence data directly from the genome is often problematic and also Tn*5* is capped by repeating units (~1.5kb in length), which means that any primer designed would bind to both ends of the Tn*5*. Therefore that primer would sequence from both ends of the transposon, producing two sets of conflicting data from one sequencing reaction. This meant that the transposon needed to be removed from the genome and have one of its ends removed.

This was accomplished by isolating the genomic DNA of the mutant (Chapter 2, section 2.6.1) and digesting it with *Eco*RI. The digested DNA was then ligated with pBluescript® II SK (Stratagene) that had also been digested with *Eco*RI (Chapter 2, section 2.6.4). When the ligation mix was transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and grown on LA plates containing kanamycin, only cells housing a pBluescript into which the region of DNA containing Tn*5* had been inserted would be able to grow. Such colonies had their plasmids isolated. As the Tn*5* transposon contains no *Eco*RI sites, it remains intact during the digestion and so its removal from the genome is dependent on the restriction sites present in the flanking DNA. This meant that, although removed from the genome of the mutant, the transposon still had both ends present and needed one of them to be removed. This was achieved by digesting the isolated pBluescript plasmid with *Bam*HI (or *Sal*I), as one site for this enzyme is present in the middle of the transposon and one is present in the vector near the insertion site. Such a digestion removes a section of DNA from the plasmid containing one of the ends of the Tn*5* (Fig. 5.18).

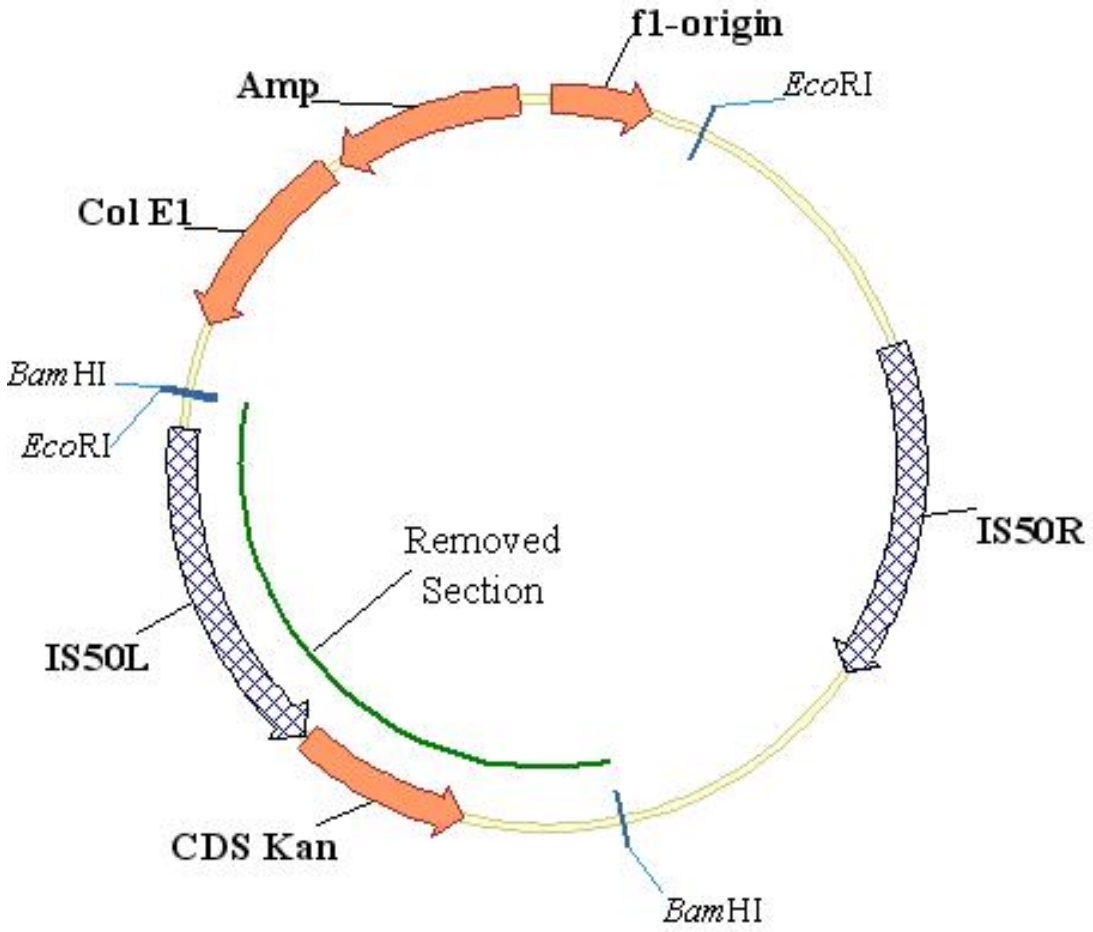


Figure 5.18. Removing Half of Tn*5* Transposon from its pBluescript Host. Figure shows the pBluescript vector containing the region of DNA from the mutant with the Tn*5*, following an *Eco*RI digest and insertion, and the section that would be removed with a *Bam*HI digest.

When this digestion was ligated, transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and grown on LA plates containing ampicillin, only a re-ligated pBluescript vector (although now containing only half the Tn*5*) was able to grow. Such colonies had their plasmids isolated and were sequenced with the IS50R primer, designed to sequence the DNA from the host genome that flanked the Tn*5* transposon.

The RU2415 data showed that the transposon had interrupted a gene (RL3429); the 9bp repeat sequence (CGTCTATGC) associated with the transposon started with the 510th base of RL3429. RL3429 shares sequence identity to a LysR-type regulator (Figs. 5.19 & 5.20) (Schell, 1993) and was near a gene (RL3431) that appeared to be involved with sugar (arabinose) efflux (Figs. 5.19 & 5.21).

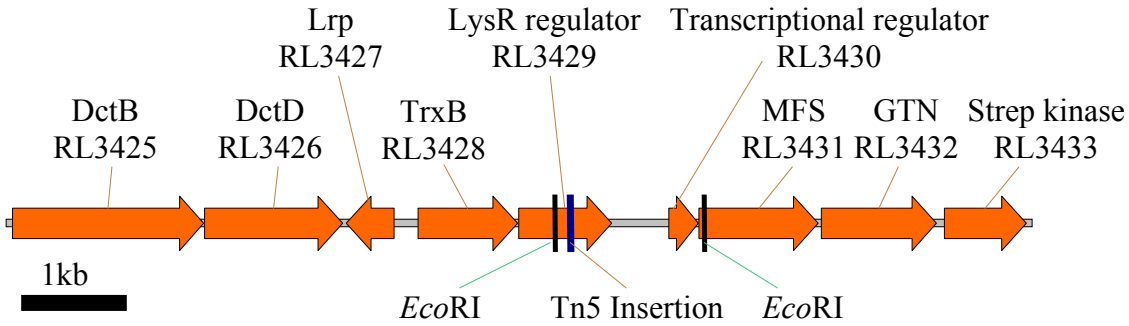


Figure 5.19. Genomic Region of RU2415 near Tn*5* Insertion. Diagram shows where Tn*5* was and the two *Eco*RI sites used to remove that region of DNA and clone it into pBluescript. ~1kb scale shown.

Transcription regulator, LysR family Atu2186 (imported) [*Agrobacterium tumefaciens* C58]
Length = 300aa

Score = 573 bits (1478), Expect = 2e-162
Identities = 279/298 (93%), Positives = 291/298 (97%), Gaps = 0/298 (0%)
Frame = +1

```
Query  16   MPLDWDKLRIFHAAAEAGSFTHAADKLHLSQSAISRQVSALEQDVGTKLFHRHARGLILT  195
            MPLDWDKLRIFHAAAEAGSFTHAADKLHLSQSAISRQVSALEQDVG KLFHRHARGLILT
Sbjct  3    MPLDWDKLRIFHAAAEAGSFTHAADKLHLSQSAISRQVSALEQDVGVKLFHRHARGLILT  62

Query  196  EQGELLYRTAHDVLLKLETVKMQLTETTEKPSGKLRVTTTVGLGQGWLTDKIQEFLQLYP  375
            EQGELLYRTAHDVLLKLETVKMQLTETTEKPSGKLRVTTTVGLGQGWLTDK+QEFLQLYP
Sbjct  63   EQGELLYRTAHDVLLKLETVKMQLTETTEKPSGKLRVTTTVGLGQGWLTDKVQEFLQLYP  122

Query  376  DVQIQLILDNEEVDVNMRHADCAIRLRQPQQSDLIQRKLFTVHMHVYAAPSYINRHGEPQ  555
            ++ IQLILDNEE+DVNMRHADCAIRLRQPQQSDLIQRKLFTVHMHVYAAPSYINRHGEPQ
Sbjct  123  EMSIQLILDNEELDVNMRHADCAIRLRQPQQSDLIQRKLFTVHMHVYAAPSYINRHGEPQ  182

Query  556  KVEDLDNHRIITFGEPAPSYLLDVNWLEVAGRSSDNKRIPHLQINSQTSIKRAALLGIGV  735
             VEDLDNHRII+FGEPAP+YLLDVNWLE AGRSSDN RIPHLQINSQTSIKRA LLGIG+
Sbjct  183  SVEDLDNHRIISFGEPAPNYLLDVNWLENAGRSSDNTRIPHLQINSQTSIKRACLLGIGI  242

Query  736  ACLPDYIVGRDPGLIQLAINADVPSFDTYFCYPDEIKNAAKLKAFRDFIVSKARNWNF  909
            ACLPDYIVGRDPGLIQL++ AD+PSFDTYFCYPDE+KNAAKLKAFRDFIV+KARNWNF
Sbjct  243  ACLPDYIVGRDPGLIQLSLAADIPSFDTYFCYPDEMKNAAKLKAFRDFIVAKARNWNF  300
```

Figure 5.20. BLAST Analysis of RL3429.

Major Facilitator Family Transporter [*Pseudomonas fluorescens*] Length=388aa

Score = 535 bits (1378), Expect = 1e-150
Identities = 276/387 (71%), Positives = 317/387 (81%), Gaps = 1/387 (0%)
Frame = +1

```
Query  1     MPLALLVLALSSFAIGTTEFVIMGLLPEVAADLSVSIPQAGWLVTGYALAVAIGAPVMAI  180
             MPL+LL+LALS+FAIGTTEFVIMGLLP+VAADL VSIP AGWLVTGYAL VAIGAP MA+
Sbjct  1     MPLSLLILALSAFAIGTTEFVIMGLLPDVAADLGVSIPGAGWLVTGYALGVAIGAPFMAL  60

Query  181   STAKLKRRTALIALMAFFIAGNLLCALASDYWVLMIARVVTALCHGAFFGIGSVVAAGLV  360
             +TA+L R+ AL+ALM  FI GNLLCALASDY VLM ARV+TALCHGAFFGIGSVVAAGLV
Sbjct  61    ATARLPRKAALVALMGIFIVGNLLCALASDYNVLMFARVITALCHGAFFGIGSVVAAGLV  120

Query  361   AEDRKARAVALMFTGLTLANALGVPLGTAIGQAYGWRATFGVVTIIGIVTISGLIAILPR  540
               +++A AVALMFTGLTLAN LGVPLGTA+GQA GWR+TF  VT+IG+V + GL+  LP
Sbjct  121   PANKRASAVALMFTGLTLANVLGVPLGTALGQAAGWRSTFWAVTVIGVVALIGLLRFLPA  180

Query  541   DKQQENGSILREIAALRNGGLWLALSTTVFFAASMFALFTYIAPLLRDVTGVSPEGVTWT  720
              + +E   +  E+AAL+  G+WL+LS T  FAASMF LFTY+APLL DVTGVSP+GVTWT
Sbjct  181   KRDEEKLDMRAELAALKGAGIWLSLSMTALFAASMFTLFTYVAPLLGDVTGVSPKGVTWT  240

Query  721   LFLIGLGLTIGNLVGGKLADWRLGATLAGVFAAIAITSIAFSYTSRFFIPAEITLFLWAM  900
             L LIGLGLT+GN++GGKLAD RLGATL GVFAA+A+ S   ++TS   IP EITLFLWA
Sbjct  241   LLLIGLGLTLGNIIGGKLADKRLGATLIGVFAAMAVVSTVLTWTSVALIPTEITLFLWAT  300

Query  901   ASFAAVPALQVGVVGFGKDAPNLVSTINIGAFNTGNALGAWVGGLVIDAGLDLTRVPLAA  1080
             ASFAAVPALQV VV FGK APNLVST+NIGAFN GNALGAWVGG VI  G  LT VPLAA
Sbjct  301   ASFAAVPALQVNVVTFGKAAPNLVSTLNIGAFNIGNALGAWVGGSVIAHGFGLTSVPLAA  360

Query  1081  ALMALIGLGATALTYLSARGRAALAPA  1161
             A +A++ L  T +T+   G A LAPA
Sbjct  361   AALAILALLVTLITFRQG-GNAELAPA  386
```

Figure 5.21. BLAST Analysis of RL3431.

As noted above, the Tn*5* was inserted into RL3429, a predicted LysR, in close proximity to RL3431, a predicted sugar efflux system. It could be that this efflux system is responsible for the removal of fructose from cells when it accumulates to such a concentration that it prevents growth and that the LysR is responsible for the regulation of this system. With the RL3429 mutated by the transposon, the efflux system may be unable to function correctly and so fructose was not expelled from cells, resulting in their death on plates containing high amounts (100mM and more) of fructose (or sucrose), but not low concentrations (<10mM).

It is worth mentioning here that pRU855 (RU1519) would only show significant induction with hyper-osmosis caused by sucrose but not by NaCl or mannitol. It could be that the genes associated with pRU855 may also play a role in sugar efflux under conditions where they are present in high concentrations.

5.3. Discussion

The initial screening of the mutants made by pK19mob insertion under severe hyper-osmotic conditions did not reveal any vital genes required in stress. There could be many reasons why a mutation in these genes did not result in an increased doubling time. It could be that 3841 contains other copies of the genes mutated. It has been observed that *R. leguminosarum* contains many copies of the same genes, e.g. the *fix* genes, *nodT* (Chapter 4) and it could be that when one of the genes is mutated another takes its place/function.

Another reason could be that, although induced under hyper-osmotic conditions, the products of the tested genes have no function relating to the optimum stress response of *R. leguminosarum*, which is why the mutants of these genes showed no real change in growth rate. This would not be the first occurrence in *Rhizobium* of a mutant having no growth phenotype or a gene being induced under a condition where it has no apparent function (Cabanes *et al*., 2000; Ampe, *et al*., 2003; Djordjevic, *et al*., 2003; Becker *et al*., 2004).

One of the generated mutants (RU2184 – mutated RL1157, a two-component response regulator) proved most interesting, as it prevented the GFP production of one of the stress-induced fusions (pRU862) under hyper-osmotic and acidic stress. RL1157 (and RL1156, a histidine kinase) did not appear to be a global regulator as its effect seemed limited to RL1155 (and perhaps neighbouring genes). However, as mentioned above, the mutants made in RL1155 and RL1157 had no change in growth or symbiotic phenotype when compared to the wild-type, leaving the real function of these genes and their role in stress response unknown.

It is worth noting here, that at this stage of the investigation the genome sequence of 3841 was incomplete and the proximity of the predicted two-component response regulator gene to the hypothetical gene encoded by RL1155 was only found by chance. However, now the sequence of *R. leguminosarum* is currently undergoing annotation and on its completion, the identification of potential regulators to stress-induced genes based on their proximity will be possible (Chapter 4, section 4.3).

Two mutants were isolated during the screening of two individual Tn*5* mutant libraries. However their growth phenotypes were not based on severe hyper-osmotic stress as was first thought. This result (as well as the data from Chapter 4 and failure to obtain any Tn*5* mutants when searching for regulatory mutants) indicates that the genome of 3841 may contain several copies of the same gene and that the presence of a 'master' regulator that controls stress response is unlikely. As already mentioned, it has already been shown that *Rhizobium* does contain many copies of the same gene, whether homologues or paralogues (Chapter 4). The presence of such genes would give the genome of 3841 a high

level of redundancy, meaning that a mutation in just one gene would not create a change in bacterial growth, as another gene could be present to take the mutant's place. This would explain why no mutants could be isolated that could not grow under severe hyper-osmosis and why the targeted mutations made with the pK19mob vector had little effect on the growth of *R. leguminosarum*.

These results lead research away from the search of a 'master' regulator and instead specific systems were investigated, as the next chapters show.

CHAPTER 6: INVESTIGATION OF TRANSPORTERS INVOLVED IN THE UPTAKE OF OSMOPROTECTANTS

6.1. Introduction

Rhizobia require the capacity to interact with their environment. One function necessary is the ability for bacteria to transfer compounds across their cell membrane; whether into the cell (such as nutrients) or out of the cell (such as waste products). This transfer can be passive, relying on concentration gradients to move compounds across the membrane; or it can be active with compounds being moved across under the organism's control, regardless of any concentration gradient. ATP-binding cassette (ABC) transporters are one example of active transporters and are found in abundance in rhizobia (Konstantinidis & Tiedje, 2004).

One of the stress-induced fusions, pRU843, isolated during this research (Chapter 3) mapped to a region of *R. leguminosarum* DNA that contained a gene with sequence identity to *proV* (Chapter 4). This gene (pRL100079) is from the ProU operon that encodes for the ABC component of a compatible solute ABC transporter (proline/glycine betaine) and homologues of this system are found in many organisms. Not only that, but this plasmid was only strongly induced under of hyper-osmotic stress, as is the ProU operon. pRL100079 and the operon to which it belongs could therefore encode a ProU-like system in 3841 and may be involved in the uptake of compatible solutes under osmotic upshift. A mutation of this gene was generated and although it was shown that the mutant did not have a loss in cell vitality under standard, stressed or in planta conditions (Chapter 5), it could still play the role in osmoprotection that its sequence identity indicates.

There are six ProU-like systems in the 3841 genome and these were studied to determine what solutes they may transport and ascertain what osmoprotectants, if any, *R. leguminosarum* may use under an osmotic upshift.

6.2. Results

6.2.1. Identifying ProU-like Systems (QATs)

As mentioned above, plasmid pRU843 is highly induced by hyper-osmotic stress and contains a genomic region which maps to a region of DNA (pRL100079), which is responsible for encoding a protein with sequence identity to ProV. It was therefore assumed that the ABC transporter that pRL100079 is part of may play a similar role to ProU, i.e. be responsible for the uptake of compatible solutes during osmotic upshift. It was fortunate that this ProU-like operon was found given the random nature of the LB3 library, and as such it was not known if this was the only transporter of its kind in *R. leguminosarum*, or one of many. However, the preliminary sequence of 3841 was recently made available, which made it possible to search for other ProU-like systems within the *R. leguminosarum* genome.

The genes from the already isolated ProU-like system (pRL100079, pRL100080 and pRL100081) were used as templates to search the 3841 preliminary genome for other similar operons. Additionally, ProU-like operons were identified from *S. meliloti* and were also used as templates, in order to increase the chances of finding other ProU-like systems. The search was successful and another five operons (in addition to the one mapped from pRU843) were identified – making six in total.

ProU is responsible for the uptake of proline and glycine betaine in *E. coli* for use as compatible solutes. However, proline has been shown not to have osmoprotective properties in *S. meliloti* (Boncompagni *et al*., 1999) although its dimethylated form, proline betaine does (Miller & Wood, 1996). Proline betaine is a quaternary amine compound (QAC) (a molecule containing a nitrogen bound to 4 functional groups), as are all betaines. For this reason, the 3841 ProU-like operons were all termed putative quaternary amine transporters (QAT) and the six systems were as a consequently named QAT1 to 6 (with the original pRU843 associated QAT being QAT6).

To establish how similar each QAT was to other known ProU-like systems, their SBP, IMP and ABC components were compared to the QAT components from different organisms. The organisms used for these comparisons were; *S. meliloti*, *M. loti*, *A. tumefaciens*, *Brucella melitensis*, *Bacillus Subtilis*, *Pseudomonas aeruginosa*, *Lactococcus lactis*, *E. coli* K-12 and the *Rhizobium* plasmid pNGR234. These organisms were selected as they were either members of the same group of bacteria as *R. leguminosarum* (the α-proteobacteria) or were from bacteria that have had a ProU-like system previously characterised.

Some of these organisms contained incomplete operons (e.g. no SBP component) whilst others had two separate genes for each IMP. Also, *L. lactis* had one gene (LL1451) which encoded for a hybrid of the SBP and IMP components. All of these genes were used during the analysis. The protein sequences for each component were entered into AlignX (InforMax) and phylogentic trees for ABCs (Fig. 6.1), IMPs (Fig 6.2) and SBPs (Fig. 6.3) were generated. The *L. lactis* SBP/IMP hybrid was entered into both the IMP and SBP analysis.

(Sequences for each organism were obtained from the following websites; http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/ – *S. meliloti*; http://www.kazusa.or.jp/rhizobase/ – *M. loti* and the *Rhizobium* plasmid pNGR234; http://turgon.vbi.vt.edu/cgi-bin/docfilter?file=userindex.html – *A. tumefaciens* and *B. melitensis*; http://www.pseudomonas.com/GenomeSearchU.asp – *P. aeruginosa*; http://genolist.pasteur.fr/SubtiList/ – *B. Subtilis*; http://biocyc.org/ECOLI/blast.html – *E. coli* K-12; http://www.expasy.org/tools/blast/?LACLA – *L. lactis*.)

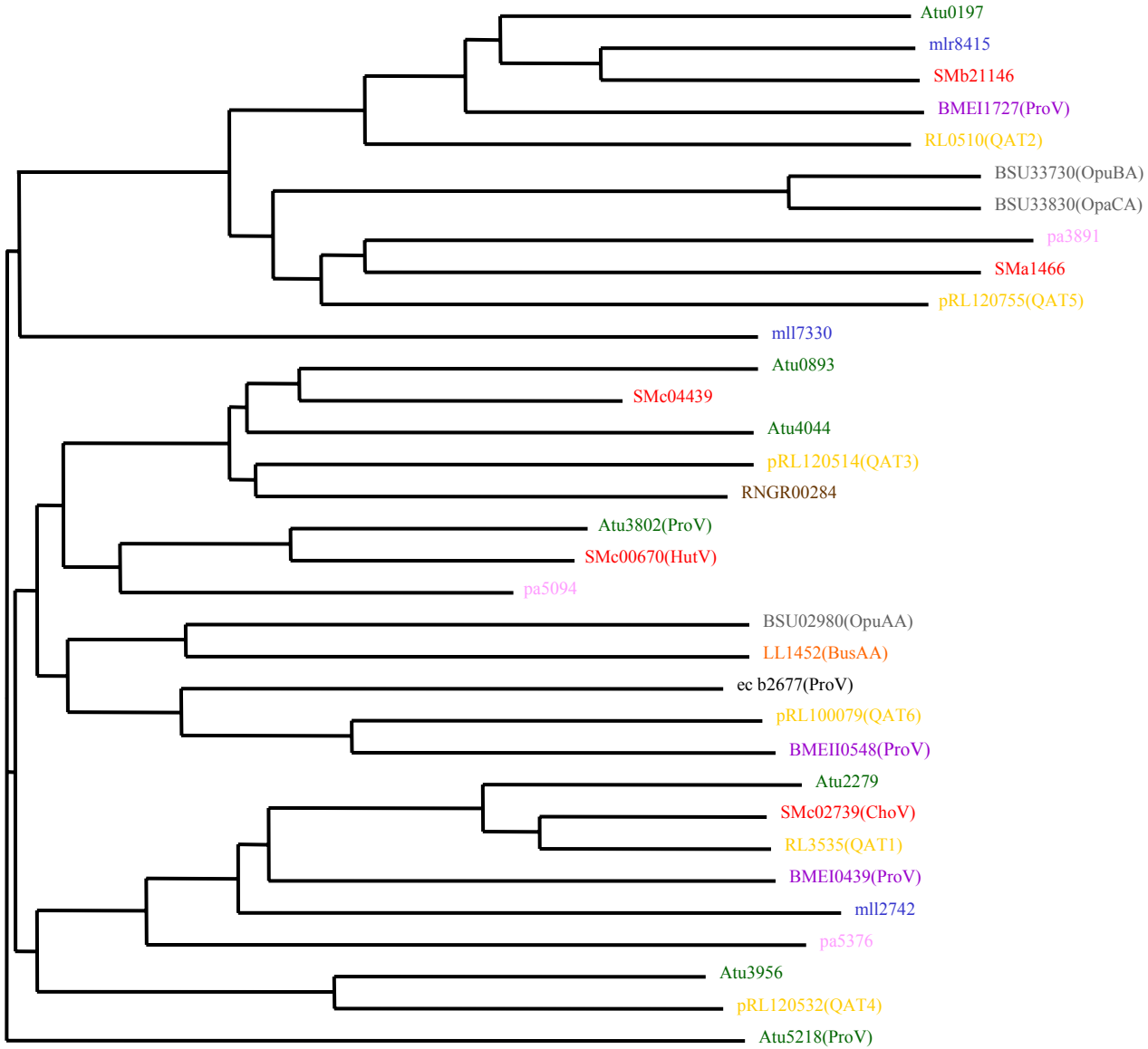


Figure 6.1. Phylogenetic Tree of the ABC Components of QATs from Various Organism. Numbers shown are the official accession numbers (prefixes indicated below) given for each gene, gene name given in brackets when known. However, many of these genes have been named following a genomic investigation and have not necessarily been characterised as having the potential function described.
Key (prefixes in brackets): Yellow (RL) – *R. leguminosarum*, Red (sM)– *S. meliloti*, Green (Atu) – *A. tumefaciens*, Purple (BMEI)– *B. melitensis*, Blue (mll) – *M. loti*, Brown (RNGR) – plasmid pNGR234, Grey (BSU) - *B. Subtilis*, Pink (pa) – *P. aeruginosa*, Orange (LL) – *L. lactis* and Black (ec) – *E. coli*.

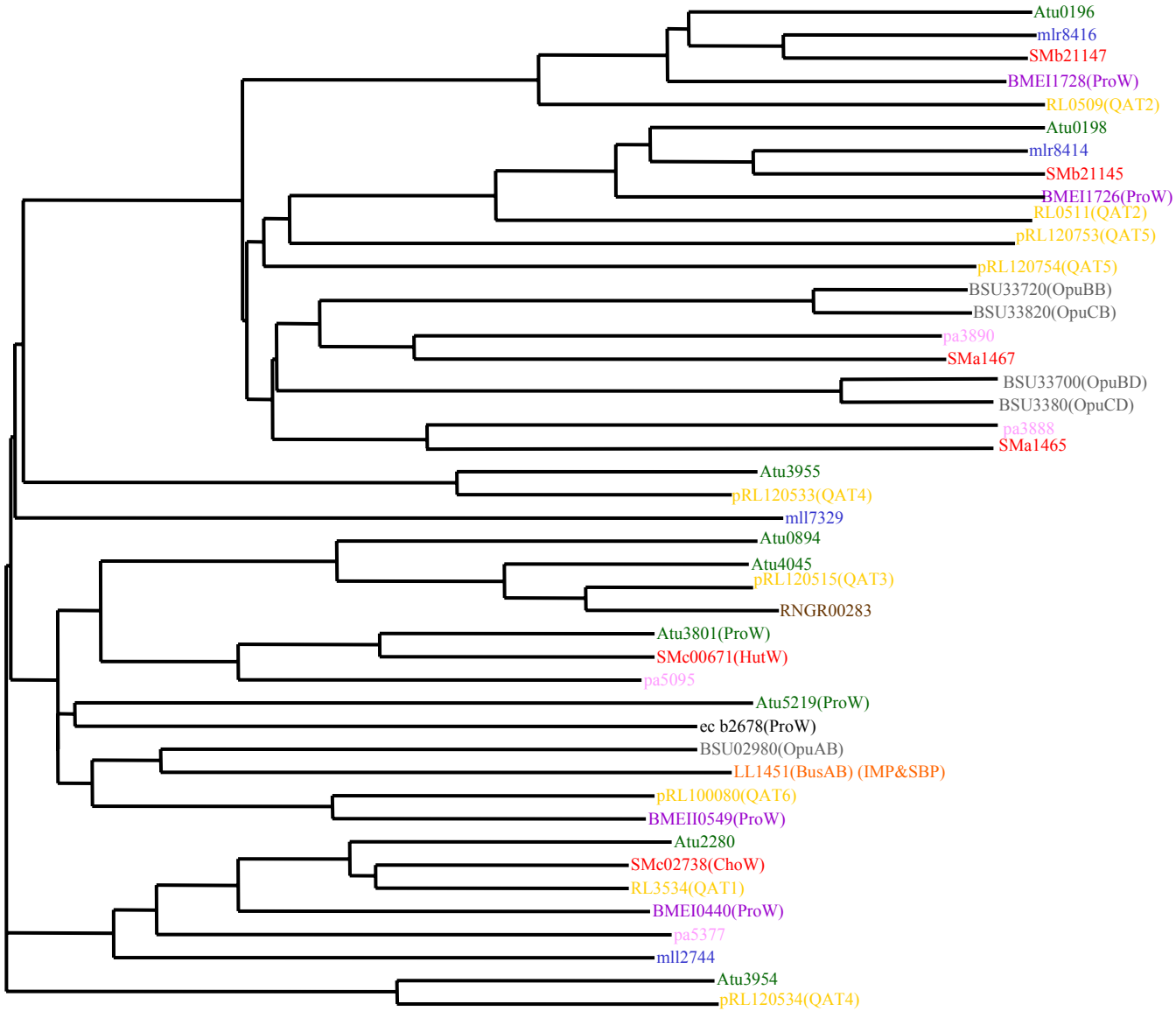


Figure 6.2. Phylogenetic Tree of the IMP Components of QATs from Various Organism. Numbers shown are the official accession numbers (prefixes indicated below) given for each gene, gene name given in brackets when known. However, many of these genes have been named following a genomic investigation and have not necessarily been characterised as having the potential function described.

Key (prefixes in brackets): Yellow (RL) – *R. leguminosarum*, Red (sM)– *S. meliloti*, Green (Atu) – *A. tumefaciens*, Purple (BMEI)– *B. melitensis*, Blue (mll) – *M. loti*, Brown (RNGR) – plasmid pNGR234, Grey (BSU) - *B. Subtilis*, Pink (pa) – *P. aeruginosa*, Orange (LL) – *L. lactis* and Black (ec) – *E. coli*.

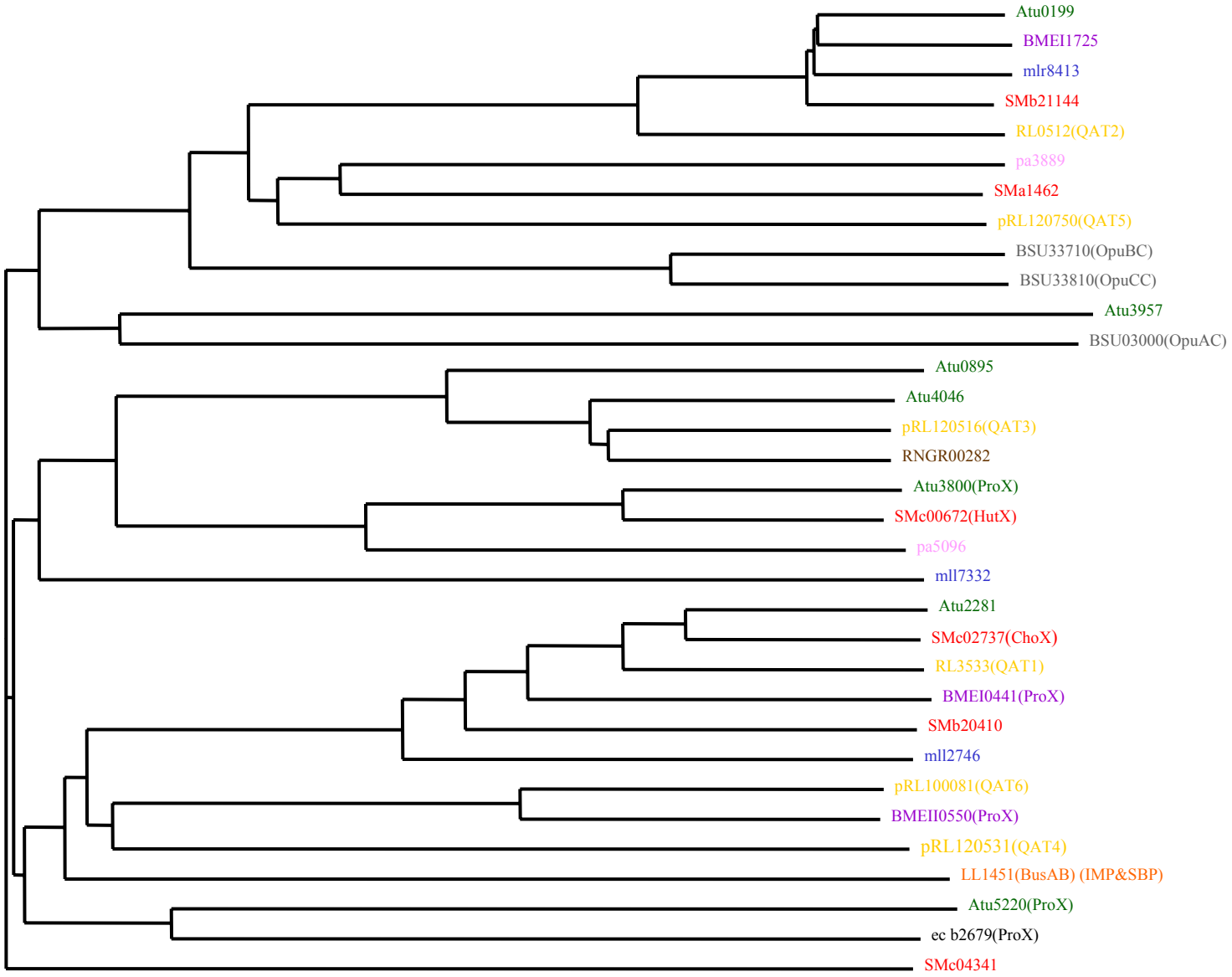


Figure 6.3. Phylogenetic Tree of the SBP Components of QATs from Various Organism. Numbers shown are the official accession numbers (prefixes indicated below) given for each gene, gene name given in brackets when known. However, many of these genes have been named following a genomic investigation and have not necessarily been characterised as having the potential function described.
Key (prefixes in brackets): Yellow (RL) – *R. leguminosarum*, Red (sM) – *S. meliloti*, Green (Atu) – *A. tumefaciens*, Purple (BMEI)– *B. melitensis*, Blue (mll) – *M. loti*, Brown (RNGR) – plasmid pNGR234, Grey (BSU) – *B. Subtilis*, Pink (pa) – *P. aeruginosa*, Orange (LL) – *L. lactis* and Black (ec) – *E. coli*.

As can be seen from Figs. 6.1 to 6.3, two ProU-like systems have already been identified and characterised in *S. meliloti*, the Hut system and the Cho system. The Hut system is responsible for high-affnity histidine uptake and is believed to play a key role in histidine metabolism in *S. meliloti* (Boncompagni *et al*., 2000). Further to this, inhibition studies have shown that Hut can transport proline and proline betaine with high-affinity and also glycine betaine, ectoine and carnitine with low-affinity, although mutant studies have shown that the SBP component of Hut, HutX, does not bind glycine betaine (Boncompagni *et al*., 2000). Hut is induced only by the presence of histidine and not by osmotic upshift, further supporting the belief that Hut is only used in a metabolic role and not in response to hyper-osmosis. The Cho system is highly specific (as shown by inhibition studies) and responsible for high-affinity choline transport. Just like the Hut system, it is only induced by the presence of the solute it transports (i.e. choline) and is not induced by hyper-osmosis (Dupont *et al*., 2004).

As yet, no complete ABC system (ProU-like or otherwise) that responds to osmotic upshift has been fully identified and characterised in rhizobia, although an osmotically-induced SBP component that binds glycine betaine has been isolated (Talibart *et al*., 1990; LeRudulier *et al*., 1991). The only system that has been found to be involved with osmoprotectant uptake, is the (analogous) ProP-like secondary transporter BetS in *S. meliloti* (Boscari *et al*., 2002), although no gene with significant sequence identity to *betS* could be found in the preliminary genomic sequence of 3841.

6.2.2. Isolation and Generation of QAT Mutants

Figures 6.4 – 6.9 show the genomic region of each of the QAT systems.

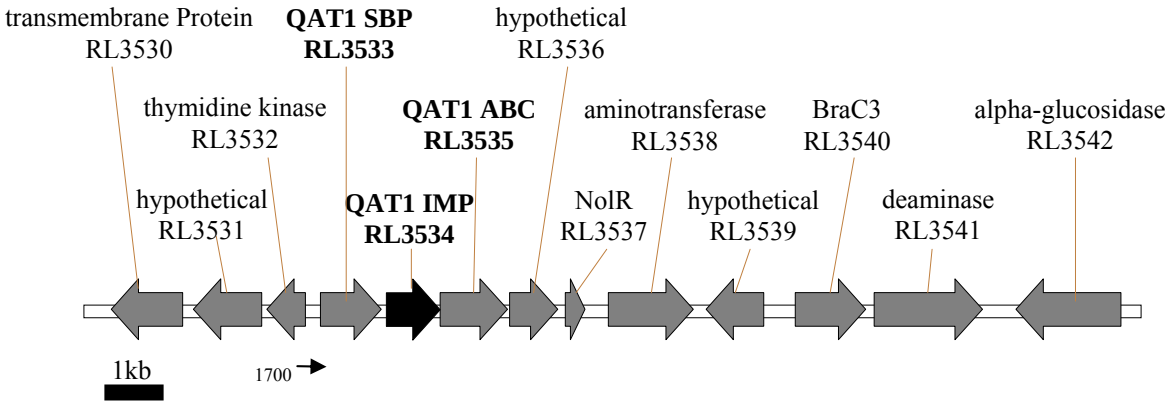


Figure 6.4. Genomic Region Surrounding QAT1 Operon. Arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown. Black arrow shows target gene (Table 6.1) and stick arrow and number shows possible promoter region (section 6.2.4).

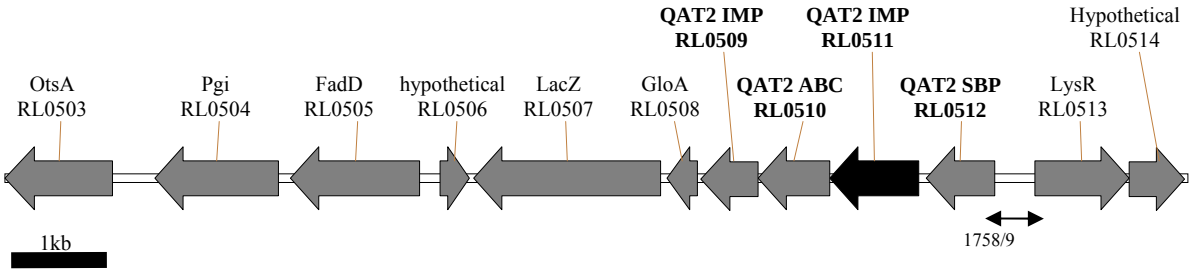


Figure 6.5. Genomic Region Surrounding QAT2 Operon. Arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown. Black arrow shows target gene (Table 6.1) and stick arrows and numbers show possible promoter region (section 6.2.4).

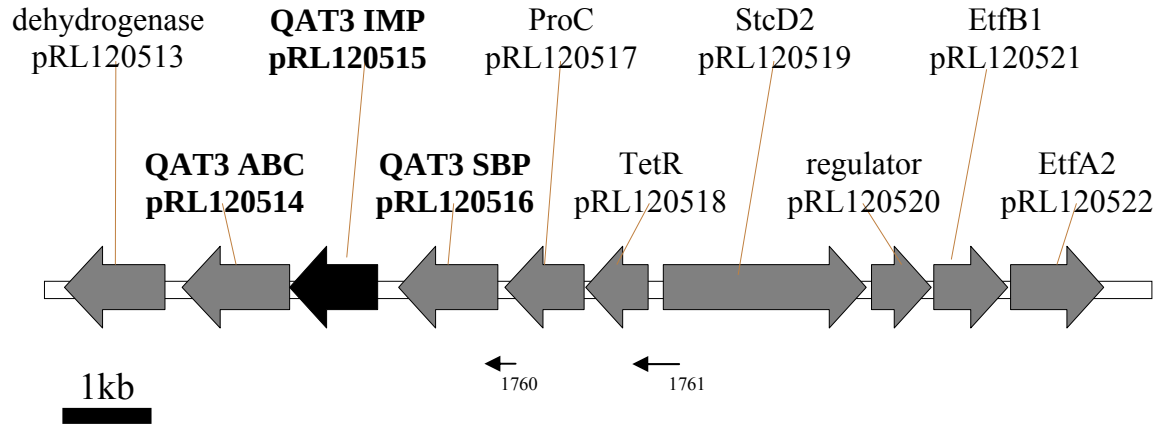


Figure 6.6. Genomic Region Surrounding QAT3 Operon. Arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown. Black arrow shows target gene (Table 6.1) and stick arrows and numbers show possible promoter region (section 6.2.4).

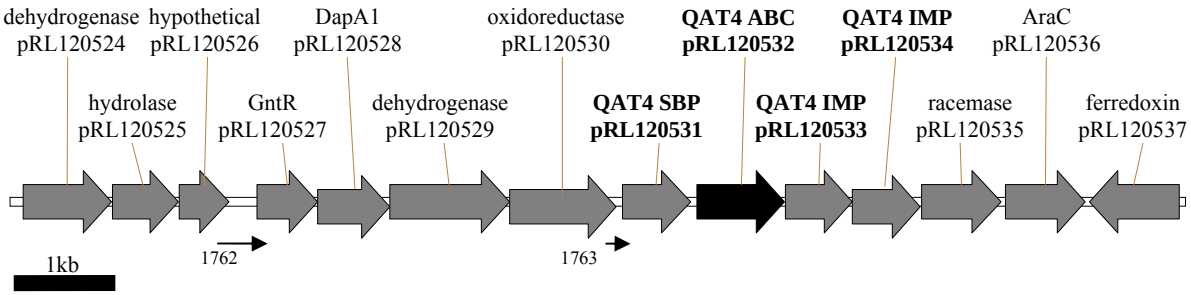


Figure 6.7. Genomic Region Surrounding QAT4 Operon. Arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown. Black arrow shows target gene (Table 6.1) and stick arrows and numbers show possible promoter region (section 6.2.4).

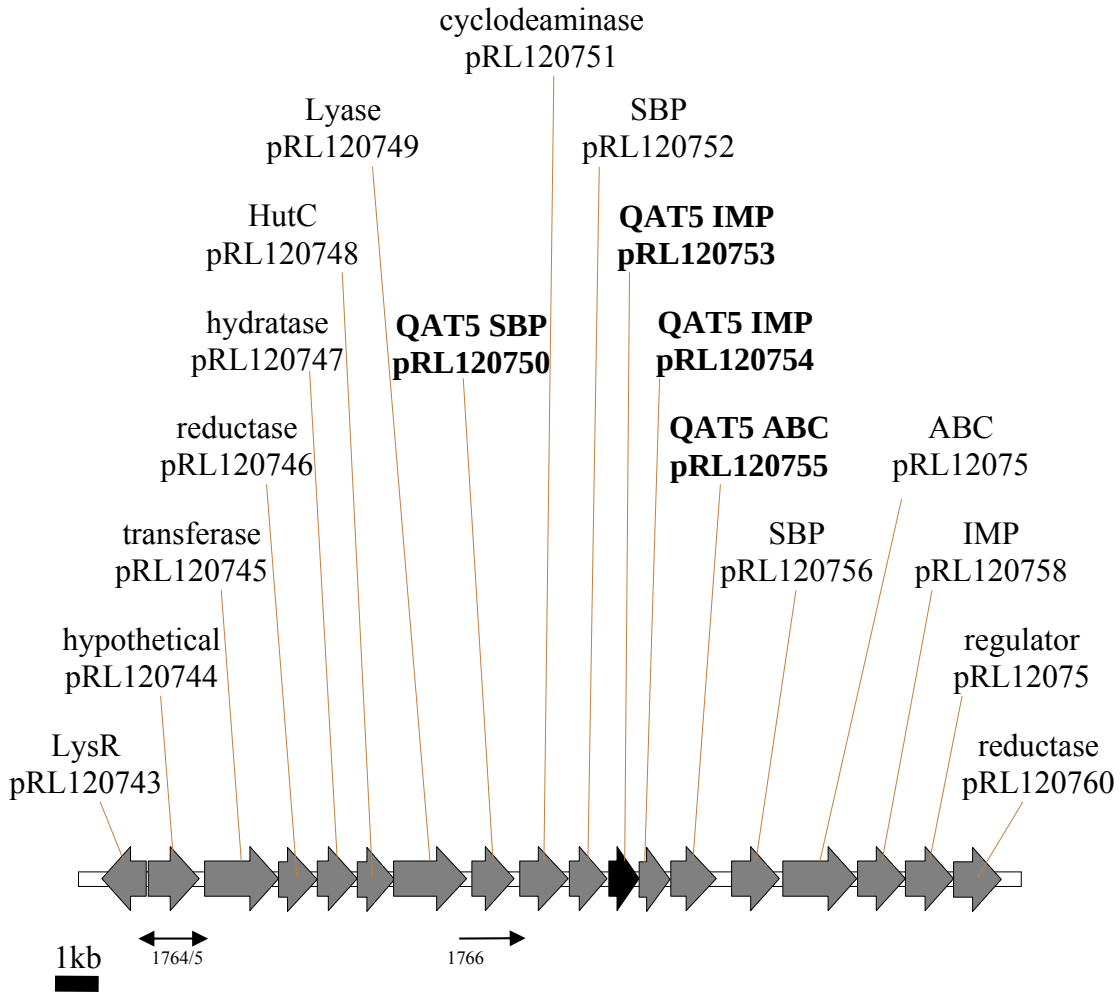


Figure 6.8. Genomic Region Surrounding QAT5 Operon. Arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown. Black arrow shows target gene (Table 6.1) and stick arrows and numbers show possible promoter region (section 6.2.4).

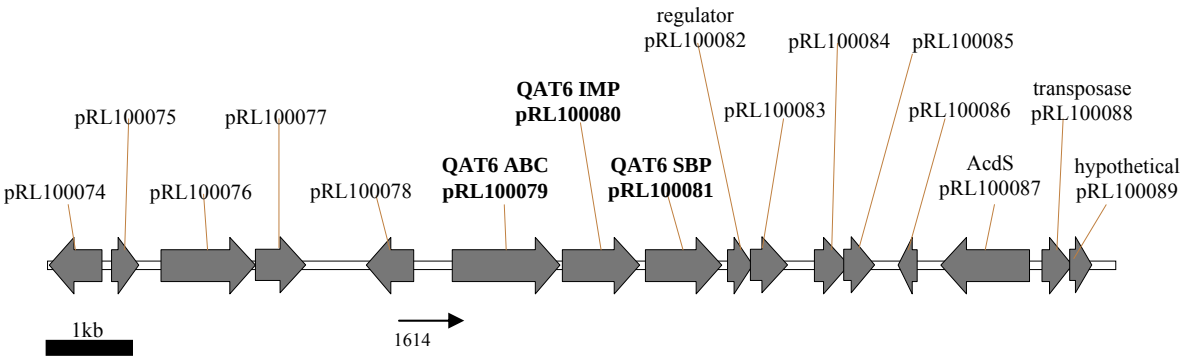


Figure 6.9. Genomic Region Surrounding QAT6 Operon. Arrows show predicted genes and black line shows ~1kb scale. Gene numbers shown are those given by the Sanger Centre. Where appropriate, potential function of the protein encoded for by each gene is also shown. Black arrow shows target gene (Table 6.1) and stick arrow and number shows possible promoter region (section 6.2.4).

Interestingly, the QAT3 and QAT4 operons were in close proximity to each other on the pRL12 plasmid (within 20 predicted genes of each other). Also the QAT5 operon was very interesting, as it was interrupted by two other genes, a predicted cyclodeaminase and another SBP component. Sequence comparisons, homology and phylogeny studies suggest that pRL120750 is the correct SBP for QAT5 and not pRL120752, which is predicted by BLAST analysis to bind to phosphate/phosphonate and not a QAC. However, it is unusual for an ABC operon to be broken up like this and to my knowledge is the first example of such a case (assuming the QAT5 operon is indeed an operon as predicted). It is possible that the second predicted SBP (pRL120752) could also use the ABC transport complex encoded by the rest of the QAT5 operon (pRL120753 – pRL120755) and this would not be the first occurrence of more then one SBP utilising the same ABC complex (Higgins & Ames, 1981); or pRL120752 could be an 'orphan' SBP associated with a different and unrelated ABC complex. The predicted cyclodeaminase encoded by pRL120751 was also interesting, as this enzyme converts ornithine to proline releasing ammonia (Fig. 6.10) (Trovato *et al*., 2001). Whilst rhizobia cannot use proline as a compatible solute, it may be possible that the cyclodeaminase could convert N-methylated ornithine to proline betaine, which can then be used by the bacteria. Alternatively it may convert other amino acids to produce a complex cyclic derivative. Cyclodeaminases are quite uncommon and are limited to specialised-niche bacteria; as such they have been isolated in *A. tumefaciens* and *S. meliloti* and in both organisms have been shown to be beneficial to the bacteria's interaction with plants (Sans *et al*., 1987; Soto *et al*., 1994).

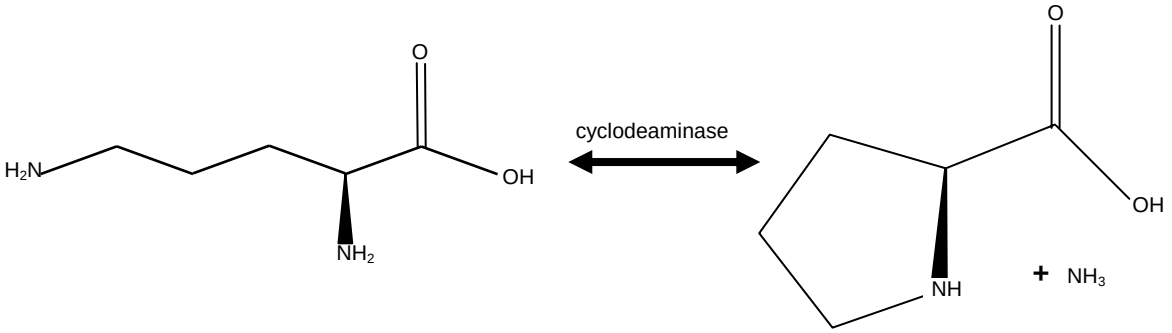


Figure 6.10. Mode of Action of Cyclodeaminase. Ornithine is converted to proline releasing ammonia

The QAT5 operon was also near to a HutC-like gene (pRL120748), which encodes a putative histidine utilization repressor. Hisitidine also contains a ring structure, like proline (Fig. 6.16) and so the cyclodeaminase encoded by pRL120751 may react with histidine as well/instead. The proximity of these two genes near the QAT5 operon suggests that it is not involved in osmoregulation but in a metabolic role, perhaps with histidine; this would not

be the first time a ProU-like system has been linked to a metabolic role with histidine in rhizobia (Boncompagni *et al*., 2000).

The presence of a ProC-like gene (pRL120517) near the QAT3 operon is also interesting, as this encodes for a putative pyrroline-5-carboxylate reductase. This enzyme reduces pyrroline to proline. Whilst rhizobia cannot use proline as a compatible solute, it may be possible that the reductase could convert N-methylated pyrroline to proline betaine, which can then be used by the bacteria. This would suggest that QAT3 may be responsible for uptake of N-methylated pyrroline, which may then be converted into proline betaine.

In order to obtain mutants for each QAT system, a Tn*5* mutant library was used (Downie *et al*., unpublished). This library is believed to cover approximately half of the 3841 genome and a PCR screening technique is used to isolate a mutation in a specific gene (see below). Once the six QAT operons had been identified, genes could be targeted for screening. Target genes were carefully selected so that a mutation would definitely cause the loss of a functioning ABC complex. To accomplish this, the first non-SBP gene in the operon was targeted as it was hoped that this would put the rest of the operon out of frame and so prevent expression of downstream genes. Targeting the first non-SBP gene in the operon would prevent at least one (if not all) of the components in the ABC complex from being encoded and so a mutant would prevent the formation of a functional complex, as Tn*5* insertions do not always effect neighbouring genes (Berg *et al*., 1980).

Primers were then designed to determine if a Tn*5* was present in the targeted gene within the mutant library (Table 6.1). A QAT6 mutant (ABC component) was already available from research performed earlier (RU2192 – Chapter 5, section 5.2.2).

Table 6.1. Target Genes for QAT Mutants. Table shows the genes targeted for each QAT system, what component of the ABC system they encode for and the primer used to isolate them.

Target Gene	QAT	Component	Target Primer
RL3534	1	IMP	RL3534lh
RL0511	2	IMP	RL0511lh
pRL120515	3	IMP	pRL120515rh
pRL120532	4	ABC	pRL120532rh
pRL120753	5	IMP	pRL120753lh

The Tn*5* mutant library used to isolate mutations in QAT genes was made up of 17,280 individual mutants that had been divided into groups of 30, each group being stocked in a well in a 96-well microtitre plate and therefore, six microtitre plates contain the whole library. Furthermore, the DNA from each pool of mutants had been isolated and

stocked in the same way and additionally, each group of DNA was pooled into rows and columns according to the layout of the six microtitre plates (Fig. 6.11).

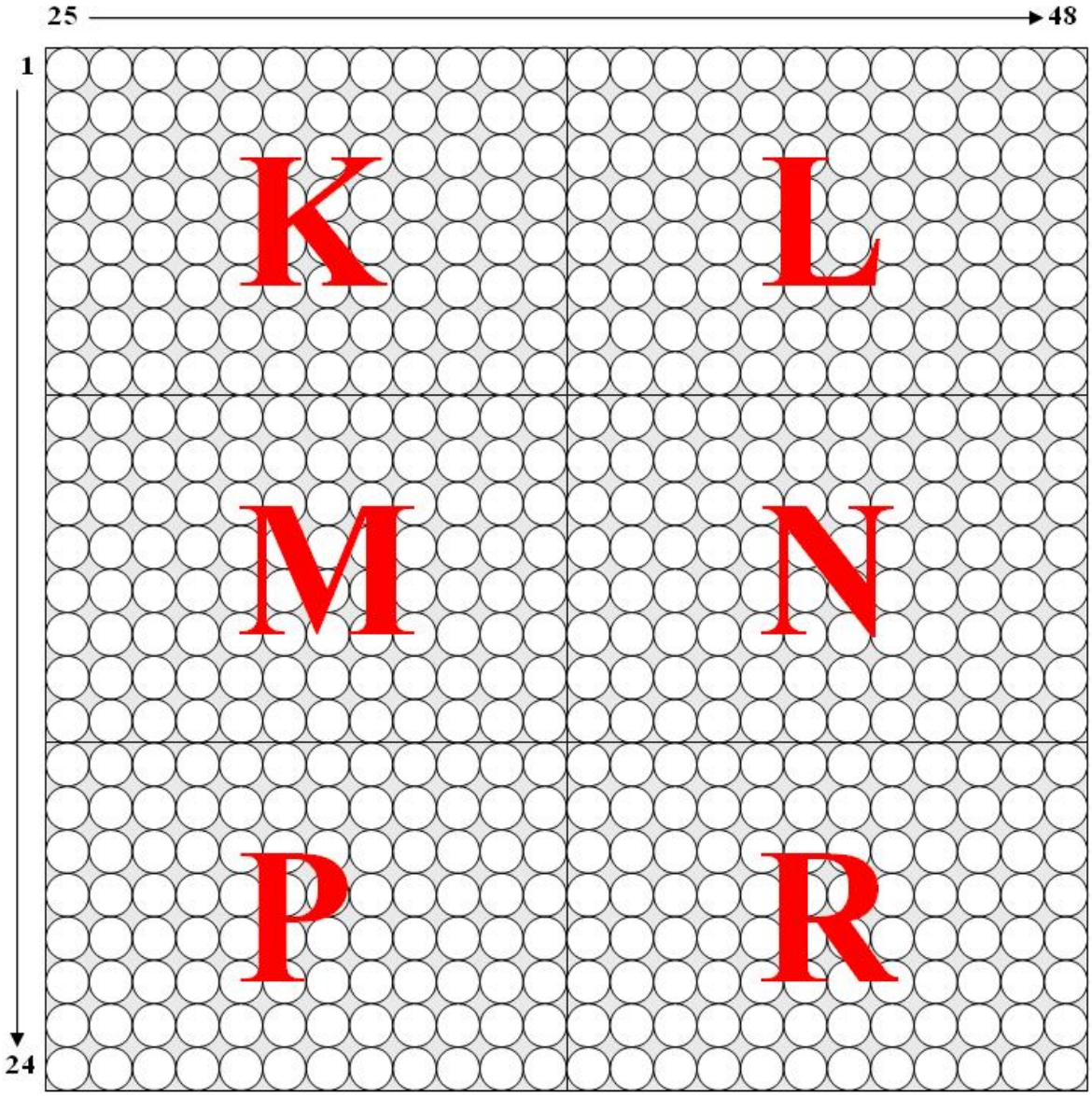


Figure 6.11. Layout of the Six 96-Well Microtitre Plates containing the Tn*5* Mutant Library. Each plate was assigned a letter and was arranged as shown. Each row and column therefore contained 24 wells, each housing 30 Tn*5* mutants.

PCR was then performed on the DNA pools from each row and column, using the target primer and the IS50 Downie primer, which is specific to the end of the Tn*5* and sequences away from the transposon. The target primer was designed so it was upstream (or downstream) and would sequence towards the targeted gene. If a Tn*5* was present in the targeted gene then a PCR product would be made with the target primer and IS50 Downie. Of course, more than one mutant could be present in the mutant library and so therefore, more than one PCR product may be made.

Each well from each microtitre plate was represented in a row and a column and it was this fact that allows a mutant to be identified and isolated (Figs. 6.12 & 6.13)

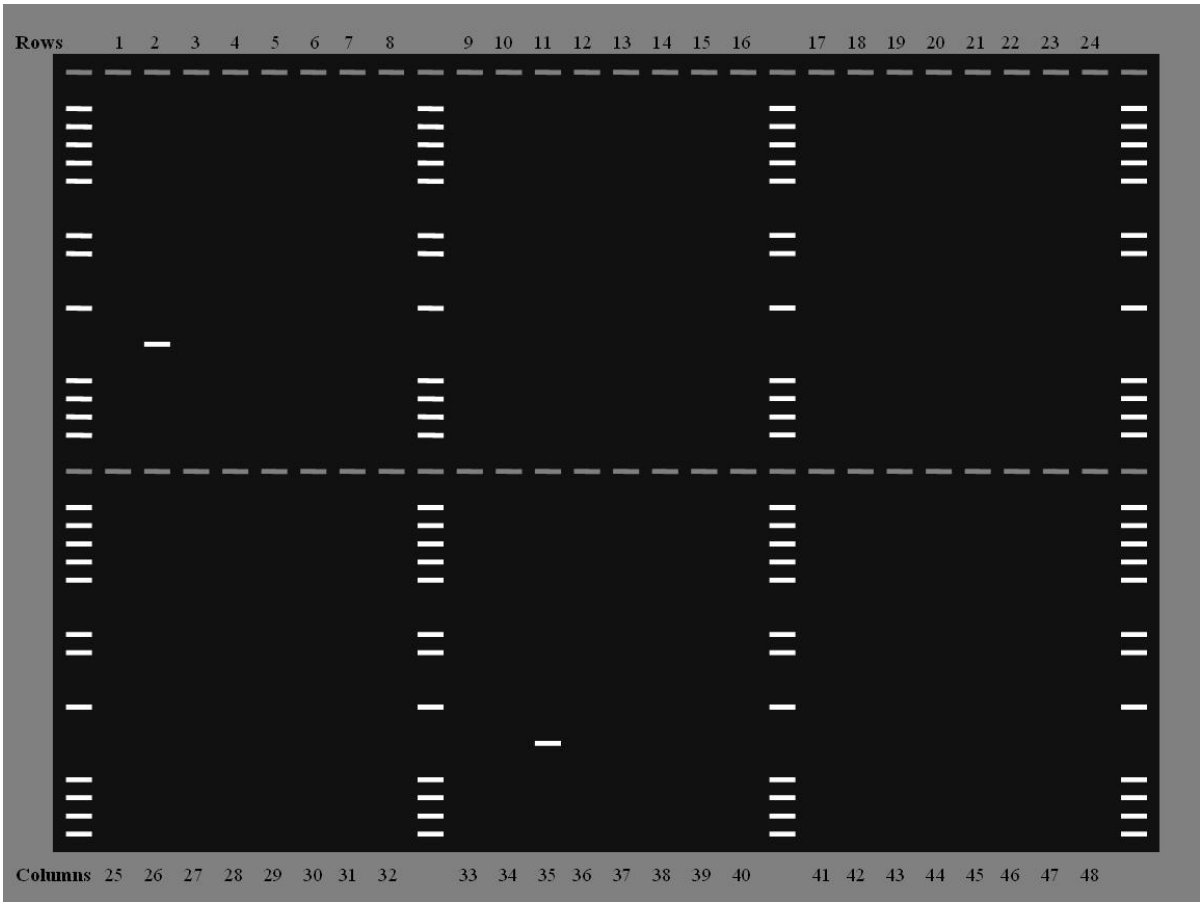


Figure 6.12. Example of Results from PCR of DNA from Rows and Columns. This is an example of the results that can be achieved from performing PCR on the DNA pools from the rows and columns. The top half of this 'gel' represents the rows, whilst the bottom half represents the columns. A PCR product of the same size is present in row 2 and column 35.

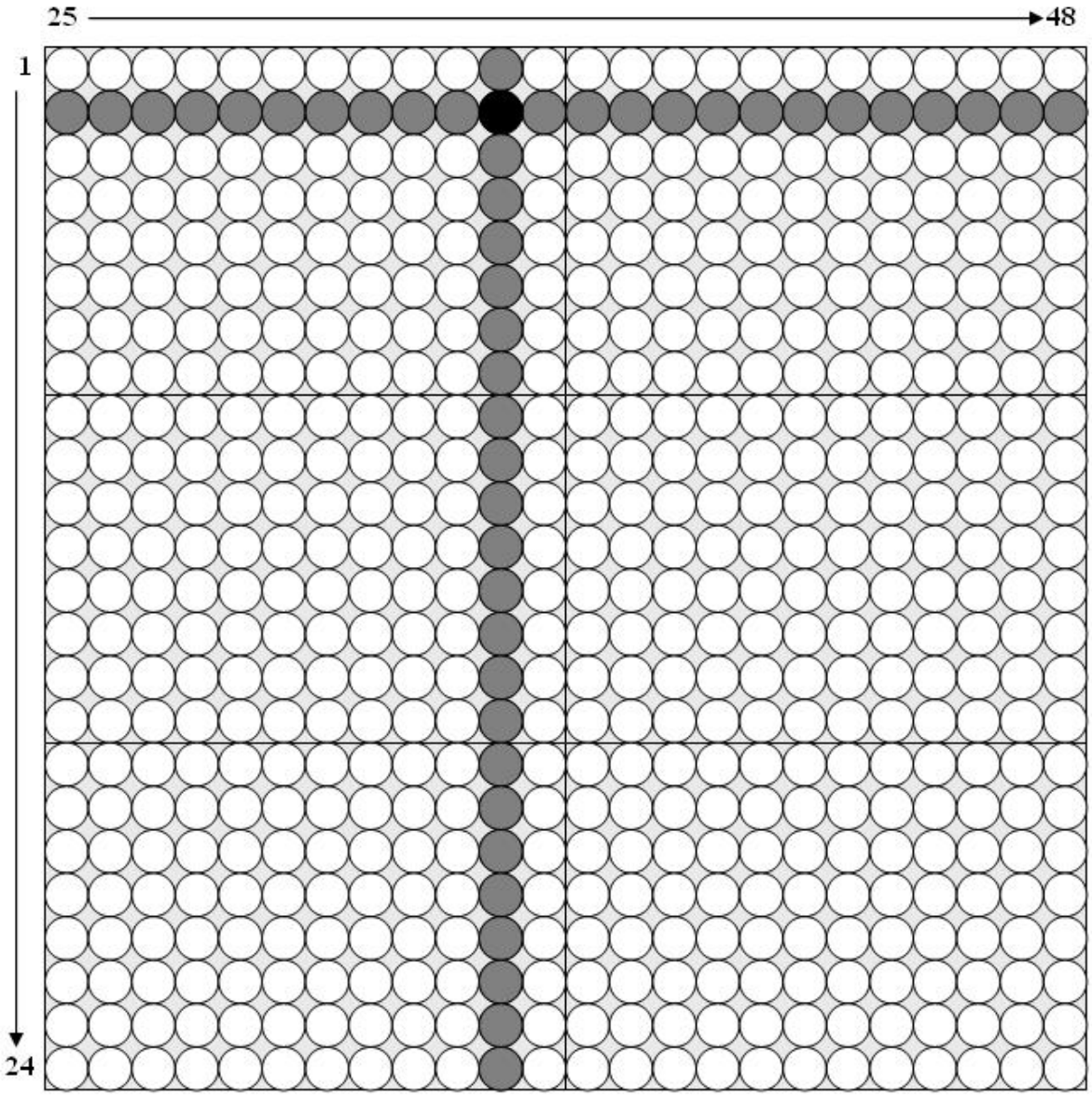


Figure 6.13. Isolating Mutant. From the 'gel' in Figure 6.12, a PCR product of similar size was found with the row 2 DNA pool and the column 35 DNA pool. The well that is present in both row 2 and column 35 is the second row (B) and eleventh column of plate K (Fig. 6.11). Therefore one of the thirty mutants within B11 of plate K contains a Tn*5* in the target gene.

The size of the PCR product obtained is then used to estimate where the Tn*5* is and that it is within the target gene. (e.g. if the product is ~700bp in length, the transposon must be ~700bp from the site where the target primer binds and this can be checked against the sequencing data for the QAT to make sure it is within the targeted gene.)

Once the well containing the correct mutant was identified, a sample of the appropriate bacterial stock was taken from the well, serially diluted and spread onto TY plates to identify which of the thirty mutants from the pool had the transposon insert. Once the bacteria had grown, colony PCR was carried out using the target primer and IS50 Downie until the target mutant was identified. This colony was then isolated, given a

unique strain number and stocked (Table 6.3). The PCR product of the IS50 Downie and target primer was sequenced (using the IS50 Downie primer) and the sequencing data was used to locate where exactly the Tn*5* transposon was in the target gene. This technique allowed mutants in QATs 1, 2 and 5, but not QATs3 and 4 to be isolated.

In order to obtain mutants in QAT3 and QAT4, the pK19mob insertion method (Chapter 5, section 5.2.2) was used. Primers specific to internal regions of pRL120515 and to pRL120352 were designed and PCR was carried out using BIO-X-ACT (Chapter 2, section 2.6.6). Products were then cloned into the pCR® 2.1-TOPO® vector (Chapter 2, section 2.6.4), transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and stocked. Restriction mapping was used to determine in which orientation the products had gone into the pCR® 2.1-TOPO® vector. The internal regions were then digested out of their pCR® 2.1-TOPO® host by *Hin*dIII/*Xba*I double digest, isolated by gel extraction (Chapter 2, section 2.6.2) and ligated in pK19mob that had been digested by the same enzymes (Chapter 2, section 2.6.4). Restriction mapping was used to determine in which orientation the products had gone into the pK19mob vector. Each pK19mob vector containing an internal region was transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and stocked. All of this data is summarised in Table 6.2.

Table 6.2. Target Genes and Primers for the Internal Regions Cloned into the pCR® 2.1-TOPO® Vector and the pK19mob Vector.

Gene	QAT	Primers	pCR® 2.1-TOPO®	pK19mob
pRL120515	3	p718 & p719	pRU1874	pRU1800
pRL120532	4	p793 & p794	pRU1785	pRU1801

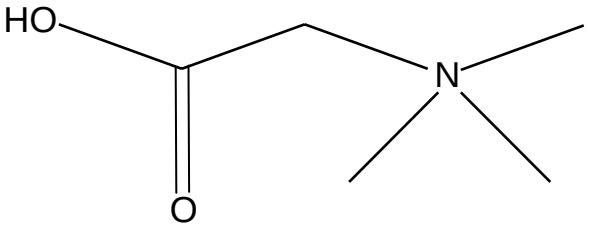
The pK19mob plasmids were then transferred into wild-type *R. leguminosarum* via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7) and the conjugation mixes were spread as serial dilutions (from 10^{-1} to 10^{-3}) on to TY plates containing streptomycin and neomycin. In order to prove the pK19mob vector had inserted correctly and the mutant had been made, colony PCR screening (Chapter 2, section 2.6.6) was used on any colonies that grew on the TY plates. (p839 & pK19/18A and p840 & pK19/18B for QAT3 and p841 & pK19/18A and p842 & pK19/18B for QAT4.) Once a mutant had been confirmed as being correct it was given a strain number and stocked (Table 6.3).

Table 6.3. Strain Numbers of QAT Mutants.

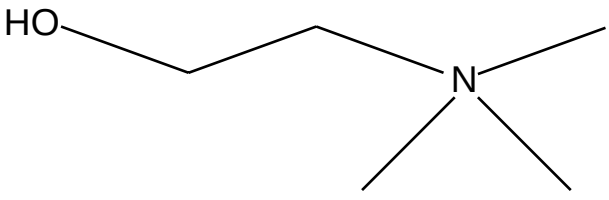
RU2410	Tn*5* transposon within QAT1 IMP gene
RU2411	Tn*5* transposon within QAT2 IMP gene
RU2496	Insertion mutant of QAT3 IMP gene
RU2497	Insertion mutant of QAT4 ABC gene
RU2412	Tn*5* transposon within QAT5 IMP gene
RU2192	Insertion mutant of QAT6 ABC gene

6.2.3. Determination of Solutes that Rescue the Growth of 3841 under Osmotic Upshift

As mentioned above, previous reports have shown that the QACs can function as osmoprotectants in rhizobia and as such can rescue growth rates of bacteria grown under hyper-osmosis. However, as different compounds affect different strains in different ways, and as most studies have been conducted with *S. meliloti*, it was unknown what compounds could act as osmoprotectants in *R. leguminosarum* 3841. To determine this, three QACs were tested, glycine betaine, choline (a precursor of glycine betaine) and carnitine, alongside trehalose and ectoine, which although not QACs have also been reported as acting as osmoprotectants in rhizobia (Breedveld *et al*., 1990; Talibart *et al*., 1994). The properties of proline betaine as osmoprotectant have been well documented, however it is an extremely difficult compound to acquire and so proline was used in its place. As previously mentioned, it is known that proline does not function as an osmoprotectant in rhizobia, so its use served as a control. Structures of the tested compounds are shown in Figure 6.14.

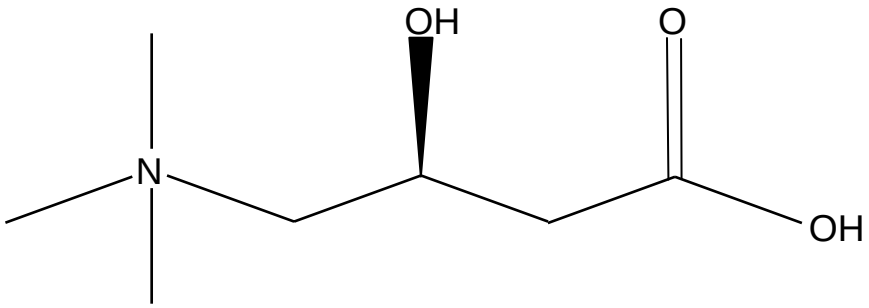


Glycine Betaine – $C_5H_{11}NO_2$

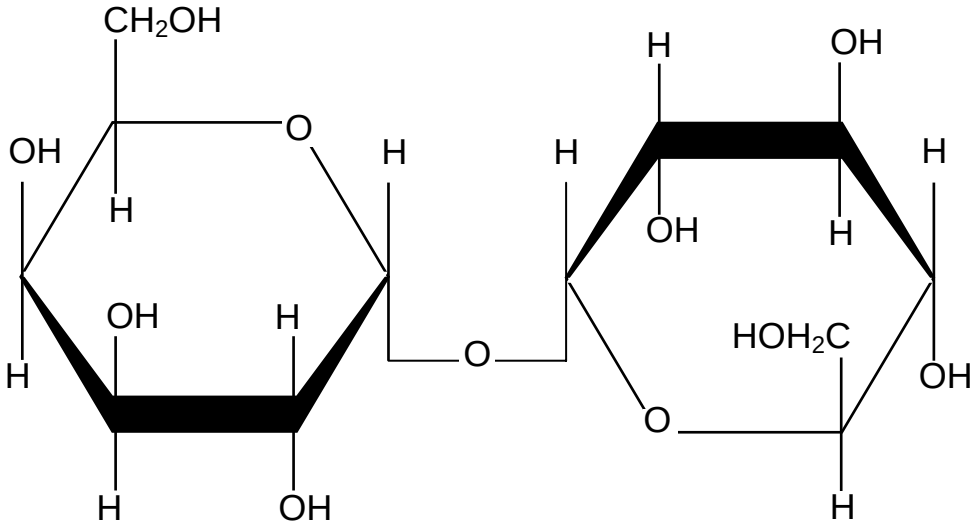


Choline – $C_5H_{14}NO$

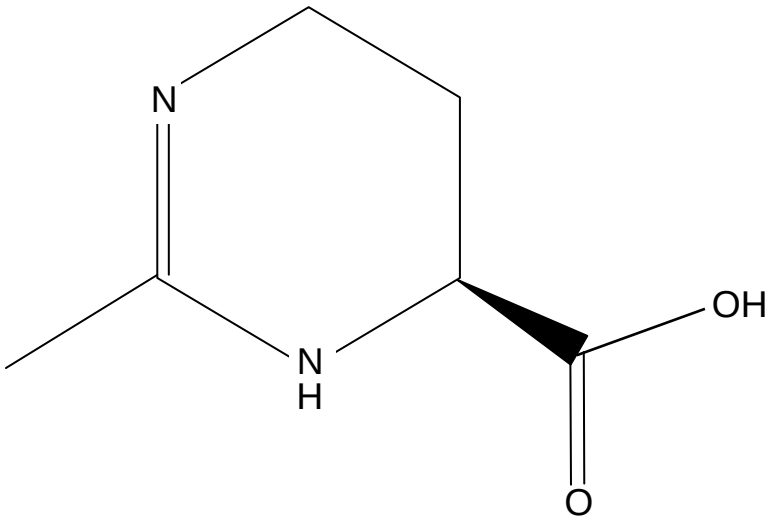
Figure 6.14. Compounds with which Mutants were Tested.



Carnitine – $C_7H_{15}NO_3$

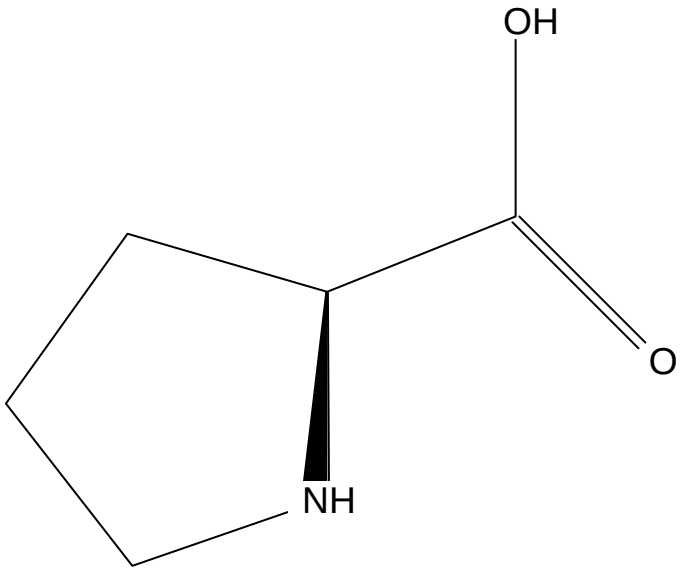


Trehalose – $C_{12}H_{22}O_{11}$



Ectoine – $C_6H_{10}N_2O_2$

Figure 6.14. (cont). Compounds with which Mutants were Tested.



Proline – $C_5H_9NO_2$

Figure 6.14. (cont.) Compounds with which Mutants were Tested.

A 3841 culture was grown up on a TY slope, washed in TY broth and resuspended to an OD_{600} = ~1. AMS (10mM glc, 10mM NH_4), with and without added osmoprotectant, was then inoculated (10µl of this suspension into 10ml of media) and cultures were left to grow. OD_{600} readings of each culture were taken at regular intervals over ~3 days and from these the mean generation times (MGT) of 3841 under each condition were calculated (Table 6.4). Osmoprotectants were present at 1mM, the concentration known to give maximum reduction of doubling time in rhizobia (Bernard *et al*. 1986).

Table 6.4. MGTs of 3841 Grown in AMS and AMS + 125mM NaCl with and without 1mM of Each Osmoprotectant. The percentage of shows the difference in growth rate between strains grown unstressed and those grown in the presence of sucrose.

Osmoprotectant	AMS	AMS + 125mM NaCl	%
None	3.9 ± 0.2	4.9 ± 0.8	125
Carnitine	3.8 ± 0.3	4.6 ± 0.5	120
Choline	3.8 ± 0.3	9.5 ± 1.4	251
Ectoine	3.7 ± 0.2	6.6 ± 1.2	182
Glycine Betaine	3.8 ± 0.2	6.2 ± 1.1	165
Proline	3.8 ± 0.2	5.3 ± 0.7	141
Trehalose	3.8 ± 0.3	5.4 ± 0.7	141

These data show that the addition of any of these compounds effect the growth of 3841 significantly when hyper-osmotically stressed. In fact it appears that the addition of choline in tandem with the osmotic upshift caused a further reduction in growth rate. One

report has showed that the addition of glycine betaine or choline did not rescue growth of any of the biovars of *R. leguminosarum* that were grown under hyper-osmosis (Boncompagni *et al*., 1999); these data confirm this and also includes other osmoprotectants. The same report also stated that although the addition of glycine betaine did not restore growth to *R. leguminosarum* grown under hyper-osmosis, cells showed stimulated glycine betaine uptake under these conditions. This led to the investigation of the QAT systems and their induction.

6.2.4. Induction of the QAT Operons

Before the QAT mutants were tested, as to what solute they could and could not transport, it was decided to ascertain what could cause their induction and so give an indication as to what they may transport. In order to monitor induction of the QATs, the promoters of each operon had to be identified. This was done by examining the surrounding area of each operon and estimating where a promoter could go. One region was chosen for QAT1, one for QAT2, two for QAT3, two for QAT4, two for QAT5 and one for QAT6 (indicated by the stick arrows on Figs. 6.4 – 6.9) and primers were designed that would amplify these region of DNA via PCR. Additionally a LysR-type regulator was found next to a potential promoter region for QAT2 and for QAT5, so PCR primers were designed that would amplify the promoter region for these genes via PCR (also indicated on Figs. 6.5 and 6.8 – appear as double headed arrows as the regions overlap those previous described above) (Table 6.5). This would show what may cause the induction of these putative regulators and so would show if they were involved in with either of the QATs.

The internal regions were amplified by PCR using BIO-X-ACT (Chapter 2, section 2.6.6) and the products were then cloned into pRU1097/D-TOPO® (Chapter 2, section 2.6.4). pRU1097/D-TOPO® contains a promoterless GFP mut 3.1 reporter gene that is used to monitor the induction of the potential promoter within the DNA insert. The pRU1097 vector was constructed 'in house' before being sent to Invitrogen where it was 'TOPO-adapted', which allows the insertion of PCR products and ligation of the vector without the need for restriction enzymes or ligases (Mauchline *et al*., in prep). One difference between pCR® 2.1-TOPO® and pRU1097/D-TOPO® is that the latter has been adapted so that PCR products can only be inserted in one orientation. To facilitate this, the 5' end of forward primer is capped with CACC, which can only bind to one end of pRU1097/D-TOPO®, forcing insertion in one orientation.

The pRU1097/D-TOPO® reactions were transformed into *E. coli* DH5α cells (Chapter 2, section 2.6.5) and spread onto LA containing gentamycin. Colonies were then

PCR screened to make sure that they contained the pRU1097/D-TOPO® vector with the relevant insert. This was done with primers p496 and p564, which bind to pRU1097/D-TOPO® either side of the insertion site. A self-ligated plasmid has a PCR product size of ~400bp, whereas it would be 400bp plus the size of any insert (Fig. 6.15). Colonies that produced a correctly sized PCR product were then confirmed as containing the right insert by restriction digests on the plasmids (Table 6.5). The plasmids were then transferred into wild-type *R. leguminosarum* via a tri-parental conjugation, using the helper plasmid pRK2013 (contained in *E. coli* 803) (Chapter 2, section 2.7) and the resulting strains were stocked.

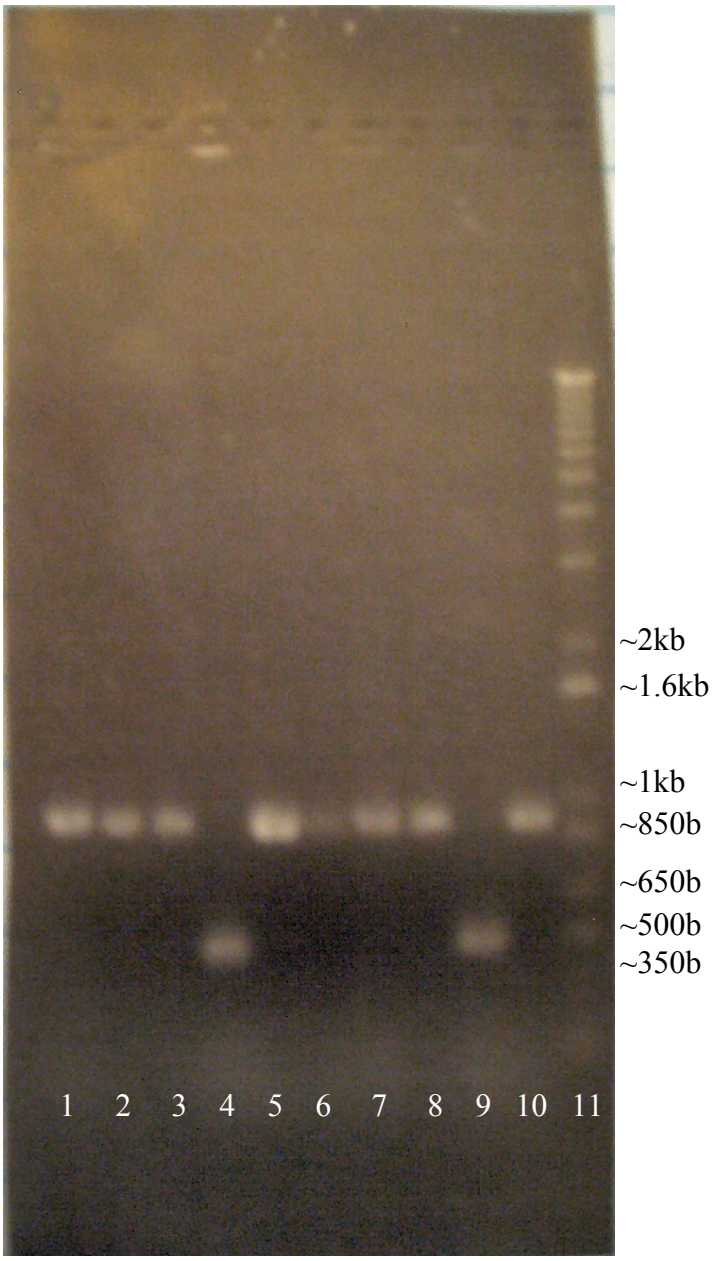


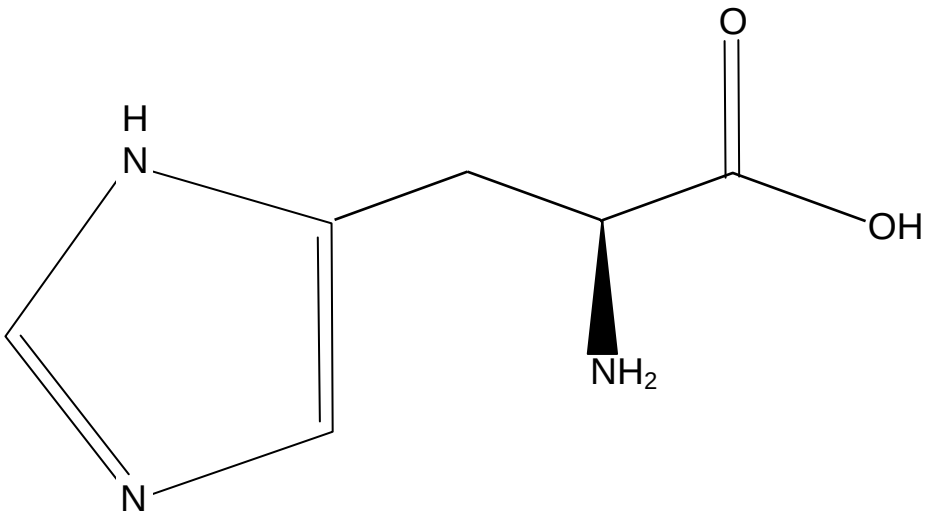
Figure 6.15. Colony PCR Screen. The 1kb ladder in lane 11 shows the PCR products in lanes 1 – 3, 5 – 8 and 10 are ~900bp, which is ~500bp bigger than the PCR products in lanes 4 and 9. This shows that 80% of the colonies tested contained the pRU1097/D-TOPO® vector with the relevant insert.

Table 6.5. Plasmid Numbers for the Regions Cloned into the pRU1097/D-TOPO® Vector. For clarification, in the cases where more then one possible promoter region was present the gene the region precedes is shown, plasmids numbers are also shown in Figs. 6.4 – 6.9.

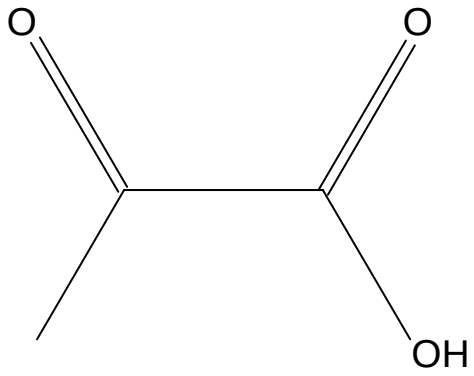
QAT	Primers	Plasmid	Gene
6	p623 &p624	pRU1614	pRL100079
1	p663 & p664	pRU1700	RL3533
2	p694 & p695	pRU1758	RL0512
2 (LysR)	p696 & p697	pRU1759	RL0513
3	p698 & p699	pRU1760	pRL120516
3	p700 & p701	pRU1761	pRL120518
4	p702 & p703	pRU1762	pRL120527
4	p704 & p705	pRU1763	pRL120531
5	p706 & p707	pRU1764	pRL120745
5 (LysR)	p708 & p709	pRU1765	pRL120743
5	p710 & p711	pRU1766	pRL120751

The promoter probes were then tested with various compounds/conditions to determine if any caused induction of the QATs. As the QATs share homology with the ProU system it was assumed that one may transport glycine betaine; however, it is known that some QAT systems are induced by (and responsible for the transport of) histidine and choline in rhizobia (Boncompagni *et al*., 2000; Dupont *et al*., 2004). It is also known from the studies conducted with pRU843 that QAT6 is induced by hyper-osmotic stress (Chapters 3 & 4) and so these compounds/conditions were also tested.

Each promoter probe was grown up on a TY slope before being resuspended and washed in AMS. A 10µl aliquot of the resuspension was then inoculated in 400µl of AMS (20mM pyruvate, 10mM NH4). The AMS then had each of the following individually added; 10mM glc, 10mM glycine betaine, 10mM choline, 10mM histidine (Fig. 6.16) and 100mM NaCl. One batch had no additions, as a control. Samples were then left to grow in 48-well microtitre plate for three days (it took this long for cells to grow to a reasonable OD_{600}) shaking at ~150rpm. Each well in the microtitre plate had a sterile glass bead added to help prevent cells from agglutinating. Microtitre plates were then read in a plate reader to record the relative fluorescence (485nm excitation; 510nm emission) and the OD_{600} of each sample (after first removing the glass bead). The relative fluorescence for each sample was then divided by its OD_{600} value to give a specific fluorescence. (Readings were also taken from an uninoculated sample, from samples inoculated with a self-ligated pRU1097/D-TOPO® vector and from a pRU1097/D-TOPO® vector with a 'stuffer' fragment inserted. These reading served as 'blanks' and were taken into account in these calculations). This experiment was carried out in triplicate and results are shown in Table 6.6.



Histidine – $C_6H_9N_3O_2$



Pyruvate – $C_3H_4O_3$

Figure 6.16. Compounds with which Promoter Probes were Tested. Compounds not present have been previously shown (Fig. 6.14).

Table 6.6. Relative Fluorescence for Promoter Probes under Various Conditions. Relative fluorescence values (Fluorescence/OD_{600}) given for pyruvate. Other values are x-fold inductions compared to pyruvate (P. fold) values. n/d = no fluorescence detected compared to 'blank' readings. If n/d recorded for pyruvate then the specific fluorescence is shown for that probe in other media. All probes grown in AMS (20mM pyruvate, 10mM NH_4) with 10mM of each solute added (100mM with NaCl).

Plasmid	QAT	Pyruvate	P. Fold	NaCl	G. Betaine	Choline	Histidine	Glucose
pRU1700	QAT1	8649	1.00	2.4	7.6	8.7	4.1	1.8
pRU1758	QAT2	22227	1.00	15.8	0.6	1.1	0.6	0.9
pRU1759	QAT2 (LysR)	9753	1.00	13.1	1.4	1.0	2.9	0.4
pRU1760	QAT3	n/d	1.00	243792	3652	1858	11664	n/d
pRU1761	QAT3	19157	1.00	n/d	1.6	0.8	0.8	1.5
pRU1762	QAT4	37871	1.00	8.1	1.4	0.8	2.3	1.6
pRU1763	QAT4	3773	1.00	19.5	0.9	0.9	3.7	n/d
pRU1764	QAT5	5946	1.00	53.5	1.6	1.0	8.5	2.4
pRU1765	QAT5 (LysR)	5724	1.00	n/d	0.2	1.6	1.0	0.3
pRU1766	QAT5	n/d	1.00	n/d	n/d	n/d	1248	153
pRU1614	QAT6	18951	1.00	9.2	1.9	1.2	2.7	2.0

From the promoter probe data it appears as though QAT1 may be a choline and/or glycine betaine transporter, as pRU1700 showed a ~8-fold induction when grown with these two compounds (Table 6.6). Homology studies (Figs. 6.1 – 6.3) indicate that out of all the *R. leguminosarum* QATs, QAT1 was most similar to the Cho system from *S. meliloti* (compared to all the other ProU-like systems studied), which supports the promoter probe data; however the Cho system in *S. meliloti* does not transport glycine betaine.

Both the QAT2 and the QAT2 (LysR) probes (pRU11758 and pRU1759 respectively) showed similar induction on NaCl (~15-fold), which suggests that the LysR may be the regulator of QAT2 and that they are induced by NaCl. Interestingly the LysR gene also showed a modest induction with histidine although QAT2 itself did not (Table 6.6). This suggests that the LysR-like gene may have another function (or regulate other genes) related to histidine, but not associated with QAT2.

Of the two QAT3 promoters tested, pRU1760 appears to contain the QAT3 promoter, whereas pRU1761 seems more likely to contain the promoter for the TetR-like regulator (pRL120518). This is inferred from the fact that very little induction was seen with pRU1761 compared to the pyruvate values, indicating it is constitutively expressed (as are most regulators); whilst pRU1760 was highly induced on NaCl.

Of the two QAT4 promoters tested, pRU1763 appears to contain the QAT4 promoter, whereas pRU1762 seems more likely to contain the promoter for the GntR-like regulator (pRL120527). As with the TetR-like gene near the QAT3 operon, this is inferred from the fact that very little induction was seen with pRU1762 compared to the pyruvate values, indicating it is constitutively expressed (as are most regulators); whilst pRU1763 was highly induced on NaCl.

The pRU1097/D-TOPO® data indicated that pRU1765 and pRU1766 did not contain a promoter, at least not one that was significantly induced under the conditions tested as little fluorescence was seem in any media (Table 6.6). Data from pRU1764 indicates that QAT5 may transport histidine, as an 8-fold induction was observed when grown with this compound, although this was modest compared induction seen with NaCl (~55-fold) (Table 6.6). If the promoter within pRU1764 is for QAT5 then it would activate the transcription of the HutC-like gene (a histidine utilization repressor) upstream of the QAT5, a gene with obvious links to histidine. However, QAT3 showed the closest identity to the Hut system in *S. meliloti* compared to all the other ProU-like systems studied and not QAT5 (Figs. 6.1 – 6.3), but neither pRU1760 or pRU1761 showed significant induction with histidine present.

QAT6 showed (pRU1614) showed a ~9-fold induction only when grown with NaCl, which came as no surprise as pRU843 had previously been characterised as inducing under hyper-osmotic conditions (Chapter 3).

This data indicates that five of the six QATs are induced by hyper-osmosis, in keeping with their identity to ProU, whilst QAT1 did not. However, from the induction patterns obtained with QAT1 appears to be more like the Cho systems, which would suggest it plays a role cell metabolism and not osmoprotection, hence it does not induce on osmotic upshift. The promoter probe data was collected from only one set of experiments and more repeats are clearly needed to confirm the recorded values; however this was not possible due to the time restrictions imposed by this project.

6.2.5. Uptake Assays with QAT Mutants

Before transport assays were carried out on the QAT mutants, uptakes were first measured on the wild-type 3841. During this set of experiments, I was fortunate enough to acquire a limited amount of ^{14}C proline betaine so that its uptake could also be measured alongside that of choline, histidine and glycine betaine. *R. leguminosarum* was grown up overnight in AMS, AMS + 100µM of the solute of interest (i.e. for choline transport, the solute was choline, etc), AMS + 100mM NaCl and AMS + 100µM solute + 100mM NaCl. (All AMS had 10mM glc and 10mM NH_4 as the carbon and nitrogen source.) Uptakes assays were carried out as described (Chapter 2, section 2.12) with 100mM NaCl being added to the RMS if cells were grown under those conditions, for each of the four solutes (Fig. 6.17). Whilst 100µM of solute may be irrelevant in the presence of 10mM glucose, it was all that was available for the proline betaine assays and so the same amount was used with choline, histidine and glycine betaine in order to allow comparisons between each set of data.

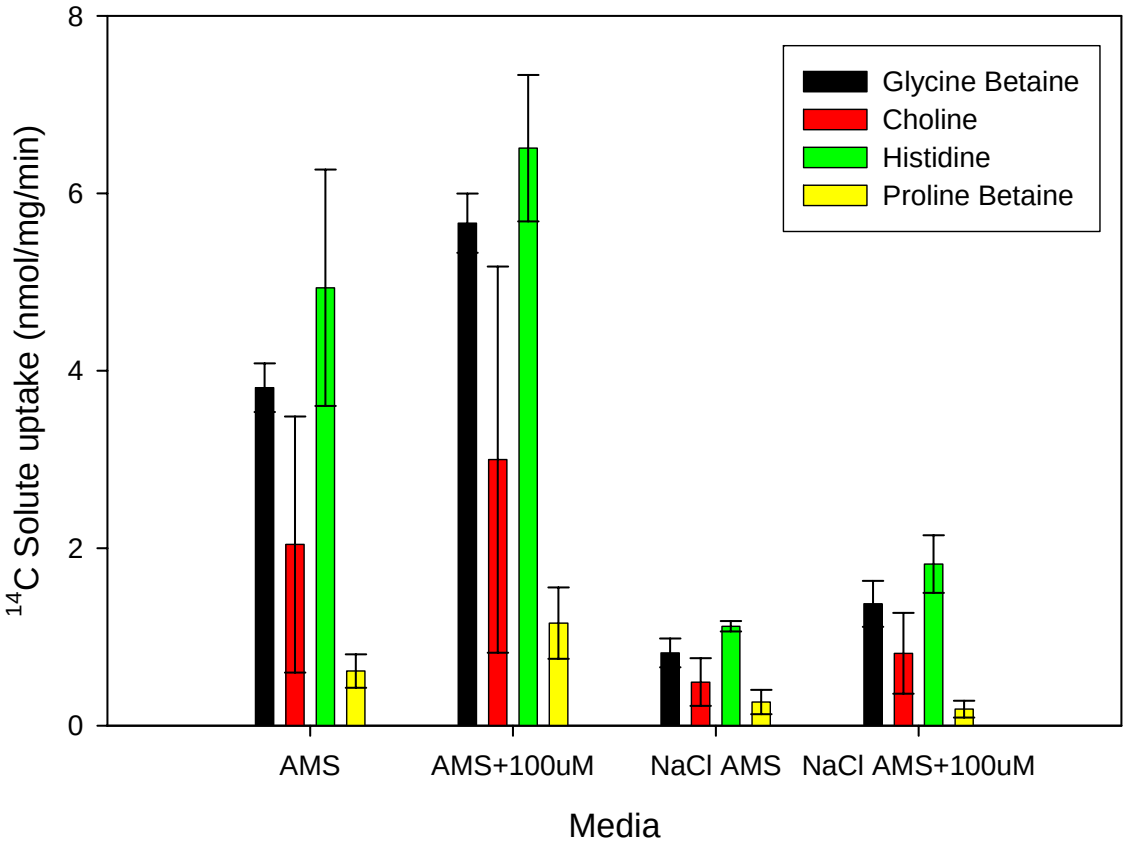


Figure 6.17. Uptake of Different Solutes with 3841. 3841 was grown up in AMS (10mM glc, 10mM NH_4) overnight plus 100µM of solute or 100mM NaCl or both or neither, as indicated. Mean uptake of each solute, as indicated, per minute, over four minutes (five minutes for proline betaine). Mean results of three experiments with standard error values. Proline betaines data is from Fig. 6.18.

Allowing *R. leguminosarum* to grow overnight in the presence of the solute of interest, increased uptake rates of that solute by approximately 30 – 50% (~90% for proline betaine) with all of the tested solutes. Interestingly, the presence of an osmotic upshift appeared to inhibit the transport of these solutes (to ~20 – 25% of unstressed values), contradicting most reports regarding the uptake of osmoprotectants in rhizobia (Bernard *et al*., 1986; Boncompagni *et al*., 1999). Dupont *et al*. (2004) reported that the Cho system is inhibited by the presence of hyper-osmosis generated by NaCl, however the addition of choline to the media restored transport rates to that seen if NaCl was not present. Whilst uptake rates in the presence of NaCl increased on the addition of solute (by ~60% - 70%) (apart from with proline betaine which decreased by ~30%), these rates were still only approximately 35 – 40% of those grown without stress.

In spite of this, the QAT mutants and their uptakes were examined. Unfortunately RU2496 (QAT3 mutant) and RU2497 (QAT4 mutant) were not available at this stage (due

to time restrictions) and so only four of the six QAT were tested. Proline betaine uptakes were carried out separately from those with choline, histidine and glycine betaine. Assays for proline betaine were carried out in the same way as above; however, as the amount of radiolabelled solute was limited only one assay was performed with each mutant under each condition (Fig. 6.18).

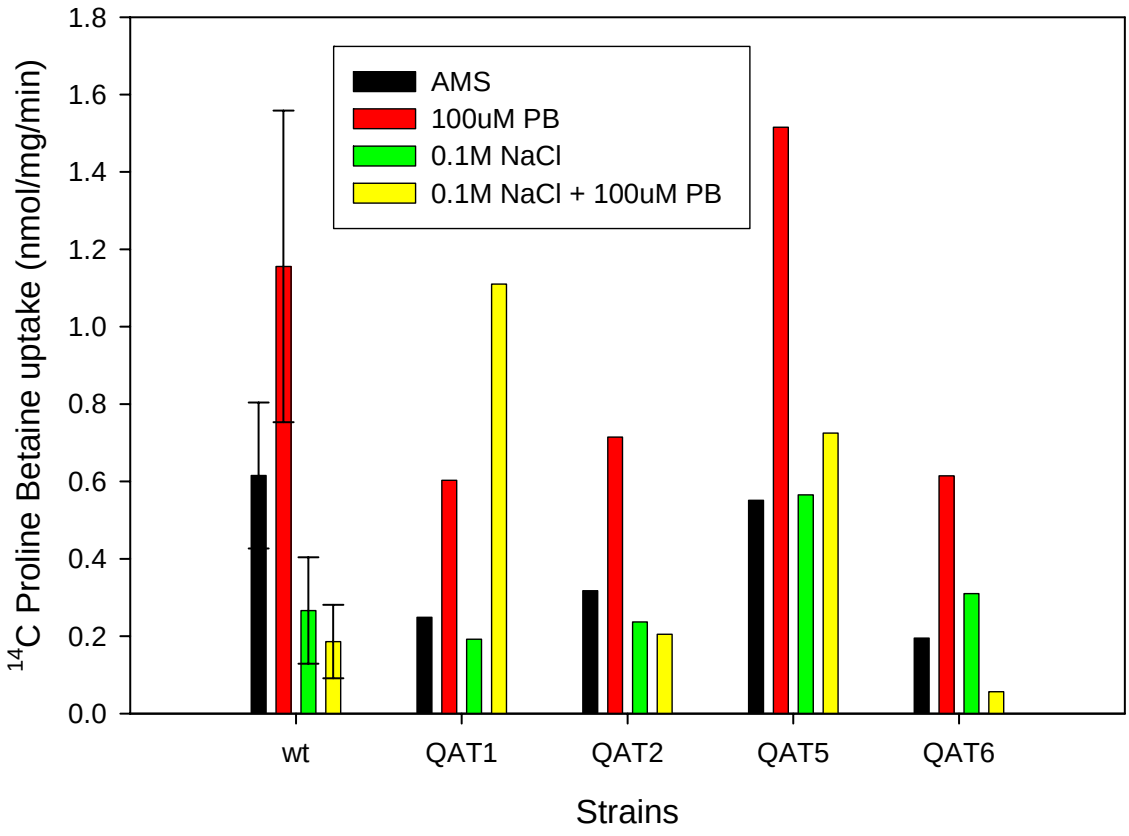


Figure 6.18. Uptake of ^{14}C Proline Betaine on 3841 and QAT Mutants. 3841 and the QAT mutants were grown up in AMS (10mM glc, 10mM NH_4) overnight plus 100µM of proline betaine or 100mM NaCl or both or neither, as indicated. Mean uptake of each solute, as indicated, per minute, over five minutes. Standard error values are shown for wild-type uptakes as these were performed in triplicate.

Although these data are only from one set of assays, they indicate that QAT1, QAT2 and QAT6 may be involved in the transport of proline betaine as the uptake rates for these mutants are lower than those observed in cells grown under some of the conditions tested. However, wild-type 3841 also had a reduction in uptake in cells grown in the presence of NaCl and so the significance of this initial data is unknown. The osmotic upshift did not seem to affect transport of proline betaine in the QAT1 and QAT5 mutants, although as mentioned, this is one set of data and so the significance of this initial result is unknown. Overall QAT6 is the most severely affected for proline betaine transport.

For the choline, histidine and glycine betaine assays, all *Rhizobium* was grown up in AMS (10mM glc, 10mM NH_4) or AMS (10mM glc, 10mM NH_4) + 100mM NaCl; no solute additions were made so that the same cultures could be used for each of the solute assays, allowing a direct comparison between results (Figs. 6.19 – 6.21).

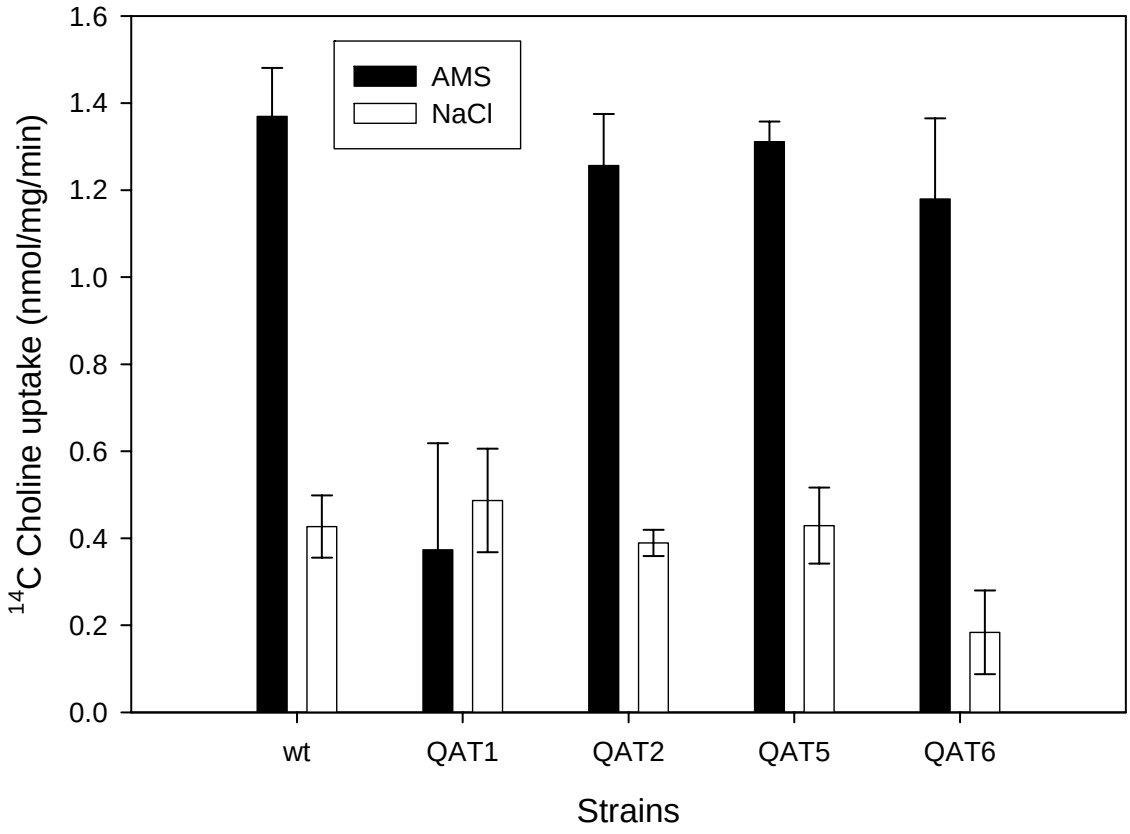


Figure 6.19. Uptake of ^{14}C Choline on 3841 and QAT Mutants. 3841 and the QAT mutants were grown up in AMS (10mM glc, 10mM NH_4) and AMS (10mM glc, 10mM NH_4) + 100mM NaCl. Mean uptake of each solute, as indicated, per minute, over four minutes. Mean results of three experiments with standard error values.

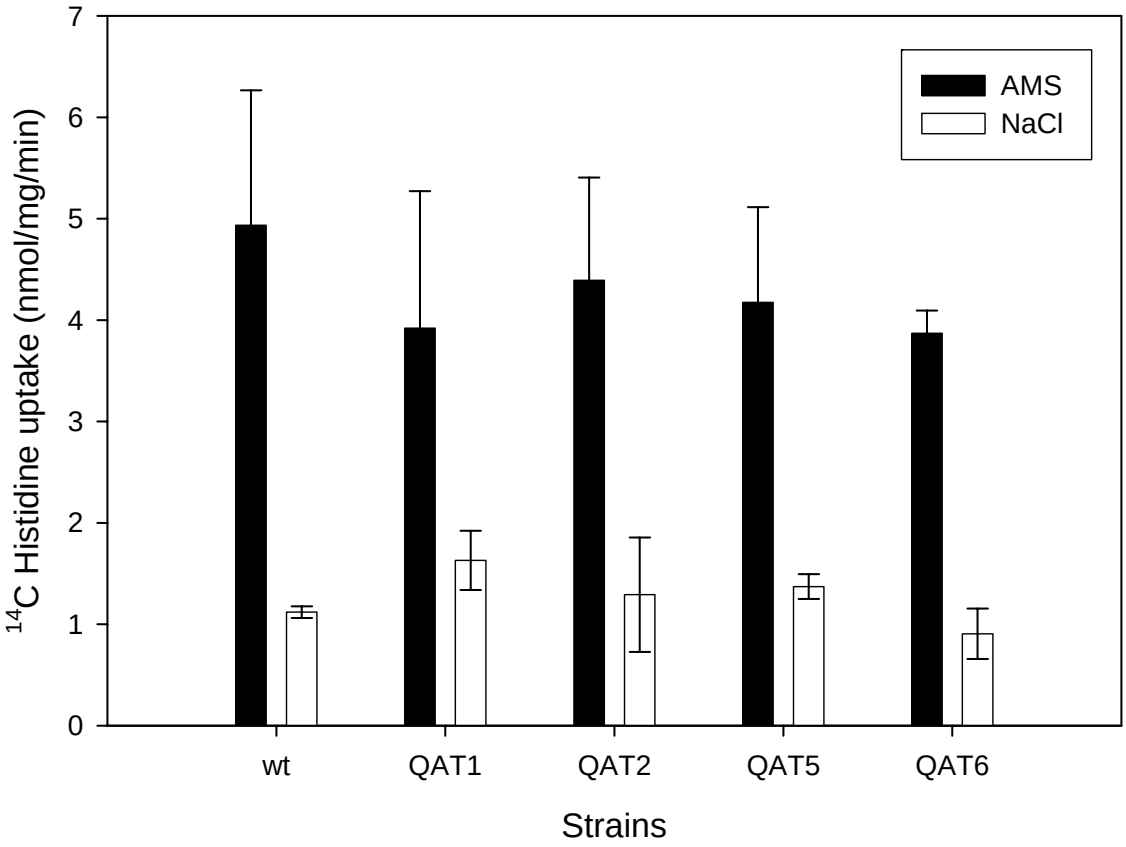


Figure 6.20. Uptake of ^{14}C Histidine on 3841 and QAT Mutants. 3841 and the QAT mutants were grown up in AMS (10mM glc, 10mM NH_4) and AMS (10mM glc, 10mM NH_4) + 100mM NaCl. Mean uptake of each solute, as indicated, per minute, over four minutes. Mean results of three experiments with standard error values.

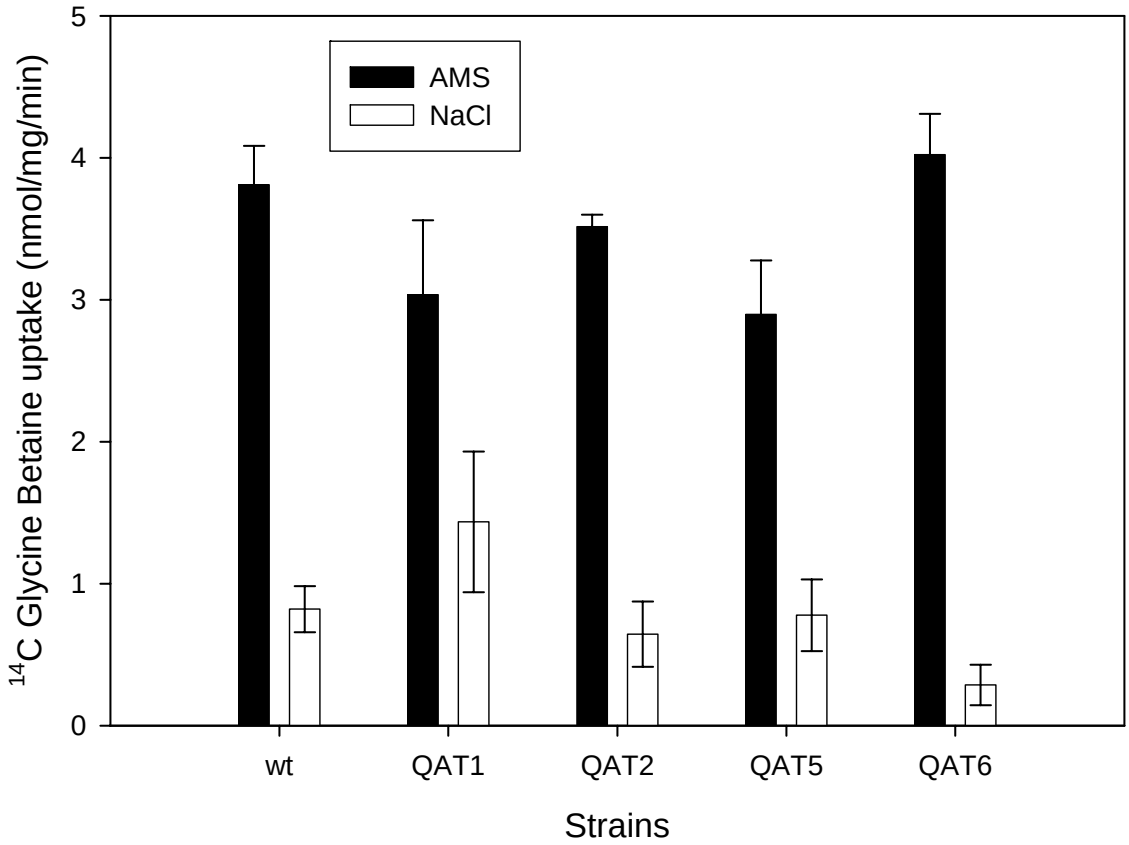


Figure 6.21. Uptake of ^{14}C Glycine Betaine on 3841 and QAT Mutants. 3841 and the QAT mutants were grown up in AMS (10mM glc, 10mM NH_4) and AMS (10mM glc, 10mM NH_4) + 100mM NaCl. Mean uptake of each solute, as indicated, per minute, over four minutes. Mean results of three experiments with standard error values.

None of QAT systems tested appear to be responsible for the transport of histidine or glycine betaine, as no significant difference was seen in uptake rates between the wild-type and any of the mutants (Figs. 6.20 and 6.21). However, QAT1 seems to be required for the uptake of choline, as its uptake rate was barely detected (<0.4nmol/mg/min) in the QAT1 mutant (compared against transport in wild-type 3841) (Fig. 6.19). As shown above, the transcription of QAT1 was induced by choline (and glycine betaine, although from Fig. 6.21 its overall transport rate is unaffected) (Table 6.6) and had a high identity to the choline transporter, the Cho system, in *S. meliloti* (Figs. 6.1 – 6.3). QAT1 therefore appears to be a homologue of the Cho system in *R. leguminosarum*.

In order to determine how important QAT1 was to *R. leguminosarum* and its growth, the QAT mutants were streaked onto AMA (10mM glc, 10mM NH_4), AMA (10mM choline), AMA (10mM histidine) and AMA (10mM glycine betaine) and left to grow (Figs. 6.22 – 6.25).

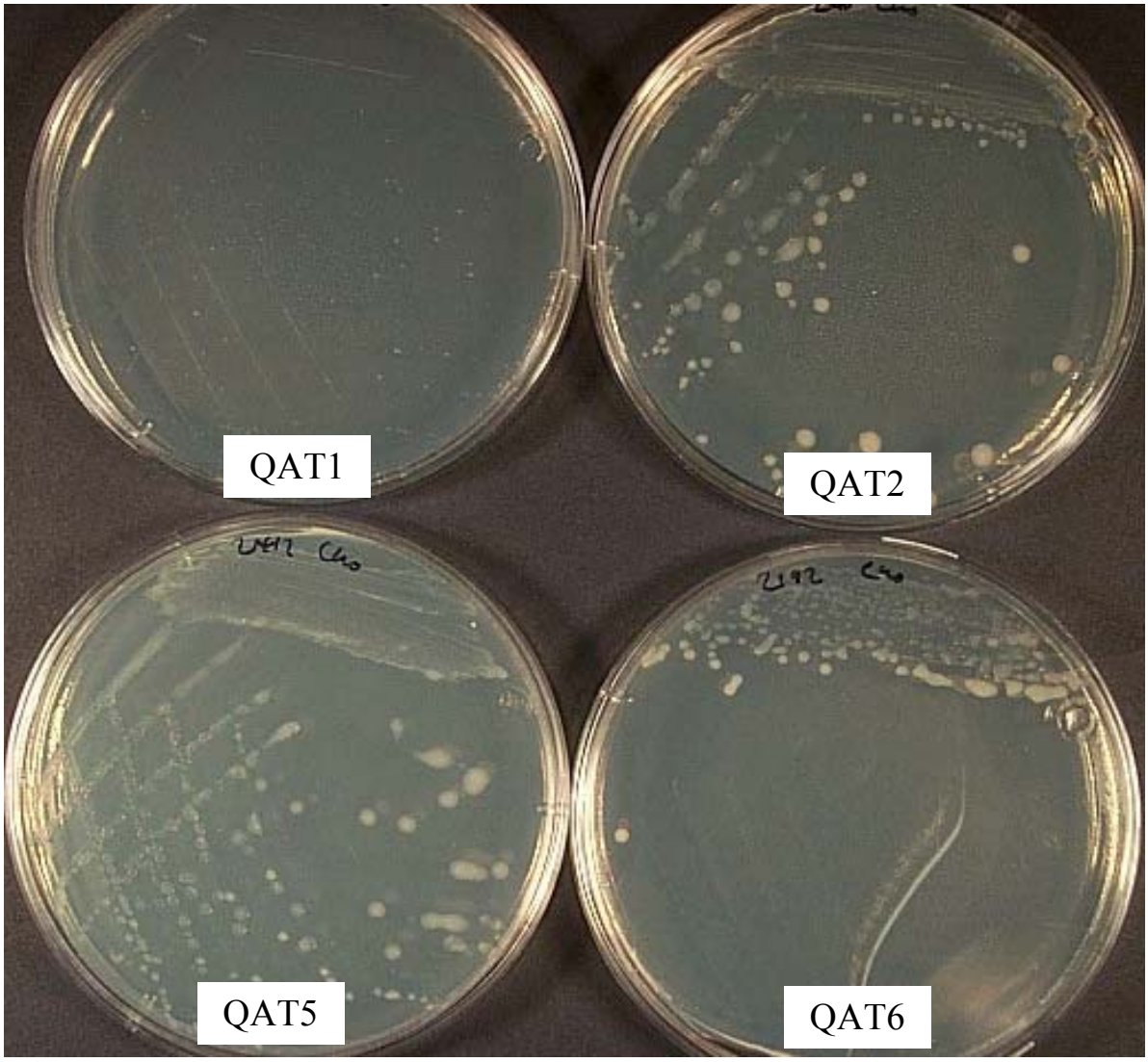


Figure 6.22. Growth of QAT Mutants on AMA (10mM Choline). As can be seen, the QAT1 mutant cannot grow when choline is provided as the carbon/nitrogen source, whilst the others can.

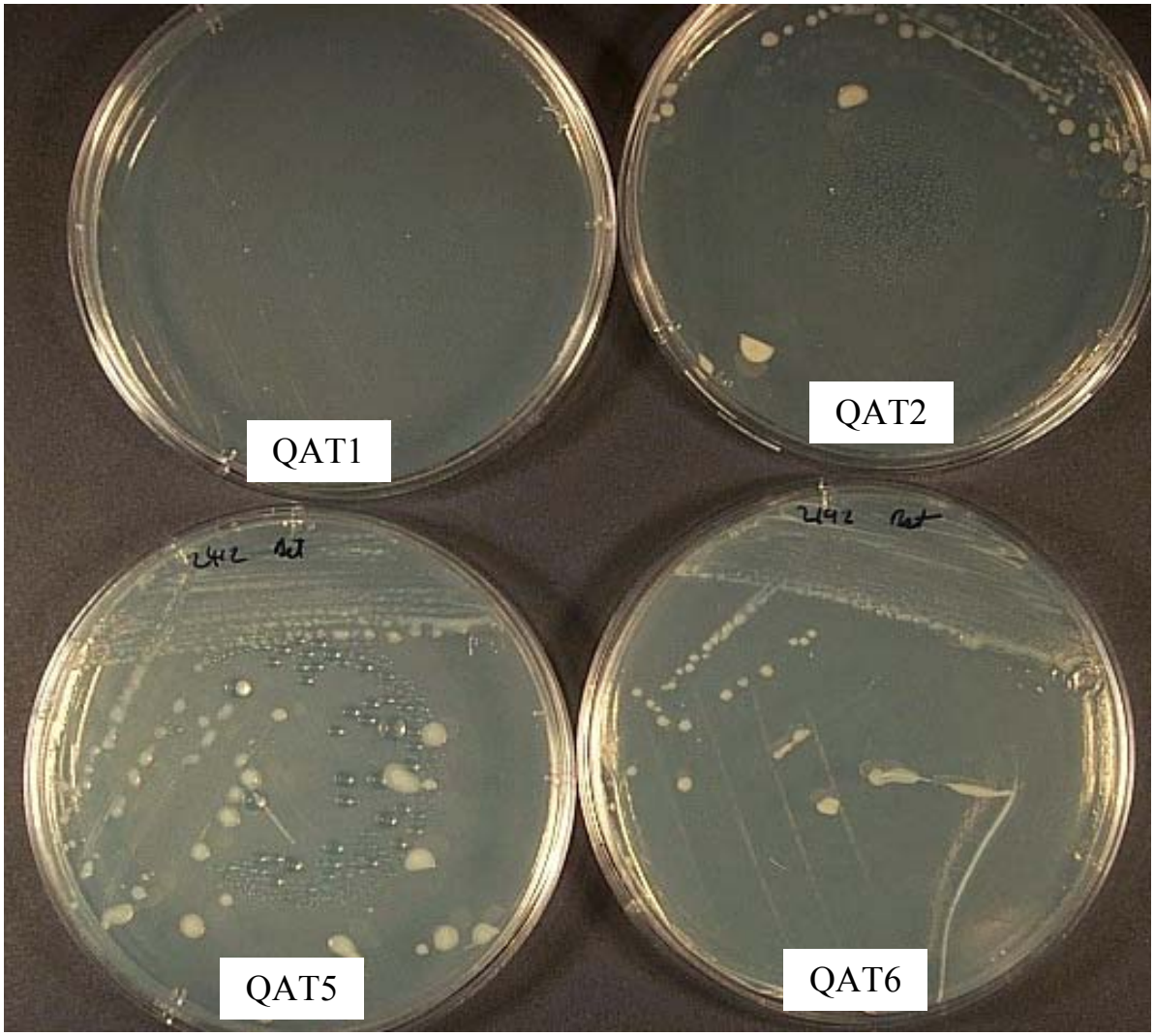


Figure 6.23. Growth of QAT Mutants on AMA (10mM Glycine Betaine). As can be seen, the QAT1 mutant cannot grow when glycine betaine is provided as the carbon/nitrogen source, whilst the others can.

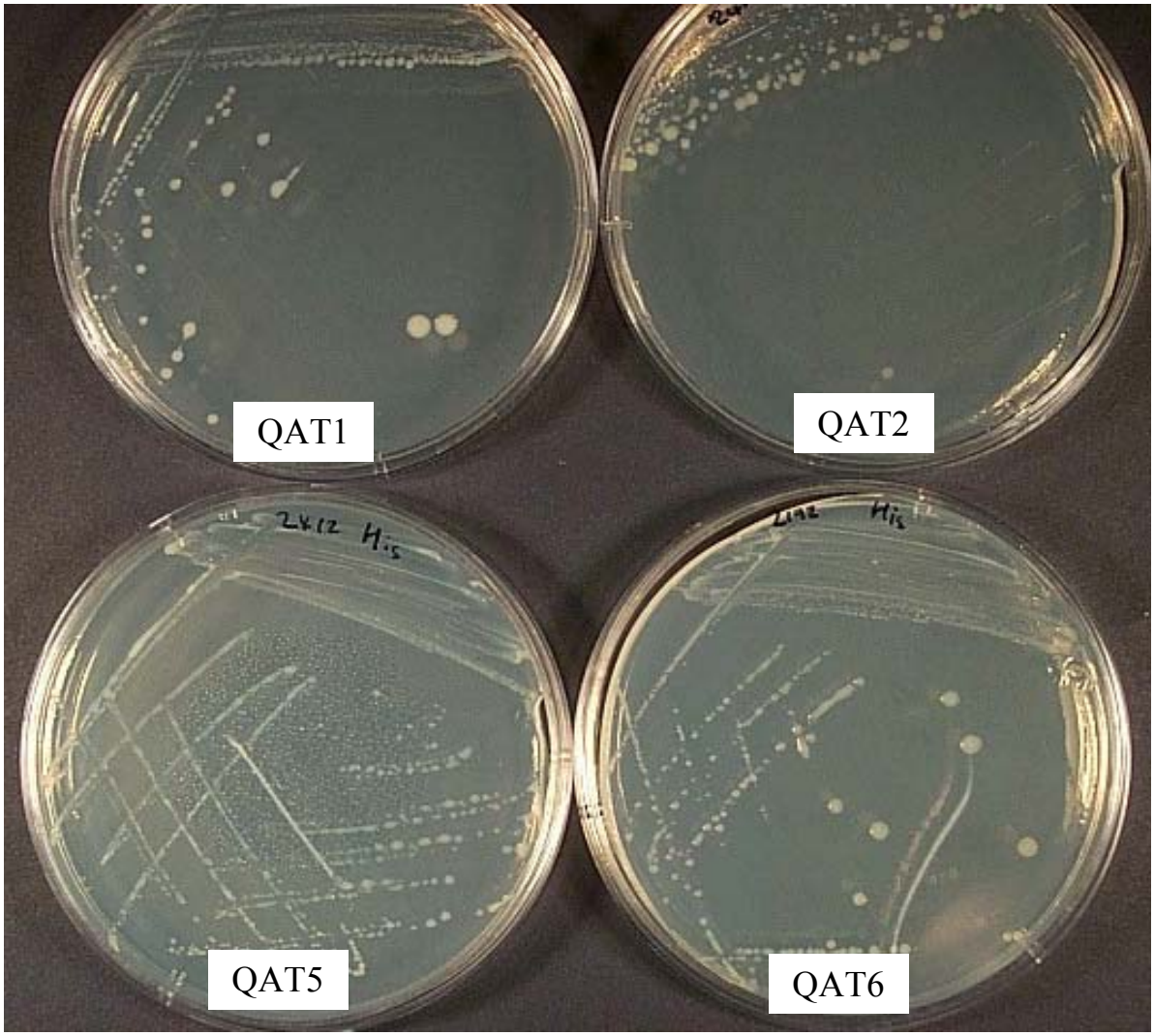


Figure 6.24. Growth of QAT Mutants on AMA (10mM Histidine). As can be seen, all QAT mutants can grow when histidine is provided as the carbon/nitrogen source.

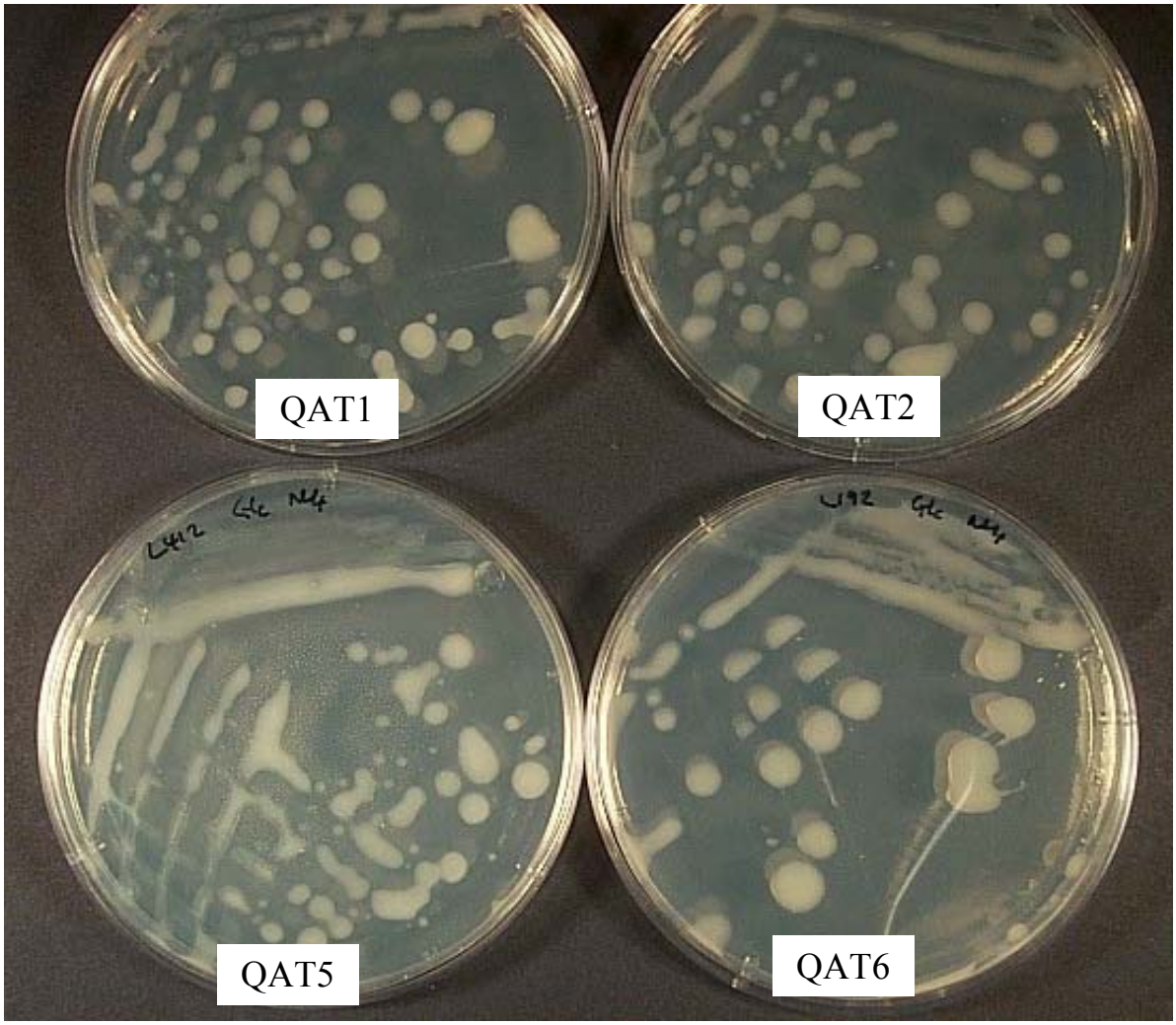


Figure 6.25. Growth of QAT Mutants on AMA (10mM Glc, 10mM NH_4). As can be seen, all QAT mutants can grow when glucose and ammonia are provided as the carbon/nitrogen source.

The QAT1 mutant can not grow when choline is the only carbon/nitrogen source (Fig. 6.22), as expected; although interestingly, it could not grow on glycine betaine either (Fig. 6.23). Although QAT1 does not appear to be responsible for the uptake of glycine betaine (Fig. 6.21), it is induced by it (Table 6.6) and cannot grow when it is supplied as the sole carbon/nitrogen source (Fig. 6.23). However, the cultures used in the uptake experiments were grown on 10mM glc and 10mM NH_4 and *R. leguminosarum* may need to be grown on glycine betaine (or choline) for uptake via QAT1 to become the predominant system. This could explain why the QAT1 mutant could not grow on glycine betaine, but was able to transport it when grown on glucose and ammonia. Due to the time constraints of this project the optimum conditions required for uptake of glycine betaine (and choline and histidine) may not have been determined and this is something that is planned for future work.

Alternatively, it may be possible that QAT1 does transport glycine betaine but another transporter that is responsible for the glycine betaine is also present in 3841. This is why uptake rates did not significantly change in the QAT1 mutant. However, if this is true, it would be expected that the QAT1 mutant could use this other transporter and so be able to grow AMA with 10mM glycine betaine, which was not the case and so this is unlikely. Choline is a precursor of glycine betaine and so a connection between the growth on the two substrates may exist, although this unknown at this time.

Interestingly, a mutant of the Cho system in *S. meliloti* was able to grow on media containing choline as the sole carbon and nitrogen source (Dupont *et al*., 2004). (It was not tested with glycine betaine as the Cho system was shown not to transport that substrate.) Uptake assays with the *S. meliloti* Cho mutant did not show a total inhibition of choline transport, suggesting that another choline transport is present in *S. meliloti* allowing a Cho mutant to grow on choline. This data indicates that whilst similar in sequence identity, QAT1 and Cho are not exact homologues as they have differences in growth phenotypes and potentially in substrate specificity.

The uptakes rates of the wild-type and QAT mutants were all inhibited by the presence of osmotic upshift, in the choline, histidine and glycine betaine experiments (Figs. 6.19 – 6.21). Again, this clearly contradicts previous reports on the uptake of osmoprotectants in rhizobia (Bernard *et al*., 1986; Boncompagni *et al*., 1999) and as such warrants more study.

6.3. Conclusion

Six ProU-like systems were identified within *R. leguminosarum* using sequence comparisons, homology and phylogeny studies and BLAST analysis. The induction of each of these systems was investigated (using promoter probes) as was their role in solute uptake (using mutational studies). From these data, a homologue to the Cho system in *S. meliloti* was identified in 3841 (QAT1). QAT1 is clearly induced by the presence of choline and glycine betaine and a mutant of this system was unable to grow in media where choline or glycine betaine was the sole carbon/nitrogen source. Furthermore, when grown on glc/NH_4, QAT1 was the main transport system for choline but not glycine betaine. This may be due to the fact that the cultures were not grown on choline or glycine betaine as the carbon/nitrogen source, due to time constraints. This may also explain why QAT5 did not appear to be responsible for the transport of histidine, although its induction pattern and neighbouring genes indicated that it may be. Repeating these assays under optimum conditions and to also include the mutants of QAT3 and QAT4 is planned for future work.

The role of osmoprotectants was also investigated during this research, but no compound could be found that functioned as a compatible solute by rescuing the growth rate of 3841 undergoing hyper-osmosis.

However, the most startling discovery of this research was the apparent inhibition of the QATs caused by osmotic upshift. Inhibition of solute transport caused by hyper-osmosis would help to explain why the tested osmoprotectants had no effect in rescuing the growth of stressed bacteria, as they would be unable to import the solutes into their cells. The effect of osmotic upshift transport systems was therefore investigated in the next chapter.

CHAPTER 7: EFFECT OF OSMOTIC UPSHIFT ON SOLUTE UPTAKE VIA ABC TRANSPORTER SYSTEMS

7.1. Introduction

It was established that *R. leguminosarum* that had been grown overnight in AMS (10mM glc, 10mM NH_4) + 100mM NaCl, then washed and resuspended in RMS + 100mM NaCl had lower solute uptake rates compared to 3841 treated in exactly the same way but without the presence NaCl at any stage (Chapter 6). It was clear from these data that the presence of NaCl used to generate an osmotic upshift was responsible for this loss in transport rates. Data reported earlier clearly showed that the transcription of a QAT ABC transporter was strongly induced under hyper-osmosis generated by sucrose, mannitol and NaCl (pRU843 – Chapters 3 & 4). It was therefore completely unexpected to observe a decrease in solute uptake rates, under conditions where this ABC transporter had been specifically induced.

The QAT systems investigated in the previous chapter were, prior to this work, uncharacterised in *R. leguminosarum* and therefore the solutes which they transported were also unknown. This made it difficult to compare the uptake data from cells that had experienced an osmotic upshift to those that had not. This analysis would have been easier if characterised ABC transporters, with a known solute/s, had been studied. For this reason, the effect of osmotic upshift on solute uptake rates was investigated in known transporters of 3841. Transporters investigated included both ABC systems and secondary permeases.

The aim of the work presented in this chapter is to determine whether an osmotic upshift is responsible for a loss in transport in *R. leguminosarum* 3841.

7.2. Results

7.2.1. Effect of Sucrose Concentration and of Exposure Time on AIB Uptake Rates

As had been seen in the previous chapter, the presence of NaCl caused a decrease in solute uptake; however, it was not known if it was the presence of NaCl itself, or the osmotic upshift it generated, which caused this decrease. It is well known that if used in equal concentrations, NaCl generates a greater stress then sucrose (ver der Heide *et al*., 2001; Chapter 5 section 5.2.3). Sucrose was therefore tested to see if, when added at the same concentration (100mM), it would have had the same effect on uptake rates as NaCl. Also, the transport systems under investigation were changed from the unknown QATs to two well-characterised ABC transporters, Aap and Bra. The amino acid permease (Aap) and the branched chain amino acid transporter (Bra) (Walshaw & Poole, 1996; Hosie *et al*., 2002a) are known to be the only two permeases in 3841 responsible for the uptake of 2-Amino-Isobutyric acid (AIB). AIB is a non-metabolisable amino acid analogue and its transport has been well characterised in *Rhizobium* (Walshaw & Poole, 1996).

Therefore, transport of AIB via Aap and Bra was investigated. As AIB uptake via Aap and Bra had been previously characterised, the only variable in this set of transport assays was the presence of an osmolyte (100mM sucrose) used to generate an osmotic upshift. This made the analysis of uptake data easier, as there were fewer unknowns present.

As mentioned above in the assays with the QATs, 100mM NaCl was added to the AMS used for overnight growth and the RMS used for washing and resuspension of cultures. To determine if the decrease in uptake rates was caused by added osmolyte in the AMS or RMS, two experiments were performed. In the first, 3841 was grown up overnight in AMS (10mM glc, 10mM NH_4) + 100mM sucrose and then washed and resuspended in standard RMS. In the second, 3841 was grown up overnight in AMS (10mM glc, 10mM NH_4) and then washed and resuspended in RMS + 100mM sucrose. This meant that these cells were in the presence of sucrose for approximately an hour (the time cultures were left to starve after washing before uptake assays are performed). These two experiments were carried out in tandem alongside a control 3841 culture, which was grown up overnight in AMS (10mM glc, 10mM NH_4) and then washed and resuspended in RMS (no sucrose at any stage). AIB uptake was measured on the three cultures (Fig. 7.1).

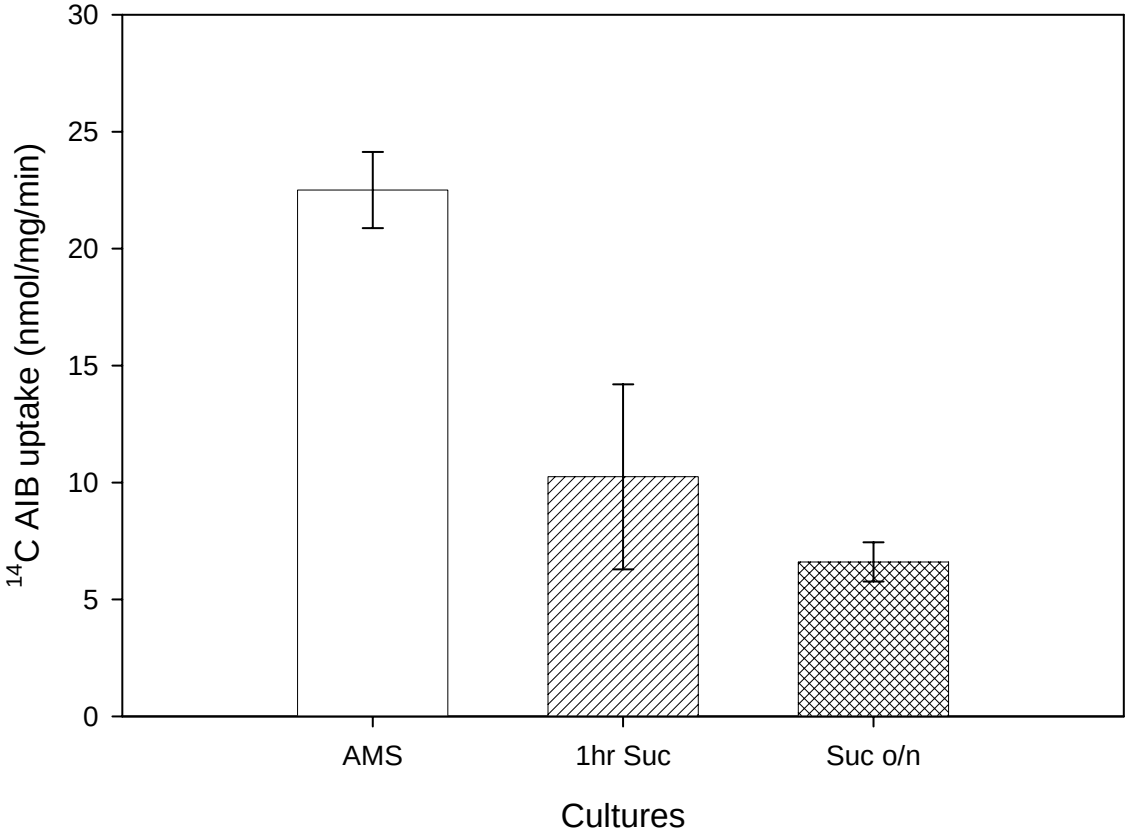


Figure 7.1. Uptake of ^{14}C AIB on Non-Stressed and Stressed (for 1 Hour and Overnight) 3841. 3841 was grown up in AMS (10mM glc, 10mM NH_4) overnight, washed & resuspended in RMS (white) or was grown up in AMS (10mM glc, 10mM NH_4) overnight, washed & resuspended in RMS + 100mM sucrose (single-hatch) or was grown up in AMS (10mM glc, 10mM NH_4) + 100mM sucrose overnight, washed & resuspended in RMS (cross-hatch). Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values.

AIB uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Hosie *et al.*, 2002a). Cells in RMS + 100mM sucrose for an hour had an AIB uptake rate which was 46% of the control culture, compared to 3841 grown overnight in AMS (10mM glc, 10mM NH_4) + 100mM sucrose which was 29% of the control. These data show that the presence of 100mM sucrose in either RMS for an hour or AMS (10mM glc, 10mM NH_4) overnight caused a decrease in AIB uptake. These data also indicate that the longer the exposure time to osmotic upshift, the greater the decrease in uptake; i.e. cells exposed to overnight shock had a greater reduction in AIB uptake rate then cells that were stressed for only an hour.

As these data show the amount of time cells were exposed to an osmotic upshift affects the uptake rate, the concentration of sucrose used was then investigated to determine if it was also significant. For this set of assays, cells were grown up overnight in AMS

(10mM glc, 10mM NH_4) then washed in RMS, but resuspended for an hour in RMS + either 100mM, 75mM, 50mM or 25mM sucrose. Cells were also resuspended in standard RMS as a control. AIB uptake of each culture was measured (Fig. 7.2).

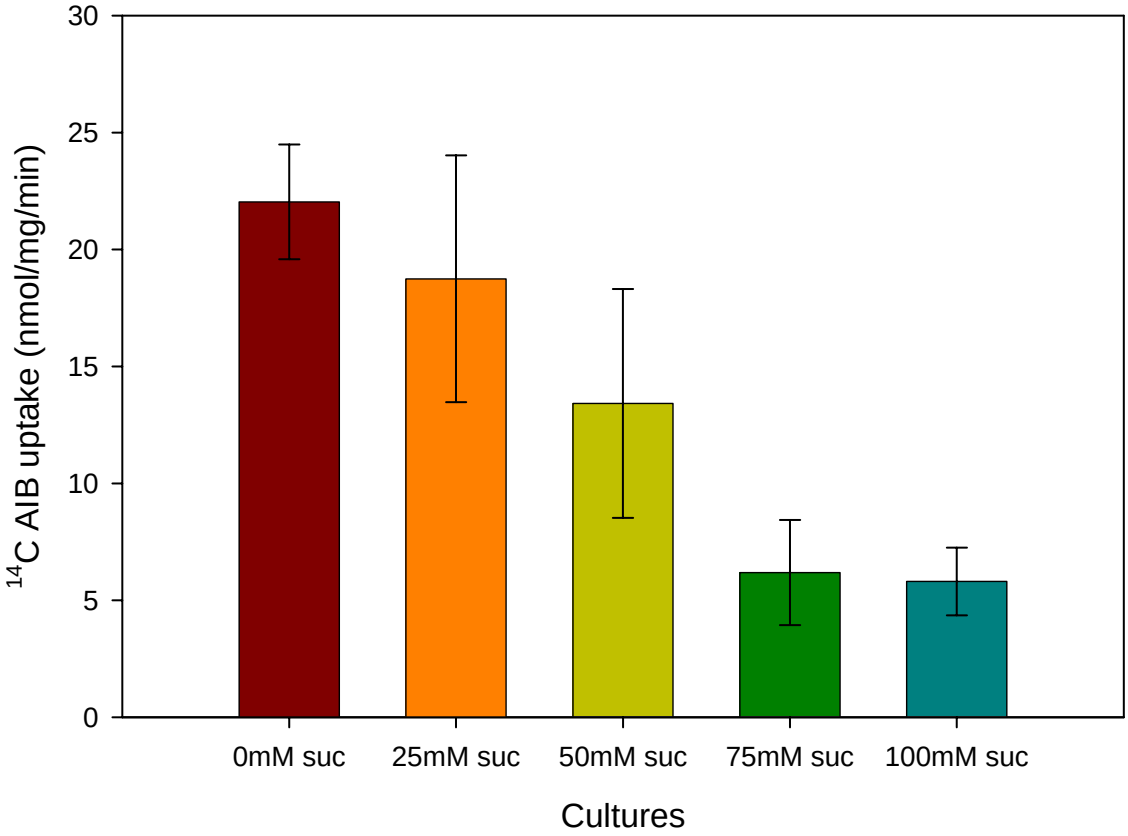


Figure 7.2. Uptake of ^{14}C AIB on 3841 Exposed to Different Concentrations of Sucrose 1 Hour before Assays. Cells were resuspended for an hour before assays were performed in RMS containing the concentration of sucrose indicated on the x-axis. Mean uptake of AIB per minute, over four minutes. Mean results of at least four experiments with standard error values.

These data clearly prove that the concentration of sucrose used also plays a role in the decrease of AIB uptake, as the greater the concentration, the greater the decrease. This means that both the amount of time cells were exposed to osmotic upshift and the concentration of osmolyte used had an effect on the uptake rates of AIB in 3841.

7.2.2. Immediate Effect of Very High Concentrations of Osmolyte on AIB Uptake Rates

At this point it is worth remembering previous uptake experiments conducted with spheroplasts and ABC transporters. In order to prevent spheroplasts from bursting (as they have no outer membrane to protect them for their environment and its osmolality) they are kept in a solution containing 20% sucrose (approx. 600mM) (Hosie *et al.*, 2002b). Transport assays are then carried out on the spheroplasts as soon as they are generated. Previous studies that investigated solute transport in spheroplasts had not observed any significant uptake through ABC systems, as no SBP components are present (see section 7.2.6). However, even if SBPs were present, could the addition of such a high concentration of sucrose have had an immediate effect on transport assays?

To determine this, cells were grown up overnight in AMS (10mM glc, 10mM NH_4), then washed and resuspended in RMS for an hour as standard. Transport assays were then carried out as usual with just one exception. Ten seconds before ^{14}C AIB was introduced to the assay mix, sucrose was added so that its final concentration was either 0mM, 100mM, 200mM, 300mM, 400mM or 500mM. This meant that sucrose was added to cultures immediately before transport assays were performed (Fig. 7.3). The assay volume was always maintained at 500µl.

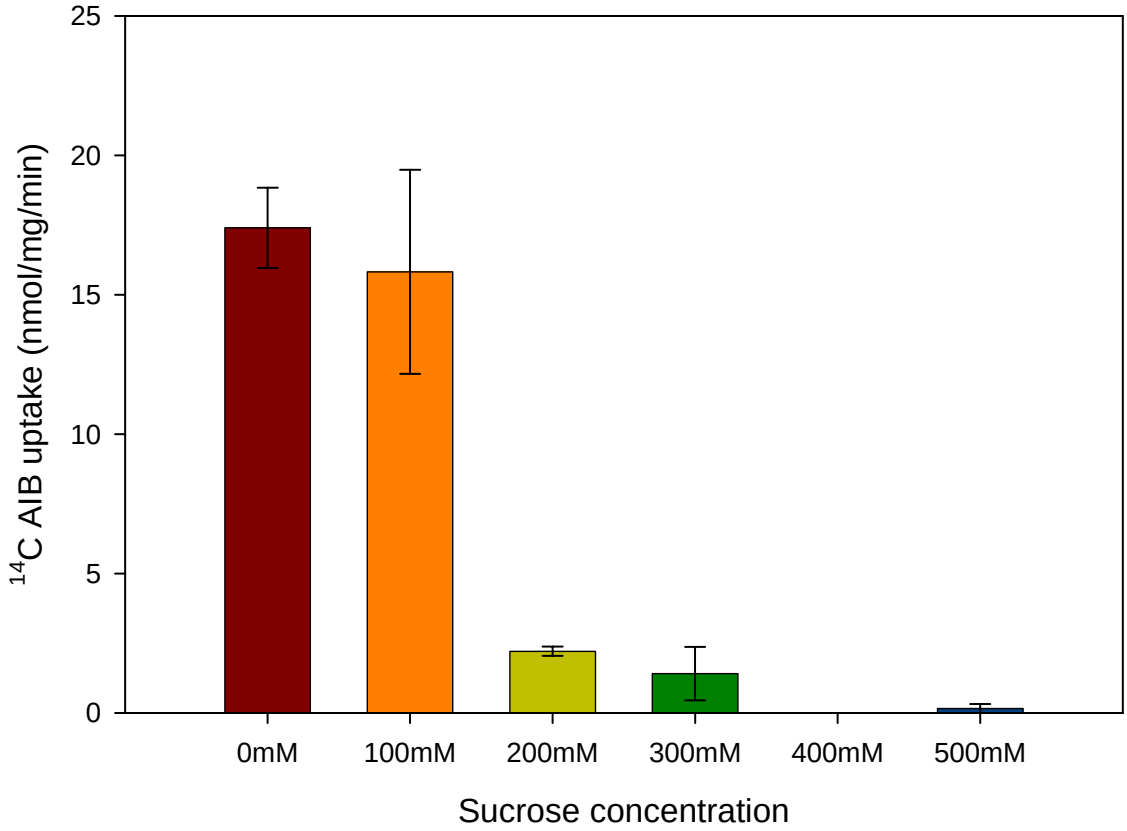


Figure 7.3. Uptake of ^{14}C AIB on 3841 Exposed to Different Concentrations of Sucrose, Ten Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values. The 0mM value is a mean of the data from this experiment and the data from Figs. 7.4 & 7.5.

These data show that AIB uptake decreased by 88% on the addition of 200mM sucrose only ten seconds before the assay was performed. It also shows that it took more than ten seconds exposure to 100mM sucrose before a significant decrease in AIB uptake is observed (decrease is only 9% compared to the previous decreases of 54% after 1 hour incubation and 71% after overnight incubation).

It was still unclear if this immediate effect was caused by a high concentration of sucrose alone. Other compounds were therefore tested to determine if they could cause a similar decrease in uptake of AIB.

The first compound to be tested was NaCl, as it was the compound used in the QAT assays (Chapter 7). It was therefore decided to repeat the previous experiment but with 100mM increments of NaCl added ten seconds before the assay instead of sucrose (Fig. 7.4).

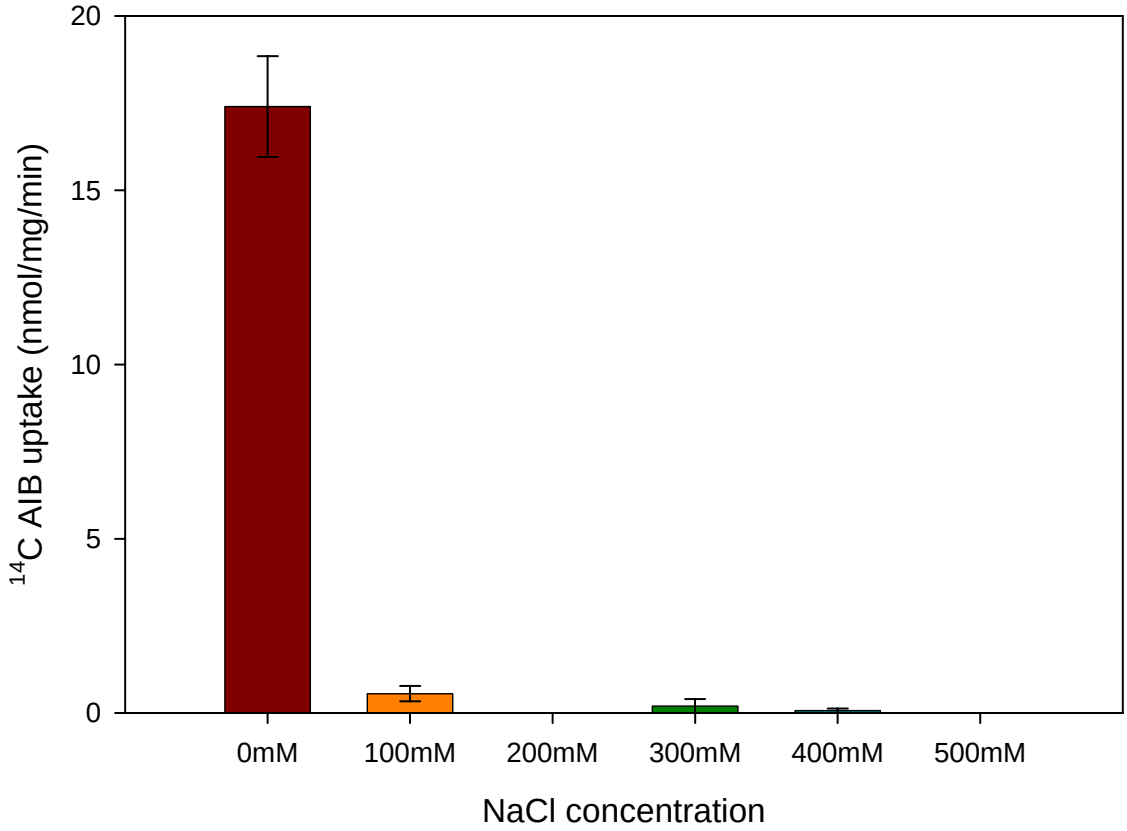


Figure 7.4. Uptake of ^{14}C AIB on 3841 Exposed to Different Concentrations of NaCl, Ten Seconds before Assays. Cells were exposed to the concentration of NaCl indicated on the x-axis immediately before assay. Mean uptake of AIB per minute, over four minutes. Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values. The 0mM value is a mean of the data from this experiment and the data from Figs. 7.3 & 7.5.

These data show a significant decrease in AIB uptake, even at 100mM (a 97% decrease). As mentioned above, NaCl generates a greater osmotic stress than sucrose at the same concentrations and a greater effect on uptake was seen here. These data imply that the decrease in uptake rates was caused by osmotic upshift, regardless of how it was generated. To prove this, the experiment was repeated with mannitol used instead of sucrose or NaCl (Fig. 7.5).

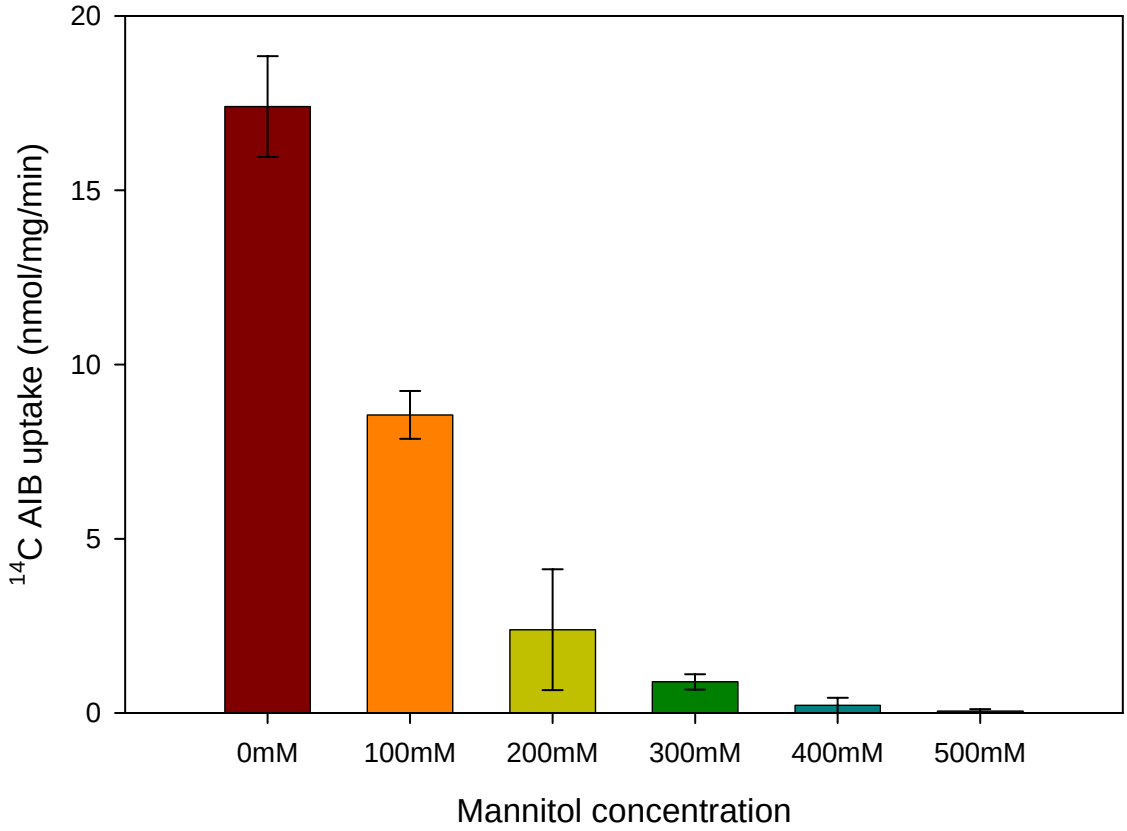


Figure 7.5. Uptake of ^{14}C AIB on 3841 Exposed to Different Concentrations of Mannitol, Ten Seconds before Assays. Cells were exposed to the concentration of mannitol indicated on the x-axis immediately before assay. Mean uptake of AIB per minute, over four minutes. Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values. The 0mM value is a mean of the data from this experiment and the data from Figs. 7.3 & 7.4.

The same effect was clearly seen again and was similar to that observed with sucrose, as mannitol caused an 86% decrease at 200mM (88% in sucrose). The data from Figures 7.3, 7.4 and 7.5 are compared in Figure 7.6.

These data clearly show that, it is not just sucrose that causes a decrease in AIB uptake, but also NaCl and mannitol. Each solute has a different threshold for significant inhibition of AIB transport, but all seem to reduce AIB uptake to less then 20% of normal when a final concentration of 200mM was used ten seconds before the assays were performed.

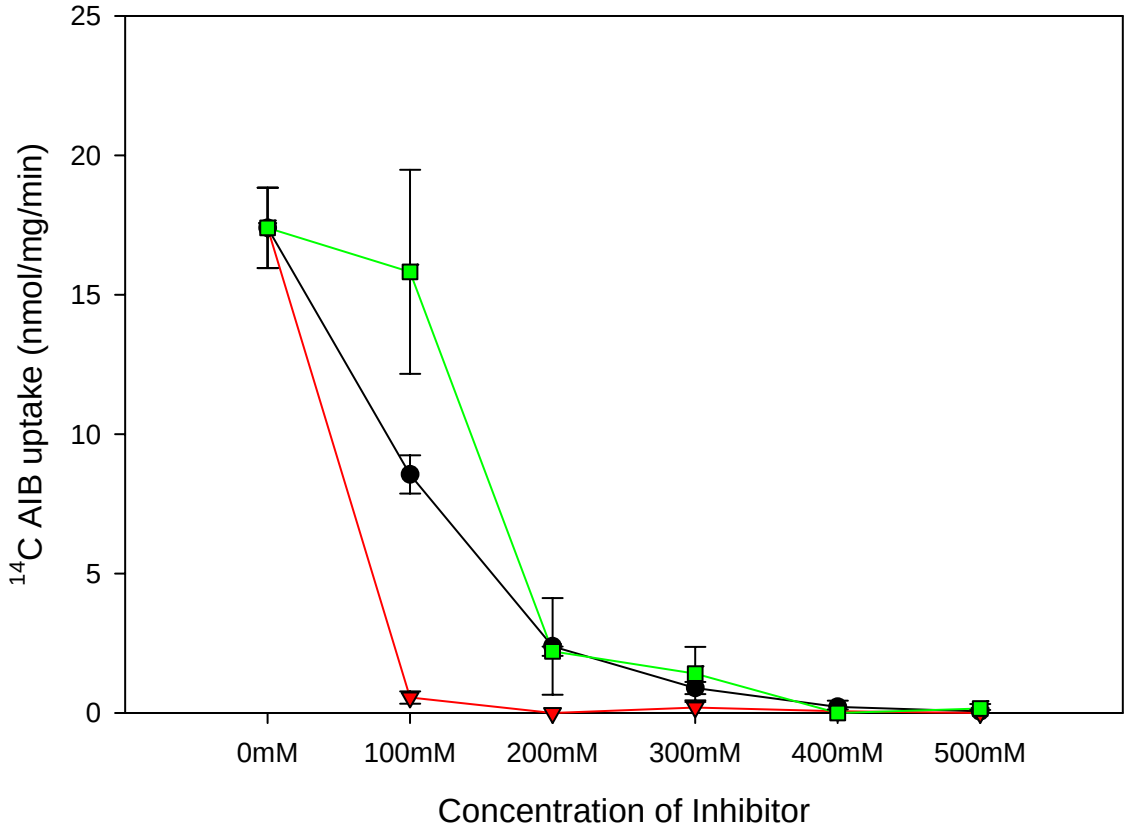
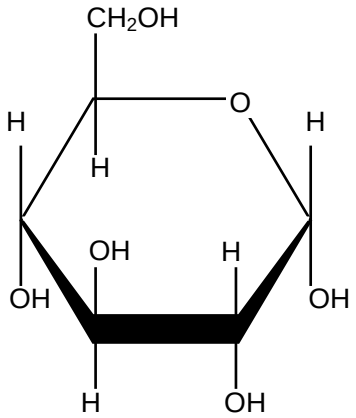
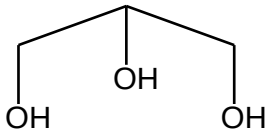


Figure 7.6. Uptake of ^{14}C AIB on 3841 Exposed to Different Concentrations of Sucrose, NaCl & Mannitol, Ten Seconds before Assays. The combined data from Figs. 7.3, 7.4 & 7.5. Cells were exposed to the concentration indicated on the x-axis of either sucrose (green squares), NaCl (red triangles) or mannitol (black circles) immediately before assay. Mean uptake of AIB per minute, over four minutes. Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values.

It was therefore decided to test the addition of some more osmolytes at 200mM immediately before transport assays were performed, to determine if they too would cause a decrease in AIB uptake rates. Potassium chloride (KCl), glycerol, glucose and polyethylene-glycerol (PEG) 200 (Fig.7.7) were all tested alongside 200mM sucrose, repeated as a control. The results from these assays (as well as those already reported above) are shown below (Fig. 7.8).

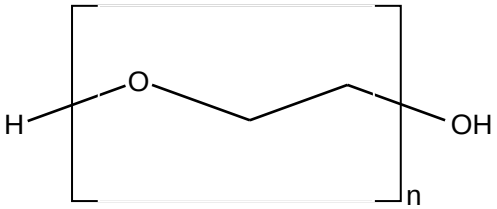


Glucose – $C_6H_{12}O_6$

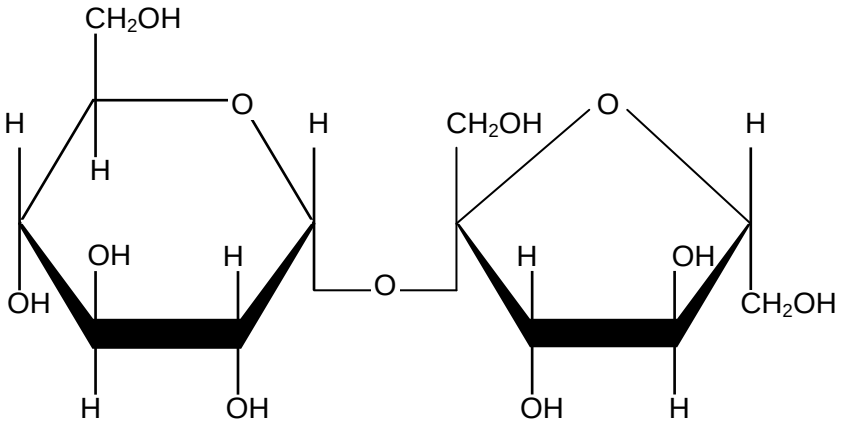


Glycerol – $C_3H_8O_3$

Mannitol – $C_6H_{14}O_6$



PEG – $H(OC_2H_4)_nOH$ (for PEG200, n~4)



Sucrose – $C_{12}H_{22}O_{11}$

Figure 7.7. Structures and Chemical Formulas of the Different Compounds Used to Inhibit Uptake Assay. Structures of NaCl and KCl are not included, as they are simple ionic compounds.

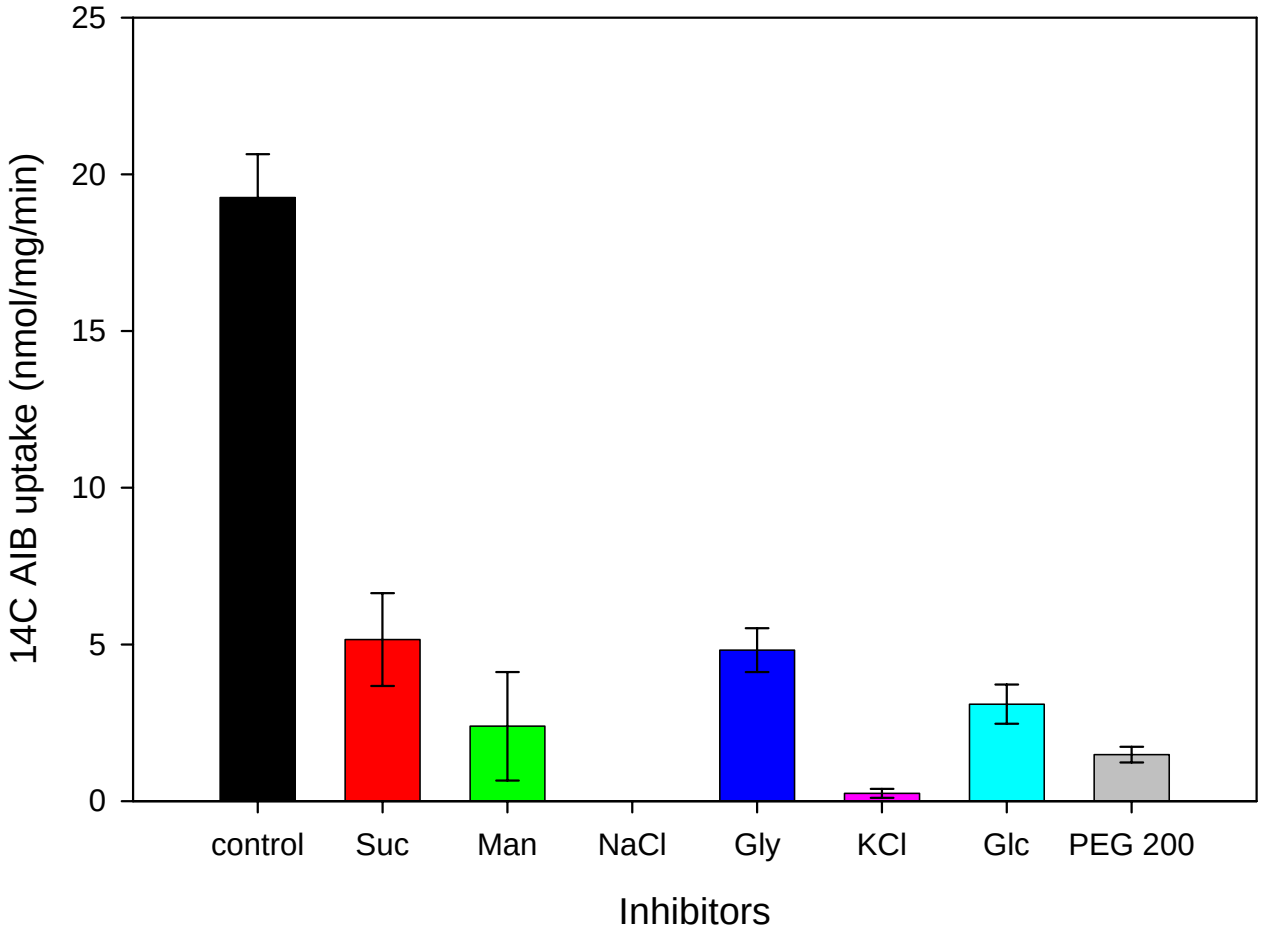


Figure 7.8. Uptake of ^{14}C AIB on 3841 Exposed to 200mM of Different Compounds Ten Seconds before Assay. Cells were exposed to 200mM of the compound indicated on the x-axis immediately before assay. Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values. (Suc = sucrose, Man = mannitol, Gly = glycerol, Glc = glucose)

As can be seen, all of the compounds tested cause a decrease in AIB uptake when added at 200mM ten seconds before assays began. This decrease was greater than 70% with all of the compounds tested.

As so many compounds were tested, all of which covered a range of size and charge, it was unlikely that this effect was caused by their molecular weight. The amount of compound added (in grams) varied between molecule used, and therefore the percentage composition was different too (Table 7.1). If the size of the molecule caused the decrease in uptake rates, then a direct relationship between molecular weight and decrease in uptake rate would have been observed but it was not (Table 7.1).

Table 7.1. Percentage Composition and AIB Uptake of Each Compound Used. Percentages are weight/volume and uptakes are given as a percentage compared to cells with no additions. Molecular weight of each compound is also shown.

Compound	Molecular Weight	Composition (w/v%)	Uptake(%)
Nothing added	-	0	100
NaCl	58.4	1.17	0
KCl	74.5	1.49	1
Glycerol	92.1	1.84	25
Glucose	180.2	3.60	16
Mannitol	182.2	3.64	12
PEG 200	~200	4.00	8
Sucrose	342.3	6.85	27

It was also unlikely that so many different compounds would cause direct transport inhibition by either blocking transporter permeases or by binding to SBPs. Eliminating these possibilities further confirmed the hypothesis that osmotic upshift some how causes a decrease in AIB uptake.

7.2.3. Immediate Effect of Osmotic Upshift on the Uptake Rates of Other Solutes

Transport studies reported above had only been conducted with one solute, AIB. For this reason, the investigation was extended to include the uptake of other solutes.

As mentioned above, AIB is known to be transported into *R. leguminosarum* by only Aap and Bra (Walshaw & Poole, 1996; Hosie *et al*., 2002a) and therefore other solutes that utilise these two systems were tested. Glutamate is one solute known to be transported via Aap and Bra (Walshaw & Poole, 1996; Hosie *et al*., 2002a) and its uptake rate was measured as AIB's was, with sucrose added (0 – 500mM in 100mM increments) immediately before the assays were performed (Fig. 7.9). Overnight cultures were grown up in AMS (10mM glc, 10mM NH_4) as before.

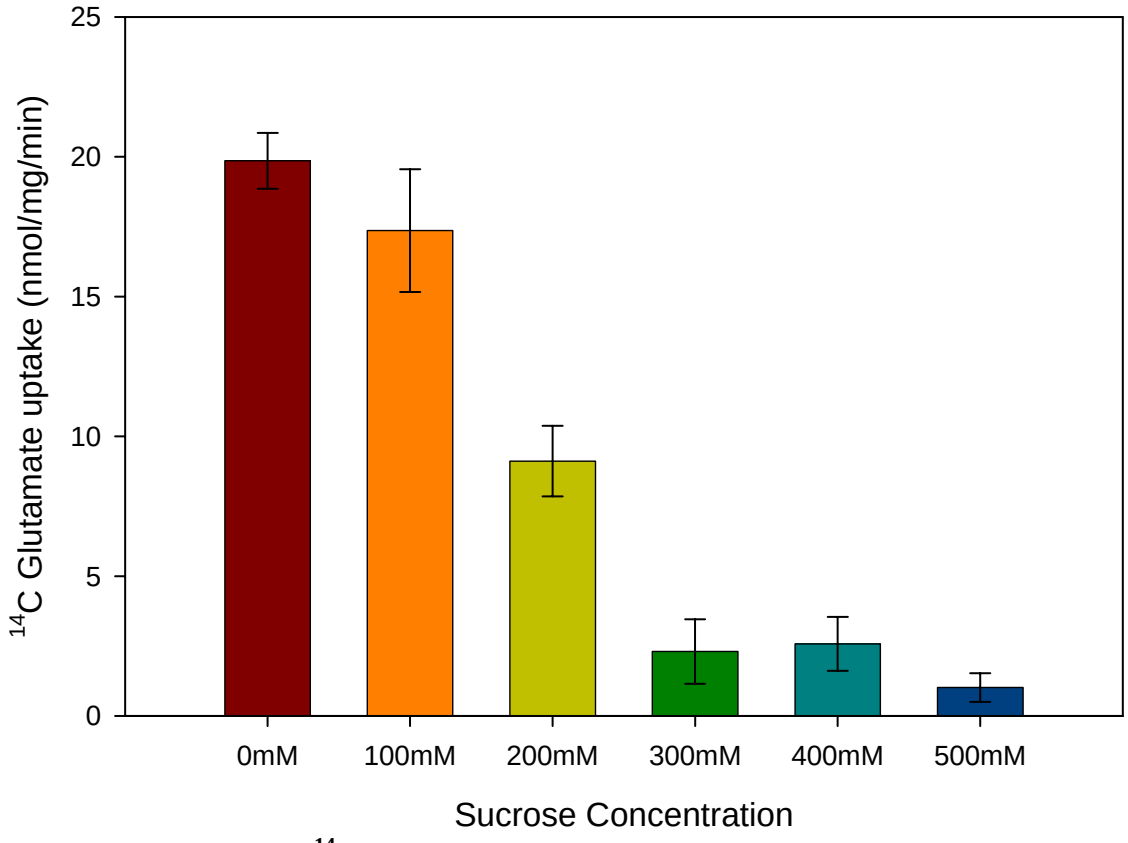


Figure 7.9. Uptake of ^{14}C Glutamate on 3841 Exposed to Different Concentrations of Sucrose 10 Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of glutamate per minute, over four minutes. Mean results of at least three experiments with standard error values.

Glutamate uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Hosie *et al*., 2002a). The pattern observed with the addition of increasing amounts of sucrose 10 seconds before transport assay was very similar to that seen with AIB uptake (Fig. 7.3).

This experiment was repeated but γ-amino-n-butryic acid (GABA) was used as the transport solute (Fig. 7.10). GABA differed slightly from glutamate and AIB, in that it is only transported into cells via the Bra system and not by Aap (Hosie *et al*., 2002a). Overnight cultures were grown up in AMS (10mM glc, 10mM NH_4) as before.

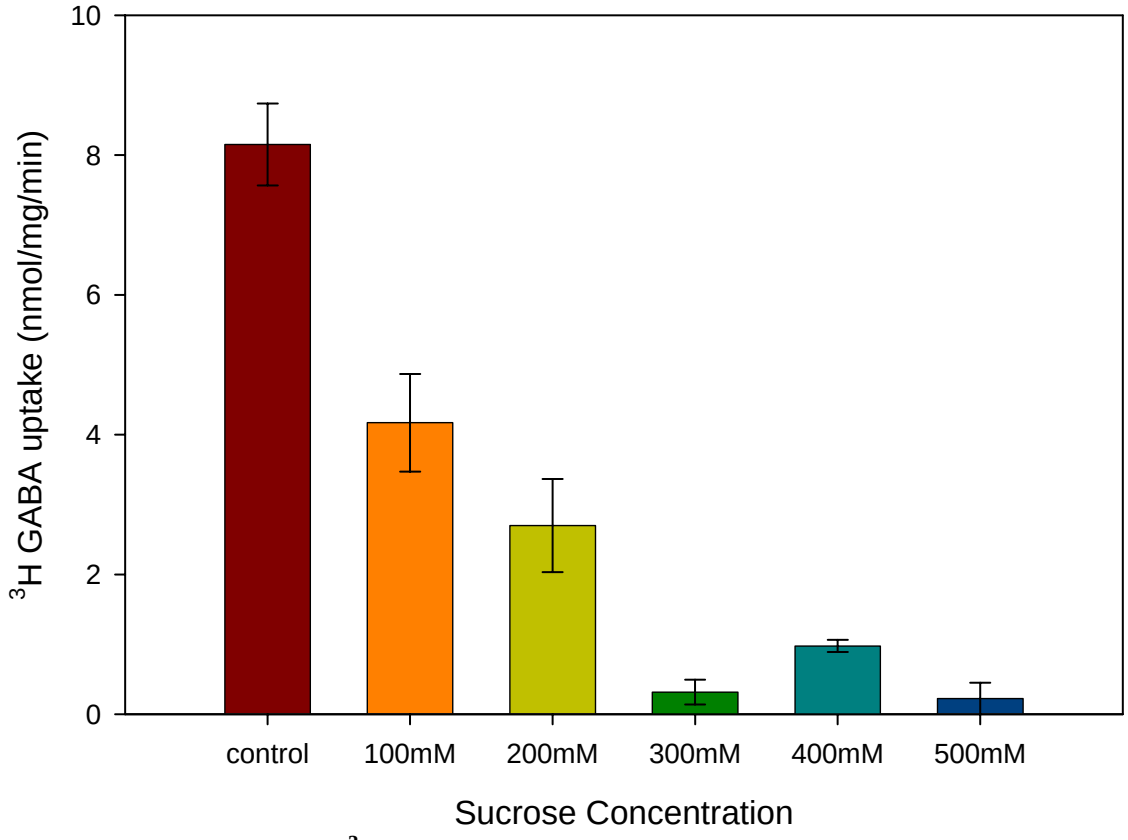


Figure 7.10. Uptake of ^{3}H GABA on 3841 Exposed to Different Concentrations of Sucrose 10 Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of GABA per minute, over four minutes. Mean results of at least three experiments with standard error values.

Again, the results obtained were similar to those observed with AIB uptakes under the same conditions. Also, the GABA uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Hosie *et al*., 2002a).

Next, alanine transport was investigated (Fig. 7.11). Its uptake also differed from AIB, as it is transported into cells via Aap and Bra and by a secondary transporter as well, the monocarboxylate transport permease (MctP) (Hosie *et al*. 2002b). MctP has a very low affinity for alanine, in comparison with the two ABC transporters. The concentration of alanine used in this experiment (25μM) was chosen to observe its uptake via Aap (K_m = 509M) and Bra (K_m = 173nM) (Hosie *et al.*, 2002a). Any uptake through MctP was insignificant, as it required a higher amount of alanine present (500μM) (K_m = 560nM)

(Hosie *et al.*, 2002b) in order to create a concentration gradient to allow significant transport. Overnight cultures were grown up in AMS (10mM glc, 10mM NH_4) as before.

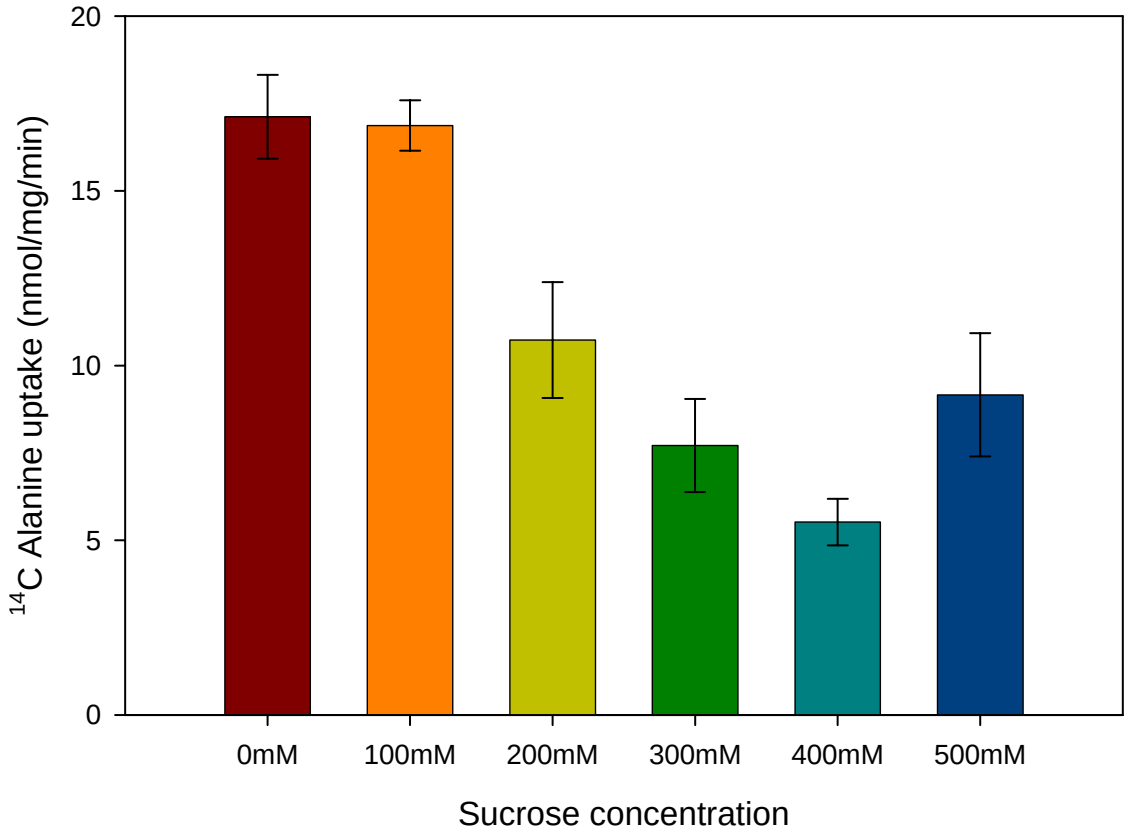


Figure 7.11. Uptake of ^{14}C Alanine (25µM) on 3841 Exposed to Different Concentrations of Sucrose 10 Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of alanine per minute, over four minutes. Mean results of at least three experiments with standard error values.

Alanine uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Hosie *et al.*, 2002a). Although uptake rates decrease with the increase of sucrose added, the drop in uptake rates was not as pronounced as it had been in the previous experiments. Also uptake rates did not continue to decrease with the increase of sucrose concentration, as uptake in the presence of 500mM sucrose was higher then those with 300mM and 400mM. This experiment was repeated several times in order to verify this was the true response and consistent results were obtained. The only difference between alanine uptake and the transport of AIB, glutamate and GABA, is that the former can also enter the cell through a secondary transporter, MctP. It is possible that the MctP may not be affected by osmotic upshift in the same way Aap and Bra are and so the results shown in Fig. 8.11 may be a combination of alanine transport by

Aap, Bra and MctP. This lead to the investigation of alanine transport via MctP in 3841 cells exposed to an osmotic upshift (section 7.2.4).

Even though the alanine results varied from those obtained with other transport solutes, it had been confirmed that the uptake of all solutes investigated were also affected by osmotic upshift immediately prior to assays, with 200mM tending to be the lowest concentration needed for >50% decrease (although only 37% decrease with alanine). The investigation was then furthered by studying two more known ABC systems in *R. leguminosarum*.

The *myo*-inositol transporter (Int) is an ABC transporter known to be highly induced when inositol is used as the sole carbon source for 3841 (Fry *et al*., 2001). This transporter is highly specific for inositol, although it is known that another constitutive low-affinity transporter also transports inositol in 3841 (Poole *et al*. 1994). However, what class of transporter this low-affinity transporter belongs to is currently unknown. For this experiment 3841 was grown up overnight in AMS (10mM inositol, 10mM NH_4) and also, to save time and resources, 200mM sucrose was the only concentration used to generate an osmotic upshift (Fig. 7.12).

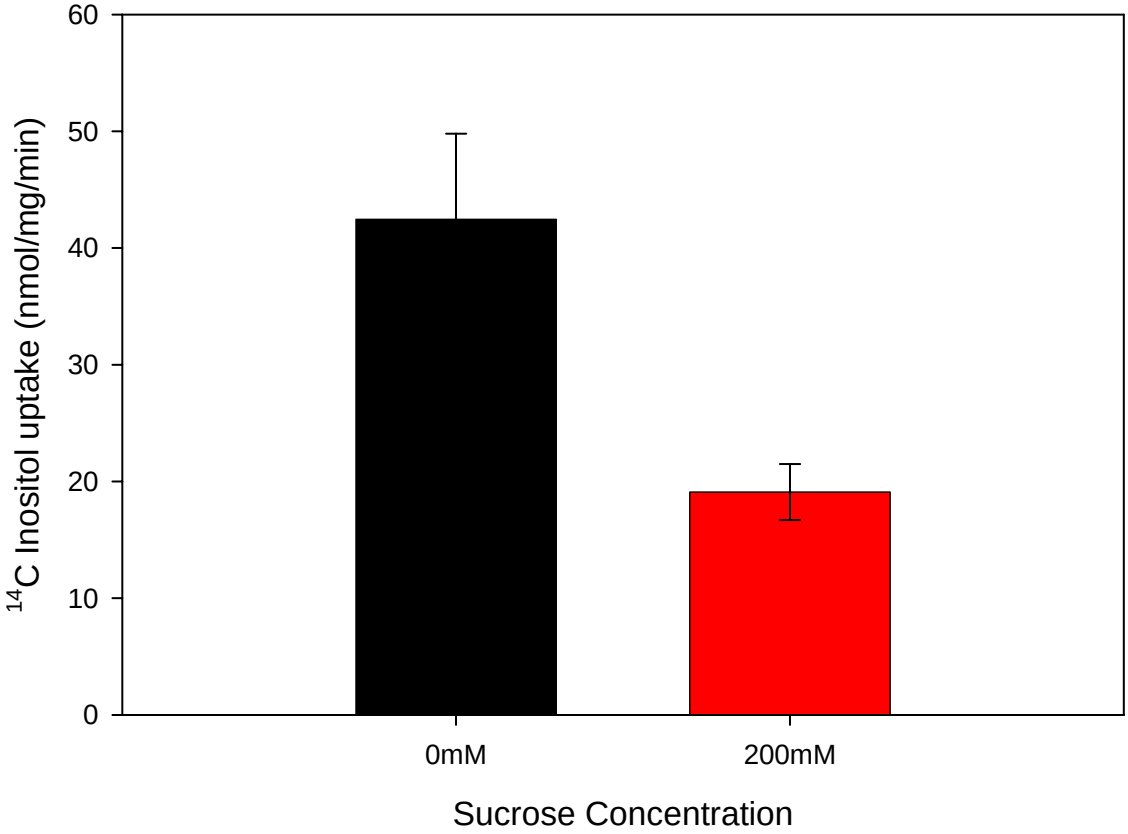


Figure 7.12. Uptake of ^{14}C Inositol on 3841 with and without 200mM Sucrose 10 Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of inositol per minute, over four minutes. Mean results of at least three experiments with standard error values.

The inositol uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Fry *et al*., 2001). The data also matches the pattern made with the previous ABC transporter results presented above; i.e. when 200mM sucrose was added ten seconds before assays, uptake was significantly reduced (<50% uptake than observed with the control cells). As nothing is known about the low-affinity inositol transporter, it is equally unknown if it was affected by sucrose. It is therefore impossible to comment any further as to the effect, if any, the low-affinity transporter had on these data.

The dipeptide permease (dpp) transporter is another ABC transporter found in *R. leguminosarum*. It is responsible for the transport of δ-Aminolevulinic acid (ALA) (Carter *et al*. 2002). As with inositol, ALA uptake was only tested with 200mM sucrose (Fig. 7.13), although overnight cultures were grown in AMS (10mM glc, 10mM NH_4) as with the other assays.

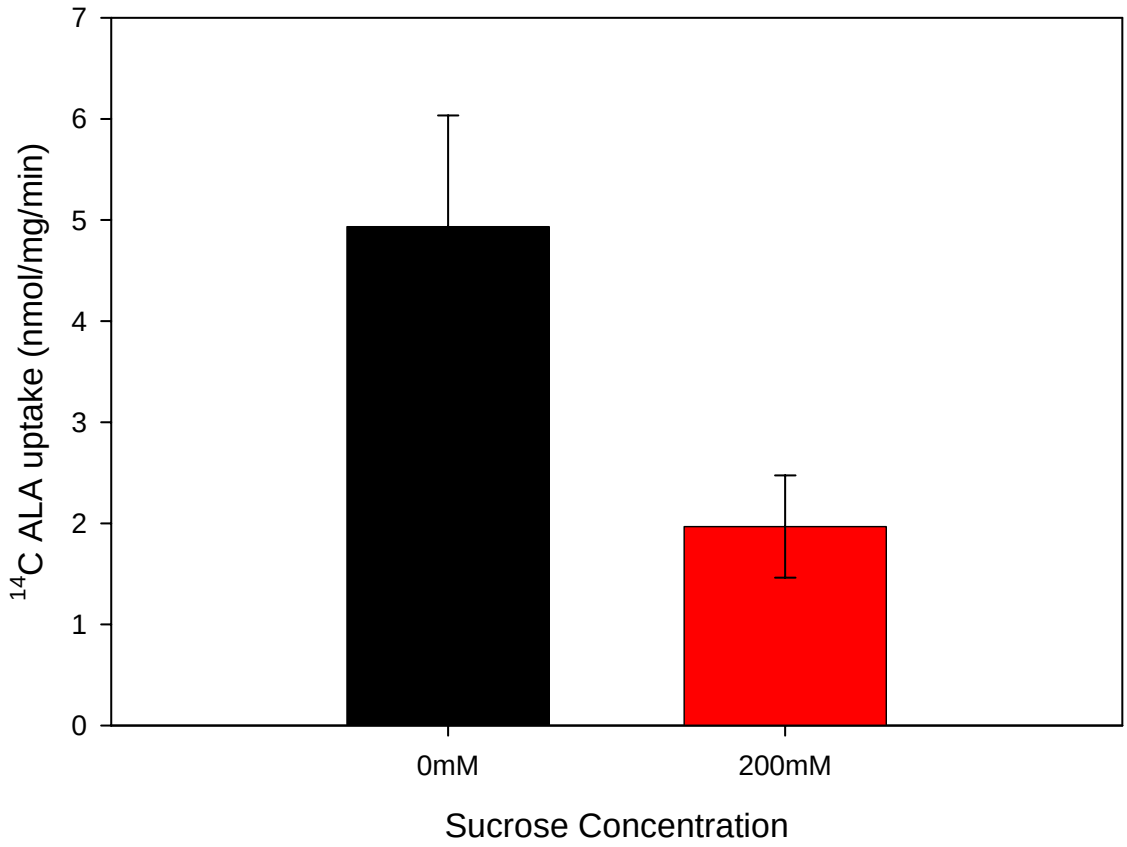


Figure 7.13. Uptake of ^{14}C ALA on 3841 with and without 200mM Sucrose 10 Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of ALA per minute, over four minutes. Mean results of at least three experiments with standard error values.

ALA uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Carter *et al*., 2002). Again, the data acquired here matches the pattern set with the previous ABC transporter results presented above; i.e. when 200mM sucrose was added ten seconds before assays, uptake was significantly reduced (<40% uptake than observed with control cells).

From all these data, it is clear that osmotic upshift (generated by many solutes) at 200mM 10 seconds prior to assays, inhibits uptake of all of the examined ABC transporters. However, as all this data had been collected only on ABC systems, it was unknown if this effect was limited to that class of transporter alone.

7.2.4. Immediate Effect of Osmotic Upshift on Uptake Rates of Solutes using Non-ABC Transporters

As reported above, alanine uptake in 3841 via Aap and Bra was investigated as part of this research, but its inhibition due to osmotic upshift was not as severe as observed with other solutes that are transported by these systems (Fig. 7.3, 7.9 – 7.11). This difference was assumed to be related to MctP, a secondary transporter that also plays a role in the transport of alanine. The monocarboxylate transport (MCT) system has been well characterised in 3841 and is know to be specific for alanine although at a much lower affinity then Aap and Bra (Hosie *et al*. 2002b). Alanine transport occurs through the MctP permease, which is always active but depends on concentration gradients. In order to observe close to maximum uptake via this transporter, alanine concentration was increased from 25µM (0.125µCi) to 500µM (0.5µCi). Uptake rates were measured as before, with and without the presence of 200mM sucrose ten seconds before assays (Fig. 7.14) and overnight cultures were grown up in AMS (10mM glc, 10mM NH_4).

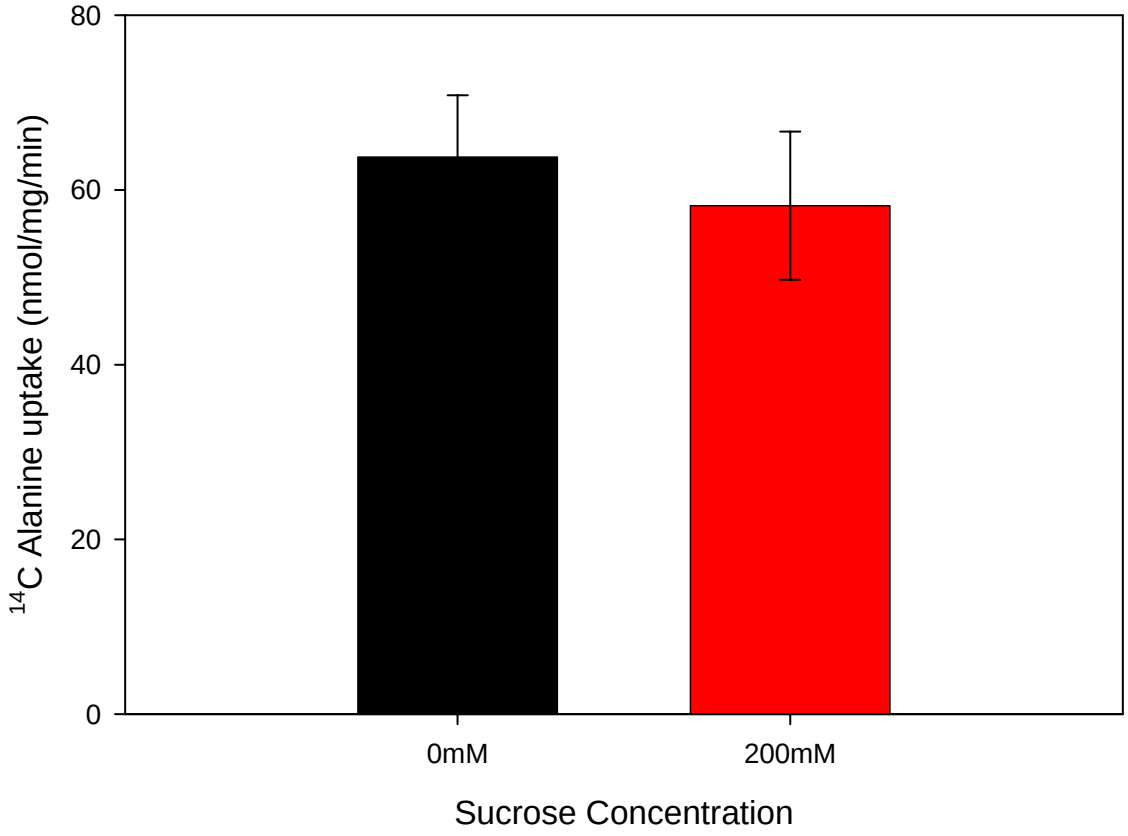


Figure 7.14. Uptake of ^{14}C Alanine (500µM) on 3841 with and without 200mM Sucrose 10 Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of alanine per minute, over four minutes. Mean results of at least three experiments with standard error values.

Uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Hosie *et al*. 2002b). No significant loss of uptake is seen in cells that undergo an osmotic upshift. This initial result indicated that MctP was not affected by osmotic upshift in the same way ABC transporters were. Although it was insignificant, a decrease in uptake of ~6nmol/mg/min was observed in cells exposed to osmotic upshift generated by 200mM. It is worth noting that while this set of assays had an elevated concentration of alanine in order to examine uptake via MctP, Aap and Bra are not inhibited by this and so could still function normally. This means that even though the ABC transporters were saturated by the conditions, they would have still contributed towards the uptake rates seen in Figure 7.14. From Figure 7.11, the control uptake rate of alanine was ~17nmol/mg/min and under osmotic upshift (200mM sucrose) it was ~11nmol/mg/min. This is a decrease of ~6nmol/mg/min, the same as was observed with alanine uptake via MctP. This data indicates that alanine transport through MctP was unaffected by the osmotic upshift caused by 200mM sucrose and the decrease seen is due to inhibition of uptake via Aap and Bra.

If MctP was unaffected by osmotic upshift, then it would have affected the alanine uptake rates via Aap and Bra, reported above (Fig. 7.11). As mentioned above, although MctP could not transport alanine at its optimum rate under the conditions used with Aap and Bra studies (25μM), it would have still made some contribution. The K_m and V_{max} have previously been reported as 560μM and 122nmol/mg/min (Hosie *et al*. 2002b). Using the following equation the uptake rate of alanine (25 μM) via MctP was calculated.

$$\text{Uptake} = \frac{V_{max} \times [\text{substrate}]}{K_m + [\text{substrate}]}$$

$$= (122 \times 25000) / (560000 + 25000)$$

$$= 5.21\text{nmol/mg/min}$$

This means that for the alanine (25 μM) uptake rates via Aap and Bra recorded above, ~5nmol/mg/min was due to transport via MctP and so was not affected by osmotic upshift. In order to take transport via MctP into account and estimate the effect of osmotic upshift on alanine uptake through just Aap and Bra, 5.21nmol/mg/min was subtracted from the rates recorded in Figure 7.11 (Fig 7.15).

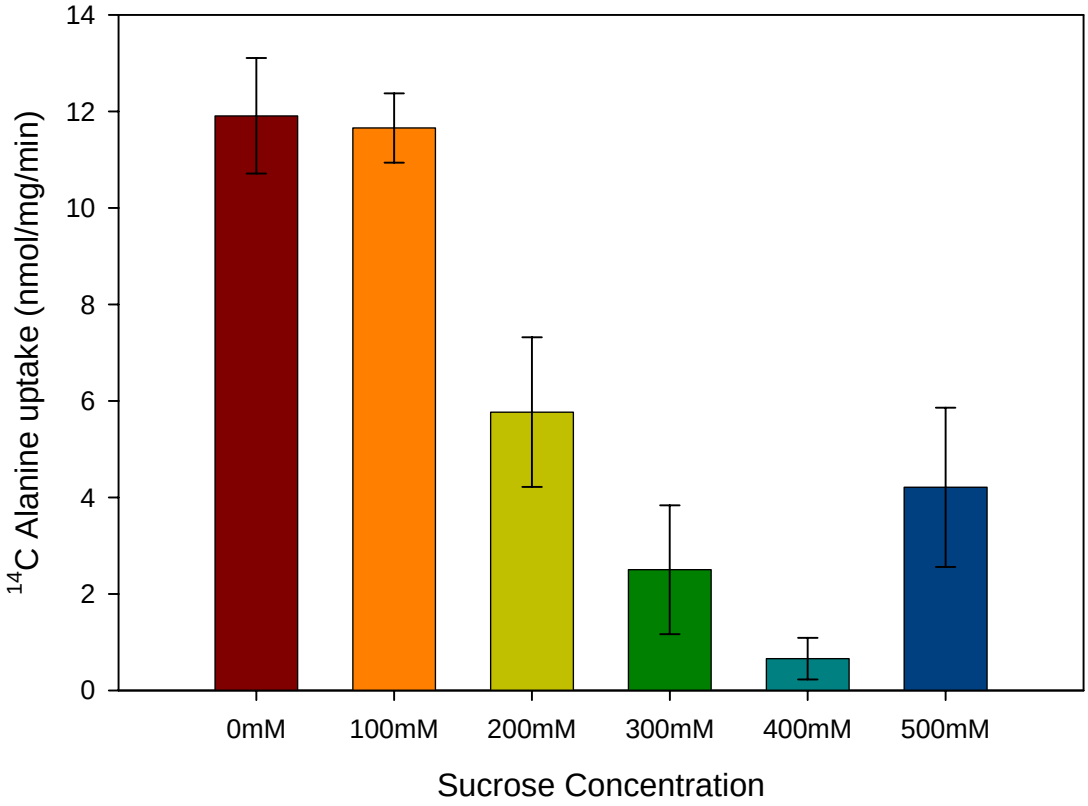


Figure 7.15. Estimated Uptake of ^{14}C Alanine (25 µM) on 3841 via just Aap and Bra Exposed to Different Concentrations of Sucrose 10 Seconds before Assays. Data from Figure 7.11 recalculated to take into account and remove alanine uptake through MctP, which was unaffected by osmotic upshift.

Even with this correction made to the data, the uptake rate of alanine in the presence of 500mM sucrose is still higher then it is at 300mM and 400mM sucrose. Despite that, this estimated data more closely resembles the pattern seen with the transport of the other solutes via Aap and Bra. Of course, this was just an estimate and the only way to determine if this hypothesis was correct, was to conduct alanine uptake assays in a MctP mutant (with 25µM/0.125µCi alanine) and in an Aap/Bra double mutant (with 500µM/0.5µCi alanine). However, suitable mutants were not available at the time of this investigation.

In order to prove that osmotic upshift had no affect on solute uptake via secondary transporters in general and not just MctP, the dicarboxylate transport (DCT) was also investigated. The DCT system system is well characterised in 3841 and is know to be specific for succinate, fumarate and L-malate (Reid & Poole, 1998). Succinate induces its own transport through the DctA permease, another secondary transporter. Succinate uptake through this permease was investigated with and without 200mM sucrose added ten seconds before assays were performed (Fig.7.16). Overnight cultures for this experiment were grown up in AMS (10mM succinate, 10mM NH_4), in order to induce the Dct system.

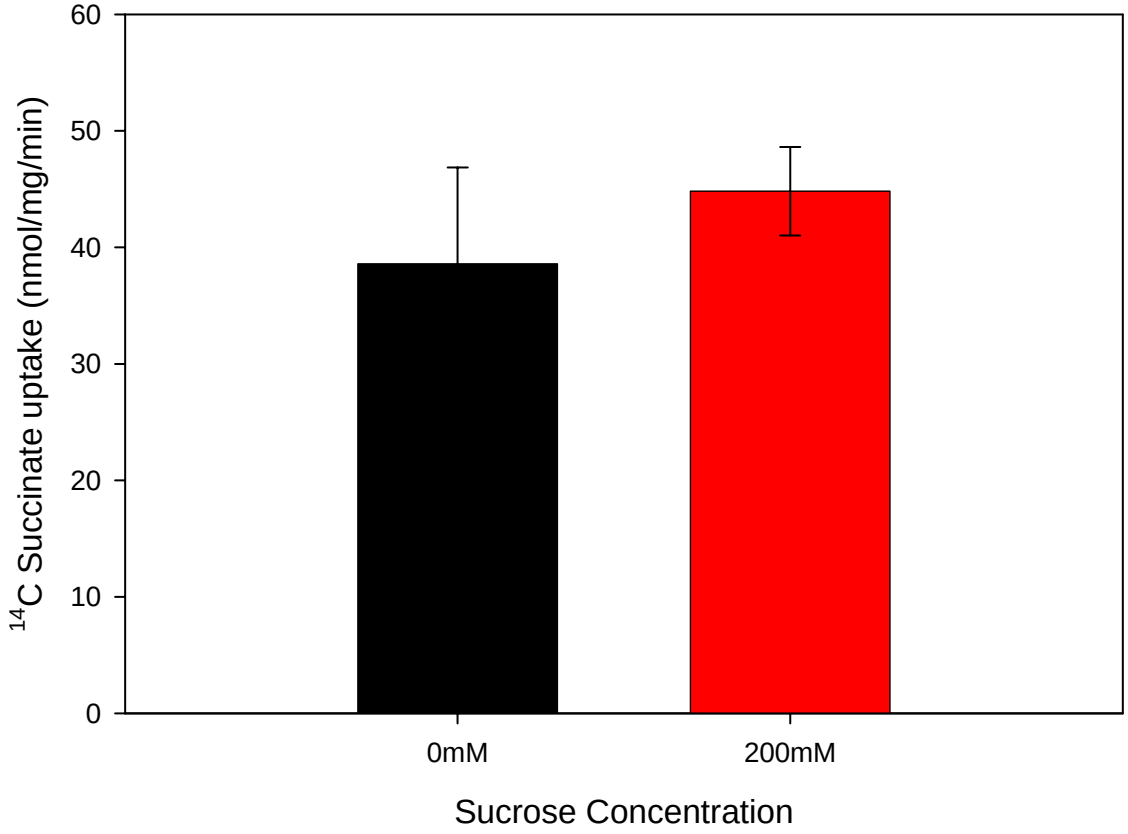


Figure 7.16. Uptake of ^{14}C Succinate on 3841 with and without 200mM Sucrose, Ten Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of succinate per minute, over four minutes. Mean results of at least three experiments with standard error values.

Uptake rates recorded from the control cultures were not significantly different to rates that had previously been reported (Reid *et al*. 1996). Just like the data obtained with alanine uptake through MctP, no significant difference was seen in succinate transport in cells that were shocked and cells that were not. This indicates that the Dct system is not inhibited by osmotic upshift generated by addition of 200mM sucrose ten seconds before assays were performed. AIB uptake assays were also conducted with these cells, i.e. grown in AMS (10mM succinate, 10mM NH_4), both the control and the culture with 200mM sucrose added ten seconds before assays were performed. As expected uptake rates were 87% decreased in cells with 200mM sucrose exposure compared to the control culture (Fig. 7.17). AIB uptake rates were lower in cells grown up in AMS (10mM succinate, 10mM NH_4) then those from obtained with cells grown up overnight in AMS (10mM glc, 10mM NH_4) (Fig. 7.3), but this too was expected as it has previously been recorded (Hosie, personal communication).

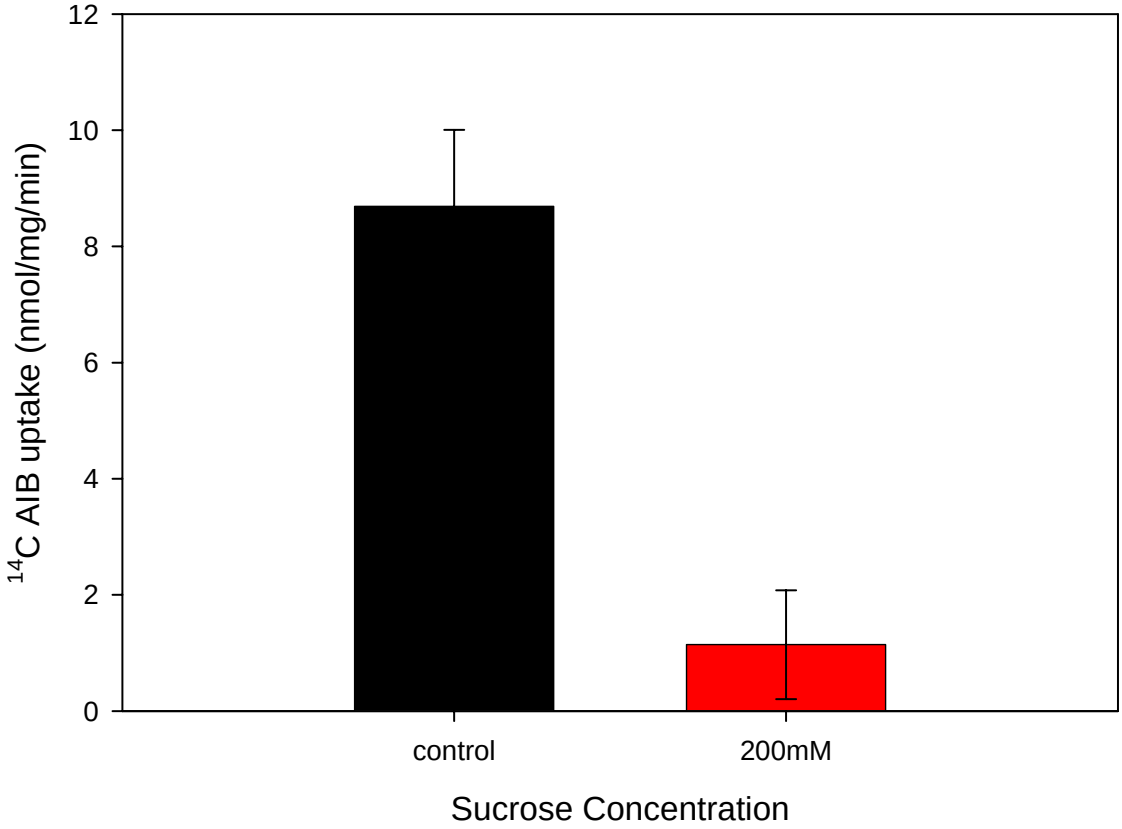


Figure 7.17. Uptake of ^{14}C AIB on 3841 with and without 200mM Sucrose, Ten Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Cells used were grown up overnight in AMS (10mM succinate, 10mM NH_4) instead of AMS (10mM glc, 10mM NH_4) as had previously been used above. Mean uptake of AIB per minute, over four minutes. Mean results of at least three experiments with standard error values.

In total, these data indicate that solute uptake through ABC systems is inhibited by osmotic upshift, whereas transport via secondary transporters is not. This fact was taken into consideration when glucose uptake was investigated.

Glucose transport in *R. leguminosarum* is believed to be carried out by at least two uptake mechanisms (de Vries *et al*., 1982). Although the class of transporters could not be identified, de Vries reported a loss in glucose transportation in cells after a 600mM sucrose shock. Due to this, glucose uptake was examined with and without 200mM sucrose added ten seconds before assays were performed (Fig.7.18). Overnight cultures were grown up in AMS (10mM glc, 10mM NH_4)

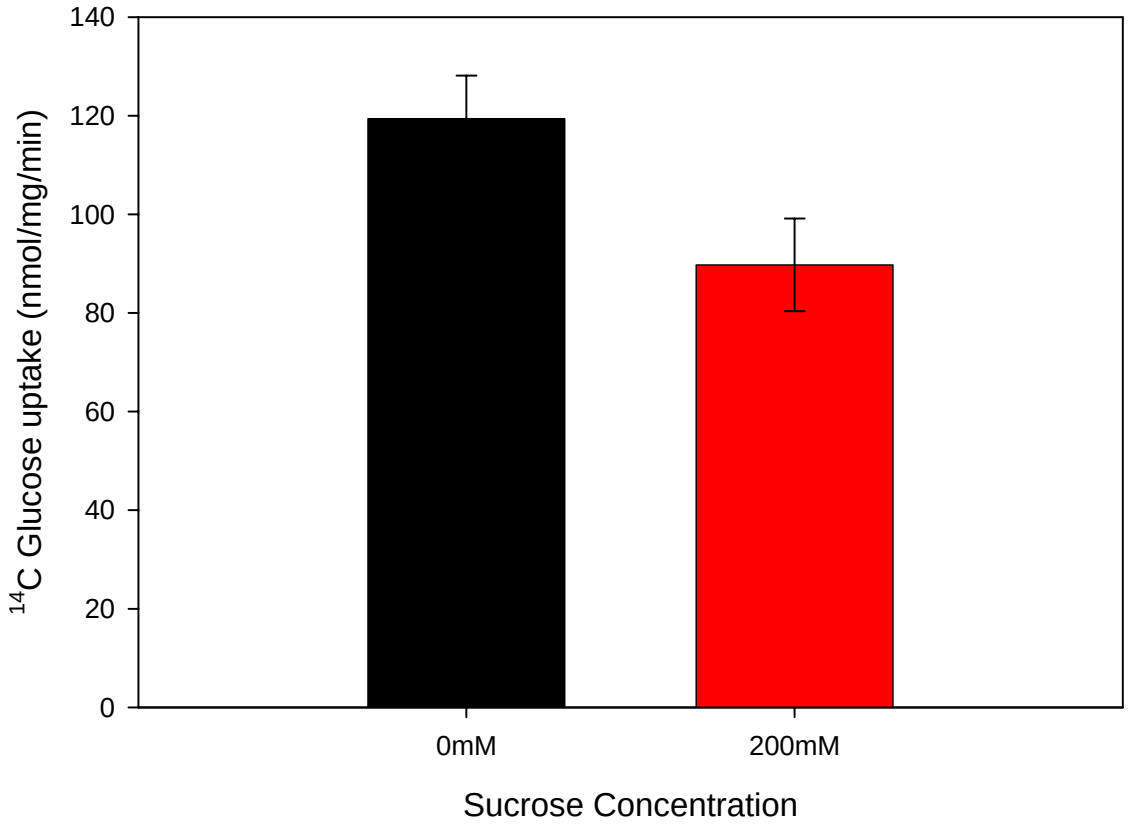


Figure 7.18. Uptake of ^{14}C Glucose on 3841 with and without 200mM Sucrose, Ten Seconds before Assays. Cells were exposed to the concentration of sucrose indicated on the x-axis immediately before assay. Mean uptake of glucose per minute, over four minutes. Mean results of at least three experiments with standard error values.

In comparison to the control culture, the rate of glucose uptake decreased by 25% when exposed to an osmotic upshift. As glucose transport is inhibited by the osmotic upshift, it indicates that its uptake is through an ABC system. However, as this decrease is less than that seen with the other solutes, it also suggests that another non-ABC system is also involved in glucose uptake in 3841. This system clearly has a higher affinity for glucose than MctP has for alanine, as there is a relatively high rate of uptake at 25µM of glucose. Further study is clearly required in order to confirm the hypothesis that glucose uptake in *R. leguminosarum* was via an ABC system and a secondary transporter.

All of the data collected throughout this research on the effect of 200mM sucrose shock has been collected and is shown in Figure 7.19.

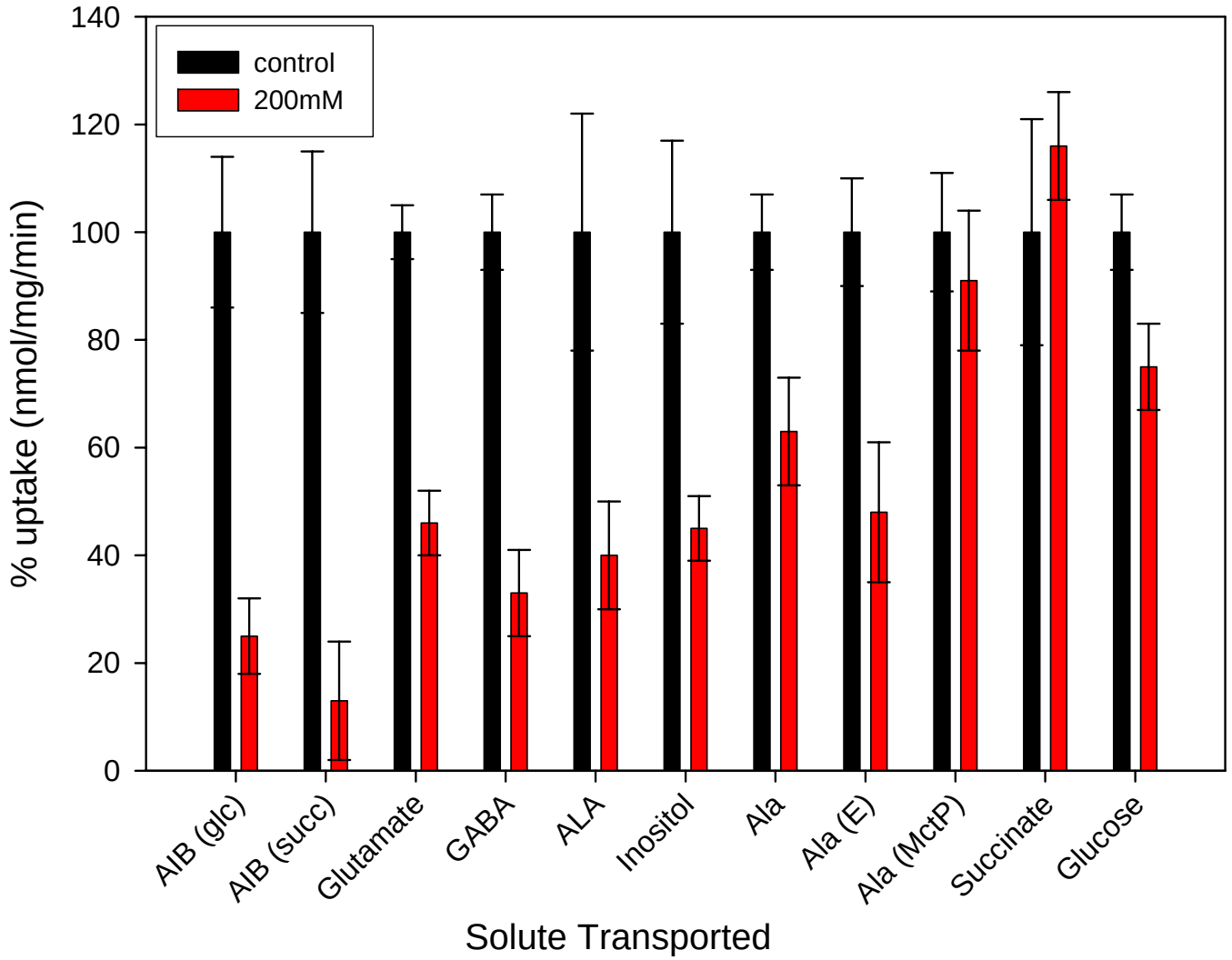


Figure 7.19. Percentage Uptake of Different Solutes with and without 200mM sucrose. Uptakes are given as a percentage compared to those from cultures with no additions (black bars) and those with osmotic upshift (red bars). All of the data was presented individually above.

Key:
AIB (glc) = AIB uptake from cultures grown overnight in AMS (10mM glc, 10mM NH_4)
AIB (succ) = AIB uptake from cultures grown overnight in AMS (10mM succinate, 10mM NH_4)
Ala = alanine uptake
Ala (E) = estimated alanine uptake via Aap & Bra with no MctP
Ala (MctP) = alanine uptake via MctP (500µM)

7.2.5. Effect of 200mM Sucrose on Cells

The next question that was addressed was; how did the osmotic upshift affect the cells and bring about the decrease in solute uptake via ABC transporter? It is well known that osmotic downshift causes the loss of SBPs from ABC systems. Could the addition of 200mM sucrose ten seconds before assays were performed have disrupted cell membranes? Such an upshift could cause cells to lose periplasmic components, such as the SBPs. Loss of SBPs would decrease the efficiency of ABC systems and may have caused the loss of solute uptake through these transporters.

In order to determine if periplasmic components are released on exposure to osmotic upshift, the method for collecting periplasmic fractions (see Chapter 2, section 2.13.1) was adapted. Two cultures of 3841 were grown up overnight in AMS (10mM glc, 10mM NH_4), spun down and washed as per the protocol. However, instead of resuspending both pellets in 10ml of Tris-HCL pH 8 with 20% sucrose and 1mg/ml of lysozyme; one was resuspended in 10ml of AMS (10mM glc, 10mM NH_4) (negative control) whilst the other was resuspended in 10ml of AMS (10mM glc, 10mM NH_4) + 200mM sucrose. Cells were left at 26°C for five minutes (as similar to conditions of cells prior to uptake assays as possible) before being spun down and the supernatant from both cultures was collected. AIB and alanine (via MctP therefore 500μM/0.5μCi alanine present) uptake assays were then carried out on the cell components of both cultures, which produced results not were not significantly different to those reported above (Fig. 7.20).

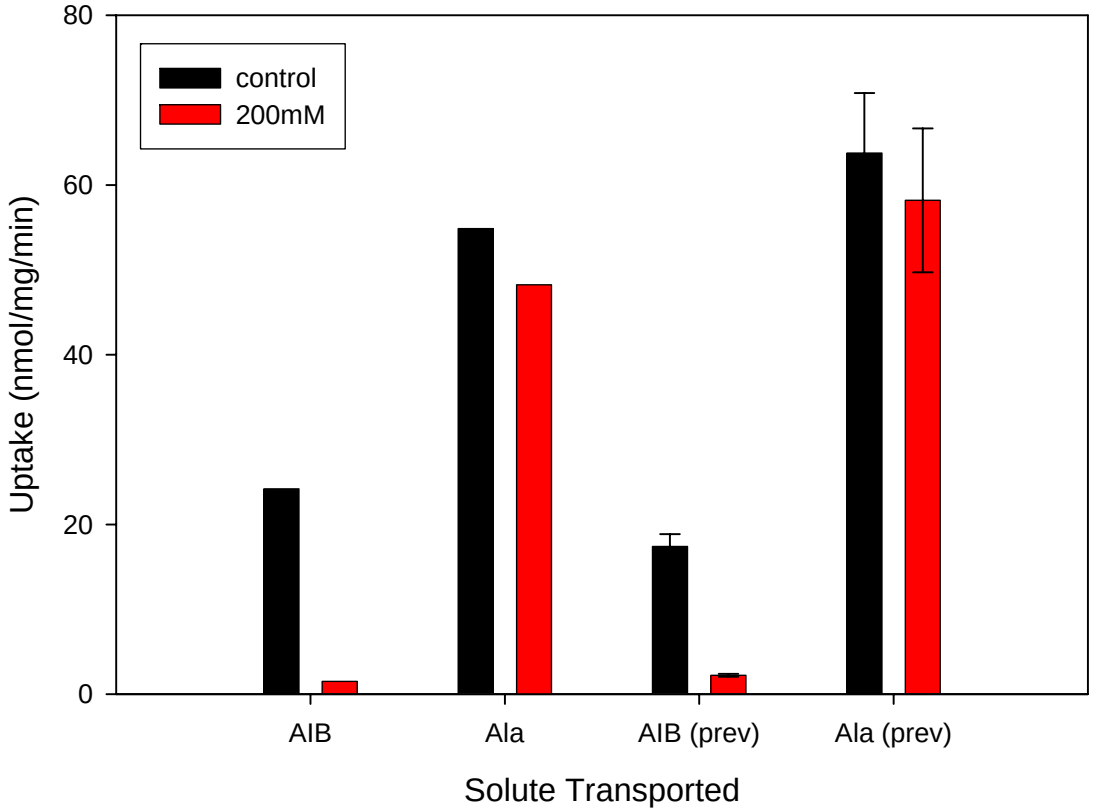


Figure 7.20. Uptake of ^{14}C AIB on 3841 with and without 200mM Sucrose, Ten Seconds before Assays. Cells were exposed to 200mM sucrose (red bars) immediately before assay or not (black bars). Mean uptakes per minute, over four minutes for AIB or alanine (Ala) via MctP. Results on right (labelled prev) are reproduced from Figs. 7.3 and 7.14.

As a positive control, periplasmic proteins were isolated from another culture grown in AMS (10mM glc, 10mM NH_4) by lysozyme / EDTA treatment (see Chapter 2, section 2.13.1). Samples from each of the three conditions were run on an SDS-PAGE gel (Fig. 7.21).

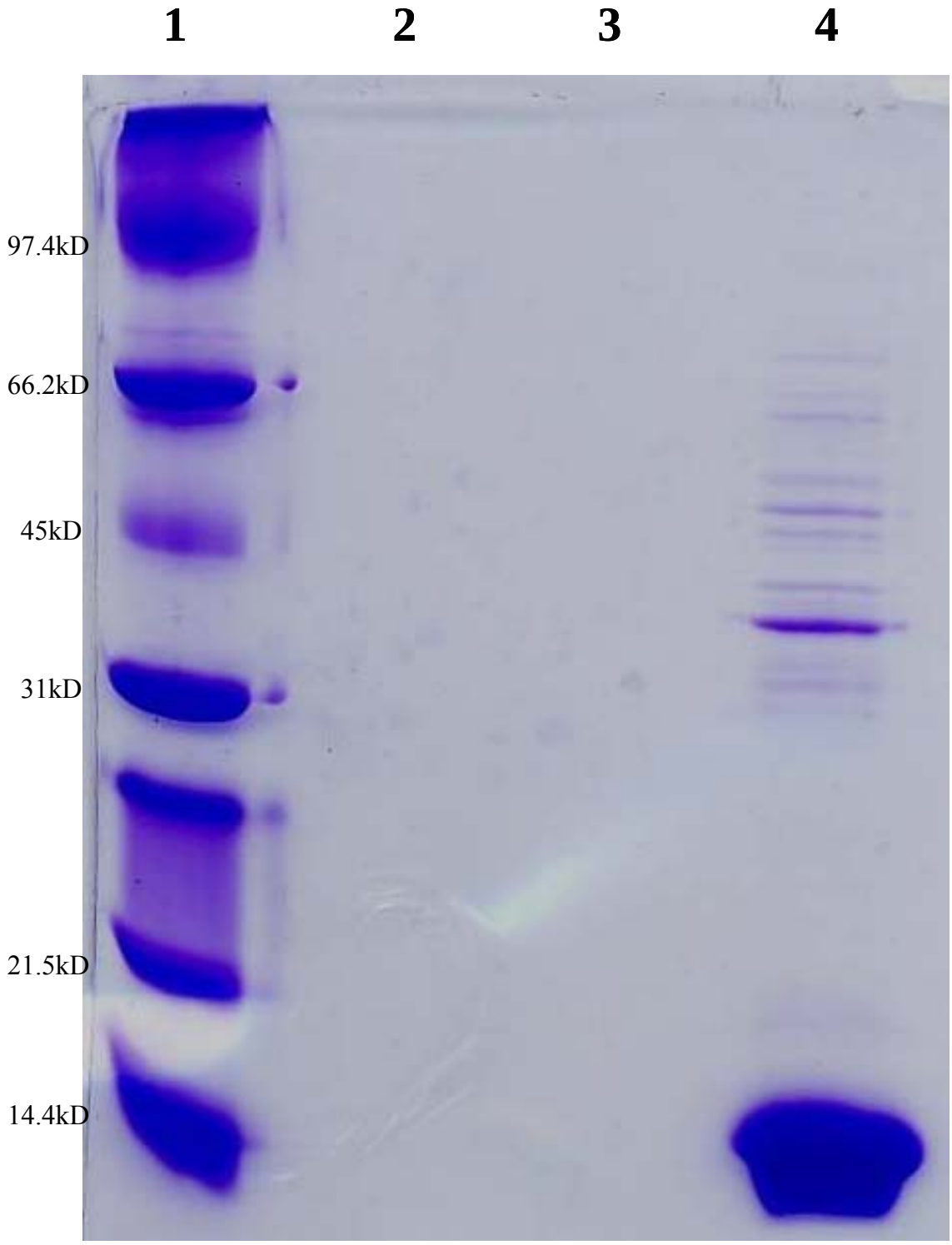


Figure 7.21. SDS-PAGE Gel of Periplasmic Fractions of 3841. Lanes: 1) Low range protein marker (sizes shown) 2) Fraction from untreated cells (negative control) 3) Fraction from cells treated with 200mM sucrose AMS 4) Fraction from cells treated with lysozyme and EDTA to release periplasmic contents (positive control).

When compared on the SDS-PAGE gel to the negative and positive controls, no proteins could be seen in the cells exposed to 200mM sucrose. It is therefore clear that 200mM osmotic shock does not cause periplasmic fractions to leech from cells.

If cell membranes remain intact when exposed to an osmotic upshift, then SBPs cannot be removed and so all of the components required for ABC transporters would be present, they just do not function as well. This posed the question of whether cells could recover on removal of the immediate hyper-osmotic shock considering none of their components are lost.

To determine this, two sets of cultures grown up overnight in AMS (10mM glc, 10mM NH_4) which were both spun down and washed as usual. However, on the final resuspension, one culture was resuspended in RMS as usual whilst the other was resuspended in RMS + 200mM sucrose. AIB uptake assays were then immediately carried out on both sets of cultures (i.e. there was no hour starvation period). As soon as assays had been carried out, both sets of cells were washed and resuspended in standard RMS, which removed the hyper-osmotic shock. In order to wash the cells as quickly as possible, 10ml of each culture was taken, split into 1ml aliquots and spun in a microcentrifuge for 5 minutes at 13,000rpm. Aliquots were washed in RMS in the before being pooled and final resuspension in standard RMS to an OD_{600} of ~1. AIB uptake assays were then carried out on both sets of cultures (approx. 20 minutes after the first set of assays). Assays were also carried out 40 minutes and 60 minutes after the initial assays were performed (Fig 7.22). In between these sets of assays, cells were left shaking at 60rpm at 28°C.

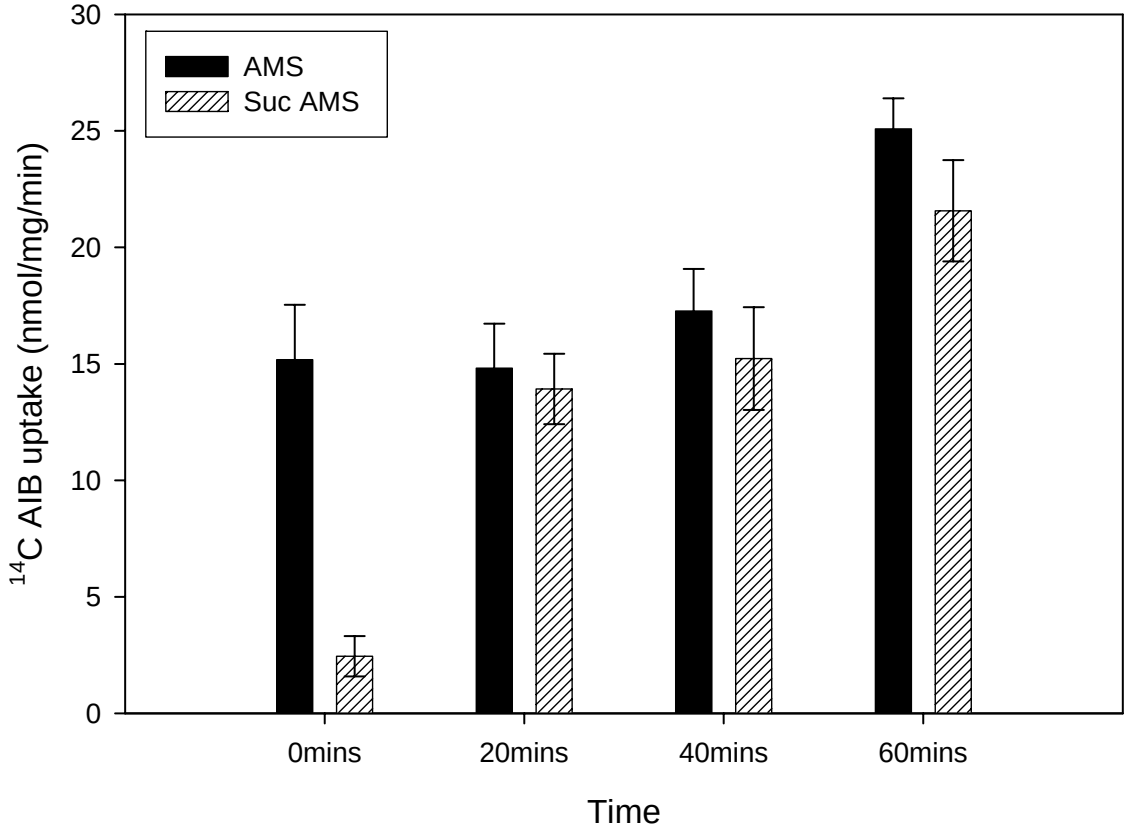


Figure 7.22. Uptakes of ^{14}C AIB on 3841 with Osmotic Upshift and after Removal of Shock. Hatched bars indicate the culture that was exposed to 200mM sucrose, black bars indicate the control culture. Osmotic upshift was only present in the 0 minute assay and was removed for subsequent assays. Assays were performed every 20 minutes, as indicated by x-axis, after the initial assay. Mean uptake of AIB per minute, over four minutes. Mean results of at least four experiments with standard error values.

These data clearly shows that AIB uptake rates are restored to normal on removal of the hyper-osmotic shock, as no significant difference is seen between the rates of the control cultures and the rates of the test cultures.

7.2.6. Spheroplast and Bacteroid Data

As mentioned above, previous reports have shown no solute uptake in spheroplasts or bacteroids via ABC systems, although it is known that these transporters are active and expressed in both (Djordjevic *et al.*, 2003; Dupont *et al.*, 2004). Upon their generation, spheroplasts are kept in media containing 20% (585mM) sucrose, because they are cells with weakened outer cell membranes and so they are stored in hyper-osmotic conditions to prevent them from bursting through osmosis. It has already been shown that an addition of 200mM sucrose ten seconds before assays were performed severely inhibited uptake through ABC systems and that 500mM sucrose almost abolished all uptake in the systems tested. These data was considered when investigating uptake via ABC systems in spheroplasts.

In work conducted by White (unpublished), three cultures were grown up in tandem overnight in AMS (10mM glc, 10mM NH_4); one was washed and resuspended in RMS (negative control), one was washed in RMS and resuspended in RMS + 20% sucrose and the third was treated to create spheroplasts, as described by Hosie *et al*. (2002b). AIB uptake was measured in each as well as alanine uptake via MctP (500μM/0.5μCi alanine present) to check the spheroplasts were still viable (Fig. 7.23).

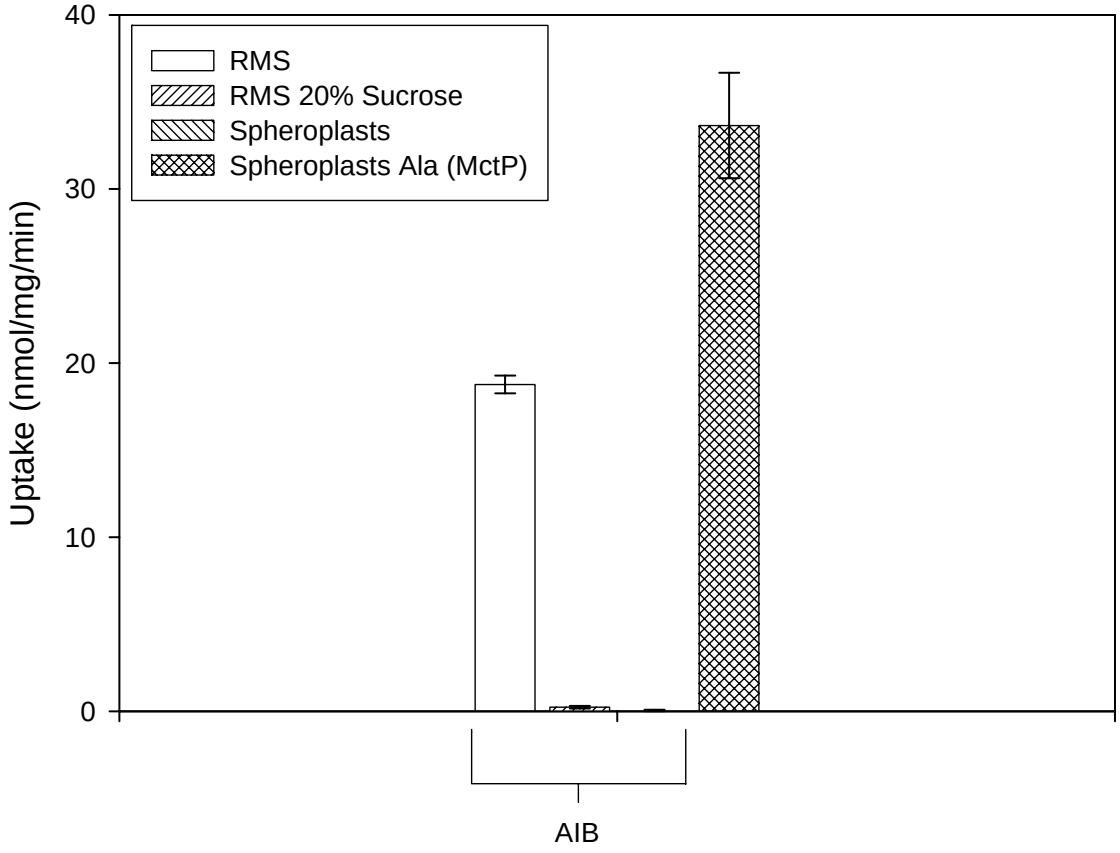


Figure 7.23. Uptake of ^{14}C AIB on Untreated 3841, Sucrose Exposed Cells and Spheroplasts. AIB uptake from cells resuspended in RMS, cells resuspended in RMS + 20% sucrose or spheroplasts (as indicated). Alanine (500μM) uptake via MctP (cross-hatched bar) shows spheroplasts had not burst. Mean uptake per minute, over four minutes. Mean results of at least three experiments with standard error values (White, unpublished).

As can be seen, the amount of sucrose used in spheroplast generation almost eradicates AIB transport, even without the presence of lysozyme. It is known that no uptake was seen via ABC transporters in spheroplasts as they had no SBP components. Whilst this is true, it can be seen here that the amount of sucrose used in spheroplast generation almost totally inhibits uptake through ABC systems in cells containing SBPs.

Upon their isolation, bacteroids are also kept in media with a high amount of sucrose (300mM) and so a pilot experiment that investigated their uptake of succinate and glutamate was performed by Hosie & Lodwig (unpublished). As outlined above, in *R. leguminosarum* succinate is transported by a secondary transporter (DctA) and glutamate is transported by the Aap and Bra ABC systems. In order to fully compare secondary transporters to ABC systems, a double knock-out mutant of Aap and Bra was made. Also, a glutamate permease gene (*gltP*) was taken from *E. coli* and cloned to the pJP2 vector (pRU976). pRU976 was transferred into the double mutant so that glutamate transport could be achieved though the

secondary transporter GltP. pJP2, not containing the *gltP* gene, was also transferred into both wild-type and the double mutant in order to show that the pJP2 was not responsible for any change in uptake rates. Five strains of *R. leguminosarum* were used in total to inoculate pea plants; wild-type, an Aap/Bra double mutant, wild-type containing pJP2, Aap/Bra double mutant containing pRU976 and Aap/Bra double mutant containing pJP2. After the pea plants were left to grow for a suitable amount of time, bacteroids were isolated from their root nodules using the methods of Lodwig *et al*. (2003). Succinate (Fig. 7.24) and glutamate (Fig. 7.25) uptake assays were performed on these isolations as described above. All this work was conducted in *R. leguminosarum* A34 not *R. leguminosarum* 3841.

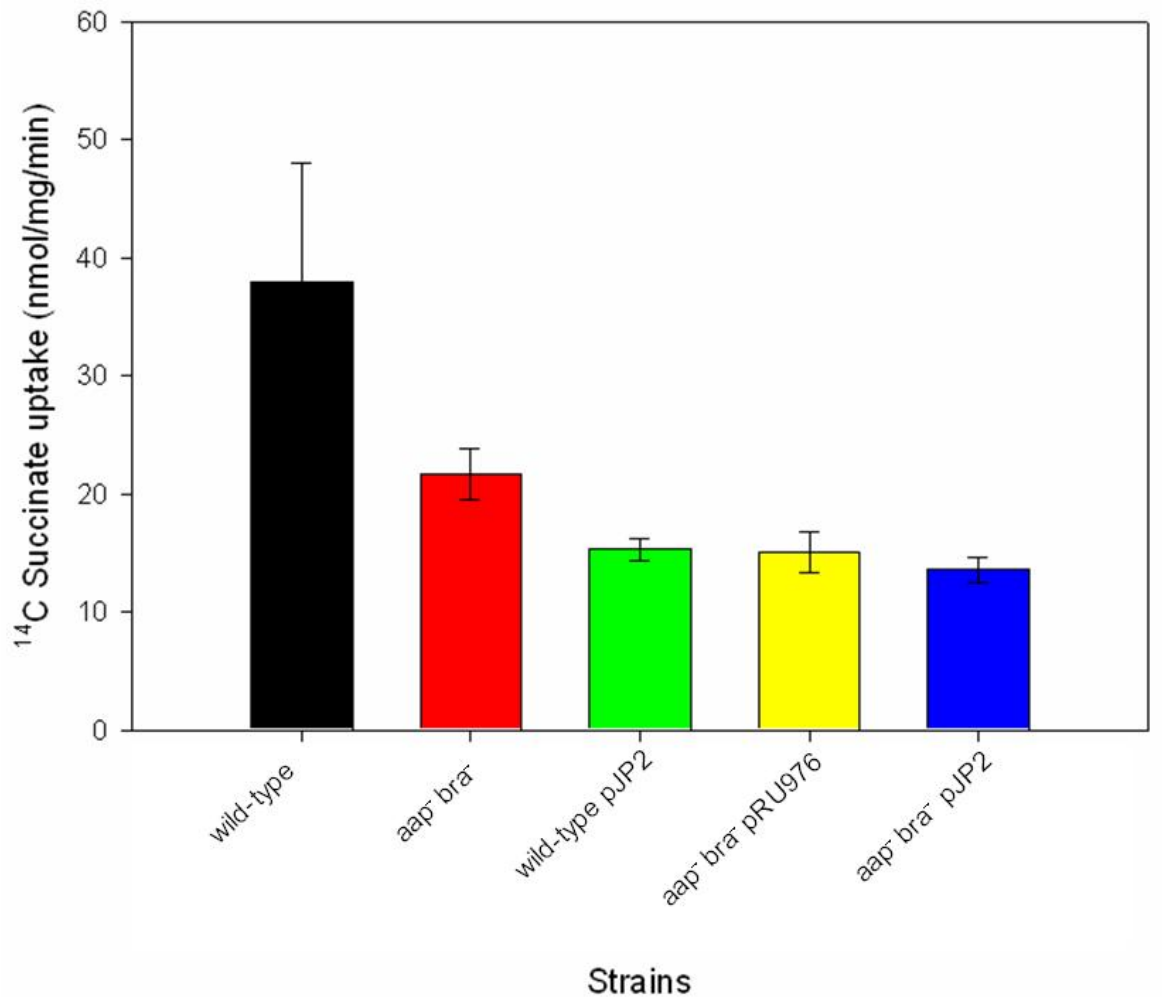


Figure 7.24. Uptake of ^{14}C Succinate on the Bacteroids Isolates of Pea Plants Inoculated with Various Strains of *R. leguminosarum*. Uptakes rates recorded from isolates of wild-type (black bar), Aap/Bra double mutant (red bar), wild-type with pJP2 (green bar), Aap/Bra double mutant with pRU976 (*gltP*) (yellow bar) and Aap/Bra double mutant with pJP2 (blue bar). Mean uptake per minute, over four minutes. Mean results of at least three experiments with standard error values (Hosie & Lodwig, unpublished).

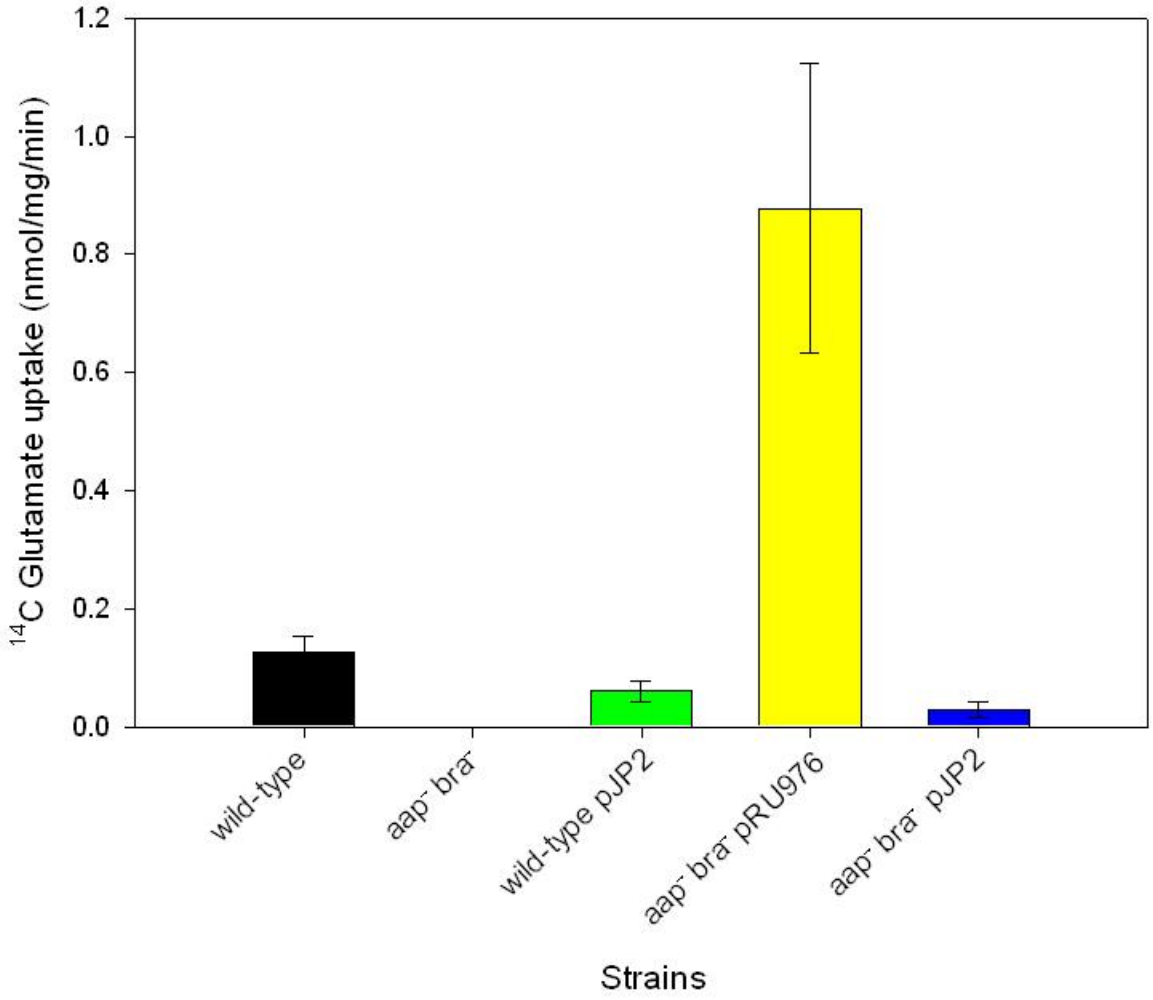


Figure 7.25. Uptake of ^{14}C Glutamate on the Bacteroids Isolates of Pea Plants Inoculated with Various Strains of *R. leguminosarum*. Uptakes rates recorded from isolates of wild-type (black bar), Aap/Bra double mutant (red bar), wild-type with pJP2 (green bar), Aap/Bra double mutant with pRU976 (*gltP*) (yellow bar) and Aap/Bra double mutant with pJP2 (blue bar). Mean uptake per minute, over four minutes. Mean results of at least three experiments with standard error values (Hosie & Lodwig, unpublished).

These data shows that succinate transport in bacteroids was not significantly affected by the loss of Aap or Bra, nor was it affected by the addition of pJP2 or pRU976. The uptake rates were lower than those reported earlier in free-living cells (Fig. 7.16), but this was attributed to the harsh nature of bacteroid isolation that may have damaged cells and so hindered solute transportation. Glutamate transport was not significantly detected in any strain apart from the one that expressed the *E. coli* GltP secondary transporter. Whilst the Aap/Bra double mutant lacked the ability to transport glutamate, the wild-type strain was fully functional but was shown to be unable to transport succinate. This indicates that uptake through ABC systems in bacteroid isolates is inhibited, but no inhibition is observed with secondary transporters. Initially, this loss of transportation was also attributed to the harsh nature of bacteroid isolation that may have damaged cells (possibly by the loss of

SBPs) and so hindered solute transportation. However, work presented here regarding sucrose inhibition of ABC systems may shed new light on these findings.

7.3. Discussion

Although de Vries (1982) reported a loss of glucose transport following high osmotic shock, this research has examined the effect in much greater detail. The data presented here clearly shows hyper-osmotic shock can cause a decrease in the rate of solute uptake. The magnitude of this effect was dependent both on the concentration of osmotic stress, and on the time of exposure. If a stress was introduced at 200mM, an immediate decrease in solute uptake of at least 50% was seen in the majority cases. Also, this effect was generated regardless of the type of compound used to produce the osmotic upshift.

Interestingly this phenomenon appeared to be specific to ABC transporters, as the MCT and DCT systems (both of which use secondary transporters) were not significantly affected by osmotic upshift. There are two major differences between secondary and ABC transporters; the latter require a SBP to function and are driven by ATP instead of by proton coupling and concentration gradients.

With so many different compounds tested, it was unlikely that the any of the molecules used to cause the osmotic upshift were directly responsible for blocking a SBP from interacting to either its solute or its IMP complex. SDS-PAGE gels showed no traces of proteins in the supernatant of cells exposed to the osmotic shock, proving that the osmotic upshift did not cause cell lysis and so SBPs were not lost from the *Rhizobium*. Furthermore, uptake rates were restored to match the rates from unstressed cells immediately on the removal of the hyper-osmotic shock proving that no component from an ABC system was lost during the osmotic upshift. *Rhizobium* must therefore respond to the hyper-osmotic shock immediately and this osmotic upshift prevented ABC systems from functioning affectively.

One possibility is that on encountering an osmotic upshift of severe enough magnitude, *R. leguminosarum* cell membranes immediately contract in an attempt to close any pores. This would prevent cells from losing water but would also cause the inner and outer membranes to come together during their constriction. This in turn would reduce the periplasmic volume and effectively trap any free proteins between the membranes. As SBPs are periplasmic, under these conditions their movements would be restricted and as a result their functionality might be compromised. This might explain why uptake rates are decreased (but not abolished) with ABC systems, but not with other classes of transporter. On removal of the hyper-osmotic stress, membranes returned to their regular state, freeing SBPs and so solute uptake via ABC transporters could return to normal, as reflected by the data presented here.

This is just a hypothesis; the inhibition may also be due to energy coupling aspect that is specific to ABC systems, e.g. ATP generation might be affected. Alternatively, SBPs may 'dock' with their appropriate IMP complex during the osmotic upshift.

An immediate cellular response to an osmotic effect has been well characterised with mechanosensitive channels, which under hypo-osmotic stress must react in milliseconds, else cells would burst with the rapid influx of water (Stokes *et al.*, 2003).

The slight differences in the decreased uptake caused by each of the compounds tested can be explained by the differences in osmolarity generated by each compound. This is because osmotic effects are colligative and depend on the total solute present. Ionic molecules such as NaCl and KCl will disassociate into their ions whilst sugars and polyols remain intact. This in effect increases the amount of solute present when ionic molecules are used and was observed with NaCl and KCl, both of which caused a greater inhibition of uptake rate at 200mM in comparison the other osmolytes tested (Fig. 7.8).

Likewise, differences in uptake rates between solutes may be explained by the number of systems capable of transport and/or the specific binding affinities between each SBP and each solute/IMP. Further protein work is clearly required to confirm this.

Although the mechanism behind the inhibition is not completely known, the effect itself is clear. This technique could prove a simple and inexpensive method for determining whether or not an uptake system is an ABC transporter, much in the same way vanadate can be used (Urbatsch *et al.*, 1995). This technique was shown when glucose uptake was investigated (Fig. 7.18) and indicated its transport by at least one ABC system and at least one secondary transporter.

These data also raised an interesting question regarding spheroplasts and bacteroid studies. All previous work with spheroplasts and bacteroids has involved resuspending them in a high percentage of sucrose, to protect them from their environment. In all of these studies no uptake through ABC systems was observed, although it was known that the transporters are expressed in both. Indeed, mutant studies have shown that not only are ABC systems active in nodules but that in bacteroids they play a crucial role in the amino acid cycle vital to *Rhizobium*-legume symbiosis (Lodwig *et al.*, 2003).

In the case of spheroplasts, SBPs would not have been present and so transport would be abolished, however, the media used in spheroplast generation contained an amount of sucrose (600mM), which would almost totally abolish transport through ABC systems regardless of the presence of SBPs. Indeed, transport rates obtained with spheroplasts and with whole cells suspended in 600mM sucrose were virtually identical.

In the case of bacteroids, SBPs should not have been lost during their isolation (although it remained a possibility) and it has previously been hypothesised that the bacteroid isolation process is so harsh on the bacteria that cell functions are greatly disrupted, which is why no transport can be observed. However the data presented here shows that the presence of sucrose in the suspension media would have caused the lack of uptake via ABC systems. Therefore, the data presented here (proving ABC systems are inhibited by high amount of sucrose) may invalidate any previous observations concerning the loss of transport via ABC systems in spheroplasts or bacteroids. It also suggests that refinements need to be made to the protocols used in spheroplast/bacteroid isolation in order to find a balance between protecting cells from osmotic pressure and to allow uptake assays to be performed uninhibited. This could prove difficult to accomplish though as a certain amount of sucrose (or any other osmolyte) is required in isolation media in order to protect cells from bursting under osmotic pressure. Only further spheroplast/bacteroid studies will help to refine the isolation techniques and uptake assays performed on these cells.

Data presented in this report appear to have created a paradox. Initial data (Chapters 3 & 4) clearly showed that an ABC system (termed QAT6) was specifically induced under hyper-osmosis. Further data (Chapter 6 and this chapter) clearly shows that transport via ABC systems is inhibited by osmotic upshift. Why would *Rhizobium* induce a system under conditions where it would be inhibited?

Another problem that arises from this data is that identifying hyper-osmotically induced uptake in *R. leguminosarum* 3841 could prove difficult, a fact proved in the previous chapter. At this time it is unknown how widespread the effect of hyper-osmotic shock on ABC transporters is and whether or not any other organisms display a similar phenotype. Many studies have demonstrated *S. meliloti*'s ability to transport osmoprotectants into its cells on exposure to hyper-osmotic media (Bernard *et al.* 1986). It may be that *S. meliloti* does not have the same hindrances as *R. leguminosarum* regarding osmotic upshift and ABC transport systems. However, further studies have determined that the primary transporter of glycine betaine and proline betaine in *S. meliloti* is the BetS transporter (Boscari *et al*. 2002). This is not an ABC system, but a secondary transporter. (As previously mentioned, no gene with significant sequence identity to *betS* could be found in the preliminary genomic sequence of 3841.) If *S. meliloti* does have the same properties as *R. leguminosarum* then the BetS would not be effected by hyper-osmosis, just as the DCT/MCT system is unaffected in 3841. Indeed, another study by the same group has shown that the Cho ABC transporter, responsible for the uptake of choline (a glycine betaine precursor), is inhibited by the presence of NaCl (Dupont *et al.*, 2004). Therefore it

is possible that in *Rhizobium*, all ABC systems are inhibited by osmotic upshift and other transporters are responsible for osmoprotectant uptake.

Clearly further study, and maybe refinements to the current protocols, is required to fully understand and investigate transport in *R. leguminosarum* under osmotic upshift.

CHAPTER 8: FINAL DISCUSSION

8.1. LB3 Screening Results and Characterisation of Fusions

Thirty-two fusions were isolated during the screening of the LB3 promoter probe library. The focus of this project was then concentrated on those fusions induced by hyper-osmotic and/or acidic stress. Obtaining sequencing data for each fusion and the release of the preliminary genome of *R. leguminosarum* 3841 allowed the genes (or operons) that were most likely induced by the stressful conditions to be identified. The sequences of each gene were then analysed using the BLAST and Pfam software packages, which allowed putative functions to be assigned to the product of each gene. This, along with the induction data for each fusion, permitted them to be assigned to a general or specific stress response. Some genes were selected as models of a specific stress response.

The predicted operon of pRL100079 to pRL100081 (fusion pRU843) is an excellent example of a hyper-osmotic stress response (induced by sucrose, NaCl or mannitol). The operon shares sequence identity with the ProU system in *E. coli*, an extremely well studied and characterised transporter involved in the uptake of compatible solutes during osmotic upshift. The discovery of this operon led to the investigation of ProU-like systems in *R. leguminosarum* (see below).

The predicted operon of pRL90174 and pRL90175 (fusion pRU845) appears to encode a novel system used by 3841 to remove acid when countering an environment with a low pH. The novel mechanism removes acid from the cell by coupling a decarboxylation pathway to PHB. PHB is usually broken down to β-hydroxybutyrate, oxidised to acetoacetate by BdhA and then converted into acetoacetyl-coA. Instead, using the putative products of pRL90174 and pRL90175, β-hydroxybutyrate could be converted to acetoacetate (via the pRL90175 product) and is then converted to acetone (via the pRL90174 product). This removes the carboxylic acid group from the compound, which would cause the internal pH of the cell to increase. PHB is abundant in *R. leguminosarum* cells and so provides a reservoir for this system to use. BLAST analysis indicated that this system is not present in other sequenced α-proteobacteria and so may be unique to 3841, but the use of decarboxylases to remove carboxylic acid groups from molecules (e.g. GABA) in response to a low pH has been reported in other bacteria (Castanie-Cornet & Foster, 2001; Hommais *et al.*, 2004).

Approximately 38% of the stress-induced genes were identified as being hypothetical; a similar result to the genome of *S. meliloti*, ~40% of which is made up of hypothetical genes. Another interesting discovery was that the genome of 3841 contains many copies of the same genes. At least three copies of a *nodT*-like gene (RL3856, pRL100178 and pRL100291), two copies of a *bdhA*-like gene (RL3569 and pRL90175) and

two copies of some of the *fix* genes (e.g. pRL90013 and pRL100293 are both *fixH*-like) are present within the genome of *R. leguminosarum*. This indicates that these copies may be paralogues with different functions, which may be induced by different conditions, or may possess a high level of redundancy (see below).

Whilst effective, this type of study is not intended to compete with microarray experiments, which will permit the induction and repression of the entire genome to be monitored in one assay. The purpose of the LB3 library was not to be exhaustive in the examination of the 3841 genome, but to provide an easy means to monitor gene expression under stressful conditions. Plasmid fusions allow many different media to be screened at once with little expense and allow the transfer of fusion to other strains (e.g. mutants) to see if their expression is altered. Microarray studies have already been successful in *S. meliloti* (Cabanes *et al.*, 2000; Ampe, *et al.*, 2003; Djordjevic, *et al.*, 2003; Rüberg *et al.*, 2003; Becker *et al.*, 2004) and are currently being developed for *R. leguminosarum*. Future work will involve using microarrays to monitor the response to hyper-osmosis and other stresses in *R. leguminosarum*.

8.2. Mutational Studies

The attempt to locate a global regulator(s) of stress responses in 3841 by screening some of the markers isolated from LB3 in a Tn*5* mutant library was unsuccessful. This may be because of the redundant nature of the *R. leguminosarum* genome (as mentioned above), or because such mutations are lethal. If, as predicted, the genome of 3841 contains more than one copy of key genes, a mutation in one of these copies will not cause a severe effect to the bacteria, as a homologue or paralogue is present to function in place of the mutant.

The theory of redundancy within *R. leguminosarum* is further supported by the studies conducted with mutations made in specific stress-induced genes. None of the mutants showed any severe growth defects when grown in standard or stressed conditions, or when grown in symbiosis with pea plants. This indicates that systems may exist within 3841 (whether analogous or homologous) that can take over from the mutated genes. Alternatively, whilst the genes mutated were shown from the data obtained from the pOT fusions to be stress-induced, they may not encode a vital system in the stress response, which is why no severe growth defects were seen.

One mutant unable to grow under hyper-osmotically stressed conditions was isolated from a Tn*5* mutant library. However, the stress response that the gene was involved with was highly specific to an excess of fructose and not hyper-osmosis itself. The mutation was in a *lysR*-like regulator gene and it was assumed that this regulator maybe

responsible for controlling this fructose response in *R. leguminosarum*. A gene encoding a MFS-like transporter, that shares sequence identity to a sugar efflux permease, is downstream of the *lysR*-like gene and may be involved in the cell's response to a high concentration of fructose. One of the fusions isolated from the LB3 library, pRU855 (RU1519), was shown to produce high levels of GFP only when sucrose was used to generate hyper-osmosis. This, along with the LysR-like mutant data, suggests that 3841 may have a different response to an osmotic upshift generated by sugars, compared to that of polyols and/or ionic molecules. Any potential differences in the response to hyper-osmosis stress generated by different classes of molecules in *R. leguminosarum* will be something that will be investigated in future work.

Whilst no change in growth phenotype was observed, one of the pK19mob generated mutants, RU2184, is in a regulatory gene involved in the stress response (RL1157). A mutation in this predicted two-component response regulator gene prevented the transcription of RL1155 (a predicted hypothetical gene) and pRU862 (GFP fusion to RL1155) no longer expressed GFP under hyper-osmotic or acidic conditions, when present in RU2184. Although mutations in RL1155 and RL1157 did not cause any growth defects to *R. leguminosarum*, to my knowledge this is the first regulator associated with a hyper-osmotic and acidic response to be identified within 3841. RU2184 was tested with all of the stress-induced fusions isolated from LB3 and it only had an effect on pRU862 and its induction, however, further studies using the DNA of RU2184 in a microarray experiment could reveal other genes that RL1157 may regulate. This is something planned for one of the first experiments once 3841 microarrays are available.

It was fortunate that RL1157 was discovered through the random nature of the LB3 library. Now the predicted genome of *R. leguminosarum* is available, potential regulator genes that are in close proximity to stress-induced genes can be located by BLAST analysis and investigated accordingly. This too is planned for the future work and potential regulator have already be identified (Chapter 4, section 4.3).

8.3. Hyper-Osmotic Uptake and QAT Systems

Six QAT systems were identified within *R. leguminosarum* using sequence comparisons, homology and phylogenetic studies and BLAST analysis. Preliminary promoter probe data indicated that five of these were induced by osmotic upshift and one was induced by choline and glycine betaine. However, uptake of solutes through ABC systems is inhibited by an osmotic upshift. The degree of inhibition is dependent on both the time of exposure to the hyper-osmosis and the severity of the osmotic upshift, but was

independent of the osmolyte used to generate the stress. Additionally this phenomenon appears to be specific to ABC transporters, as the MCT and DCT systems (both of which use secondary transporters) were not significantly affected by osmotic upshift. The fact that five of the six QAT systems appear to be induced under conditions, in which transport is inhibited, is somewhat of a paradox and future work intends to further characterise the induction patterns of each of the QAT systems to determine what is occurring within cells under hyper-osmosis.

Whilst the mechanism behind the inhibition is not completely known, the effect itself is clear and raises many questions with regard to previous studies involving spheroplasts and bacteroids isolated from *R. leguminosarum*. All previous work with these types of cells has involved resuspending them in a high percentage of sucrose, to protect them from lysis. In all of these studies, no uptake through ABC systems was observed, although it was known that the transporters are expressed in both spheroplasts and bacteroids. Whilst the SBP component would not be present in sheroplasts, which would inhibit uptake of solutes through ABC transporters, bacteroids should still have the SBP components present. However, if transport assays were carried out in the presence of sucrose, as previous reported for all bacteroid studies, then uptake would be inhibited by the osmotic upshift. This would explain why no uptake has been observed in previous bacteroid studies and may invalidate any previous work. Future work will concentrate on refining the protocols used in spheroplast/bacteroid isolation to try and find a balance between protecting cells from cell lysis and to allow uptake assays to be performed uninhibited. However, this may be difficult to achieve as a specific amount of sucrose (or any other osmolyte) is required in isolation media in order to protect cells, such as bacteroids, from bursting due to their fragile outer membrane. This amount may prove to be too high, inhibiting any uptake activity seen via ABC transporters.

This data also raises further questions as to how 3841 counters an osmotic upshift. None of the compatible solutes tested contributed to the growth of 3841 under hyper-osmosis and the systems used by other bacteria to transport these osmoprotectants into cells are inhibited in *R. leguminosarum* by an osmotic upshift. How does *R. leguminosarum* respond to hyper-osmotic conditions? Studies in *S. meliloti* have shown that the BetS transporter, a secondary transporter, is responsible for the uptake of betaines as compatible solutes, but no gene with significant sequence identity to *betS* could be found in the preliminary genomic sequence of 3841. Perhaps rhizobia use secondary transporters, and not ABC systems, when acquiring compatible solutes under an osmotic upshift and clearly

further study is required to fully understand and investigate how *R. leguminosarum* responds to hyper-osmosis. This too, is planned for future work.

The QAT studies did reveal a homologue to the Cho system in *S. meliloti* in 3841 (QAT1). QAT1 is clearly induced by the presence of choline and glycine betaine and a mutant of this system was unable to grow in media where choline or glycine betaine was the sole carbon/nitrogen source. However, studies indicated that after growth on AMA (10mM glc, 10mM NH_4) QAT1 was responsible for the transport of choline but not obviously glycine betaine. This may be because, due to time constraints, the transport assays were not performed in cells grown on choline or glycine betaine as the carbon/nitrogen source. Future work is planned which will identify the optimum growth conditions for the induction of the QAT1 system. This work will also involve studies on QAT3 and QAT4, which were not conducted during this project because mutants in them were not made in time for analysis. The promoter probe data for the QAT systems will also be repeated, allowing a more conclusive overview of induction patterns to be concluded.

8.4. Conclusion

The purpose of this study was to try and lay down foundations for the genetic analysis of stress in *R. leguminosarum* 3841. This study has provided a number of tools and approaches for future, as outlined above. Not only have a number of important systems been identified and model gene fusions been isolated but the precise growth conditions needed for array analysis have been established. As such this should provide a powerful set of approaches for examining the stress responses of *R. leguminosarum* in both the laboratory, soil and rhizosphere environments.

REFERENCES

Agre, P., Brown, D. & Nielsen, S. (1995). Aquaporin Water Channels: Unanswered Questions and Unresolved Controversies. *Current Opinions in Cell Biology*. **7.** 472 – 483.

Albright, L. M., Huala, E. & Ausubel, F. M. (1989). Prokaryotic Signal Transduction Mediated by Sensor and Regulator Protein Pairs. *Annual Review of Genetics*. **23.** 311 – 336.

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. *Molecular Microbiology*. **36.** 508 – 515.

Allaway, D., Schofield, N. A., Leonard, M. E., Gilardoni, L., Finan, T. M. & Poole, P. S. (2001). Use of Differential Fluorescence Induction and Optical Trapping to Isolate Environmentally Induced Genes. *Environmental Microbiology*. **3.** 397 – 406.

Altschul, S. H., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acid Research*. **25.** 3389 – 3402.

Amarger, N., Macheret, V. & Laguerre, G. (1997). *Rhizobium gallicum* sp. nov. and *Rhizobium giardinii* sp. nov., from *Phaseolus vulgaris* Nodules. *International Journal of Systematic Bacteriology*. 47. 996 – 1006.

Ampe, F., Kiss, E., Sabourdy, F. & Batut, J. (2003). Transcriptiome Analysis of *Sinorhizobium meliloti* during Symbiosis. *Genome Biology*. **4.** R15.

Aneja, P. & Charles, T. C. (1999). Poly-3-Hydroxybutyrate Degradation in *Rhizobium* (*Sinorhizobium*) *meliloti*: Isolation and Characterisation of a Gene Encoding 3-Hydroxybutyrate Dehydrogenase. *Journal of Bacteriology*. **181.** 849 – 857.

Appleby, C. A. (1984). Leghemoglobin and *Rhizobium* Respiration. *Annual Review of Plant Physiology*. **35.** 443 – 478.

Arnold, W., Rump, A., Klipp, W., Priefer, U. B.& Pühler, A. (1988). Nucleotide Sequence of a 24,206 Base-Pair DNA Fragment Carrying the Entire Nitrogen Fixation Gene Cluster of *Klebsiella pneumoniae*. *Journal of Molecular Biology*. **203.** 715 – 738.

Balestrasse, K. B., Gardley, L., Gallego, S. M. & Tomaro, M. L. (2001). Response of Antioxidant Defence System in Soybean Nodules and Roots Subjected to Cadmium Stress. *Australian Journal of Plant Physiology*. **28.** 497 – 504.

Barra, L., Pica, N., Gouffi, K., Walker, G. C., Blanco, C. & Trautwetter, A. (2003). Glucose-6-Phosphate Dehydrogenase is Required for Sucrose and Trehalose to be Efficient Osmoprotectants in *Sinorhizobium meliloti*. *FEMS Microbiology Letters.* **229.** 183 – 188.

Batut, J., Daven-Mingot, M. L., David, M., Jacobs, J., Garnerone, A.-M. & Kahn, D. (1989). *fixK*, a Gene Homologous with *fnr* and *crp* from *Escherichia coli*, Regulates Nitrogen Fixation Genes Both Positively and Negatively in *Rhizobium meliloti*. *EMBO Journal*. **8.** 1279 – 1286.

Bearson, B. L., Wilson, L. & Foster, J. W. (1998). A Low pH-Inducible, PhoPQ-Dependent Acid Tolerance Response Protects *Salmonella typhimurium* against Inorganic Acid Stress. *Journal of Bacteriology*. **180.** 2409 – 2417.

Becker, A., Berges, H., Krol, E., Bruand, C., Ruberg, S., Capela, D., Lauber, E., Meilhoc, E., Ampe, F., de Bruijn, F. J., Fourment, J., Francez-Charlot, A., Kahn, D., Kuster, H., Liebe, C., Puhler, A., Weidner, S. & Batut, J. (2004). Global Changes in Gene Expression in *Sinorhizobium meliloti* 1021 under Microoxic and Symbiotic Conditions. *Molecular Plant-Microbe Interactions*. **17.** 292 – 303.

Becker, L. A., Bang, I.-S., Crouch, M.-L. & Fang, F. C. (2005). Compensatory Role of PspA, a Member of the Phage Shock Protein Operon, in *rpoE* Mutant *Salmonella enterica* serovar Thyphimurium. *Molecular Microbiology*. **56.** 1004 – 1016.

Benhizia, Y., **Benhizia, H.**, **Benguedouar, A.**, **Muresu, R.**, **Giacomini, A.** & **Squartini A**. **(2004).** Gamma Proteobacteria can Nodulate Legumes of the Genus *Hedysarum*. *Systematic and Applied Microbiology*. **27.** 462 – 468.

Beringer, J. E., Beynon, J. L., Buchanan-Wollaston, A. V. & Johnston, A. W. B. (1978). Transfer of the Drug Resistance Transposon Tn5 to *Rhizobium*. *Nature*. **276.** 633 – 634.

Berg, D. E., Weiss, A. & Crossland, L. (1980). Polarity of Tn*5* insertion mutations in *Escherichia coli*. *Journal of Bacteriology*. **142.** 439 – 446.

Bergersen, F. (1965). Ammonia – An Early Stable Product of Nitrogen Fixation by Soybean Root Nodules. *Australian Journal of Biological Sciences.* **18.** 1 - 9.

Bernard, T., Pocard, J.-A., Perroud, B. & Le Rudulier, D. (1986). Variations in the Response of Salt-Stressed *Rhizobium* Strains to Betaines. *Archives of Microbiology*. **143.** 359 – 364.

Bhat, U. R., & Carlson, R. W. (1992). Chemical Characterization of pH-Dependent Structural Epitopes of Lipopolysaccharides from *Rhizobium leguminosarum* biovar *phaseoli*. *Journal of Bacteriology*. **174.** 2230 – 2235.

Boivin C. & Giraud E. (1999). Molecular Symbiotic Characterization of Rhizobia: Toward a Polyphasic Approach using Nod Factors and *nod* Genes, p. 295 – 299. *In* E. Martínez-Romero & G. Hernández (ed.), *Highlights of Nitrogen Fixation Research*.

Boncompagni, E., Østerås, M., Poggi, M.-C. & Le Rudulier, D. (1999). Occurrence of Choline and Glycine Betaine Uptake Metabolism in the Family *Rhizobiaceace* and Their in Osmoprotection. *Applied and Environmental Microbiology*. **65.** 2072 – 2077.

Bonhivers, M., Carbrey, J. M., Gould, S. J. & Agre, P. (1998). Aquaporins in *Saccharomyces*. Genetic and Functional Distinction between Laboratory and Wild-type Strains. *Journal of Biological Chemistry*. **273.** 27565 – 27572.

Boscari, A., Mandon, K., Dupont, L., Poggi, M.-C. & Le Rudulier, D. (2002). BetS Is a Major Glycine Betaine/Proline Betaine Transporter Required for Early Osmotic Adjustment in *Sinorhizobium meliloti*. *Journal of Bacterology*. **184.** 2654 – 2663.

Botsford, J. L. & Lewis, T. A. (1990). Osmoregulation in *Rhizobium meliloti*: Production of Glutamic Acid in Response to Osmotic Stress. *Applied and Environmental Microbiology*. **56.** 488 – 494.

Boumahdi, M., Mary, P. & Hornez, J.-P. (2001). Changes in Fatty Acid Composition and Degree of Unsaturation of (Brady)Rhizobia as a Response to Phases of Growth, Reduced Water Activities and Mild Desiccation. *Antonie van Leeuwenhoek*. **79.** 73 – 79.

Braun, A. D. (1997). Microbial Water Stress. *Bacteriology Review*. **40.** 803 – 846.

Breedveld, M. W., Dijkema, C., Zevenhuizen, L. P. T. M. & Zehnder, A. J. B. (1993). Response of Intercellular Carbohydrates to NaCl Shock in *Rhizobium leguminosarum* biovar *trifolii* TA-1 and *Rhizobium meliloti* SU-47. *Journal of General Microbiology*. **139** 3157 – 3163.

Breedveld, M. W., Zevenhuizen, L. P. T. M. & Zehnder, A. J. B. (1990). Osmotically Induced Oligo- and Polysaccharide Synthesis by *Rhizobium meliloti* SU-47. *Journal of General Microbiology*. **136.** 2511 – 2519.

Bremer, E. & Krämer, R. (2000). Coping with Osmotic Challenges: Osmoregulation through Accumulation and Release of Compatible Solutes in Bacteria, p. 79 – 97. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*.

Brhada, F., Poggi, M. C., Van de Sype, G. & LeRudulier, D. (2001). Osmoprotection Mechanisms in Rhizobia Isolated from *Vicia faba* var. *majoe* and *Cicer arietinum*. *Agronomie*. **21.** 583 – 590.

Brown, C. M. & Dilworth, M. J. (1975). Ammonia Assimilation by *Rhizobium* Cultures and Bacteroids. *Journal of General Microbiology*. **86.** 39 – 48.

Buchanan-Wollaston, V. (1979). Generalized Transduction in *Rhizobium leguminosarum*. *Journal of General Microbiology*. **112.** 135 – 142.

Burn, J. Rossen, L. & Johnston, A. W. B. (1987). Four Classes of Mutations in the *nodD* Gene of *Rhizobium leguminosarum* biovar *viciae* that Affect its Ability to Autoregulate and/or Activate Other *nod* Genes in the presence of Flavonoid Inducers. *Genes and Development*. **1.** 456 – 464.

Cabanes, D., Boistard, P. & Batut, J. (2000). Identification of *Sinorhizobium meliloti* Genes Regulated during Symbiosis. *Journal of Bacteriology*. **182.** 3632 – 3637.

Calamita, G., Bishai, W. R., **Preston, G. M., Guggino, W. B.** & **Agre, P. (1995).** Molecular Cloning and Characterization of AqpZ, a Water Channel from *Escherichia coli*. *Journal of Biological Chemistry*. **270.** 29063 – 29066.

Callaham, D. A. & Torrey, J. G. (1981). The Structural Basis for Infection of Root Hairs of *Trifolium repens* by *Rhizobium*. *Canadian Journal of Botany.* **59.** 1647 – 1664.

Cardenas, L., Dominguez, J., Santana, O. & Quinto, C. **(1996).** The Role of *nodI* and *nodJ* Genes in the Transport of Nod Metabolites in *Rhizobium etli*. *Gene*. **173.** 183 – 187.

Carter, R. A., Yeoman, K. H., Klein, A., Hosie, A. H. F., Sawers, G., Poole, P. S. & Johnson, A. W. B. (2002). *dpp* Genes of *Rhizobium leguminosarum* Specify Uptake of δ-Aminolevulinic Acid. *Molecular Plant-Microbe Interactions*. **15.** 69 – 74.

Castanie-Cornet, M. P. & Foster, J. W. (2001). *Escherichia coli* Acid Resistance: cAMP Receptor Protein and a 20bp *cis*-Acting Sequence Control pH and Stationary Phase Expression of the *gadA* and *gadBC* Glutamate Decarboxylase Genes. *Microbiology.* **147.** 709 – 715.

Chen, H., Higgins, J., Kondorosi, E., Kondorosi, A., Djordjevic, M. A., Weinman, J. J. & Rolfe, B. G. (2000). Identification of *nolR*-Regulated Proteins in *Sinorhizobium meliloti* using Proteome Analysis. *Electrophoresis.* **21.** 3823 – 3832.

Chen, H., Richardson, A. E. & Rolfe, B. G. (1993). Studies of the Physiology and Genetic Basis of Acid Tolerance in *Rhizobium leguminosarum* biovar *trifolii*. *Applied and Environmental Microbiology*. **59.** 1798 – 1804.

Chen, W.-M., Moulin, L., Bontemps, C., Vandamme, P., Béna, G.,& Boivin-Masson, C. (2003). Legume Symbiotic Nitrogen Fixation by ß-Proteobacteria Is Widespread in Nature. *Journal of Bacteriology*. **185.** 7266 – 7272.

Chen, W.-P. & Kuo, T.-T. (1993). A Simple and Rapid Method for the Preparation of Gram-Negative Bacterial Genomic DNA. *Nucleic Acids Research*. **21.** 2260.

Chen, W. X., Ly, G. S., Qi, Y. L., Wang, E. T., Yuan, H. L. & Li, J. L. (1991). *Rhizobium huakuii* sp. nov. Isolated from the Root Nodules of *Astragalus sinicus*. *International Journal of Systematic Bacteriology*. **41.** 275 – 280.

Chen, W. X., Tan, Z. Y., Gao, J. L., Li, Y. & Wang, E. T. (1997). *Rhizobium hainanense* sp. nov., Isolated from Tropical Legumes. *International Journal of Systematic Bacteriology*. **47.** 870 – 873.

Chen, W. X., Wang, E. T., Wang, S. Y., Ly, Y. B., Chen, X. Q. & Li, Y. (1995) Characteristics of *Rhizobium tianshanense* sp. nov., A Moderately and Slowly Growing Root Nodule Bacterium Isolated from an Acid Saline Environment in Xingjian, People's Republic of China. *International Journal of Systematic Bacteriology*. **45.** 153 – 159.

Chen, W. X., Yan, G. H. & Li, J. L. (1988). Numerical Taxonomic Study of Fast-Growing Soybean Rhizobia and a Proposal that *Rhizobium fredii* be Assigned to *Sinorhizobium* gen. nov. *International Journal of Systematic Bacteriology*. **38.** 392 – 397.

Correa, O. S., Rivas, E. A. & Barneix, A. J. (1999). Cellular Envelopes and Tolerance to Acid pH in *Mesorhizobium loti*. *Current Microbiology*. **38.** 329 – 334.

Copeland, L. Vella, J. & Hong, Z. (1989). Enzymes of Carbohydrate Metabolism in Soybean Nodules. *Phytochemistry.* **28.** 57 – 61.

Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1995). Improved Green Fluorescent Protein by Molecular Evolution using DNA Shuffling. *Nature Biotechnology*. **14**. 315 – 319.

Crockford, A. J., Davis, G. A. & Williams, H. D. (1995). Evidence for Cell-Dependent Regulation of Catalase Activity in *Rhizobium leguminosarum* bv. *phaseoli*. *Microbiology*. **141.** 843 – 851.

Csonka, L. N. & Epstein, W. (1996). Osmoregulation, p. 1210 – 1223. *In* F. C. Neidhardt (ed.), *E. coli and Salmonella: Cellular and Molecular Biology*. ASM Press.

Culham, D. E., Lasby, B., Marangoni, A. G., Milner, J. L., Steer, B. A., van Nues, R. W. & Wood, J. M. (1993). Isolation and Sequencing of *Escherichia coli* Gene *proP* Reveals Unusual Structural Features of the Osmoregulatory Proline/Betaine Transporter, ProP. *Journal of Molecular Biology*. **229.** 268 – 276.

Cunningham, S. D. & Munns, D. N. (1984). The Correlation of the Exopolysaccharide Production and Acid-Tolerance in *Rhizobium*. *Soil Science Society of America Journal*. **48.** 1273 – 1276.

Dangeard, P.A. (1926). Recherches sur les Tubercules Radicaux des Légumineuses. *Le Botaniste*. **16.** 270.

Davey, M. E. & de Bruijn, F. J. (2000). A Homologue of the Tryptophan-Rich Sensory TspO and FixL Regulate a Novel Nutrient Deprivation-Induced *Sinorhizobium meliloti* Locus. *Applied and Environmental Microbiology*. **66.** 5353 – 5359.

Dean, D. R. & Jacobson, M. R. (1992). Biochemical Genetics of Nitrogenase, p. 763 – 834. In G. Stacey, R. H. Burris & H. J. Evans (ed.), *Biological Nitrogen Fixation.*

Debruijn, F. J. & Downie, J. A. (1991). Biochemical and Molecular Studies of Symbiotic Nitrogen-Fixation. *Current Opinion in Biotechnology.* **2.** 184 – 192.

de Lajudie, P., Laurent-Fulele, E., Willems, A., Torck, U., Coopman, R., Collins, M. D., Kersters, K., Dreyfus, D. & Gillis, M. (1998a). Description of *Allorhizobium undicola* gen. nov., sp. nov. for Nitrogen-Fixing Bacteria Efficiently Nodulating *Neptunia natans* in Senegal. *International Journal of Systematic Bacteriology*. **48.** 1277 – 1290.

de Lajudie, P., Willems, A., Nick, G., Moreira, F., Molouba, F., Hoste, B., Torck, U., Neyra, M., Collins, M. D., Lindström, K., Dreyfus, B. & Gillis, M. (1998b). Characterization of Tropical Tree Rhizobia and Description of *Mesorhizobium plurifarium* sp. nov. *International Journal of Systematic Bacteriology*. **48.** 369 – 382.

de Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrojuan, G., Neyra, M., Collins, M. D., Dreyfus, B. L., Kersters, K. & Gillis, M., (1994). Polyphasic Taxonomy of Rhizobia. Emendation of the Genus *Sinorhizobium* and Description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *International Journal of Systematic Bacteriology*. **44.** 715 – 733.

Delamarche, C., Thomas, D., Rolland, J.-P., Froger, A., Gouranton, J., Svelto, M., Agre, P. & Calamita, C. (1999). Visualisation of AqpZ-Mediated Water Permeability in *Escherichia coli* by Cryoelectron Microscopy. *Journal of Bacteriology*. **181.** 4193 – 4197.

Demozay, D., Rocchi, S., Mas, J.-C., Grillo, S., Pirola, L., Chavey, C. & van Obberghen, E. (2004). Fatty Aldehyde Dehydrogenase: Potential Role in Oxidative Stress Protection and Regulation of its Gene Expression by Insulin. *Journal of Biological Chemistry*. **279.** 6261 – 6270.

de Vries, G. E., van Brussel, A. A. N. & Quispel, A. (1982). Mechanism and Regulation of Glucose Transport in *Rhizobium leguminosarum*. *Journal of Bacteriology*. **149.** 872 – 879.

Djordjevic, M. A., Chen, H. C., Natera, S., van Noorden, G., Menzel, C., Taylor, S., Renard, C., Geiger, O., the *Sinorhizobium meliloti* DNA Sequencing Consortium & Weiller, G. F. (2003). A Global Analysis of Protein Expression Profiles in *Sinorhizobium meliloti*: Discovery of New Genes for Nodule Occupancy and Stress Adaptation. *Molecular Plant-Microbe Interactions*. **16.** 508 – 524.

Djordjevic, M. A., Schofield, P. R., Ridge, R. W., Morrison, N. A., Bassam, B. J., Plazinski, J., Watson, J. M. & Rolfe, B. G. (1985). *Rhizobium* Nodulation Genes Involved in Root Hair Curling (hac) are Functionally Conserved. *Plant Molecular Biology*. **4.** 147 – 160.

Downie, J. A. & Johnston, A. W. B. (1988). Nodulation of Legumes by *Rhizobium*. *Plant*, *Cell and Environment.* **11.** 403 – 412.

Downie, J. A. & Surin, B. P. (1990). Either of Two *nod* Gene Loci can Complement the Nodulation Defect of a *nod* Deletion Mutant of *Rhizobium leguminosarum* bv *viciae*. *Molecular and General Genetics.* **222.** 81 – 86.

Dreyfus, B. L., Garcia, J. L. & Gillis, M. (1988). Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., A Stem Nodulating Nitrogen-Fixing Bacterium Isolated from *Sesbania rostrata*. *International Journal of Systematic Bacteriology*. **38.** 89 – 98.

Dupont, L., Garcia, I., Poggi, M.-C., Alloing, G., Mandon, K. & Le Rudulier, D. (2004). The *Sinorhizobium meliloti* ABC Transporter Cho is Highly Specific for Choline and Expressed in Bacteroids from *Medicago sativa* Nodules. *Journal of Bacteriology*. **186.** 5988 – 5996.

Evans, I. J. & Downie, J. A. **(1986).** The *nodI* Product of *Rhizobium leguminosarum* is Closely Related to ATP-Binding Bacterial Transport Proteins: Nucleotide Sequence of the *nodI* and *nodJ* genes. *Gene*. **43.** 95 – 101.

Faucher, C., Camut, H., Dénarié, J. & Truchet, G. (1989). The *nodH* and *nodQ* Host Range Genes of *Rhizobium meliloti* Behave as Avirulence Genes in *R. leguminosarum* bv. *viciae* and Determine Changes in the Production of Plant Specific Extracellular Signals. *Molecular Plant-Microbe Interactions*. **2.** 291 – 300.

Ferenci, T. (2001). Hungry Bacteria – Definition and Properties of a Nutritional State. *Environmental Microbiology*. **3.** 605 – 611.

Figurski, D. H. & Helinski, D. R. (1979). Replication of an Origin-containing Derivative Plasmid RK2 Dependant on a Plasmid Fraction provided in Trans. *Proceedings in National Academy of Sciences of the United States of America*. **76.** 1648 – 1657.

Filler, W. A., Kemp, R. M., Ng, J. C., Hawkes, T. R., Dixon, R. A. & Smith, B. E. (1986). The *nifH* Gene Product is Required for the Synthesis or Stability of the 274 Iron-Molybdenum Cofactor of Nitrogenase from *Klebsiella pneumoniae*. *European Journal of Biochemistry*. **160.** 371 – 377.

Fischer, H. M. (1994). Genetic Regulation of Nitrogen Fixation in *Rhizobia*. *Microbiology Review*. **58.** 352 – 386.

Fisher, R. F., Tu, J. K. & Long, S. R. (1985). Conserved Nodulation genes in *Rhizobium meliloti* and *Rhizobium trifolii*. *Applied and Environmental Microbiology*. **49.** 1432 – 1435.

Forst, S., Delgado, J. & Inouye, M. (1989). Phosphorylation of OmpR by the Osmosensor EnvZ Modulates Expression of the *ompF* and *ompC* Genes in *Escherichia coli*. *Proceedings in National Academy of Sciences of the United States of America*. **86.** 6052 – 6056.

Foster, J. W. (1993). The Acid Tolerance Response of *Salmonella typhimurium* Involves Transient Synthesis of Key Acid Shock Proteins. *Journal of Bacteriology*. **175.** 1981 – 1987.

Foster, J. W. (2000). Microbial Responses to Acid Stress, p. 99 – 115. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Fougère, F. & Le Rudulier, D. (1990a). Uptake of Glycine Betaine and its Analogues by Bacteroids of *Rhizobium meliloti*. *Journal of General Microbiology*. **136.** 157 – 163.

Fougère, F. & Le Rudulier, D. (1990b). Glycine Betaine Biosynthesis and Catabolism in Bacteroids of *Rhizobium meliloti*: Effects of Salt Stress. *Journal of General Microbiology*. **136.** 2503 – 2510.

Foussard, M., Garnerone, A.-M., Ni, F., Soupene, E., Boistard, P. & Batut, J. (1997). Negative Autoregulation of the *Rhizobium meliloti fixK* gene is Indirect and Requires a Newly Identified Regulator, FixT. *Molecular Microbiology*. **25.** 27 – 37.

Freidman, A. M., Long, S. R., Brown, S. E., Buikema, W. J. & Ausubel. (1982). Construction of a Broad Host Range Cosmid Cloning Vector and its Use in the Genetic Analysis of *Rhizobium meliloti*. *Gene*. **18.** 289 – 296.

Fry, J., Wood, M. & Poole, P. S. (2001). Investigation of *myo*-Inositol Catabolism in *Rhizobium leguminosarum* bv. *viciae* and its Effect on Nodulation Competitiveness. *Molecular Plant-Microbe Interactions*. **14.** 1016 – 1025.

Fujihara, S. & Yoneyama, T. (1993). Effects of pH and Osmotic Stress on Cellular Polyamine Contents in the Soybean *Rhizobia fredii* P220 and *Bradyrhizobium japonicum* A1017. *Applied and Environmental Microbiology*. **59.** 1104 – 1109.

Fuller, F., Kunstner, P. W., Nguyen, T. & Verma, D. P. S. (1983). Soybean Nodulin Genes: Analysis of cDNA Clones Reveals Several Major Tissue-Specific Sequences in Nitrogen-Fixing Root Nodules. *Proceedings of the National Academy of Science of the USA*. **80.** 2594 – 2598.

Galibert, F., Finan, T. M., Long, S. R., Pühler, A., Abola, P., Ampe, F., Balroy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dréano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thébault, P., Vandenbol, M., Vorhölter, F.-J., Weidner, S., Wells, D. H., Wong, K., Yeh, K. C., Batut, J. (2001). The Composite Genome of the Legume Symbiont *Sinorhizobium meliloti*. *Science*. **29.** 668 – 672.

Galinski, E. A. & Trüper, H. G. (1994). Microbial Behaviour in Salt-Stressed Ecosystems. *FEMS Microbiology Reviews*. **15.** 95 – 108.

Gallie, D. R. (1991). The Cap and Poly(A) Tail Function Synergistically to Regulate mRNA Translational Efficiency. *Genes and Development*. **5.** 2108 – 2116.

Gao, J. L., **Turner, S. L.**, **Kan, F. L.**, **Wang, E. T.**, **Tan, Z. Y.**, **Qiu, Y. H.**, **Gu, J.**, **Terefework, Z.**, **Young, J. P.**, **Lindström, K.** & **Chen, W. X. (2003).** *Mesorhizobium septentrionale* sp. nov. and *Mesorhizobium temperatum* sp. nov., Isolated from *Astragalus adsurgens* Growing in the Northern Regions of China. *International Journal of Systematic and Evolutionary Microbiology*. **54.** 2003 – 2012.

Glenn, A. R. & Dilworth, M. J. (1994). The Life of Root Nodule Bacteria in the Acidic Underground. *FEMS Microbiology Letters*. **123.** 1 – 10.

Glenn, A. R., Poole, P. S. & Hudman, J. F. (1980). Succinate Uptake by Free-Living and Bacteroid Forms of *Rhizobium leguminosarum*. *Journal of General Microbiology*. **119.** 267 – 271.

Gloux, K. & Le Rudulier, D. (1989). Transport and Catabolism of Proline Betaine in Salt-Stressed *Rhizobium meliloti*. *Archives of Microbiology*. **151.** 143 – 148.

Gonzalez-Gonzalez, R., Botsford, J. L. & Lewis, T. (1990). Osmoregulation in *Rhizobium meliloti*: Charaterization of Enzymes Involved in Glutamate Synthesis. *Canadian Journal of Microbiology.* **36.** 469 – 474.

Gonzalez-Pasayo, R. & Martinez-Romero, E. (2000). Multiresistance genes of Rhizobium etli CFN42. *Molecular Plant-Microbe Interactions*. **13.** 572 – 577.

Gouffi, K., Bernard, T. & Blanco, C. (2000). Osmoprotection by Pipecolic Acid in *Sinorhizobium meliloti*: Specific Effects of D and L Isomers. *Applied and Environmental Microbiology*. **66.** 2358 – 2364.

Gouffi, K., Pica, N., Pichereau, V. & Blanco, B. (1999). Disaccharides as a New Class of Nonaccumulated Osmoprotectants for *Sinorhizobium meliloti*. *Applied and Environmental Microbiology*. **65.** 1491 – 1500.

Graham, P. H., Sadowsky, M. J., Keyser, H. H., Barnet, Y. M., Bradley, R. S., Cooper, J. E., De Ley, D. J., Jarvis, B. D. W., Roslycky, E. B., Stijdom, B. W., & Young, J. P. W. (1991). Proposed Minimal Standards for the Description of New Genera and Species of Root- and Stem-Nodulating Bacteria. *International Journal of Systematic Bacteriology*. **41.** 582 – 587.

Gustafson, A. M., O'Connell, K. P. & Thomashow, M. F. (2002). Regulation of *Sinorhizobium meliloti* 1021 *rrnA*-Reporter Gene Fusions in Response to Cold Shock. *Canadian Journal of Microbiology.* **48.** 821 – 830.

Haaker, H. & Klugkist, J. (1987). The Bioenergetics of Electron Transfer to Nitrogenase. *FEMS Microbiology Reviews*. **46.** 57 – 71.

Halbleib, C. M. & Ludden, P. W. (2000). Regulation of Biological Nitrogen Fixation. *Journal of Nutrition.* **130.** 1081 – 1084.

Hall, J. L. (2002). Cellular Mechanisms for Heavy Metal Detoxification and Tolerance. *Journal of Experimental Botany*. **53.** 1 -11.

Hansen, M. C., Palmer (Jr), R. J., Udsen, C., White, D. C. & Molin, S. (2001). Assessment of GFP Fluorescence in Cells of *Streptococcus gordonii* under Conditions of Low pH and Low Oxygen Concentration. *Microbiology*. **147.** 1383 – 1391.

Haydon, D. J. & Guest, J. R. (1991). A New Family of Bacterial Regulatory Proteins. *FEMS Microbiology Letters*. **79.** 291 – 296.

Hecker, M., Schumann, W. & Völker, U. (1996). Heat-Shock and General Stress Response in *Bacillus subtilis*. *Molecular Microbiology*. **19.** 417 – 428.

Hecker, M. & Völker, U. (1998). Non-Specific, General and Multiple Stress Resistance of Growth Restricted *Bacillus subtilis* Cells by the Expression of the σ^B Regulon. *Molecular Microbiology*. **29.** 1129 – 1136.

Heppel, L. A. (1969). The Effect of Osmotic Shock on Release of Bacterial Proteins and on Active Transport. *The Journal of General Physiology*. **54.** 95 – 113.

Hengge-Aronis, R. (1999). Interplay of Global Regulators and Cell Physiology in the General Stress Response of *Escherichia coli*. *Current Opinions in Microbiology*. **2.** 148 – 152.

Hengge-Aronis,R. (2000). The General Stress Response in *Escherichia coli*, p. 161 - 178. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Higgins, C. F. & Ames, G. F.-L. (1981). Two Periplasmic Transport Proteins which Interact with a Common Membrane Receptor Show Extensive Homology: Complete Nucleotide Sequences. *Proceedings in National Academy of Sciences of the United States of America*. **78.** 6038 – 6042.

Higgins, C. F. & Linton, K. J. (2004). The ATP Switch Model for ABC Transporters. *Nature Structural and Molecular Biology*. **11.** 918 – 926.

Hofmann, K., Heinz, E. B., Charles, T. C., Hoppert, M., Liebl, W. & Streit W. R. (2000). *Sinorhizobium meliloti* Strain 1021 *bioS* and *bdhA* Gene Transcriptions are Both Affected by Biotin Available in Defined Medium. *FEMS Microbiology Letters*. **182.** 41 – 44.

Hommais, F., Krin, E., Coppée, J.-Y., Lacroix, C., Yeramian, E., Danchin, A. & Bertin, P. (2004). GadE (YhiE): A Novel Activator Involved in the Response to Acid Environment in *Escherichia coli*. *Microbiology*. **150.** 61 – 72.

Hong, G.-F., Burn, J. E. & Johnston, A. W. (1987). Evidence that DNA Involved in the Expression of Nodulation (*nod*) Genes in *Rhizobium* binds to the Product of the Regulatory Gene *nodD*. *Nucleic Acids Research*. **15.** 9677 – 9690.

Honma, M. A., Asomaning, M. & Ausubel, F. M. (1990). *Rhizobium meliloti nodD* Genes Mediate Host-Specific Activation of *nodABC*. *Journal of Bacteriology*. **172.** 901 – 911.

Hosie, A. H. F., Allaway, D., Dunsby, H. A., Galloway, C. S. & Poole, P. S. (2002a). *Rhizobium leguminosarum* has a Second General Amino Acid Permease with Unusually Broad Solute Specificity and High Similarity to Branched Chain Amino Acid Transporters (Bra/LIV) of the ABC Family. *Journal of Bacteriology*. **184.** 4071 – 4080.

Hosie, A. H. F., Allaway, D. & Poole, P. S. (2002b). A Monocarboxylate Permease of *Rhizobium leguminosarum* is the First Member of a New Subfamily of Transporters. *Journal of Bacteriology*. **184.** 5436 – 5448.

Howard, K. S., McLean, P. A., Hansen, F. B., Lemley, P. V., Koblan, K. S., Orme-Johnson, W. H. (1986). *Klebsiella pneumoniae nifM* Gene Product is Required for Stabilisation and Activation of Nitrogenase Iron Protein in *Escherichia coli*. *Journal of Biological Chemistry*. **261.** 772 – 778.

Imperial, J., Ugalde, R. A., Shah, V. K. & Brill, W. J. (1984). Role of the *nifQ* Gene Product in the Incorporation of Molybdenum into Nitrogenase in *Klebsiella pneumoniae*. *Journal of Bacteriology*. **158.** 187 – 194.

Jamet, A., Sigaud, S., Van de Sype, G., Puppo, A. & Hérouart, D. (2003). Expression of the Bacteril Catalase Genes during *Sinorhizobium meliloti-Medicago sativa* Symbiosis and their Crucial Role during the Infection Process. *Molecular Plant-Microbe Interactions*. **16.** 217 – 225.

Jarvis, B. D. W., Pankhurst, C. E. & Patel, J. J. (1982). *Rhizobium loti* A New Species of Legume Root Nodule Bacteria. *International Journal of Systematic Bacteriology*. **32.** 378 – 380.

Jarvis, B. D. W., van Berkum, P., Chen, W. X., Nour, S. M., Fernández, M. P., Cleyet-Mare, J.-C. & Gillis, M. (1997). Transfer of *Rhizobium loti*, *Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to *Mesorhizobium* gen. nov. *International Journal of Systematic Bacteriology*. **47.** 895 – 898.

Jefferson, R. A., Burgess, S. M. & Hirsh, D. (1986). 3-Glucuronidase from *Escherichia coli* as a Gene-Fusion Marker. *Proceedings in National Academy of Sciences of the United States of America*. **83.** 8447 – 8451.

Jennings, M. P., Hood, D. W., Peak, I. R., Virji, M. & Moxon, E. R. (1995). Molecular Analysis of a Locus for the Biosynthesis and Phase-Variable Expression of the Lacto-N-Neotetraose Terminal Lipopolysaccharide Structure in *Neisseria meningitidis*. *Molecular Microbiology*. **18.** 729 – 740.

Jiang, W., Hou, Y. & Inouye, M. (1997). CspA, the Major Cold-Shock Protein of *Escherichia coli*, is an RNA Chaperone. *Journal of Bacteriology*. **272.** 196 – 202.

Jordan, D. C. (1984). Family III. *Rhizobiaceae* Conn 1938, p. 234 – 236. *In* N. R. Krieg & J. C. Holt (ed.), *Bergey's Manual of Systematic Bacteriology*.

Kannenberg, E. L. & Brewin, N. J. (1989). Expression of a Cell Surface Antigen from *Rhizobium leguminosarum* 3841 is Regulated by Oxygen and pH. *Journal of Bacteriology*. **171.** 4543 – 4548.

Kahn, D., David, M., Domergue, O., Daveran, M. L., Ghai, J., Hirsch, P. R. & Batut, J. (1989). *Rhizobium meliloti fixGHI* Sequence Predicts Involvement of a Specific Cation Pump in Symbiotic Nitrogen Fixation. *Journal of Bacteriology*. **171.** 929 – 939.

Karunakaran, R., Mauchline, T. H., Hosie, A. H. F. & Poole, P. S. (2005). A Family of Promoter Probe Vectors Incorporating Autofluorescent and Chromogenic Reporter Proteins for Studying Gene Expression in Gram-Negative Bacteria. *Microbiology.* **151.** 3249 – 3256.

Keen, N. T., Tamaki, S., Kobayashi, D, & Trollinger, D. (1988). Improved broad-host-range plasmids for DNA cloning in gram-negative bacteria. *Gene*. **70.** 191 – 197.

Keyser, H. H. & Munns, D. N. (1979). Tolerance of Rhizobia to Acidity, Aluminium and Phosphate. *Soil Science Society of America Journal*. **43.** 519 – 523.

Kidd, P. S., Llugany, M., Poschenrieder, C., Gunsé, B. & Barceló, J. (2001). The Role of Root Exudates in Aluminium Resistance and Silicon-Induced Amelioration of Aluminium Toxicity in Three Varieties of Maize (*Zea mays* L.) *Journal of Experimental Biology*. **52.** 1339 – 1352.

Kim, C. H., **Newton, W. E. & Dean, D. R. (1995).** Role of the MoFe Protein Alpha-Subunit Histidine-195 Residue in FeMo-Cofactor Binding and Nitrogenase Catalysis. *Biochemistry.* **34.** 2798 – 2808.

Kiss, E., Huguet, T., Poinsot, V. & Batut, J. (2004a). The *typA* is Required for Stress Adaptation as well as for Symbiosis of *Sinorhizobium meliloti* 1021 with Certain *Medicago truncatula* Lines. **17.** 235 – 244.

Kiss, S. A., Stefanovits-Banyai, E. & Takacs-Hajos, M. (2004b). Magnesium-Content of Rhizobium Nodules in Different Plants: The Importance of Magnesium in Nitrogen-Fixation of Nodules. *Journal of American College of Nutrition*. **23.** 751S - 753S.

Kneen, M., Farinas, J. Li, Y. & Verkman, A. S. (1998). Green Fluorescent Protein as a Noninvasive Intracellular pH Indicator. *Biophysical Journal*. **74.** 1591 – 1599.

Kondorosi, E., Banfalvi, Z. & Kondorosi, A. (1984). Physical and Genetic-Analysis of a Symbiotic Region of *Rhizobium meliloti* - Identification of Nodulation Genes. *Molecular and General Genetics.* **193.** 445 – 452.

Konstantinidis, K. T. & Tiedje, J. M. (2004). Trends between Gene Content and Genome Size in Prokaryotic Species with Larger Genomes. *Proceedings of the National Academy of Sciences of the United States of America*. **101.** 3160 – 3165.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. (2000). Crystal Structure of the Bacterial Membrane Protein TolC Central to Multidrug Efflux and Protein Export. *Nature*. **405.** 914 – 919.

Kurchak, O. N., Provorov, N. A. & Simarov, B. V. (2001). Plasmid pSym1-32 of *Rhizobium leguminosarum* bv. *viceae* Controlling Nitrogen Fixation Activity, Effectiveness of Symbiosis, Competitiveness and Acid Tolerance. *Russian Journal of Genetics*. **37.** 1025 – 1031.

Kuykendall, L. M., Saxena, B., Devine, T. E. & Udell, S. E. (1992). Genetic Diversity in *Bradyrhizobium japonicum* Jordan 1982 and a Proposal for *Bradyrhizobium elkanii* sp. nov. *Canadian Journal of Microbiology*. **38.** 501 – 503.

Laemmli, U. K. (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*. **227.** 680 – 685.

Laeremans, T. & Vanderleyden, J. (1998). Review: Infection and Nodulation Signalling in *Rhizobium-Phaseolus vulgaris* Symbiosis. *World Journal of Microbiology and Biotechnology*. **14.** 787 – 808.

LeRudlier, D., Gloux, K. & Riou, N. (1991). Identification of an Osmotically Induced Periplasmic Glycine Betaine-Binding Protein from *Rhizobium meliloti*. *Biochimica et Biophysica Acta.* **1061.** 197 – 205.

Leyer, G. J. & Johnson, E. A. (1993). Acid Adaptation Induces Cross Protection Against Environmental Stresses in *Salmonella typhimurium*. *Applied and Environmental Microbiology*. **59.** 1842.

Li, Y., Moe, P. C., Chandrasekaran, S., Booth, I. R. & Blount, P. (2002). Ionic Regulation of MscK, a Mechanosenstive Channel from *Escherichia coli*. *The EMBO Journal.* **21.** 5323 – 5330.

Lindström, K. (1989). *Rhizobium galegae*, A New Species of Legume Root Nodule Bacteria. *International Journal of Systematic Bacteriology*. **39.** 365 – 367.

Llopis, J., McCaffery, J., Miyawaki, A., Farquhar, M. G. & Tsien, R. Y. (1998). Measurement of Cytosolic, Mitochondrial, and Golgi pH in Single Living Cells with Green Fluorescent Proteins. *Proceedings of the National Academy of Sciences of the United States of America*. **95.** 6803 – 6808.

Lodwig, E. M., Hosie, A. H. F., Bourdès, 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.** 722 – 726.

Lodwig, E. M. & Poole, P. S. (2003). Metabolism of *Rhizobium* Bacteroids. *Critical Reviews in Plant Science*. **22.** 37 – 79.

Lomovskaya, O. & Lewis, K. (1992). *emr*, An *Escherichia coli* Locus for Multidrug Resistance**.** *Proceedings of the National Academy of Sciences of the United States of America*. **89.** 8938 – 8942.

Long, S. R. (1989). *Rhizobium* Genetics. *Annual Review of Genetics*. **23.** 483 – 506.

Lopez, O., Morera, C., Miranda-Rios, J., Girard, L., Romero, D. & Soberón, M. (2001). Regulation of Gene Expression in Response to Oxygen in *Rhizobium etli*: Role of FnrN in *fixNOQP* Expression and in Symbiotic Nitrogen Fixation. *Journal of Bacteriology*. **183.** 6999 – 7006.

López-García, S. L., Vázquez, T. E. E., Favelukes, G. & Lodeiro, A. R. (2001). Improved Soybean Root Association of N-Starved *Bradyrhizobium japonicum*. *Journal of Bacteriology*. **183.** 7241 – 7252.

Lortet, G., Méar, N., Lorquin, J., Dreyfus, B., de Lajudie, P., Rosenberg, C. & Boivin, C. (1996). Nod Factor Thin-Layer Chromatography Profiling as a Tool to Characterize Symbiotic Specificity of Rhizobial Strains: Application to *Sinorhizobium saheli*, *S. teranga* and *Rhizobium* sp. Strains Isolated from *Acacia* and *Sesbania*. *Molecular Plant-Microbe Interactions*. **9.** 736 – 747.

Mandon, K., Østerås, M., Boncompagni, E., Trinchant, J. C., Spennato, G., Poggi, M. C. & LeRudulier, D. (2003). The *Sinorhizobium meliloti* Glycine Betaine Biosynthetic Genes (*betICBA*) are Induced by Choline and Highly Expressed in Bacteroids. *Molecular Plant-Microbe Interactions*. **16.** 709 – 719.

Martínez-Romero, E., Segovia, L., Martins, F., Franco, A. A., Graham, P. & Pardo, M. A. (1991). *Rhizobium tropici*, A Novel Species Nodulating *Phaseolus vulgaris* L. Beans and *Leucaena* sp. Trees. *International Journal of Systematic Bacteriology*. **41.** 417 – 426.

Michiels, J., Verreth, C. & Vanderleyden, J. (1994). Effects of Temperature Stress on Bean-Nodulating *Rhizobium* strains. *Applied and Environmental Microbiology*. **60.** 1206 – 1212.

Miller, J. H. (1972). *Experiments in molecular genetics*. New York: Cold Spring Harbor Laboratory Press.

Miller, K. J. & Wood, J. M. (1996). Osmoadaptation by Rhizosphere Bacteria. *Annual Reviews of Microbiology*. **50.** 101 – 136.

Miranda-Rios, J., Morera, C., Taboada, H., Davalos,A., Encarnacion, S., Mora, J. & Soberon, M. (1997). Expression of Thiamin Biosynthetic Genes (*thiCOGE*) and Production of Symbiotic Terminal Oxidase cbb3 in *Rhizobium etli*. *Journal of Bacteriology*. **179.** 6887 – 6893.

Moulin, L., Munive, A., Dreyfus, B., & Boivin-Masson, C. 2001. Nodulation of Legumes by Members of the β-Subclass of Proteobacteria. *Nature*. **411.** 948 – 950.

Münchbach, M., Nocker, A. & Narberhaus, F. (1999). Multiple Small Heat Shock Proteins in *Rhizobia*. *Journal of Bacteriology*. **181.** 83 – 90.

Natera, S. H. A., Guerreiro, N. & Djordjevic, M. A. (2000). Proteome Analysis of Differentially Displayed Proteins as a Tool for the Investigation of Symbiosis. *Molecular Plant-Microbe Interactions*. **13.** 995 – 1009.

Neidle, E. L. & Kaplan, S. (1992). *Rhodobacter sphaeroides rdxA*, a Homolog of *Rhizobium meliloti fixG*, Encodes a Membrane Protein which May Bind Cytoplasmic [4Fe-4S] Clusters. *Journal of Bacteriology*. **174.** 6444 – 6454.

Nelson, D. L. & Cox, M. M. (2000). *Lehninger Principles of Biochemistry: Third Edition*. Worth Publishers, New York.

Netzer, W. J. & Hartl, F. U. (1998). Protein Folding in the Cytosol: Chaperonin-Dependant and -Independent Mechanisms. *Trends in Biochemical Science*. **23.** 68 – 73.

Nick, G., de Lajudie, P., Eardly, B. D., Suomalainen, S., Paulin, L., Zhang, X., Gillis, M. & Lindström, K. (1999). *Sinorhizobium arboris* sp. nov. and *Sinorhizobium kostiense* sp. nov., Isolated from Leguminous Trees in Sudan and Kenya. *International Journal of Systematic Bacteriology*. **49.** 1359 – 1368.

Nogales, J., Campos, R., BenAbdelkhalek, H., Olivares, J., Lluch, C. & Sanjuan, J. (2002). *Rhizobium tropici* Genes Involved in Free-Living Salt Tolerance are Required for the Establishment of Efficient Nitrogen-Fixing Symbiosis with *Phaseolus vulgaris*. *Molecular Plant-Microbe Interactions*. **15.** 225 – 232.

Nour, S. M., Cleyet-Marel, J.-C., Normand, P., Fernández, M. P. (1995). Genomic Heterogeneity of Strains Nodulating Chickpeas (*Cicer arietinum* L.) and Description of *Rhizobium mediterraneum* sp. nov. *International Journal of Systematic Bacteriology*. **45.** 640 – 648.

Nour, S. M., Fernández, M. P., Normand, P. & Cleyet-Marel, J.-C. (1994). *Rhizobium ciceri* sp. nov. Consisting of Strains that Nodulate Chickpeas (*Cicer arietinum* L.) *International Journal of Systematic Bacteriology*. **44.** 511 – 522.

O'Brian, M. R. (1996). Heme Synthesis in the *Rhizobium*-Legume Symbiosis: A Palette for Bacterial and Eukaryotic Pigments. *Journal of Bacteriology.* **178.** 2471 – 2478.

O'Connell, K., Gustafson, A. M., Lehmann, M. D. & Thomashow, M. F. (2000). Identification of Cold Shock Gene Loci in *Sinorhizobium meliloti* by using a *luxAB* Reporter Transposon. *Applied and Environmental Microbiology*. **66.** 401 – 405.

Ogasawara, M., Suzuki, T., Mutoh, I., Annapurna, K., Arora, N. K., Nishimura, Y. & Maheshwari, D. K. (2003). *Sinorhizobium indiaense sp*. nov. and *Sinorhizobium abri* sp. nov. Isolated from Tropical Legumes, *Sesbania rostrata* and *Abrus precatorius*, Respectively. *Symbiosis*. **34.** 53 – 68.

O'Hara, G. W., Goss, T. J., Dilworth, M. J. & Glenn, A. R. (1989). Maintenance of Intercellular pH and Acid Tolerance in *Rhizobium meliloti*. *Applied and Environmental Microbiology*. **55.** 1870 – 1876.

Outten, F. W., Outten, C. E. & O'Halloran, T. V. (2000). Metalloregulatory Systems at the Interface between Bacterial Metal Homeostasis and Resistance, p. 145 –157. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Panoff, J.-M., Corroler, D., Thammavongs, B. & Boutibonnes, P. (1997). Differentiation between Cold Shock Proteins and Cold Acclimation Proteins in a Mesophilic Gram-Positive Bacterium, *Enterococcus faecalis* JH2-2. *Journal of Bacteriology*. **179.** 4451 – 4454.

Pao, S. S., Paulsen, I. T. & Saier (Jr), M. H. (1998). Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews*. **62.** 1 – 34.

Patschkowski, T., Bates, D. M. & Kiley, P. J. (2000). Mechanisms for Sensing and Responding to Oxygen Deprivation, p. 61 – 78. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Patschkowski, T., Schluter, A. & Priefer, U. B. (1996). *Rhizobium leguminosarum* bv. *viciae* Contains a Second *fnr*/*fixK*-like Gene and an Unusual *fixL* Homologue. *Molecular Microbiology*. **21.** 267 – 280.

Peng, G. X., Tan, Z. Y., Wang, E. T., Reinhold-Hurek, B., Chen, W. F. & Chen, W. X. (2002). Identification of Isolates from Soybean Nodules in Xinjiang Region as *Sinorhizobium xinjiangense* and Genetic Differentiation of *S. xinjiangense* from *Sinorhizobium fredii*. *International Journal of Systematic and Evolutionary Microbiology*. 52. 457 – 462.

Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J. D., Völker, U. & Hecker, M. (2001). Global Analysis of the General Stress Response of *Bacillus subtilis*. *Journal of Bacteriology*. **183.** 5617 – 5631.

Petronilli, V. & Ames, G. F.-L. (1991). Binding Protein-Independent Histidine Permease Mutants: Uncoupling of ATP Hydrolysis from Transmembrane Signalling. *The Journal of Biological Chemistry*. **266.** 16293 – 16296.

Phadtare, S., Yamanaka, K. & Inouye, M. (2000). The Cold Shock Response, p. 33 – 45. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Pichereau, V., Pocard, J.-A., Hamelin, J., Blanco, C. & Bernard, T. (1998). Differential Effects of Dimethylsulfoniopropionate, Dimethylsulfonioacetate, and Other S-Methylated Compounds on the Growth of *Sinorhizobium meliloti* at Low and High Osmolarities. *Applied and Environmental Microbiology*. **64.** 1420 – 1429.

Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. (1994) *myo*-Inositol Catabolism and Catabolite Regulation in *Rhizobium leguminosarum* bv *viciae*. *Microbiology*. **140.** 2787 – 2795.

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. *Journal of General Microbiology*. **131.** 1441 – 1448.

Poolman, B., Blount, P., Folgering, J. H. A., Friesen. R. H. E., Moe, P. C. & van der Heide, T. (2002). How do Membrane Proteins Sense Water Stress? *Molecular Microbiology*. **44.** 899 – 902.

Poolman, B. & Glaasker, E. (1998). Regulation of Compatible Solute Accumulation in Bacteria. *Molecular Microbiology*. **29.** 397 – 407.

Potts, M. (1994). Desiccation Tolerance of Prokaryotes. *Microbiology Review*. **58.** 755 – 805.

Prell, J., Boesten, B., Poole, P. & Priefer, U. B. (2002). The *Rhizobium leguminosarum* bv. *Viciae* VF39 γ-aminobutyrate (GABA) aminotransferase gene (*gabT*) is Induced by GABA and Highly Expressed in Bacteroids. *Microbiology*. **148.** 615 – 623.

Preisig, O., Zufferey, R. & Hennecke, H. (1996). The *Bradyrhizobium japonicum fixGHIS* Genes are Required for the Formation of the High-Affinity cbb3-type Cytochrome Oxidase. *Archives of Microbiology*. **165.** 297 – 305.

Preston, G. M., Jung, J. S., Guggino, W. B. & Agre, P. (1994). Membrane Topology of Aquaporin CHIP. Analysis of Functional Epitope-Scanning Mutants by Vectorial Proteolysis. *Journal of Biological Chemistry*. **269.** 1668 – 1673.

Priefer, U. B., Aurag, J., Boesten, B., Bouchmouch, I., Defez, R., Filali-Maltouf, A., Miklis, M., Moawad, H., Mouhsine, B., Prell, J., Schlüter, A. & Senatore, B. (2001). Characterisation of *Phaseolus* Symbionts Isolated from Mediterranean Soils and Analysis of Genetic Factors Related to pH Tolerance. *Journal of Biotechnology*. **91.** 223 – 236.

Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., & Rolfe, B. G. (1986). Flavones Induce Expression of Nodulation Genes in *Rhizobium*. *Nature.* **323.** 632 – 635.

Reed, J. W. & Walker, G. C. (1991a). Acidic Conditions Permit Effective Nodulation of Alfalfa by Invasion-Deficient *Rhizobium meliloti* ExoD Mutants. *Genes and Development*. **5.** 2274 – 2287.

Reed, J. W. & Walker, G. C. (1991b). The *exoD* Gene of *Rhizobium meliloti* Encodes a Novel Function Needed for Alfalfa Nodule Invasion. *Journal of Bacteriology*. **173.** 664 – 677.

Reeve, W. G., Tiwari, R. P., Kale, N. B., Dilworth, M. J. & Glenn, A. R. (2002). ActP Controls Copper Homeostasis in *Rhizobium leguminosarum* bv. *Viciae* and *Sinorhizobium meliloti* Preventing Low pH-Induced Copper Toxicity. *Molecular Micorbiology*. **43.** 981 – 991.

Reeve, W. G., Tiwari, R. P., Wong, C. M., Dilworth, M. J. & Glenn, A. R. (1998). The Transcriptional Regulator Gene *phrR* in *Sinorhizobium meliloti* WSM419 is Regulated by Low pH and Other Stresses. *Microbiology*. **144.** 3335 – 3342.

Reid, C. J. & Poole, P. S. (1998). Roles of DctA and DctB in Signal Detection by the Dicarboxylic Acid Transport System of *Rhizobium leguminosarum*. *Journal of Bacteriology*. **180.** 2660 – 2669.

Reid, C. J., Walshaw, D. L. & Poole, P. S. (1996). Aspartate Transport by the Dct System in *Rhizobium leguminosarum* Negatively Affects Nitrgoen-Regulated Operons. *Microbiology*. **142.** 2603 – 2612.

Richardson, A. E. & Simpson, R. J. (1989). Acid-Tolerance and Symbiotic effectiveness of *Rhizobium trifolii* Associated with a *Trifolium subterraneum* L.-Based Pasture Growing in an Acid Soil. *Soil Biology and Biochemistry*. **21.** 87 –95.

Richardson, A. E., Simpson, R. J., Djordjevic, M. A. & Rolfe, B. G. (1988). Expression of Nodulation Genes in *Rhizobium leguminosarum* bivoar *trifolii* is Affected by Low pH and by Ca and Al Ions. *Applied and Environmental Microbiology*. **54.** 2541 – 2548.

Riccillo, P. M., Muglia, C. J., de Bruijn, F. J., Roe, A. J., Booth, I. R. & Aguilar, O. M. (2000). Glutathione is Involved in Environmental Stress Responses in *Rhizobium tropici*, Including Acid Tolerance. *Journal of Bacteriology*. **182.** 1748 – 1753.

Rivilla, R. & Downie, J. A. (1994). Identification of a *Rhizobium leguminosarum* Gene Homologous to NodT but Located Outside the Symbiotic Plasmid. *Gene*. **144.** 87 – 91.

Rivilla, R., Sutton, J. M. & Downie, J. A. (1995). Rhizobium leguminosarum NodT is related to a family of outer-membrane transport proteins that includes TolC, PrtF, CyaE and AprF. *Gene*. **161.** 27 – 31.

Robertson, J. G., Lyttleton, P., Bullivant, S. & Grayston, G. F. (1978). Membranes in Lupin Root Nodules. I. The Role of Golgi Bodies in the Biogenesis of Infection Threads and Peribacteroid Membranes. *Journal of Cell Science*. **30.** 129 – 149.

Robinson, A. C., Dean, D. R. & Burgess, B. K. (1987). Iron-Molybdenum Co-Factor Biosynthesis in *Azotobacter vinelandii* Requires the Iron Protein of Nitrogenase. *Journal of Biological Chemistry*. **262.** 14327 – 14332.

Roche, P., Debelle, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Denarie, J. & Prome, J. C. (1991). Molecular Basis of Symbiotic Host Specificity in *Rhizobium meliloti*: *nodH* and *nodPQ* Genes Encode the Sulfation of Lipooligosaccharide Signals. *Cell*. **67.** 1131 – 1143.

Rome, S., Fernández, M. P., Brunel, B., Normand, P., Cleyet-Marel, J.-C. (1996). *Sinorhizobium medicae* sp. nov., Isolated from Annual *Medicago* spp. *International Journal of Systematic Bacteriology*. **46.** 972 – 980.

Rossen, L., Shearman, C. A., Johnston, A. W. B. & Downie, J. A. (1985). The *nodD* Gene of *Rhizobium leguminosarum* is Autoregulatory and in the Presence of Plant Exudate Induces the *nodABC* Genes. *The EMBO Journal*. **4.** 3369 – 3373.

Rüberg, S., Tian, Z.-X., Krol, E., Linke, B., Meyer, F., Wang, Y., Pühler, A., Weidner, S. & Becker, A. (2003). Construction and Validation of a *Sinorhizobium meliloti* Whole Genome DNA Microarray: Genome-Wide Profiling of Osmoadaptive Gene Expressioni. *Journal of Biochemistry.* **106.** 255 – 268.

Saier, M. H. (2000). A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. *Microbiology and Molecular Biology Reviews*. **64.** 354 – 411.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual (Second Edition)*. New York: Cold Spring Harbor Laboratory Press.

Sans, N., Schroder, G. & Schroder, J. (1987). The Noc Region of Ti Plasmid C58 Codes for Arginase and Ornithine Cyclodeaminase. *European Journal of Biochemistry*. **167.** 81 – 87.

Santos, R., Hérouart, D., Puppo, A. & Touati, D. (2000). Critical Protective Role of Bacterial Superoxide Dismutase in *Rhizobium*-Legume Symbiosis. *Molecular Microbiology*. **38.** 750 – 759.

Savoure, A., Sallaud, C., El-Turk, J., Zuanazzi, J., Ratet, P., Schultze, M., Kondorosi, A., Esnault, R. & Kondorosi, E. (1997). Distinct Response of *Medicago* Suspension Cultures and Roots to Nod Factors and Chitin Oligomers in the Elicitation of Defense-Related Responses. *Plant Journal*. **11.** 277 – 287.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G. & Pühler, A. (1994). Small Mobilizable Multi-Purpose Cloning Vectors derived from the *Escherichia coli* Plasmids pK18 and pK19: Selection of Defined Deletions in the Chromosome of *Corynebacterium glutamicum*. *Gene*. **145**. 69 – 73.

Schell, M. A. (1993). Molecular Biology of the LysR Family of Transcriptional Regulators. *Annual Review of Microbiology*. **47.** 597 – 626.

Schlaman, H. R. M., Horvath, B., Vijgenboom, E., Okker, R. J. H. & Lugtenberg, B. J. J. (1991). Suppression of Nodulation Gene Expression in Bacteroids of *Rhizobium leguminosarum* bv. *viciae*. *Journal of Bacteriology*. **173.** 4277 – 4287.

Schluter, A., Patschkowski, T., Quandt, J. Selinger, L. B., Weidner, S., Kramer, M., Zhou, L., Hynes, M. F. & Priefer, U. B. (1997). Functional and Regulatory Analysis of the Two Copies of the *fixNOQP* Operon of *Rhizobium leguminosarum* Strain VF39. *Molecular Plant-Microbe Interactions*. **10.** 605 – 616.

Schofield, N. A. (1995). Development of Optical Trapping for the Isolation of Environmentally Regulated Genes. *Ph.D. Thesis*, *University of Reading*.

Scholla, M. H. & Elkan, G. H. (1984). *Rhizobium fredii* sp. nov. A Fast Growing Species that Effectively Nodulates Soybeans. *International Journal of Systematic and Bacteriology*. **34.** 484 – 486.

Segovia, L., **Young, J. P. & Martínez-Romero, E**. **(1993).** Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* Type I Strains as *Rhizobium etli* sp. nov. *International Journal of Systematic Bacteriology*. **43.** 374 – 377.

Shearman, C. A., Rossen, L., Johnston, A. W. B. & Downie, J. A. **(1986).** The *Rhizobium leguminosarum* Nodulation Gene *nodF* Encodes a Polypeptide Similar to Acyl-Carrier Protein and is Regulated by *nodD* Plus a Factor in Pea Root Exudate. *The EMBO Journal*. **5.** 647 – 652.

Shi, W., Li, C., Louise, C. J. & Adler, J. (1993). Mechanism of Adverse Conditions Causing Lack of Flagella in *Escherichia coli*. *Journal of Bacteriology*. **175.** 2236 – 2240.

Sigaud, S., Becquet, V., Frendo, P., Puppo, A. & Hérquart, D. (1999). Differential Regulation of Two Divergent *Sinorhizobium meliloti* Genes for HPII-Like Catalases during Free-Living Growth and Protective Role of Both Catalases during Symbiosis. *Journal of Bacteriology*. **181.** 2634 – 2639.

Simon, R., Priefer, U. & Pühler, A. (1983). A Broad Host-Range Mobilization System for *in vivo* Genetic Engineering: Transposon Mutagenesis of Gram-Negative Bacteria. *Biotechnology*. **1.** 784 – 791.

Singh, S., Kayastha, A. M., Asthana, R. K., Srivastava, P. K. & Singh, S. P. (2001). Response of *Rhizobium leguminosarum* to Nickel Stress. *World Journal of Microbiology & Biotechnology***. 17.** 667 – 672.

Slattery, J. F., Coventry, D. R. & Slattery, W. J. (2001). Rhizobial Ecology as Effected by the Soil Environment. *Australian Journal of Experimental Agriculture*. **41.** 289 – 298.

Smith, M. N., Crane, R. A., Keates, R. A. B. & Wood, J. M. (2004). Overexpression, Purification, and Characterisation of ProQ, a Posttranslational Regulator for Osmoregulatory Transporter ProP of *Escherichia coli*. *Biochemistry*. **43.** 12979 – 12989.

Smith, L. T. & Smith, G. M. (1989). An Osmoregulated Dipeptide in Stressed *Rhizobium meliloti*. *Journal of Bacteriology*. **171.** 4714 – 4717.

Smulski, D. R., Huang, L. L., McCluskey, M. P., Gladnick Reeve, M. J., Vollmer, A. C., van Dyk, T. K. & LaRossa, R. A. (2001). Combined, Functional Genomic-Biochemical Approach to Intermediary Metabolism: Interaction of Acivicin, a Glutamine Amidotransferase Inhibitor, with *Escherichia coli* K-12. *Journal of Bacteriology*. **183.** 3353 – 3364.

Soto, M. J., Zorzano, A., Garcia-Rodriguez, F. M., Mercado-Blanco, J., Lopez-Lara, I. M., Olivares, J. & Toro, N. (1994). Identification of a Novel *Rhizobium meliloti* Nodulation Efficiency *nfe* Gene Homolog of *Agrobacterium* Ornithine Cyclodeaminase. **7.** 703 – 707.

Soupène, E., Foussard, M., Boistard, P. Truchet, G. & Batut, J. (1995). Oxygen as a Key Developmental Regulator of *Rhizobium meliloti* N_2-Fixation Gene Expression within the Alfalfa Root Nodule. *Proceedings of the National Academy of Sciences of the United States of America*. **92.** 3759 – 3763.

Spaink, H. P. (1996). Regulation of Plant Morphogenesis by Lipo-Chitin Oligosaccharides. *Critical Reviews in Plant Sciences.* **15.** 559 – 582.

Spaink, H. P., Sheeley, D. M., van Brussel, A. A. N., Glushka, J., York, W. S., Tak, T., Geiger, O,. Kennedy, E. P., Reinhold, V. N. & Lugtenberg, B. J. J. **(1991).** A Novel Highly Unsaturated Fatty Acid Moiety of Lipo-oligosaccharide Signals Determines Host Specificity of *Rhizobium*. *Nature*. **354.** 125 – 130.

Spaink, H. P., Weinman, J., Djordjevic, D. A., Wijffelman, C. A., Okker, R. J. H. & Lugtenberg, B. J. J. (1989). Genetic Analysis and Cellular Localisation of the *Rhizobium* Host Specificity-Determining NodE Protein. *EMBO Journal*. **8.** 2811 – 2818.

Squartini, A., Struffi, P., Doring, H., Selenska-Pobell, S., Tola, E., Giacomini, A., Vendramin, E., Velázquez, E., Mateos, P. F., Martínez-Molina, E., Dazzo, F. B., Casella, S. & Nuti, M. P. (2002). *Rhizobium sullae* sp. nov. (formerly '*Rhizobium hedysari*'), the Root-Nodule Microsymbiont of *Hedysarum coronarium* L. *International Journal of Systematic and Evolutionary Microbiology*. **52.** 1267 – 1276.

Stokes, N. R., Murray, H. D., Subramaniam, C., Gourse, R. L., Louis, P., Bartlett, W., Miller, S. & Booth, I. R. (2003). A Role for Mechanosensitive Channels in Survival of Stationary Phase: Regulation of Channel Expression by RpoS. *Proceedings of the National Academy of Sciences of the United States of America*. **100.** 15959 – 15964.

Storz, G. & Zheng, M. (2000). Oxidative Stress, p. 47 – 59. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Summers, M. L., Elkins, J. G., Elliot, B. A. & McDermott, T. R. (1998). Expression and Regulation of Phosphate Stress Inducible Genes in *Sinorhizobium meliloti*. *Molecular Plant-Microbe Interactions*. **11.** 1094 – 1101.

Talibart, R., Jebbar, M., Gouesbet, G., Himdi-Kabbab, S., Wróblewski, H., Blanco, C. & Bernard, T. (1994). Osmoadaptation in *Rhizobia*: Ectoine-Induced Salt Tolerance. *Journal of Bacteriology*. **176.** 5210 – 5217.

Talibart, R., LeHénaff, M., Bernard, T. & Wróblewski, H. (1990). Identification of Bacterial Periplasmic Glycine Betaine-Binding Protein after Electrophoresis and Affinity Labelling. *Journal of Biochemical and Biophysical Methods.* **21.** 155 – 164.

Tan, Z. Y., Kan, F. L., Peng, G. X., Wang, E. T., Reinhold-Hurek, B. & Chen, W. X. (2001). *Rhizobium yanglingense* sp. nov. Isolated from Arid and Semi-Arid Regions in China. *International Journal of Systematic and Evolutionary Microbiology*. **51.** 909 – 914.

Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A. & Takahashi, H. (1993) Heterogeneity of the Principal Sigma Factor in Escherichia coli: The *rpoS* Gene Product Sigma 38, is a Second Principal Sigma Factor of RNA Polymerase in Stationary-Phase *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*. **90.** 3511 – 3515.

Tao, H., Brewin, N. J. & Noel, K. D. (1992). *Rhizobium leguminosarum* CFN42 Lipopolysaccharide Antigenic Changes Induced by Environmental Conditions. *Journal of Bacteriology*. **174.** 2222 – 2229.

Thorne, S. H. & Williams, H. D. (1997). Adaptation to Nutrient Starvation in *Rhizobium leguminosarum* bv. *phaseoli*: Analysis of Survival, Stress Resistance and Changes in Macromolecular Synthesis during Entry to and Exit from Stationary Phase. *Journal of Bacteriology*. **179.** 6894 – 6901.

Tighe, S. W., de Lajudie, P., Dipietro, K., Lindström, K., Nick, G. & Jarvis, B. D. W. (2000). Analysis of Cellular Fatty Acids and Phenotypic Relationships of *Agrobacterium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* Species using the Sherlock Microbial identification System. *International Journal of Systematic and Evolutionary Microbiology*. **50.** 787 – 801.

Tiwari, R. P., Reeve, W. G., Dilworth, M. J. & Glenn, A. R. (1996a). An Essential Role for *actA* in Acid Tolerance of *Rhizobium meliloti*. *Microbiology*. **142.** 601 – 610.

Tiwari, R. P., Reeve, W. G., Dilworth, M. J. & Glenn, A. R. (1996b). Acid Tolerance in *Rhizobium meliloti* strain WSM419 Involves a Two-Component Sensor-Regulator System. *Microbiology*. **142.** 1693 – 1704.

Todd, J. D., Wexler, M., Sawers, G., Yeoman. K. H., Poole, P. S. & Johnston, A. W. B. (2002). RirA, an Iron-Responsive Regulator in the Symbiotic Bacterium *Rhizobium leguminosarum*. *Microbiology*. **148.** 4059 – 4071.

Toledo, I., Lloret, L. & Martínez-Romero, E. (2003). *Sinorhizobium americanus*, sp. nov., A New *Sinorhizobium* Species Nodulating Native *Acacia* spp. in Mexico. *Systematic and Applied Microbiology*. **26**. 54 – 64.

Trovato, M., Maras, B., Linhares, F. & Constantino, P. (2001). The Plant Oncogene *rolD* Encodes a Functional Ornithine Cyclodeaminase. *Proceedings of the National Academy of Sciences of the United States of America*. **98.** 13449 – 13453.

Trzebiatowski, J. R., Ragatz, D. M. & de Bruijn, F. J. (2001). Isolation and Regulation of *Sinorhizobium meliloti* 1021 Loci Induced by Oxygen Limitation. *Applied and Environmental Microbiology*. **67.** 3728 – 3731.

Urbatsch, I. L., Sankaran, B., Weber, J. & Senior A. E. (1995). P-Glycoprotein is Stably Inhibited by Vanadate-Induced Trapping of Nucleotide at a Single Catalytic Site. *Journal of Biological Chemistry*. **270.** 19383 – 19390.

van Berkum, P., Beyene, D., Bao, G., Campbell, T. A. & Eardly, B. D. (1987). *Rhizobium mongolense* sp. nov. is One of Three Rhizobial Genotypes Identified which Nodulate and Form Nitrogen-Fixing Symbioses with *Medicago ruthenica* [(L.) Ledebour]. *International Journal of Systematic Bacteriology*. **48.** 13 – 22.

van der Heide, T., Stuart, M. C. A. & Poolman, B. (2001). On the Osmotic Signal and Osmosensing Mechanism of an ABC Transport System for Glycine Betaine. *The EMBO Journal*. **20.** 7022 – 7032.

van Rhijn, P. & Vanderleyden, J. (1995). The *Rhizobium*-Plant Symbiosis. *Microbiological Reviews*. **59.** 124 – 142.

Velázquez, E., Igual, J. M., Willems, A., Fernández, M. P., Muñoz, E., Mateos, P. F., Abril, A., Toro, N., Normand, P., Cervantes, E., Gillis, M. & Martínez-Molina, E. (2001). *Mesorhizobium chacoense* sp. nov., A Novel Species that Nodulates *Prosopis alba* in the Chaco Arido Region (Argentina). *International Journal of Systematic and Evolutionary Microbiology*. **51.** 1011 – 1021.

Ventosa, A., Nieto, J. J. & Oren A. (1998). Biology of Moderately Halophilic Aerobic Bacteria. *Microbiology and Molecular Biology Review*. **62.** 504 – 544.

Wallington, E. J. & Lund, P. A. (1994). *Rhizobium leguminosarum* Contains Multiple Chaperonin (*cpn60*) Genes. *Microbiology*. **140.** 113 – 122.

Walshaw, D. L. (1995). The General Amino Acid Permease of *Rhizobium leguminosarum* biovar *viciae*. *Ph.D. Thesis*, *University of Reading*.

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. *Molecular Microbiology*. **21.** 1239 – 1252.

Wang, E. T., Rogel, M. A., García-de los Santos, A., Martínez-Romero, J., Cevallos, M. A., & Martínez-Romero, E. (1999a). *Rhizobium etli* bv. *mimosae*, A Novel biovar Isolated from *Mimosa affinis*. *International Journal of Systematic Bacteriology*. **49.** 1479 – 1491.

Wang, E. T., Tan, Z. Y., Willems, A. Y., Fernández-López, M., Reinhold-Hurek, B. & Martínez-Romero, E. (2002). *Sinorhizobium morelense* sp. nov., A *Leucaena leucocephala*-Associated Bacterium that is Highly Resistant to Multiple Antibiotics. *International Journal of Systematic and Evolutionary Microbiology*. **52.** 1687 – 1693.

Wang, E. T., van Berkum, P., Beyene, D., Sui, X. H., Dorado, O. Chen, W. X. & Martínez-Romero, **E. (1998).** *Rhizobium huautlense* sp. nov., A Symbiont of *Sesbania herbacea* that has a Close Phylogenetic Relationship with *Rhizobium galegae*. *International Journal of Systematic Bacteriology*. **48.** 687 – 699.

Wang, E. T., van Berkum, P., Sui, X. H., Beyene, D., Chen, W.X. & Martínez-Romero E. (1999b). Diversity of Rhizobia Associated with *Amorpha fruticosa* Isolated from Chinese Soils and Description of *Mesorhizobium amorphae* sp. nov. *International Journal of Systematic Bacteriology*. **49.** 51 – 65.

Watkin, E. L. J., O'Hara, G. W. & Glenn, A. R. (2003). Physiological Responses to Acid Stress of an Acid-Soil Tolerant and an Acid-soil Sensitive strain of *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biology and Biochemistry*. **35.** 621 – 624.

Watson, R. J., Heys, R., Martin, T. & Savard, M. (2001). *Sinorhizobium meliloti* Cells Require Biotin and either Cobalt or Methionine for Growth. *Applied and Environmental Microbiology*. **6.** 3767 – 3770.

Wei, G. H., Wang, E. T., Tan, Z. Y. Zhu, M. E. & Chen, W. X. (2002). *Rhizobium indigoferae* sp. nov. and *Sinorhizobium kummerowiae* sp. nov., Respectively Isolated from *Indigofera* spp. and *Kummerowia stipulacea*. *International Journal of Systematic and Evolutionary Microbiology*. **52.** 2231 – 2239.

Weiner, J. H., & Heppel, L. A. (1971). A Binding Protein for Glutamine and its Relation to Active Transport in *Escherichia coli*. *Journal of Biological Chemistry*. **246.** 6933 – 6941.

Wells, D. H. & Long, S. R. (2002). The *Sinorhizobium meliloti* Stringent Response Affects Multiple Aspects of Symbiosis. *Molecular Microbiology*. **43.** 1115 – 1127.

Wells, D. H. & Long, S. R. (2002). The *Sinorhizobium meliloti* Stringent Response Affects Multiple Aspects of Symbiosis. *Molecular Microbiology*. **43.** 1115 – 1127.

Willems, A., Fernández-López, M., Muñoz-Adelantado, E., Goris, J., De Vos, P., Martínez-Romero, E., Toro, N. & Gillis, M. (2003). Description of New *Ensifer* Strains from Nodules and Proposal to Transfer *Ensifer adhaerens* Casida 1982 to *Sinorhizobium* as *Sinorhizobium adhaerens* comb. nov. Request for an Opinion. *International Journal of Systematic and Evolutionary Microbiology*. **53.** 1207 – 1217.

Wilson, K. J., Parra, A. & Botero, L. (1999). Application of the GUS Marker Gene Technique to High-Throughput Screening of Rhizobial Competition. *Canadian Journal of Microbiology*. **45.** 678 – 685.

Wood, J. M. (1999). Osmosensing by Bacteria: Signals and Membrane-Based Sensors. *Microbiology and Molecular Biology Reviews.* **63.** 230 – 262.

Xi, C., Schoeters, E., van der Leyden, J. & Michels, J. (2000). Symbiosis-Specific Expression of *Rhizobium etli casA* Encoding a Secreted Calmodulin-Related Protein. *Proceedings of the National Academy of Sciences of the United States of America*. **97.** 11114 – 11119.

Xu, L. M., Ge, C., Cui, Z., Li, J. & Fan, H. (1995). *Bradyrhizobium liaoningense* sp. nov. Isolated from the Root Nodules of Soybean. *International Journal of Systematic Bacteriology*. **45.** 706 – 711.

Yao, P. Y. & Vincent, J. M. (1969). Host Specificity in the Root Hair "Curling Factor" of *Rhizobium* sp. *Australian Journal of Biological Sciences.* **22.** 413 – 423.

Yao, Z. Y., Kan, F. L., Wang, E. T., Wei, G. H. & Chen, W. X. (2002). Characterization of Rhizobia that Nodulate Legume Species of the Genus *Lespedeza* and description of *Bradyrhizobium yuanmingense* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*. **52.** 2219 – 2230.

Young, M. (2003). The Genus Name Ensifer Casida 1982 Takes Priority over *Sinorhizobium* Chen *et al*. 1988, and *Sinorhizobium morelense* Wang *et al*. 2002 is a Later Synonym of *Ensifer adhaerens* Casida 1982. Is the Combination '*Sinorhizobium adhaerens*' (Casida 1982) Willems *et al*. 2003 Legitimate? Request for an Opinion. *International Journal of Systematic and Evolutionary Microbiology*. **53.** 2107 – 2110.

Yura, T., Kanemori, M. & Morita, M. T. (2000). The Heat Shock Response: Regulation and Function, p. 3 – 18. *In* G. Storz & R. Hengge-Aronis (ed.), *Bacterial Stress Response*. ASM Press, Washington, D.C.

Zahran, H. H. (1999). *Rhizobium*-Legume Symbiosis and Nitrogen Fixation under Severe Conditions and in an Arid Climate. *Microbiology and Molecular Biology Reviews*. **63.** 968 – 989.

Zahran, H. H., Räsänen, L. A., Karsisto, M. & Lindström, K. (1994). Alteration of Lipopolysaccharide and Protein Profiles in SDS-PAGE of Rhizobia by Osmotic and Heat Stress. *World Journal of Microbiology and Biotechnology*. **10.** 100 – 105.

Zakhia, F. & de Lajudie, P. (2001). Taxonomy of Rhizobia. *Agronomie*. **21.** 569 – 576.

Zgurskaya, H. I., Keyhan, M. & Matin, A. (1997). The Sigma S Level in Starving *Escherichia coli* Cells Increases Solely as a Result of its Increased Stability, Despite Decreased Synthesis. *Molecular Microbiology*. **24.** 643 – 651.

APPENDIX: PLASMID INSERT SEQUENCES

pRU843

```
   1  caggcgcacgtcacttgtaaccggcagcgtccccggggtaacgctgagcg
  51  caataagctgggcgaggatcgaaaagacttgattggtgttttcggcgccc
 101  ctctctgggacgcaataagccgagccacggtagactgtggcggcaaagca
 151  gttgtgcgatcgtccccgagtaactttgataatcgatgtgtcggtcgccc
 201  gccgaagggcctcctccgtctggagcctgaaggcattgggggtggatcgc
 251  atcaatgctccaagataacgaaagacagcaaaaggcgagcgcagggaaat
 301  ggagaaggtgaaggaagcttcaccttctatatcgttgtgaaaccttacta
 351  cgaggtccttcggcgtcgcgcccgtacccgaccagttcgcgctgcagatc
 401  tggccagggccctgacggactaggttgcgcttgaaggcgtcgtcgctgtg
 451  ttcgggtcggtccctgatttcatgatcacgatcgatgccaacttctggat
 501  gtctgcaagcgacgctcaggcccgcgccctcatggagcatgagctatcgc
 551  actgtgcgcaggagctcgacgacttcggagccccgaagttccgcaagcgt
 601  cggattcacccgaagccgcgcacaacgccagggtcgcagaagcccaaaac
 651  agcctatttttgcgccatcacgatgacccgctggcaataaatttcatttt
 701  tttatggacatcgagcgtatatccacctacataacgttcaatgcctccgc
 751  cgaagcgctggtttcgaatggaacatgaagcgccgaacgcagacgttact
 801  atcgatgattgccattatggcaaagccatacttggatggccagatacgtg
 851  attaaatcaatagcgcttttctacattggaaacacacccgacacttggaa
 901  attcacagtaaatactgccgtctcgacaagcatcggctaaaaaaggagaa
 951  tgcctagcctcttgagctcggcgcccaaattatatgaattctgcaactgg
1001  aaaagacgcgccgacacaatcggcaacgaccaagatcagcctcaagaaca
1051  tctacaaagtcttcggcgagcatccgaaaaaagccttcgcactactgcgg
1101  gccggaaagaccaagtctgaaattcatgcagcaacgggctgctcgatcgg
1151  cgtcaacgacgccagtttcgatattcgtgccggcgagatattcgtgatca
1201  tgggcctgtcgggctccggcaaatcaacgctgctgcgccttctcaaccga
1251  ctgatcgaaccgagttccggatcgatcgagatcgatggccgcgacatcac
1301  cggcatgtcgcgcagcgagctgatcgcgcttcgccgccgcgacatcagca
1351  tggtcttccaatccgttgcactgctgccgaaccgcacggtcctcaacaac
1401  gccgccttcggtctcgaagtcgccggtgtcggcgaggccggccgcaagca
1451  gaaggcactggcggcgctgaaggccgttggcctagatggctatgccgaca
1501  gccgcccggaccagctctccggtggcatgaagcagcgcgtcggccttgcc
1551  cgcgcactggccagcgagccgacgatcctgctgatggacgaagccttctc
1601  ggccctcgacccgctgatccgcacggaaatgcaggacgaactggttcgcc
1651  tgcagtcggaacacagccgcaccatcgttttcgtcagccacgatctcgac
1701  gaggcgatgcgcatcggcgaccggatctgcatcatgcagaacggcaatgt
1751  cgtgcaggtcggcgcgcccgatgaaatcgtcacccagccggccaacgact
1801  atgttcgctcgttcttccgcaatgtcgatgtcgcccatgtcttcaaggcc
1851  ggcgatgttgcccgcaagtcgcaggtgacgatc
```

pRU844

```
   1  cgcacgcgctgaagaacctcgaagaccagattgcccatctgtcggcgctg
  51  atgagcgcgccgcgcgaaagcgccgtaataccccccgaactcgaccagcg
 101  catgggcgcgatcgaagactatatggcgacgagcgatgaatatatcatcg
 151  aggcggcacgccaggcggcagaggccgtcgtcgaagcctattcgcgccat
 201  ggcggtccacagggtgtcatgcccgccgccgacatgtcggcgctgacggc
 251  gctggccgaggatctgcgccacctcgaggatctcagccgcgacagcgagg
 301  agcgtacgcacaagactttccaggcgctgcacgagacgctggtgcatatc
 351  gccgaccgcctcgacggcatgggagatcgcggccggccggccgcccagat
 401  gccgatcgccgacgtcgatttcgatgtcgatccctatgcgttgatggtgg
 451  ccgaagccgacatgaacaggacgccggccgctgccccgacagcgaaggct
 501  tctcctgttatccgcacggccgaggtcgcagccgaagctgccgccccggc
 551  acaagccaccgccccggcacaaaccaacgccatgagcggaacgagcgcga
 601  tcgccatcgaagccgcgaccaggacgacggagacagcagcgacatcggca
 651  gggacacgggcaccggcaaaggccagcctgctcgccaacctcggcaagcg
 701  gctgctgcccgccaagaaggccgagagccggacgaccgaacggccgatga
 751  tcgaccccgcgccgtcgatcgatcccaccgacgtggtgccgacggatgcg
 801  gcaaacgaactgctcgagccgggctcgggcgcgcctgacgtgaagaagat
 851  ccttgaacgggttcgcgccagccagagtgcggcgcgcggcaaacctgccg
 901  gcgaaaccgatcgcgccgattatatcgctgccgcgcgccgtgcggcgcag
 951  gcggccgcgatggaggtggatgccaatccgaagcaggcagccggcaaggc
1001  cgagaagaaaggcgcgaatgccgacaaggccggcaaggcttccgacaaga
1051  cgggcaagaccagcgccttttcgcgctaccggcggccgatcctgctggcg
```

```
1101  gtcggcgccgtgctgctcgcgatcatggcctttccgcttgcccgaacgct
1151  gacaacaggcgagcgcgcgccgcagccgccagcagaagtctcggcgctga
1201  ccggcgcggaaaacccggcgccggccttgcctgacgcgacacccgcccag
1251  ccggatgccgccgctctcggaacgaccgctccggcagcggatgccacgcc
1301  gcccgtcgccgaacaggcgcagccggaagcgaccccgcctgacgccggcg
1351  agcatctgaccgacgtgacgccgcttgacggcgaaggggctgcgaccctt
1401  gcccctcccggtccgtcgggtgcggcgcaggagacatcgggcttcgttcc
1451  gaacgcaccggctgcggcagcacccgcaccggccatcaccgttcccgaca
1501  ccgtccagccgaaatcgcttgccgatgcggcgagcggcggcgatgcgctg
1551  gcgctgttcgagatcggcgcgcgttattcggacggccgcaacggcatgac
1601  ggtcgatcagaagcaggcggcgagctggtaccagcttgcggccgacaagg
1651  gctttgcgccggccgaataccgcctcggcagcatgtacgaaaagggcaac
1701  ggcgtcgaacgcgacatcgccaaggcgaagggtttttacgagcaggcggc
1751  aaaccagggcaatgccagcgccatgcacaatctcgccgtgctctacgcct
1801  ccggcgcgctcggccagcaggattatgcgacggccgcctcgtggttcacc
1851  aaggctgcgaacctcggcatcaccgacagccagttcaacctggcgatcct
1901  ctgcgcgcgcggcaacggggttccggcggatctcgaagagtcctacaagt
1951  ggtttgcgatcgccgccaagggcggcgacaaggatgcggcgcagaagcgc
2001  gacgaagtggccaaagccatgaagcccgaccagctcgaacgggcgcgcgc
2051  caaggccgatctctggaaggcggagccggtcgaccatcgcacgaacgcca
2101  tcgacatccccgacgaatgggccggcaccggcgccaagacggcgagcgtc
2151  gacatgaagaaggcgatccgcaacatccaggcgatcctcaacaataacgg
2201  cttcgacgccggcgtgccggacggcgaaatgggcgcgaagaccgtgaccg
2251  cgatcaaaaacttccagaagtcggtcgggcaggagccggacggcaaggtg
2301  accgatgcgaccgtgaaggcgctgctcgaacgcaacaaacagggcagcaa
2351  ggcaatctgaggcccgcgtcgcatcgacgcaggagtgtctctctattgcg
2401  ggcgatgaaaaacgcctggcggcactatgccgcaaggctttttttcaattt
2451  tgccgtcgcctgtcacaaaacctgaaaaaagccggattatgcccggacat
2501  atcgatc
```

pRU845

```
   1  tcggcgccccgtgattcctcagataaagctccagtgcgtcgaaatagctg
  51  aaggcgctttccgaacggacaaagcgcaactgcatcaacctgcccgtcgc
 101  atcatcgacgaacaccagcagcgagcaggccggcccaagatcttcgaacc
 151  agcgatgctcggacccgtcgatctgcaccagctcgccataggcttcacgc
 201  cgcaagcggggctgatgaaacgtccgccgctgcttgcgcgacagccacag
 251  gcccgcatccaccatccagctgcgcaacgtctcgcgcgacacccgcaatc
 301  catcgcgctcggcaagcttctcggccgccaatgtcggcccgaagtccgca
 351  tagcgttcgccaaccagcgtcaccgcgtaatctcgaacaccgtcactgat
 401  gcggttgttcgacggccggccgatcgctttgtgccggatcgacgccgcac
 451  cacccgtgttgatccgctccagcagacgacgcacctggcgcgtactcaga
 501  tcaagcgcatgtgccgccgacaccatcgtcatccggccggcgatcacttt
 551  cgatagaatctcgatccgctgcagatcacgctcgctcatcgctatcagtc
 601  ccatccgcaatctccaccgtcatcaaacccggggagtgtgacattccaac
 651  tttgcagaaacaggacacttcaactttgcggctacaatcaaaatttgcgg
 701  cgtaacggatattatggaacacaagagtgctcctgtctcgaagcgagact
 751  tcattgattgaaccagctgacgacgctggtcgaacggagtagcaagcctg
 801  aaatgaaggattggcctggcgtgggtgaattcgcagtctacatagctgca
 851  tggccagcgttgtcacattttgttgcattgcaaagatgccctcggcagga
 901  ctaactaaaggcctcaatggttgcgaagaatgcgccgctttgactagcaa
 951  gcctctgaaatgacgtagtatatgattacccagaagccttcctgcattac
1001  tgaaacttaatctgccccgtgcagctcgatggccttctccgtgttatttg
1051  ttagcagctgcagcatttgaggaaaagcgtcccgtgagtactattccgac
1101  agagaaaaccgtgctggtgtttcagggcggtggtgctctgggggcctacc
1151  aggccggcgcttacgaagctttgcatgaagctggcatccgccccgactgg
1201  ctggccgggatatctatcggttcgatcaactcagccatcattgccggaag
1251  ccccgtcgatcagcgtgtcgataacctaaggaccttctggcatcgggtct
1301  cgtccggactgcccggccattttctcggcaatggcaatgcgatgcgcaaa
1351  tggttcaacgaatcctccgcttttctggggtctttgaccggcgttcccgg
1401  ctttttcacgccgcgcgtctttgcgccatggaacatcccgggcgatccga
1451  tggccgcgatcagcctttacgacaccgcgccgttgcaggagaccctcgcc
1501  ggtctcgtcgatttcgacctgatcaattccggtgccatccgtctcagcct
1551  cggcgctgtcgatgttgttagcggcaacttcaattatttcgacaacctcg
```

```
1601  attgtacttttcgcccaaacacgtggccgcctccggggcgttgccgccg
1651  ggttttgcgcccgtcgagatcgacggccgctattattgggatggaggtat
1701  cgtttccaacacgccgctgcagcgcattctcggtggtagcgagttggaga
1751  cagatctctgcatcttccaggtcgatttgttcagcgccaagggcgtgctg
1801  ccgaaggatgtctttgacgtcgatgcccgcgagaaggaaatccgcttttc
1851  cagccgcacgcggctcaacaccgaccagttccgcaagcttcagtcggtgc
1901  gcatggcggccaagaggctgactgaaaagctgccgcctgagctgaaggac
1951  gatcccgatgcgaggctccttgaaaagatcggcaatgattgcgccgtgac
2001  catggtgcacctgatctatcggcatgccgcctatgaaaccggatctaagg
2051  actatgagttttcgcggctatccgtggaggaacactggaaggccggccat
2101  gacgatgtggtcgagacgctgaaccatccagattggctcaaccgaacgcg
2151  tccgaccaatgggatacggattttcgaccttgccgagcaacgcctgcgtg
2201  ggaaaaagccatgaaaattgaagatgtcgttcgcaatgcctttgccatgc
2251  cgcttaccagtccatcctacccgccagggccttaccggttcgtcaatcgc
2301  gaatatatgat
```

pRU846

```
   1  tctcggcatgcgcctgcatatcgggttccggaatatccttgccgagcgtc
  51  gccgcgatggtgtcggcgatcgaggccgaaaagaaatcgtagatctcgat
 101  catcgcccggttgcccgaagcggcgatgacggccttgtggaaggcgaggt
 151  cgcgttcgatgaaggcggcctggttgccgccgtcgtaattgccgcgttcg
 201  gcaagcagcctgcgaagcccggcaacaatctccggcgtcttgcggatcgc
 251  tgtcagccgggcggcctcgacgtcgagcgccaggcgcgcctcgaactggt
 301  cacgcaggctggcgcgccgcgccatcgtcagcggtcggccggcatcgctg
 351  gtggcgcggacataggtgcccgaaccctgccttgtctcgagataaccctg
 401  agagacgagaacgcgcaccgcctcgcgcaccgtgccgcggctgacggaca
 451  gcatggccgacagggaggcctcattcggcaatttctcgccgaccgcccag
 501  cgtttgccgagaatgtcgccgcggatcgcctcgatcgcctcatcggcgag
 551  attggtcttgctgagcgctctcatcccactactcataaagtcatctgatg
 601  actttatgagtagtgggatttcgatctttcgtcaacgcccggcatcgaga
 651  caataaaggcatcaagacaataaaaaagggccgcctcagcgacccctttcc
 701  atgagtttggcctcaaactggcctggactgtctatgcagccccaggaagg
 751  cggggattagagcgagggctgcctcgccttcgtcacctcttccatgcccg
 801  ttacgaggtccgaaatctcattagacattggatcaccttccttttcgttct
 851  gttgctgatgacttaaaagtaggccgattcttcgcagatgcaagagagaa
 901  atgcgttcagtgcgcagctttgatgaaggtgccgttctgcagttcgcgca
 951  tcgcctgcatgatttcctcgcgcgaattcatcacgatcgggccgtgccag
1001  gcaaccggctcctcgatcggcttgccggagacaagcaggaaacggatgcc
1051  ctggtcgcctgcctgcaccgtcacttcgtcgccgctgtcgaagaccacga
1101  gcgtgcggttgccggacatgtcgcggatattcaattcctcgccgtcgact
1151  tccttctcgacgcgcacgccgaagggtttcgatgcatcgcggaaggtgcc
1201  ggagccggcgaagatataggcgaaggcattgcggtaggtatcgaccggca
1251  ggcgcttcttccggccgggcggcaccgaaacgtcgagatagatcggctcg
1301  gcggcgacgccatcgaccggaccggccttgccccagaaatcgccgcagat
1351  cacccgcaccgccgtgccgtcatcatcgacgacgatagggatatcgatgg
1401  atttgacgtcctggtagcgcggcggcgtcatcttcagcgaggagggcaag
1451  ttcgcccagagctggaagccgtgcatacggccggcaaaatcgcccttcgg
1501  catctcctggtgcatgatgccgctgccggccgtcatccactggatgtcgc
1551  cggcgccgagcaggccgctattgcccaagctgtcgccgtgctcgacggtg
1601  ccggcgagcacataggtgatcgtctcgatgccgcgatgcggatgccaggg
1651  aaaaccgcggatatattccgcgggattgtcgttgcggaaatcatccatca
1701  tcaggaagggatcggtcatcgcgggatcgccgaagccgaagacgcggtgc
1751  agcttgacgccggcgccttccatggtcggtgtggcagtactttcatgctt
1801  gacggggcggatcgacatggcgtcctcctaggggttcaacttggcggcga
1851  ctatagccgcccggctgccgggcgcgaagtctgcgggacagacaacgcag
1901  tgttcacaaaagaacacagcgtttcgccgaatttgaagaggtcagacgtc
1951  aagaaatcgtcatgtttcacttgcaaaccggcagggcgttaagctgttat
2001  tagcgacgatattcgacaattcagttgtcgtttcagcattttccgccgat
2051  gcggcgggttccggagcaaagagccatgtgtcgctgggcagcctatcgcg
2101  gagacccccctctatctcgaggagttggtatcctcgcccgcccattccctg
2151  atcgagcagtcgcattgcgccacccgcgccaagaccgcgacgaatggcga
2201  tggcttcggcattgcctggtatggcgatagacccgagcccggccgctacc
2251  gcgatatcctgcccgcctggtcggattgcaatctgaagagcctggcgcgg
```

2301 cagatc

pRU848

```
   1 tcagggaaaaacccaggatcgagtggcagggccgctttcgaccgtccctc
  51 agtggcgagggtgtctttccccgtccgcaccagccaaccttgccaatctg
 101 gctgggtgttggcgggacgcccgaatccttcatccgcgccggaacacttg
 151 gcttgccgctgatggtcgccatcatcggcggtagtttcgagcggtttctg
 201 ccgttggtcgaactctatcgacgtgcatggttggccgcggaccaccccag
 251 cgaggaacagagagtcggcgtgcacgctctcggctttgtcggggagaccg
 301 acaaggcagccagggaagcgttcttccccggctggttccagatgttcacg
 351 gacgctggacgggaaaggggctggccaccaccgacgcaggcgcaattcga
 401 cttcatgtgcggtccgaacggcgcgtttctggttggcagcccgcagacgg
 451 ttgcggcaaagatcgagcatattaacaggacattgggaggggtttctcgc
 501 ctcaccttccagatgagtaccgccgccttggagaccgaagcaatgcgccg
 551 ttcgatcgaactccttggcacccacgtcgcaccgatcgttcgcgcacacc
 601 gtctcggccggagccgttgagcgcgcctacattgatcggcgccgctctta
 651 ctcagagatcagtctctggtaaccggccagcatcttcgaatgctcgaccg
 701 acccgtgatcgtgaagctgtcgccaggacccgttcctgcgcaccaatcac
 751 tgcgtgtgatgtcctacacattgttataccagggtgccgcggcagatctc
 801 cggaaggagacgcggcggccttcatcaaactggtcgttgtgaactgccgg
 851 gtgaaggtttgcgttgcggcagactgttccccgcgggtaggtagaattga
 901 caaaagttcccatcagaagggcgtgcgccgatcgagaacgtcaagcgccg
 951 ttaaatagagctcacgagtcagcgtttctgcgcctgtcctcaaggcgagt
1001 tttgcggcctgtcccgaccacagattcgtgaaatcgtcgcgcgagtccgc
1051 ctccgccttcgctctgataggggcaagcgcggcggccgccgttggaaatg
1101 caggaacgccctccgagattgggccaagctcgcgcacgagccgattgacc
1151 acgctcctcgccggacgaccggtaaagacgttcgcgatcgcagtgccgtc
1201 gtctctcgaggcggccagcgcctcagcgtggacgtcgggaatgcgtgcct
1251 ccggacagaacaggtatgccgttccgatctggacggcggaggcgcccagc
1301 atcaatgcggcagccactccccgtccgtcggcaatgccgcccgccgcaat
1351 gactggaacacgaactgcgtcgacgacctgtggtacgagcgccatcgtgc
1401 caacctgcgtcgccatgttatctgtcagaaagttgcctcggtggcctccc
1451 gcctcgaaacccatcgcaatgacggcatcgactccgtttgcttccaacca
1501 tactgcctcggcgacggtggtcgcggaggaaacgactttggcgccggtgg
1551 ccttcacccgctgtacgagctcccttgccggaagcccgaaatgaaagctc
1601 accacctcaggcctcaaatcctccacgacatcgcagaaggcggggtcgaa
1651 aggcgcacgcacgcctcccgcgactatcgtagccggatcgagatcggctt
1701 cggcatagtatggtgcaagcagtgctctccagcgcgcctgcgctgacgga
1751 tcgtctgccggcgaaaggtgcgcgaaaaagttcacgttgacgggcgcctt
1801 ggtcgccgcgcggatctcccccaaggccttgcgaagcccctcgacgttca
1851 acaaggcgctcggcagcgaacccagccctccggccttggcggtggctacc
1901 accatgtctacggtcgtcgccccgccatgggcgcctgtatgatgggaag
1951 ctcgatcccgaacagatcgaggatgcggcgatccggccagcgagtcatgt
2001 gaactccaatcattcgtggtgctgctggggtgatacctccggccattcct
2051 cacgtctattgtaccttggtgtgcccaaggcggcaccccgatccgaaggt
2101 atcggagacaccttcgtacaatatcgtggcacggttttcggcgcgattgt
2151 cccggcagagcggaggacgatgcagaccgtccggctgcttccggcacgca
2201 gccggtcaacgcatcaatcgagggagcacgcgccgcaaacgaacgaaggc
2251 tacccgcgtcgccgagccggaggcaggcatgagcgtgcataaaaccgggg
2301 ttatcaccggggcttcgcagggaatcggagcgggactggtctacggccgt
2351 cagcatccggcggtcgcaatgcgcgcgccgactttggcacgatttgaaaa
2401 gtaatacttttaatgagcgaagggttggcatacctgccgtcgatgcctat
2451 gtgagatagacagccatacgccgacggtttacgcaaagcctccattcagc
2501 atgcgagggccatggcttctcattttgaaaggatcca
```

pRU849

```
   1 tcccggccactcagcgacgaggcaggacccgaccatggccgattaccttc
  51 tggaagcaagctggagcccgatcgagggcagtttcgggcgcctcaccttc
 101 acgcttttcaatctttcgaccgagccgctgtccggcttctcgctcgccta
 151 tacgtcggagacgcgggttgccgacaaacatgtctgtgacggcggcagcc
 201 tcaagcggcaggtcgcgcatttccatgaattcctgccgcccgaggggctg
 251 agcgtgccgcccggcgggcgctggcggttcacggtcgagggactgaccag
 301 ggcgccgaaacatatgacatcaggcgtcaagtcggcctatctgacacttg
```

```
 351  gcgacgggcgccacgtttctgttggtttcggcgatctcatgctcgaaggc
 401  cgggatggtggcgtggcgccgcccattctgccgccgggccgggccgagga
 451  accctattcgctgctgccctggccgctggcgctcgggctgaaggcgggag
 501  agctgccggtcgtcctttatccggccgagccgacgcggcccgatgcggtc
 551  aaggcgctctcgctggttctggagctctatcagcggctctatccggccga
 601  caatgtgccgttttccctcagtgccgtcgaaggcgggcgggccattcgtt
 651  tcgtcatcgaatcgtcgatcgccgccttcgcttacgaattgcgttttacg
 701  gcgcatgagatcatgctttcgagtgccgatgccgccgggcggcactacgg
 751  gctgatcagcctggcgcaattgctgcacggcgctcgcgccgatcgcgagc
 801  gcttcaaattccccagtttcggcacgatcgccgaccagccgcgttatgac
 851  tggcgcggctgccatctcgatgtgtccaggcagttttatccggtggcgga
 901  cgtcgtgcggctgattgatattctcgcctggaacaagctcaacatcttcc
 951  actggcatctggccgatgacgaagcgtggcggctggagatcaaggcctat
1001  ccggcgctgacggagatcggcgcccggcgtgggccggatgaagtgctcgt
1051  gccgcagctcggcgacggggcgcaaacgcgctccggccattacacacagc
1101  aggacgccaggcggatcgttgcgcatgcagcctcgttgcatatcgaagtg
1151  gtgccggaaatcgatattcccggccatagcatggcgacactgttctcact
1201  gcccgagctcgtcgacggccaggaggcgccggacagttaccgctcggtgc
1251  agggttatccgaacaacgccctcaatccggcggtggaattcacctatgaa
1301  tttctcggcaaggtgttcgacgaaatggtgacgctgtttcccggcgaata
1351  tctccatatcggcggcgacgaagtggcgcatggctcctggctttcctcgc
1401  cgctttgcaaggcactgatggagagggagaaacttaccggcaccgccgag
1451  ctgcaatcctatttcctgaagcgtatcaaagccatgctgtccgagcgcgg
1501  caagaaactcgtcggttggaacgaggtctcgcatgggggcggggtcgatc
1551  gcgacggcacgctgctgatggcctgggaaaagcccgaggtcggtatcgag
1601  ctcgcacagcaaggttacgacgtggtgatgacaccggggcaagcctatta
1651  tctcgacatggcgcaggcggaggcctgggccgagccgggcgccagctggg
1701  cgggccacgcaccgccggaacatacctacgcttacgaggccgagggcgag
1751  ctgccggaagtgctaagggagaagatgcgcggtatccaggcctgcatctg
1801  gaccgagaacttcctctcgcgcgcctatttcaaccggctggtttttccgc
1851  gcctcccggccgtcgccgaagccgcctggacgccttctgcgcgcaaggac
1901  tgggatcgcttcgcagcgatcgtgcgaatgtggccggtgctttaagctgc
1951  cgcctcggccttcagccccaaaagcgctgccccgatcaggcctggttcga
2001  tgcggcattcgctgcgcaccaccagcgggcggtcgaacttgcgcaggatg
2051  cgggcgcgcacagcgtggtcgagttcggcaagcagcggctcgacattgga
2101  aagcccgccgccgaccggcacgatggtcgcgccggtgatgttgatcgcca
2151  gcgccaagggcgaggcgacgagatcgacatagacgtcgatcgtccgtgtc
2201  gccttctcctctccatggcgccactggccgatgatctcttcgctggcaag
2251  gtcgagatcgtgcagcgtcttgtgcaggcgctccaggccgcgggcgccgc
2301  cgaccgtgtcgacgcagcccctctggccgcagccgcaggcataggcggga
2351  atggcgacgggcggatcgccggccgcagatgcgatgatcggcccgtgacc
2401  ccattcgccggcaaagccgccggcctcgttgacgaggcgtccgtcggcaa
2451  ccaggccgccgccgacgccggtgccgaggatggcgccgaagacgatgcgg
2501  tggccgcggccggcgccgagaccggcttcggccatcgcgaaacagtcggc
2551  gtcgttggcgatcagcaccggcaggccgagctcggcttcgaggtcggctg
2601  cgagcgtgcggtcgtggatgcagggaatgttggcgcagatcaggcgctgc
2651  gtatcaggatcgacgacaccggcgatggaaaggacgatgcggctcggctg
2701  ctcaccggtctcggcgatgatgtcgcgcagggtttcgacgaaagcggcga
2751  aatcatccctgggtgtcgggcggcggccgaggggaaggatgtctgtctcg
2801  gagcgggcgatgccgcccttgatggcggagccgccgatgtcgaatgaaat
2851  gatcattgccactctgccgtttctctggtctcccgcgtctgcttgcattg
2901  ttggtggggaatttcaagcactcattcggcggtggtgtcaccggctttat
2951  ggtcgtgccgtgtgccacccccctgttctgccggacatctctcccacagg
3001  tggggagatcggataggcgggacgggatcttccattccgtcatggcagac
3051  gttggagaaatctggtgagcccaagcccataggctccgatcacagccgca
3101  atttctaaccccgcaagtcattcaattggcatattgcatcgccagaatcc
3151  ggccctacactttcatcaactgccggacccacgaccatcgatacgagaga
3201  aaacagtaccatccgaagagcacgtcccaaaggccgccgcctctcatcta
3251  ttttgcgacagatggccgccgatcagagccgcgaacggatctcgatcggc
3301  gatatattcgacacgatgggtgacagggcgatc
```

pRU850

```
   1  tcaagcgccgcgacagcgtctaccaggccggccgcatgaaaggcccctgg
```

```
  51  ttcaagtggaagcgcaacccctataatgtcgatgcggtgctgatgtatgc
 101  ccagcgtggccacggcaagcgctccagttactattccgatttcaccttcg
 151  gcgtctgggccgatgacgaggacggcgaacagctggtgcccgtcggcaag
 201  gcctatttcggcttcaccgatgccgagctcgaggtgctcgacaaatttgt
 251  gcgcaacaataccaccgagcgtttcggcccggtgcgcgcggtgcgcgccg
 301  acaaggatttcggcttcgtcgtcgaagtggccttcgagggcatcaaccgt
 351  tcgacccggcacaaatcgggggtcgccatgcgcttcccccgcatcgcccg
 401  ccttcgccccgacaaaccctcctacgaggccgaccggctgcgcacgctga
 451  tcgccatgatcgaggccaagccggcgtgaagttgcaagccctgcgcaaca
 501  atgacgcagggcaatattttcaacagcatgtgtattggcggcaacgatct
 551  tccggtgcagattctgcgtgccggggaaatgatggatgctgcgttatgag
 601  ctcgcgtgaacgaaattccctttcctccagcgccgtgccgtgctggcgg
 651  gtctcgccggtgcacttatcctgccgcgcatggcggccgctttcgatgtc
 701  ccggacgagccgcgccttgccaagcgcgactatgccaaggtccgccacca
 751  gttccgcaccaagcttttgcagaagggcccggcgcccgacaaatacgaac
 801  cgctcaatgcgcctgccgatgccgacaagatcttctaccgctcgggttac
 851  ggcgagctggaactggcggcctgggtctcgaaatacaagcgcgagcgcgc
 901  cgccaggccggcggtgctcttcctgcacggtggcaacgccatgggcatcg
 951  gccactggcagctgatgaagccctatatggatgctggttacgtcgtgatg
1001  atgccgtcgttgcgcggtgaaaacggccagatgggcaatttctccggctt
1051  ctacgacgaggtcgacgacgtgctcgccgccaccgagcgcctggcgcatc
1101  tgccgggggtcgatcccgagcgcctgttcatcgccggccatagcatcggc
1151  ggcacgctgaccatgctgacggcgatgaccacccacaaattccgcgccgc
1201  tgcaccgatttcaggcaaccccgatgccttccgcttcttcaaccgctatc
1251  cgcaggacatccgtttcgacgatagcaacgcgcatgaattcgaggtgcgt
1301  tcggcgctgtgttacgcccacagcttcaaatgcccggtccgcgtggtcca
1351  cggcaccgaggagccgcatttcaacgaccgcgccgatctgctggcccgcc
1401  gcgcccgcggcgccggcgttcatatcgaaacggaaaccgtcgccggcaat
1451  cacacctcggcgctgccggccgagatcgaacagagcatccgcttctttca
1501  cggagtggcggcctgagtttcgccgttaccggcgaaagcgattaagtgtg
1551  acggatgattgatc
```

pRU853

```
   1  tcgaaacgcgcgaatgaagctagatgctggtcccttgagctttgccaagg
  51  tcaggaagcgtgtcgtcaaacattcgatctggctcctgatcggcgtcgtc
 101  acgggcggagcgtggatcttttattttgccgacgcgccgagcctgcttgt
 151  ttcgctgttcaccggccgcgcgcctgcagccgcctacaccacggtcgcca
 201  tccttaccgcaacgacgtatgtgctcggcggtctcatgcgcgagcaggtg
 251  tgtacttacatgtgtccgtggccacgcatccagggcgcaatgctcgacga
 301  aaattctctcgtcgtcacctacaatgactggcggggcgagcaacggtcgc
 351  gtcacgccaagaaggctcaagtcaagggcctgccggtcggggattgcgtg
 401  gattgcaatgcctgcgtggcggtgtgtccgatgggaatcgacattcgcga
 451  cggacagcagatggagtgcatcacatgcgcgctctgcatcgacgcctgcg
 501  acggtgtcatggacaagctcggaaagccccgcggcctgattgcctacgcc
 551  acgctgagcgaatactcgagcaacatgtcgcttgccacggatgaaggacg
 601  aacggccgtccagccgtccagggttcgaaacgaggatgggaccctcgttc
 651  cggcgatccggcacttcaactggcgtatcatctttcgtccgagaaccgtg
 701  ttttacgcagtcgcctgggcgtcggtcggcgtggccatgctcgtccatct
 751  cgcctttcgggagcgccttgaactcaacgtcgttcacgaccgaaaccccc
 801  aatatgttctggaaagcgacggctctctgcggaatggctacacgcttcgt
 851  gtcctgaacatggtgccgacgccgagggatgtgaacataagcctggtcgg
 901  gctggaggggcgacgatgcgcatccccgagttcggcaaggatgacgctc
 951  gcagctttacagtccatgccgaacccgacgcggccacgacgctcaaggtc
1001  ttcgttacacgcaagcctaccggagccgcgatcaa
```

pRU854

```
   1  tcggattttcggttcggtgacctcaaggcagacaaggcgaccgtgtttct
  51  ggtcttgccgccagacaggcttgcggcctattcgcgttggctgcgtttgc
 101  tggtggcccagagcctgaccgagatggcgcggaccgcaccctcgccgcat
 151  ccatccgcgccgcctgttctctaccttctcgacgagtttgccgcccttgg
 201  tcatctggcgccgatcgaacgcgccatgggattgatggcaggctacggtg
 251  tgcaactctggccaatcgtccaggacattcatcagctgcgcgccacctat
```

```
 301  cgccagcgcgccggcacgtttctgtccaacgccggcgtgctgcaagtgtt
 351  cggggtcaatgatcacgacagcgcccgtctggtctccgatctgcttggcc
 401  aggagaccgtggtgttcaacacggcggcccgtgcactcgattcggaaaag
 451  agcggcctctcctttgccgagcagcatgtcggccgcccgctgctgacgcc
 501  cgacgaagtccgcaatatgcatgccgagaccgagctgctgttcatcgccg
 551  gccagcggccgatcgttgcgacaaagttgcgatattatgccgatccggaa
 601  ttttccggattgttcactgcatcacagatctgacgaggcgacatcaccaa
 651  cgtgaatcgatcgggcctcatcgattcacaaatgtcgttggacgccgccg
 701  cctggtctttgcgattctgccggcatgacacgagagccctatcgtctcta
 751  catcgaacgtatcgatccatcgaagaacatggctcgttattacgcgctgt
 801  cgatcgagccgaacctcttcggcgggacatcgctggtgcgcagctgggga
 851  cgcatcggcagccgcgggcagcagaagatccacgtcttcgacagcgaggc
 901  aaaagccgtcgatctcctgctcacccttttacgcaagaagcggtcgcgag
 951  gctatcgggtcatgcggtgatcggtatccgcaggagacagttgagtgaag
1001  gaagagacggctcacgcctcctcctcctcgtcgagttccggctgcaagga
1051  atgcaggtagtcgacggcgcgctgcgcatgtgctgccgccgagaagatgg
1101  cacgcttgtcgtggtggagaaccttgagccagccatcgagatagcctgcg
1151  tggtcggggcgtggttcgagctctgggacgatgccaagatcggcgcagag
1201  aaacgcgctgccgagttcggcaatcagttcttcgcgggcgcggtcgctcc
1251  ggtcttttgcgtagcggctgagatcgcgatcaagccgtcgcggtgcagcg
1301  gtccagtgggtgatctcgtggctgaggatggcgacataggcggcatcgtc
1351  cctaaaggcatccaggcatggcatctggatgtaatcacgctgagcggcgt
1401  aataggcagctgtccctccatgccggatcagcgcgccggtattggcaaag
1451  aagcggccggcatcgccgatgcggctcagcggatcctggtgaggcgcgac
1501  gtcgtcaaagcagccgtcgagaccggcgatctggtcggtgttgaacaccg
1551  tgtaggctttgaggaatggaatatcctgttcgatctcggtgcctgtatcg
1601  gtcgtctctgtgcggatgaaggaactggcaaagacaatggtcgtgccggt
1651  ttcgcccttgcgaacggctccgccgagctcgattgcctggcggaacgtca
1701  tccagcgtgatgaggcaaagccgcgggcgatggcttgcgaccacagcagc
1751  aggacgttgatcccggaatagggttggccgttgtggcgcagtggccggct
1801  gacctcggctgtcgcatgtcccgtcgtccagggtttggtccatggccgca
1851  cgccctgttcgaggtcggcgatgatcgtgttggtgatgcggctgtagata
1901  tcggatcg
```

pRU855

```
   1  tcccgacaaggcggatgcggcggtgcgccgcttcatggcgccggttcttg
  51  ccgcgctcgagcggatgatcctgtcatgaacacgcctgccaaccatgccg
 101  ccagcgcgctcgacgccgccgacgaattgctgatcgatgccgaggccggc
 151  agccttccgccgccgcaggccacgccggtgcggggagctgcgggtcgttt
 201  tctgctgggcgcaagcctggtgctgatcgccttcaacctgcgcccggtct
 251  tctccagcgcctcggcgctgctgccggaaatccgctcggagctcggcctc
 301  agcgcgctgggcgccagcctgctgacgacgctgccggtggtctgcctcgg
 351  cgccttctcgccgcttgcgccccgtcttgcccagcgcttcggaacggaac
 401  gcacgctgctcggcgttctcctgcttctggcgctcggcacgggtttgcgc
 451  gggctttcctccgtgccgctgctcttcatcggtacggcccttgccggcgc
 501  ctgtatcgccgtcggcaacgtgctgctgccggggttggtgaagcgggatt
 551  tcgccgggcgcgccgcgctgatgaccggcttctatacgatggcgctctgc
 601  gccggggcggcaagtgccgccgggctgacgctgccgatcgagcacgcgct
 651  cggcggctctctcgagggcgccctcgccgtctgggcgctgcctgccctca
 701  tcgtcagcctgctctggctgcctgaggttcttcgcagcggcaggcaggca
 751  agacggaacggcttccacgtcaaaggcctctggcgcgaccggctcgcctg
 801  gcaggtgacgctgttcatggggctgcaatcggcacttgcctattgcgtct
 851  tcggctggctggttccgatcttgcgcgaacgcgggctcgacggcgtcacc
 901  gccggggcgatcgtttccctgtcagtgatggtgcaagcggcatcctgcct
 951  gatcgtgccgcatatcgccgtccgcggtcgcgaccagcggctgatcaatg
1001  cgagcctctgcggcgtcgccgtcaccgccctgctcggcctgctcttcgca
1051  ccgctttcgacggtctggctatgggcggtgctgcagggcatcggccaggg
1101  tgggctgatcgcggcggccatgacgacgatcgtgctgcgctcgcgcgatc
1151  ccgatgtcgccgcccatctttccggcatggcgcaatgcgtcggctacctg
1201  ctcgccgccatcggcccgctcatcgtcggcctcatccgcggttggaccgg
1251  cagcttttcctggtgcgcagccctgttcgtcgcactcggactgggcgcgg
1301  cgatcaacggctggaaggccggccgggcggtcgagatcaatgtccatgcc
1351  gccgaaaaggacggttgatccgcccgctcacctgtccatcaccatctctt
```

```
1401  gatgccggtcgtcggccgcggatgctacgctatctcagcctctcatcccg
1451  accgccggaggaaacatcttgcgcatacgaaccctgtttctgctcgctac
1501  gtttatggcgtcgctggcacctgccctgtcccatgcacaggaaagcaatc
1551  aaccgagggcaaatgtccagagcggcgcgcgcgacggcgtgctgaagctg
1601  ttgccgcccgattccgtcaccgagcatgcgctgacgatcggcgaccggaa
1651  gtttgcctataccgccaccgccggcacgctggatctcttcggccaggacg
1701  gcgcgcagaccggcgcgatcttctacaccgcttatgtcgcaagggatagc
1751  ggggcgaaccggcctctgacctttgcctttaacggcggaccgggtgccgc
1801  ttccgcttatctccatctcgggctggtcgggccgaaggtgctcgatttcg
1851  ggccggatggacgtgacggcgccaatgcgaaactcgtcgacaacccgcaa
1901  agctggctcgatttcaccgatctcgtgctgatcgatccgatcggcaccgg
1951  ctggagccgcacgacaaaggcggatgacgccaactattacaacgtcgacg
2001  ccgacgcccagagcattgccaaggcgatcgcgctctatgtcgcgcacaac
2051  aaccgctccacctcgccaaaatatcttctcggcgaaagttatggcggctt
2101  ccgcgccgccaaggtcgcctccgtgctgcagcaaagccagggcatcatcg
2151  tcgccggcgcagtgatgctttcacccttgatcgagggccagctgatgttc
2201  aatgccgatcagttcgcgctcggcgcggcactggagctgccgtcgctcgc
2251  cgccgccgaactcgaccggcacaaggcctttgacgaagagaagcagaagg
2301  aggccgaaaccttcgcgctcggagactatctgacgacgctcgccgggccg
2351  ccgccgacgggggccgatgccgccgccttctacggcaggatcg
```

pRU857

```
   1  tcgcctcgaactcgacctcgaattccggcgggttgaagatgacgcccaaa
  51  gccgccttggtgctcgcccgaatcttgcccagccgggaacgatagctgta
 101  ggcgttgcggtagcgcaccatgcgggcatgcaggccggcggaaaactggt
 151  ggatgaagggccggccgttggcgctggcgatatcgacattgtcgacctcg
 201  ccggaggcaagcacatcgagcgcctgccagatatcgagcggcacacgcag
 251  cgagcgggcaaagaggttcatcgtgccggctggcacgacgccgagcgcaa
 301  ttccgtttttccaggcaatcgatgccgccgccgaaatcgtgccatcgccg
 351  ccgccggcaacgatgccgtcgatatcgtcgcgcttggccgcccgctccat
 401  ggcaggaacgatttcctttcctgagaagacgatggcgtcgaaatcgtgcc
 451  ccgcttcgcggaaagccgtctccgccctttttttcataggccagcatgtcg
 501  gtggtcttgaaggttccgccgtcgcgattgaaaaagcctacaagcttcat
 551  aagagctcccgtctctggcgcggaaatgctcctgccttgcccctgagatg
 601  gttgcggtatcgccgatttcaagcgcaaccttctttcctgacaaagggca
 651  gaaatgcaaaaattgcaatagacgtatgcacatagcaggacgctgcactg
 701  caaaaaatgcactcgacggcaggccggacctcacggatagatgctctatc
 751  ggagagatcgtctccacctcccgccatttccaccgatcggcttgcggccg
 801  cgcctgcgcgccagggccattgccaagcagaaaccgcctcgaggagctct
 851  cgtgagtgagattgccgccaatatttcagttgatagccgggatcaagagc
 901  tgccgtcggcaatgaccgtggcactggtccagctggcgctcgcctgcggc
 951  ggcttcggcatcggcaccggcgaattcgcgatcatgggattgctgcccaa
1001  tgtcgcccagaccttctcggtcacgacgccgcaagccggctacgtcatca
1051  gcgcctatgcgctcggtgtcgtcgtcggcgcacctgttattgccgtgctc
1101  gccgcgaagatggcgcgccgcacgctgcttttgaccctgatgctgatctt
1151  tgccgccggcaatatatcgagcgccatggcgccgacattcgaaagcttca
1201  cgctgctgcgcttcgtaagtggcctgccgcatggcgcctatttcggcgtc
1251  gcggcccttgtcgccgcctcgatggtgccggcgcatcgccgcgcccgtgc
1301  cgttggccgcgtcatgctcggcctgaccgttgccacacttctcggcacgc
1351  cttttacgaccttcttcggccagtcgctcgactggcaggtggcgtttttc
1401  tcggtcggcgtgctcggcctgctgaccgtggtgctgatctggttctacgt
1451  tccgaaggacagggtttccgcggaagcgggcttcctgcgagaactcggcg
1501  ccttccgccggccgcaggtgtggctgacgctcggcatcgccgccgtcggt
1551  tacggcggcatgttcgcgatgttcagctatatcgcctcgacgacgactga
1601  ggtggcgatgctgccggaaacagccgttccgatcatgctggtcctcttcg
1651  gcgtcggcatgaatgcgggcaatttcatcggctcgtggctcgccgacaaa
1701  tcgcttctcggcacgatcggcgggtcgctgatctataatatcgtcgtgct
1751  gaccaccttctcgctgaccgccgccaatccctatttgctcggtctctccg
1801  tcttcctcgtcggctgcggctttgccgccggcccggcgctgcagacgcgg
1851  ctgatggatgtcgccgccgacgcgcagacgcttgcggccgcgtccaacca
1901  ttccgccttcaacatcgccaatgcgatcggcgcctggctcggcggcctcg
1951  tcatcgccgggggttacggttttgcggcgaccggttatgtcggcgcggcg
2001  ctgtccttcctcggcctgttcgtctttgcagcctctctacgcctcgagcg
```

```
2051  ccgcgatcggagcgcgcaggccgtctgaggccaggcgcagctcgcggcct
2101  tccgggctgcagaagacataatcgtcaccccaggcggccaatgcctggat
2151  gacgggctcgagcgtcatgccgagcggcgtcattgcatattcgacccttg
2201  gcggcacaaccggatagacggtgcgtgataccaggccagactcttccagt
2251  tcacgcagctgcttggtcagcatgcgctgggtcacggccggcagtttacg
2301  ccgcaattcgttgaagcgcagggttccttgcatcagatgaaacaggatc
```

pRU858

```
   1  cgagcgagcagtagaccgcgcgcacccccttgctgttccagcggccgcca
  51  acccgataggcgccttctccgctatcccaagtcggcgcatggacggcctg
 101  gtcgagccgccacgcgaccagttcggtcccgccgagcgcgatcggcaaaa
 151  gggtcatgcatagacgccgtattccaggcgatcgagatgatcctcgacaa
 201  gctcgacgccggcgggcgtcccgagaagatcgatcgggcggcgctgctca
 251  agtccgaccgccggacgctccagccattgttcggcttctgcctgcgatcc
 301  gaagatatccgttgccttgcgaggatctcggcgaacttccacgctcggc
 351  cgctctgttcctggttgagcggcttggaaggcgtgtccttccgccgctgc
 401  caggtgcgcaagctcatgcccaccgccttttcgagggactcgttcttgcc
 451  gaggacgacgagagtaccgacaagatgatccagtgccgatgcaggcaggc
 501  cgtgcaggagcagctcatgggcgtcgagggcattggtcaggctgcgcgac
 551  agaatgcgcgagccgccgagcaatgcttctatcttttgcagttcaccacc
 601  gcccgctgctgctaaggaaggtactttgccgcgacagccatgatctttc
 651  tcccgcgtcatctgtcgcttattatatgtcatgtgtcgcgacaaatttca
 701  agatctgacgggcgtattcggcgtatcggccgcggaaccgtgcctttctc
 751  gctgccggcgccaaacaagcgggcaggcggcgatggcaccgcctgccaag
 801  ggcagcttgcagcgtattgaaatgccgcgatccccactgctttaagaggc
 851  tgggttgcggcgctccaattccataagcgcgcggatggcaggcagggctg
 901  ccgctcccgggattgtcagtcttttggctggtcggcgctggcagcaaga
 951  tgggctcgaccgaatgtttcggtgccatcagctgtgcaagaatgtcgtcg
1001  gccgagtggctcgaactgccaatgccggcctcgcgcagctgacgatcgag
1051  gtcgctgccgttttccagcgcggccagttcagcaccggcttcgatgcggc
1101  cggcagtgatttcctggcgtttttttgatacgctcgagactttcgacggcg
1151  ctaccgaggcgggtattgatgcccgactggctatgggcgatcgccgactg
1201  agcgcgcagcagcgactcattggctttgacattctccacctcacgcttca
1251  tctgagcaatgcgcgcttcggtctcctggatcgtgcgcagcatttgctgc
1301  tggcggggcagcagccggtcgagctcttcctggtcgcgctgcacttcggc
1351  tcgggtaccggcaatgcgctgagcgagaccctgagcgagatccgtgcggc
1401  ccgcggagatcgcagcgcgagcgctgtcggtatccttggtctccttggcg
1451  cggttttcttcaagtcggcgcatgacgctcttgttggacgccatgatgcc
1501  ggccagatcg
```

pRU859

```
   1  tcctcgacgcgtcgaaatccgcgtcattctccagagctgtcacctctgcc
  51  acagtcagaccagtggtttctgcaacgtcatccagcgaatagccgcgagc
 101  taggcgagcttctacaagcagagcaccaacggacggggcaacttgggatt
 151  tcgtaatgttcgagaccatgttcatgtcgtttaaggctccttttggcggc
 201  cggaaaacgtgcatccgttttctagttccaccgctgccgaccacgtccgg
 251  cgacggacagtttcttcgaaatccgcgggcgctctttctatcagtttgtg
 301  cgacggagtataggtctgacgcgatgaaggacggacaacccttgcatggc
 351  agccgacaagaggtgacccgctcccaactgcggcgcaaactcgcgacggc
 401  cgataaacagcttcctatctggcaggagcgttcatgccggtgggaagctc
 451  gtggcgagtgatgtactgatcgttctcaagcggggagcgacggacaaaca
 501  attcgagagttgcctcggcgtgatacgctcaagggccgcaggttcggcca
 551  gagcttgtgtgaaatctcctggctctgttgatcggccgtcgtcttgaaga
 601  aatcgccggagaacggcgcgtcgagatcaagcgcgacctgaggtcaagga
 651  cccttcatcgtcttggcaccctaagggttttgcggaacttcgagcgcccc
 701  aaaagaggatgtgagcaagcccgcattgcgcggcattcgcgcctcagttc
 751  accaagtcgttgacgttggcaggcgtgatgagctgaaacggaatgtagcc
 801  cttcttttcgacctttcgcccttagcaggcgtgagtgcggcatcgaacg
 851  atgccttgccctggccggcggcatcctggaacacggcaaagtcgagatcg
 901  cccgcctgctggcggcaagcgcgtcctgcgtggcgtcaacgccgccgacg
 951  acgacctgggtcatatcccttgccggtctttgaaacgtctgatcacgccg
1001  tgtctgcatggcggcagacggctttcgaccgtgcccagtcaagcatgccg
1051  tagagaacggcgcaaccgatggcgacttcggtctgctgaccaggccgcga
```

```
1101  cctggcccgaaggcgccgtcctatgatggccttgtatcgcccgatgaccg
1151  tttcgaccaggaagagtttagcgtagcggttggggacctattccccaag
1201  cgccaagttatgaagtgcgaaatacacaataatatatgaaataatatttt
1251  tataactcgcggctttctcccacaagtgggcgcatctgtcaaatatgaaa
1301  agatatactataaaatatacgcaacacactgttgcgctgcttgttatttt
1351  catgaagtgtcgtgtcctgaggagctacggtagcaacggaaattctttcc
1401  cgacgaatgctgagatc
```

pRU861

```
   1  tctcccgcaccggtcggtgcttgagatccgccgtggacaaaaccctttcg
  51  gagattaccgagggaggcgaagtgcaccgcgagcttcacgcagccgacac
 101  gatccggacaatcgaacgctcgcacctgctgttggtgggacggcgctttc
 151  cggaaaggcttcggctgcgtgttcgatggccttgctcccagaagtgaata
 201  ttcgtcatcatcggctcatcgcttcgacatggcagccgcgatatttaata
 251  tgaagggattacctctcaggcgtgttacggtctctctgtcgggctgagcc
 301  ttctggccggttggcattgttgccgagtgatgacttgctccgcgagtctc
 351  cggcccatgagggcctctccatgcaaaagattcccactgtgcatttcgag
 401  gcggcgagcactcttgccgtggtggtatcgtccaatgaaccgctgctttt
 451  cctgtccgacgaccagaaggtcattgcggccagcgcatccttctgccggg
 501  tcttcgacatcgaccccaagaccgtttgcggccaaagcctcagcgacatt
 551  ggaaacggcgaatgggcgatgccccagcttgcgtcactgctgacggccac
 601  cgcctcgggaagtgccaccatcgaggcctacgaaatcgactttcaacggt
 651  caaaccagaaagcacggcgcttggtcgtcaatgcacggacgctcgatgac
 701  ggcgacatcgaccatattcgcctgcttctggccatcaccgacgtgaccga
 751  tgtgcgcgccgaagctcggctgaaagatgatctcgttcgcgacaacgcga
 801  tcctgctgcaggaagtccagcacagggtcgcaaacagccttcagatcatt
 851  gccagcgttctcatgcagagcgcccggcgggtccagtcggaggaagcgcg
 901  cgggcatctccacaacgcccatcaccgggtcatgtcgattgcggccctgc
 951  aacgccaactctccatgtccaacggcggcaaggttgaactccgtacctat
1001  ttcactcagctttgccaaagcctcggcgcgtcgatgattgctgaccccga
1051  aaggctttcgattcaggtgatggtcgacgacagcgccgtggaggcggatg
1101  tttccgtacgcttggggcttgtcgtgaccgagcttgccatcaacgccctc
1151  aagcacgcctttcccgttgaacgaccaggcgcaatcgtcatcgcctatcg
1201  atcatccggcaaggattggactctctccgttaccgacaacggcatcggca
1251  tgcctgcaggtcgcgatgcaccaaaggcgggactggggaccggcattgtg
1301  gaggcgctggcaaagagcctgcatggtgacattcagttaagcgatgcggg
1351  tcccggcaccgccgttacaatccgccaccgggaaagcgccggtttgcgaa
1401  ccgacctttctacggccgcttcggaaccgaatggcttgcgaaacgctatc
1451  cctgcggagcccacgcggctcgcacctccaatctgaccgagtgagcaatg
1501  aattatggcaaagcggtcgtcctggtcgtcgaagacagcacgattattcg
1551  gatgagtgccgtcgatctggtattctctgcgggttacgaagcgcttgagg
1601  cgtgcgatgcggacgaggcaattcgtatccttgtgtcgcgaaacgacatc
1651  gatctggttttcaccgatgttcagatgcccggcacgatggacggcatcaa
1701  actatcccattacatccgaaatcggtggccgcctgtcaggctgatcgttg
1751  cctccggcgcggcaatcctcgaagagagcgatctccccatgggaagcaga
1801  gccttctcgaaaccctacgacaaccacgcaattaccgatgcaatggctca
1851  tctgctgtcgatcgggaacacgcctgagctggtaagttaagtgtttttcg
1901  acacccctcgggtcgagccgcgggtgagacgatccgccatccgctgcgcc
1951  gacgaggagtttggcggaaagcaccgccccaccctcgctcatcccaggcc
2001  tcagcccgggatgacggagggtgggtaaaaatatgcagatgaagcgaacc
2051  atgtttgtatccagccgacttgatgcgaggctgatccgggcgtagttcag
2101  gacaatagccacccctctccttcgacaccgatctgaaacgaacccatggcg
2151  cactccctcaatacagccccaacgcccaccggcggccccaaccagccgaa
2201  actcgatacggacgacatctggcaggcccgcgtcgatcttgccgcctgct
2251  tccggatggcggcaaggctcggcatggaggaagggatctgcaaccacttc
2301  tcggccgtcgttcctggatatgacgacctgttcatcgtcaacccttatgg
2351  ctacgctttcgccgaactgacggcctcgatgctgttgatctgcgattttc
2401  acggcaacgtggtgtcgggcagcgggcagcccgaggccaccgccttttac
2451  atccacgccaggattcacaagaacattccccgcgccaaggcggcattcca
2501  cacccatatgccttatgccaccgccctttcgatgacggagggcgatcccc
2551  tgatctttgccgggcagaccgcattgaaattctatggacgcacggccgtc
2601  gaccgcaactataacggcctggcgctcgacgcgcgcgagggcgatcggat
2651  cgcggccgcgatcggcgatgccgacatcgttttcatgaagcaccacggcg
```

```
2701  tgatggtctgcgcacctaacatcgccgaggcctgggacgatctctattac
2751  ctcgagcgcgcctgcgaagtgcagacgctcgccctgtcgaccggacgcga
2801  agtccttgccgtcgcgcctgagattgcggacgcggcgtaccggcaaatgc
2851  gcgaaggcgacccggagtcggcccggctccatctggagtccgtcaaacgc
2901  gcgcttgatcgcagcgagcccgaatacaagcgctgagccgcagctcgcgg
2951  tctagggcgtgaactcataaagacgcagccatagtcggatggaggcgacg
3001  ggggttcgctcgccaatggttgcatcttctgataagatcagcctcacata
3051  tctggaggcctcgcatgccctgcagcagcgcgctcgctccgttcgtcatc
3101  gccgctctcgtggccctcggtccactgccggctggagcgcagaacaatgc
3151  tgcgccaccgccggcggcgatctcccacaaggccaagctttgccagatgt
3201  gtcacggtaggaaaggtctgccgactgttaagaacacgccgatcatcgcc
3251  ggccagcacgagtcctctttgctgataggtttgcaggaatatcgcaatgg
3301  agcgcgcaccgacgagctcatgggccggatcgcgaagaatctttccgacg
3351  acgacatgaaggcgcttgcggcctacttctccgcactgccttggcccgcc
3401  taccgcgagccggccgacgcagcgagcattacccgcttccaagcgctcga
3451  cgtcgagaagaagtgcacctcgtgccaccgggagggtttcgtcggatacg
3501  ccaacactccgcgcgtcgcaaatcagaagctggactacctgatcaaaacc
3551  ctctccgactttcacgataacaagtgcccgaacatgccccgcatgacggc
3601  cttggtgcgcaatttgtcggccgacgacatcgctgccatggctcactatc
3651  ttgccggcctctgacgatcagttcgggggtcggcgatcgaccggataggg
3701  ccgccgcacggcgctctaggcgtgtgccggccgccaatgcggataggtga
3751  cataggcggtgagcgcgccctcctcgtgcgagcggatggtcaccgtttca
3801  cctttgggcatgtagacgatctcgccggggccggcggtgaccgtttcccc
3851  gtccgtggagaccgaaagccggccttccagaacgatcatgacgtcgtcaa
3901  ccgccattgtctcggtcaggctctgaccgggtgcgtagcgtccatagccg
3951  atggtgatcggccctccgtgccgctcatcgaccagattgccgacggaaat
4001  gtctgcctcctgtccgggggaacgctccaacgacgcatcggcgatcgtga
4051  acttacgcatcttcatcgtcttcctccattgcagtgtgcagaggtagcgc
4101  ggaaatttcacacgccaagacggtgcaaacgctcgaggaggaggttcccg
4151  gtcttatggtttcctcgccatagagaacggcgggtggagcatggcaacgc
4201  gccatatcgccggcggccgcgccaccgtcagatgcaccctctatatggct
4251  acgctctcggccatccgttgcaacccggccatcaagaccttccacaaaag
4301  gctgcgcgacgcgggcaaaccgcccaaggtcgccatcgccgccgccatgc
4351  gaaagctcatcatcatgatcaacactattcttagaagacgaaccccatgg
4401  aaccagccccaacaacacggttgctgaggtgcgcaggccaaggccggagc
4451  ctcgaaggacgagtgccggtggcgacacggttgcgtctgaagccgctccc
4501  cacccagcaccttaccaattcgtaagttgtcgccaagccgtcctctcgct
4551  agatctcgccatccacgcatgccgggaatgatcgggctgatgcgaatcct
4601  gcttatcgaggacgataccaagacgtccgattatatcgccaagggctttt
4651  ccgaggccgggcatgtctgtgatgtggtcggcgacggccgggacgggctg
4701  tttcaggcgcagcgcgaggcctatgacgtcatcgtcgtcgatcgcatgct
4751  gccgggtctcgatggactggcgatcgtgcgttcgcttagggctgccaagg
4801  tcggcacgtcggcgctgttcctgacgtcgatcggcggcgtcgacgatcgg
4851  gtcgaggggctggaggcgggcggcgacgattatctggtcaagccctttgc
4901  cttctccgagctgatggcgcgcgtcaatgcgcttggccggcggccgccgg
4951  tgcaggagcagcggacggtgctgaaggttgccgatctcgagctcgatctg
5001  atccggcgggaggcccgccgcgccggccaggtgatcgagctgcagccgcg
5051  cgaattcaccctgctcgaggtgctgatgcgcggcgaaggccgggtcatca
5101  ccaagacgatgctgctggagcgggtctgggatttccacttcgatccgaag
5151  accagcgtcgtcgagacccatatcagccggctgagagccaaggtcgacaa
5201  gccgttccagatccagcttctgcatacggtccgcaacaccggatacagcc
5251  tgcatgcgcctcgctagcctgcgcaggagcacgccgttccgccttgccgt
5301  caccttcggcgtgctcttcgtcgtcgccttcatcctgagcggcgcgatca
5351  tctatcacatgctgcagctcggccttgcgcgtgatctcgaacagtcgctc
5401  aacgagatgaattcgcttattatctcgacctacgagcccaacgacaccga
5451  agatctgatcaacacgctgaacaattatgcaagcttccaatcgacctccg
5501  acgggctttattcgctgacggatgcaggcgggcgcaaacttgccggcaat
5551  ttcgccgcaccccgcatcccgaacggcgtctataccgtcacctccaggga
5601  tgtcgggctgaaggggcatgaacgctaccggatgcaggtgtcgaccatcg
5651  gcccctacagcctggtggtggcggaaaacttcaacgatgtcgacgagatg
5701  ctgcggatcgtgctggtcagcttcgaatgggcggccgcgatcg
```

pRU862

```
   1  tcataccaatccgtaagaatgccgccccggccgcccgaccgccgcacttt
  51  acgaaaatgtatggttccagcgaggtcccggtgaggtttggagcgcagct
 101  tggcgccatcattgccaagtcgaggtcagctctctcatgaacatgcgcgc
 151  cgcacctgccgcccttatgccggcccttcatccggtcagcctgcgggtga
 201  acacctacctgatcaatctcgatcgtgcgcccttgcgcaggtttcgaatg
 251  gaacgcctgctggcaagcttcggccttgccttcgagcgcgtggcggcggt
 301  cgatggggcagggctgagcctgccgcatccgggcttcgacgacgcagcct
 351  atctcagccggcacggccgccggccgaacccttcgagatcggc
```

pRU863

```
   1  tcttccatgtcgatgaaaacggtccgttctcggaagcatccgggttcgtc
  51  agcacactcgccccgctcaggatcgctcccaaggcgacgttcaggaagaa
 101  ccgccatagcgccctggttggttccggtctcgacgtgtggctgaccgaga
 151  acggcgtccgccgtatcctcgtctccggcatccgcaccgaacagtgctgc
 201  gagaccacgacccgccatgcctcggatctcggctatcaagtcgactatgt
 251  cggcgaagcgaccctgacctttccaatgaccgacgtcacaggacgcacct
 301  ggagcgcgagggaaatccgggaccggaccgaactggtcctgtcgggccgt
 351  ttcgctcggatcgcgaccgtcgaacaggccttggccgggcgcgaaaaatc
 401  gctcgccgcatgaccgacgccgcgcagcccttcgatatcccggtcttcgt
 451  cgtcgtgccgccgcgcgtgctgctgctcgacgtggccggcccgatcgaag
 501  tgctgcgcaaggcgaacctcgaacagcacacggtgcgttttaccgtgacc
 551  tatatcggcccatcggcgacggtcggcagctcgatcggtcttgccgttac
 601  gggcgtcgccgcgttacccgagcgcttgcccgataacgcgcttgtcgtca
 651  ttgccggcagcgccgatgccccgatggataacaaccgcccgtgggacgaa
 701  caggaacgcgccgagcaggctgccatcgtcgcatggctgaagcgcgccat
 751  tcttccgggaattcgtctggtctcgatctgctcgggcgcattgctcgctg
 801  ccgaagcaggcatgctcgacggccgcgactgcaccacccaccacgcctgc
 851  atagaggatctggtgagactcgcacccaccgcgcgcgtccgggacaaccg
 901  gctctatgtcgaggatggagaccgcctcaccagcgccggcatcaccgccg
 951  gcatagatctcatgctgcatatcgttgccgaagcggcgggacatgcctgc
1001  gcgcttgcggttgcgcgatatcttgtcgtctatctcaggcgcggcggctc
1051  ggatccgcagctttcgccctggctcgaaggtcgcaaccatatccacccgg
1101  tcattcaccgcgcgcaggatgcggtggtcgccaacccctctcaggattgg
1151  tcggtcgcctcgctcgcgcgcctcagcggtgccagcccgcgcaacctttc
1201  acggctgttcaacgagcagacgggcatgagcgttacggatttcgtcaacc
1251  ttatgcgcgtggctcttgcccgcgagatgctcgccggttcacggctggac
1301  atggaggccgtcgcgatgcgcgccggcttcggctcggcgcggcagctccg
1351  ccgcgcctggaaccgccttaatgacggcccgccgagcgcggcacggtcaa
1401  ggctgcccgtcggttcatagactgctcgccgccggccgccagatcgggat
1451  cacggtcgcggacatattctctcgcaattttcacgttgaagcgcttgccg
1501  cgatggttgccgcatgcgctcggaaagagatccgtctgggaaatacatct
1551  gtggaagcgacggcgcatctggcttaagaaatgactgtgcgcccgttgct
1601  cccgcgatgagggatgcatggatggaattcgacgggctgcttacccatcc
1651  gcgactgaaagctagttgaaagcttcgacgtcataggaaggggcagcatc
1701  gtggaggagacacgggtcaaggtggcgagacgacgatgctagcaaaccac
1751  aggaggagatttccatgcgttcttcacgcagccttttcacaccgtcgct
1801  ttttcggcgcttctcgccgcagcatctttcgcgaccagcgccgcacatgc
1851  ggccgataagatc
```

pRU865

```
   1  tcatcttcctgacgtcgatggatatttcgggcaccgaaaaggaattcgcg
  51  gcgctgaacggccacgcgcatctgatgaagccggcgcgcgccaacgtgct
 101  gcgcaacaccgtcgtcgaagtggtgcgcgccagccgcgtcaagcaggctt
 151  ccgaggccgacattgcccggctgcagacggaagcggctgtgccggcgcca
 201  gcgcctatgccgcagaaacgggccgcggaattcgtcgacgtgctcgttgc
 251  cgaggacaacgaggtcaaccagatcgtcttcacccagatcctgcaaggca
 301  caggtctttccttcctcgtcgtcgacaatggcgaggaggcggtcgccgcc
 351  tgggagcggcatacgccgcgcatcatcatgatggacgtctcgatgcccgt
 401  catgaacggccatcaagccacccagacgatccgcgaacgcgaaaaggggc
 451  aggggcaccgggtgccgatcatcggcgtcaccgcccatgcgctcgaaagc
 501  gaccgcgaactctgcctcgatgccggcatggacgactacatgtcgaagcc
 551  gatcagtcccgaactgctcgaggaaaaaatccgccagtggctcggaacga
 601  gcgagcagcagccggagcgcaccagctactgacctccggataatttttat
```

```
 651  gaacggcgatggttggtgggtagctcttcagcgcctaatatttttacgcc
 701  gcagagcatttcgcgcttgccggcaaagcccttgcggcgctcgacggcga
 751  agcccgcggcgatgagattgcggcgcacgaagccggccgccgcataggtg
 801  gcgaaggtgccaccggctgcgcttttctcgcagacgagccgcatcagttg
 851  ctccgaccacatatcgctgttgcgcgaggggcgaagccgtcgagatacc
 901  aggcgtcgaagccgggcgtggctgccgccacgccctccaatgcatggccg
 951  catacgacgctgagcgtcgtctgggcatcgagatcgagtcgcacgatacc
1001  ggccggtatttccggccaagccgcggtcagcacctgacgttcggcatcga
1051  tctccggccagtgcgacagcgcccggccgatttcttcgccgcgcatcggg
1101  tggagttcgaaggaggtgaaatgcaggtgctggccgtctgcacggtggag
1151  cttccattgccgccaggtttcggtgaagttcaggccggtgccgaagccga
1201  gttcgccgatcaggaagtgcttccgcccgctccatcgttccggcaggccg
1251  ttgccggcgaggaaaacgtggccgcattccagccgcccgtcggtctggca
1301  ataaaaatgatcgccaaaggcggtggaatagggcatatcgccgtcgcgcc
1351  attcgagcggctgcggcgcgcctgcgccaatctgatcggggttcacgtct
1401  gtcatggaaaaagccgatagtcctgcgccggtctcaggtcaatcgtcctg
1451  cgaattgctgatcgtcggcggcggcatcatgggtctctgggtggccgtcc
1501  atgccgaacgccgtggtatccgtacctttgtcgcggacgccggcagccta
1551  ggcggaggtgcgagcggcggcctgctcggggcgctgatgccgcacatgcc
1601  ggatcg
```

pRU866

```
    1  tccagatgaccggcagcctgctgccctattattccgagcggcggcacagc
   51  agcgacgattcgacattgccggtcgaggtctttgccgccgtcacccgaag
  101  catggcctatatcgtgctaccgatgagcctcgggcttgccgcgatctcca
  151  gcgaattggtgattgtggtgttcggcgaagcgttccgccgcagcgggacg
  201  gtggtggcactgcttgcgctcgtcgctcccgcctataccttcatgcagat
  251  cctcagcctctacctgctgtcgatggacagggcccgctcccgcctcaaca
  301  tcagtgtgataggtggcctactgatggtagcgggttgtttactgatcgta
  351  cctaggcttgcggccgagggcgccgcaatcgtgcgcatcctcgtattcgt
  401  tgcgatgtcgatgatgatgatcagacagacaggattcggatcccagcttt
  451  cgggtctctacgcaagcctgacgaaggtgacgctcgcctccgtcctctgc
  501  gcttgcggagcgacttccgtgctggaattcgtccagggtccggccggatt
  551  ggtcggcgcgatcatcgccggcgcattcgcctatttcgcagcactccggg
  601  tgctgcgcgccgtgccgggcgaggatgtcgaagtcatgcgctccattctc
  651  gagaagatgccgtccctgctgcggcgaccggtcggccacgcgatcaattt
  701  cattgcgccgcggcttcccggcgatcccgatcgcgccaaggtcgcgcccg
  751  gcgaattctcgctcgaaccggccgagggcgcgggacgcagcgccgccctg
  801  cccgtcgtcttcgacggcacgattgggctgttcatgcctgaaaatcctct
  851  cgccaaaaaacgttcggccgccgtgctcttcgtcagcccctggggtttcg
  901  aggagatgtgcagccgcaaattcttccgcgtcgcggccgagcatttctcg
  951  gatatcggcgtgccgagcctgcgcttcgactatcgcggcactggcgacgc
 1001  gctcgatttcgacgcactgccggcgaggctggaaacctgggaagattcga
 1051  tccgtgcggccaccgacaagctgaagtcgctgagcggctgtgaccgcatc
 1101  atcctcatcgcacagggccttggtgcgacgcttgcccatcgcgtcggttc
 1151  ctcgatcgagggcgtcgacagcctcgtcatgctggcgccggtgctgagcg
 1201  gccgggcctatctgcgcgaactcaacatgtggtccaagatcatcgatgcc
 1251  gatctcggcctcggcaaagagcatatccagactgccaaggtgcagattgc
 1301  cgggctcgtcatgcccgaagagatcgccgccgagctcggcaagctcaaca
 1351  tcacctcgccgcaggggctcgcaacctcccgctacctgatcctcgaacgt
 1401  cctgccaaagccgaggataccggctttgccgatgcgctgaaggcgcttgg
 1451  cgccgatgtcgagcagaaggctttcgaaggctatgacgaactcgccacca
 1501  acccgctgtttgccaagacgccgatgactgtcgtggcgctgctgacggcg
 1551  tggctggagacaaggacgacggagacatccgccgcccattcgtcggcagc
 1601  gatcgacaacccgccgcttgccggcgacggttttgcggaaacgccggtcc
 1651  gtttcggaagccataatcatctggtcggcgtcgtcagccggccgctcggc
 1701  gagatcaa
```

pRU867

```
    1  tccgccatggcattgcgcaggggaagcaggagcgctgccgcaacaagagc
   51  ggcagcgaatattctgggcggttttctcaacgatgcattccttgtggtcg
  101  aaggcattggtgccttttcccgagctgtgccaactcgatgaaggattgcg
  151  gctcatgtggaggcaacccgatcctggctcgcggtgatgtcatcgcagga
```

```
 201  taagtaacggggtgaacgttaagtgcggtcaagttagtgttaagttaggt
 251  aaaacccggtcgcgcggctgagtttgtgctgtatggcttacccggatcaa
 301  tgacggacgaagtgaatgaacaaggttctcctcatcgacgacgatgccga
 351  gctgacgacgcttctgcaggaatatctggtcgaagaaggatatgacgtcg
 401  tcacggatacggacgggcgcgccgccattgcggcggcggccggcaatacg
 451  gtcgatattatcgtgctcgacatcatgatgcctcggatgaacgggatcga
 501  ggttctgcagaggatcaggaagctcagccaggtcccggtcctgatgctga
 551  ccgcaaggggcgatgacgtcgacaggatatcgggtctcaatctgggcgcc
 601  gacgactatgtgccgaagccatgctcgccgggcgaacttgcggcgagact
 651  gcgcgctatcctccgccgggcaggacagccggcggccggtacttcgaccg
 701  acacgataagggcgggaaaactcgtgattcatccgggcagcaggatcgcc
 751  gaatggcgcggcgaaagcttggaactgaccggcacggaattcagcctgat
 801  cgaggtcctcgcccgcagcgctggccagctcgtgtcgaagcaggacattt
 851  cgaagcgggccttcggcaagccgctcaccccattcgatcgccgtatcgac
 901  gtccatatcagcagcgttcgtcagaagctcggactgagggaggacggcca
 951  atcctggatccagtccgttcgcggccagggctatcaacttctcgtggact
1001  gaccatgccccggcttttctggaaattcttcacgacgatctggttgacga
1051  tggcggcgacagtcggcgtgatcatcctgctcgtcaattttctccaagag
1101  gttccttttgcgcgcgaactggaggaagagcggcgagcgatcgccctgaa
1151  cctcaccgcaaacgtgctcgccaaagatggcgaagatgccgccgcgcatt
1201  tcgtgcgcacaaatgaagagacgctaccggctggtctgacgatctccaaa
1251  actgcgaaagccgatgcttgtgcggttccaaagaccgtcgacacaagatc
1301  cgtccaggaggacggggtttgctatcagatttccctgccggtcctggcga
1351  cgttcaccttcgagaacttcggcccgttcttgccatggctcacgatcctg
1401  atttcgagcacgatatcggcaggcgcactcgcccgatacctcattcgtcc
1451  cgtcgttcatctgcgcgatggcttgagcgcgctcgcccatggccgcttcg
1501  acttccgcattggtgacaagatggccggccgaaaggacgaggttaccgcg
1551  cttgcacatgatttcgattccagcgccgcccggcttcaggagctgcagga
1601  tgcgcagcagcggttgttccatgatgtctcacacgaactgcgttcgcctt
1651  tgtcccgcctgcaggccgctgtcggcgtcctccggcagagcccggcgaaa
1701  ctcgccgccatgctggaccgcatggaccgagaggtcgaacggcttgatgc
1751  cctggtcggcgaagttctgacgcttgcaaggttgactgccggatccagcc
1801  ggccgctgaaaacgcacactcttgatgtcatcgaactcctgaacgaaatc
1851  ctcagcgacgcggcattcgaagcccaggcacgggaggtctcgatcacaac
1901  cagtgtcgagggcaccttccgcgccgaggtcgaaggtgagctggtctaca
1951  gggcgctcgaaaacgtcgtgcgcaacgccgtcaaatacacggccgagcat
2001  tcgcgcatatcagtagcttgcgaggcgacggccgaccgcctcgaaatctg
2051  cgtcacggatcaggggccaggtgtcaggcgggacgaactcgaacggatct
2101  tccagccgttctcacgtgggacagaggccgtgccaagaggcggatacggc
2151  cttggtctggcaatcaccaggcaggccatcgaacgccacggtgggcgtgt
2201  gtatgcgtcattgcctgatacaggtggcctggcaatcaccctggagcttc
2251  ccaggaagccgataccctatggcttggcggacgaacacgcctgaggaccg
2301  ctcgacgaagcttgaggcggtggtcggtcatcccgaagtaaaaacgacgg
2351  gcgcgagggaggcgggcgccgaagctttgcttttgaaacttcgcgccgga
2401  tgttccgcgtcaccggcatgggtgcgtttgcaatcatcattgacaacatc
2451  accatcggggcgcgacggtcgcgaagggcggcaaccgcgccatcgcacat
2501  tgcggtaggtcaaggcggccgccaagccggtccggtatcgtgcatgctct
2551  tctagctggagccttacacatgacgatccccgccaagtcgatgatgacaa
2601  cggatttagttacggtgtcgcctgaggccaccgtggccgaggccgcccgg
2651  tgcatgcttatccatcacgtcaccgccgtaccggtggtggacgccgataa
2701  ccggccgcttggattggtcagcgaaggcgacgtgatgcgccacttcggct
2751  cgcagtttcaaagtgagcgggcgcaatggctacgcatgctggcggaaggc
2801  gagacgctcgccccggagttcctcgccgaaatccgcctcaaccagcagca
2851  cgtccgcgagattatgcacacggccatcatctccgctggtgaggaagcct
2901  cgcttgcggagttggctgatctgatgctgaaacatgggatcaagcgcgtc
2951  cctatcctgcgtgacggcgtgttggttggcatcgtcagccgcgccgacgt
3001  ggttcgggccgtcgtcgaaaaactggacgatttgcttgagccgacggact
3051  agcaccggaagctggtcggccggaccccatcatccaccagcgcagcggcc
3101  ggtcggcgaccaacagcagcggcacgacaaaaatcagcggggcgtagggg
3151  tggcacgccttgatgcctctcatgttgagcagcatctgcattgtcgattt
3201  gaggcgatagttcgctggtggaaggcgggcagcgccctggagccgctcga
3251  cgatctgaccgacgaggagatcgcgactaagctgccggtgcatcttcgct
3301  atcttcccgacgcgctcgcggcatgaatgagaaaatggcgatgccttcct
3351  tccatccggactcgatgcagctcttcctcgttgaggacgaggccgcgcag
```

```
3401  ctctgctgaaccgagatgtgcgacgagacccggattttacccaactttac
3451  agtgtctttacagccattaacgttggcgccgctaagcctgaacctacgat
3501  cccggttcagaaagctcgccagtgagtccgtcgctctgcacgaaaacatc
3551  atcgctccgctacgccgcatcggcgatcacgcttctgcttgcgggctgcg
3601  tcagcggcccggatcacgctcccccacagatgccgcttcctgccaaattc
3651  caagagggcgggagcaagagcaacggcgatgtggtggcggcgcagtggtg
3701  gacggcctatcgggacaaacagctcgacggcctggtggctcatggcctca
3751  gcgagaaccttgatgtcctgcaggcgctcgagcgcatcaattcggcttcc
3801  gccaatgttacggttgccggtgccggtggcctgccgagcctcgatgtcgg
3851  cgcatcacacaccgtgtacggtgagaagggttcgcagcgtacgaccatcg
3901  gcaccaagaacacgacaggcggcgaggccagcctttcctggctgctcgac
3951  ttcttcggccagtaccgtcgctccaaggagagtgcaatcgcctcgcttgg
4001  ggccgcctatgcaacggccgacgatgccaagctcacctttctgaaggacc
4051  tcgtctccagttatgtcgacgcccgctattatcagcagcgcatcgcgctt
4101  tcgcaggcgaacttgaaatcccgtcaggagacctacgaactcacgcaact
4151  gcagctcaaggcaggtgcggcctctcgccttgacgtcgttcaggccgaag
4201  gccttgttcagtcgacgaaggccgacatccctggtctcgaacagagcttc
4251  accgtttcggctcaccacatcgccttgctgctcgggatgccggcggcctc
4301  gttgatgaatgagctgcagaggagtacaggccagccggtcttccgcggcg
4351  acattcgagccggcattccggccgacctcatccgcaaccgtcccgatatc
4401  cgcaaggcagaacgcgacctggcggcggccgtcgccgatatcggtgctgc
4451  cgaagcccagctttatccatcgatc
```

pRU868

```
   1  tccccgtagactcagccttgtcgagcgccatcagcgccggcacggcctcc
  51  tcgaggctgatgcggcgtccgatcagcttctgcggtgcgatctttccggc
 101  cgaaagcatcgacagcatggcgtcgtatcgccatgcctgcatgccgtggc
 151  tgccgtagatttccagctcttggccgatcacctgtgccatcgggatttgc
 201  ggcgtcgcatgctcgccgagcatcaaccccacctgcacatgccgtccgcg
 251  ccggcgcagattcttgatcgagttgaagcaggtgacgggatggccgagcg
 301  cgtcgatcgagacatgcgcgccgcccttggtgatctcgcgcaccgcctcc
 351  gccacatcggcgacgcctgacgcgttgaccgtcgccaccgccccgcactc
 401  ccgggcaaaggcgagcttctcctcggaaatgtcgatgccgatcgcattcg
 451  ccccaagcgccgtggcgatcatgatcgccgacaggccgacaccgccgcag
 501  ccatgcactgagatccattcgccgggcctggtgcgcgcctggtcggcgac
 551  ggcgcgaaacgaggtggcgaagcggcagccgaggcttgccgccgtcgcat
 601  cgtcgatcgtatcgggcaggtgcaccagattggtatcggcataatcgatc
 651  gccacatattcggcaaacgacccccaatgggtgaagcccggctggaactg
 701  gttcgggcagacctgctggttgccggaatggcactccccgcaatggccgc
 751  agccggaaacgaagggcacggtcacccggtcgcccaccttgaagcgcatc
 801  acgccgcggccggtggcgacgatccgcccggcaagctcatgtcccggcac
 851  atgcggcaggcggatatccgggtcgtggcccatccagccgtgccagtcac
 901  tgcggcagagcccgctggccccgacggatatgacgacaccgtcctctgtc
 951  ggagtcggatcggcaacggtgcggatttcgggcgcctgttcgaaggcttc
1001  gtagaacatggctttcatcggcgtcttcccttgctgcttatccgctgttg
1051  cgcccatcatcgcatgcggcagccagccattccaacattccattccatca
1101  tttttgctgatgcacccatcgatgaagctgccgttgctgcgcacgaatat
1151  tgcggcctggcgccattccccggcattttcttgcttccgtgtttccgctc
1201  gggatttgtccttgacccggcttgccgcggcggtttttccttctccggct
1251  tcaggaggaaaccatcatggccgtcgatacatcaccccgttcaaccacct
1301  ggactcatgtcgacggagagtggctccccggcaatccgccgctgatcggg
1351  cctacctcgcatgccatgtggctgggctcgacggtcttcgacggcgcccg
1401  ctggttcgacggcatcgcgccggatctcgacctgcactgccagcgcatca
1451  accgctcggcgctcgccatgggcctgaagccggtgaaatcagccgaggag
1501  attgtggcacttgcctgggaaggcgttgcgaaattcgatggcgccacggc
1551  gatttatatcaagccgatgtattggggcgagcacggttcgccgggcagcg
1601  tcgtcgccgtcgacggggaatcgacccgcttcgcgctctgcctgttcgag
1651  gcgccgatgggcggccatggcggcacaagcctcacagtctcgccctatcg
1701  ccgcccgtcgccggaaacggcgatgaccgaggcgaagaccggctcgctct
1751  atccaaacagcggccgcatgatcgccgaagcgcgcagccgcggtttcgac
1801  aatgcgctggtgcgcgacctgaacggcaacgtcgtcgagaccgcttcctc
1851  gaacgtcttcatggtcaaggacggcgtggtgatgacgcccgccgccaacc
1901  gcaccttcctcgccggcattacgcgttcccgcgtcatcggcctgctgcgc
```

```
1951  aaggccggtttcgacgtgcatgaagcagcgctctccgtcgaggatttcct
2001  gcaggccgacgagatcttcaccaccggcaactattccaaagtcgtcggcg
2051  tcacccgcctcgacgaccgcaatctccaggaaggcccggtcacccgcaag
2101  gcgctggaactttacatggactgggccttcggccgcagcgagagcgagga
2151  gtgaggagaggtcccgcgaagcgggagcaatcgatccagtgaatcgattg
2201  caacgacgaacgcccgaagcgcaagcgcagggctgcagcggttcggcgga
2251  tcagagcccagcccgcccctcatccgcctgccggcaccttctccccgctt
2301  gcggggcgaaggggatagccgcagcctctcagttcacgcaagggtcgtc
2351  aaaagtactcatctcgcctgtagaaattgccgtagaggtgcgtcagcacc
2401  gaattggctgccaatttttcttcatcgctgtcgccataaaccagtttggg
2451  cggatcgctttcgctaaagatgagagcgacaaaatgattgtaacccgcgc
2501  cggcgatccctgtcgaaacgtcgcgcccctcgatgatcgtaccgttctgt
2551  gcagagccactcgcggctgccgacgcttcgaaacgcgatttattgaaaat
2601  gacgtaccatagccaaggatcgaacaaaggcacggtgatcatcgtcaaac
2651  cggaacaatgatgatagcaataagacgactcaccattccacggatacgac
2701  gacgaaataacgcgactaatgcgccgataaaccaaacgaacagtatgccc
2751  gcgaccgagagaaagagcagcggccctcccaatagatccaagagtcctcg
2801  tagcccattggccgaatcaagcggccaaaaatccaaggcgcgatggtaag
2851  ctgcacacgcgatgccaagaaaagcggcgatgaccattcgccgaacattc
2901  caaatatctatgggcacagccgacacaccatcctgttgtttttcatagtc
2951  gaatggtgtccagcgaagtcgaggcatttttggggccgaacggcattgct
3001  gccgaattccgctgtgttgcagtgaagggtgacctccaatgtaaaattac
3051  ccctcacgtcatccacttccgatgctcagcaaatctctccgccaagagat
3101  cgatgaacagccgcaccttcgtcggcagatggctgcggttggggtagacg
3151  gcgttgatgttgaattccaccggccggtagtccggcatgatcttcaccag
3201  ccgcccttcggcgatgtcgtcgaacaccacgaaactcggcgccaggaaga
3251  tgccccgcccggtcagcgtcaggaaacgcagcatctcggcgctgttggag
3301  acgacgttgccgctgatcttgacgctctccttgttcccctcgccatcctc
3351  gaagcgccattcgtcgccataggggtaataggcatattgcaggcaattgt
3401  gatcggcgacctcggccggcgtcttcggcatcggatggctctggaaatag
3451  gccggcgagcagacgagcatatggcgccagggtgtcagcttgcgggcgac
3501  gagtgatgaatccggcggcggtacggtgcgcatcaccaggtcgtagccat
3551  cctcgatcatgtcgaccatccgttcgccgacgctgaaatcgagtgagatc
3601  gacggataaagctccataaattcgctgacgacgggcagcaggaaacgcac
3651  gatggcactgctggtatagaccttcagcgtgccccgcggcgtcgtgctca
3701  gcgcgccggccgtccggtccgcctcatcgagttcggccaggatctgcgac
3751  gagcgctcataataatatttgccggtctccgtcaggctgaccttgcgcgt
3801  cgtgcggttaagcagccgcacgccgagccggtcctccagcgattgcacgt
3851  gattgccgaccatggtgacggacatgttgagccggcgcgcggcggcggaa
3901  aaaccgccgcattccaccacgcggccaaacaccatgaggctggttagtcg
3951  atccatattgccctcggattatccgctgagagttgatgatccttcccgat
4001  ttaagcagattatcaaaatgaatgtcagggtgcattttccttcgcatcga
4051  acagggcctctcgcagaggaggccaacggtgaaggagaatgacgatggtc
4101  gagttgccccgcaaagacgttttcgaaagtgcgagggaggctgagcagat
4151  cctggccgaggaagccgccagggcacccgccgtcgaggcgcgccccatgc
4201  ctgtctccgaagcgccggttgcggacgcaccggtcgcggatgctcccaaa
4251  aagactggccgtcgcatcgtcaagcgcgccgtcatcgccgccgtcctgct
4301  tgccggcgtcgctttcgctggcgacttcggctaccgctactggacggtcg
4351  gccgcttcatcgaatccaccgacgatgcctatgtgaaagccgattacacc
4401  accgtcgccccgaaggtcgccggctatatcaaggcagtgctcgtcaacga
4451  caatgacgcggtcaaggccggccaggttctcgcccgtatcgacgaccgcg
4501  actttcaggccgcattgtcgcaggcgaaggccgatgtgaaggcggccgag
4551  gccgccatcaccaatatcgacgcccagatctctctgcagcaatcggtgat
4601  cgagcaggccaaggcgacggtcgatgcctcgcaggcctcgttcgattttg
4651  ccgtgtccgatgccgcccgctcggcccgcctgatc
```

pRU869

```
   1  tcccttctcgctcgaaacggtgcgggcaaccatgggctcggtctttgccg
  51  ttccgctcgcgcgggcgacgccagaggaattcatcgcctggcggaaatcg
 101  gccggcgtttccgtcgtcgcgacacatcttgccggggcggtcgattaccg
 151  gacgatcgactaccggaaaaagcccgtcgtgctgttgatgggcaacgaac
 201  aatccggcctgcccgagcagctggccagggaggccgacgcgcttgcccgt
 251  attccgcagcagggccgcgccgattcactgaacctcgccgttgccaccgc
```

```
 301  cgtcatgcttttcgaggcgcgccgccatctcctctcacttgccgagggca
 351  aatgaccgaacagacgcatgcacgcccggcgctgttttcgcgcccggcgc
 401  cgatcctcttcttcatcgtcgtcgccgttctgatcgaccaggccgtcaag
 451  atcgccgtcgatcattacctgccgctgcaggaggcggtgcccgtcgttcc
 501  gatgctggcgctctaccgcacctacaatctcggcgtcgccttttcgatgc
 551  tatcaggcatggatggctggttcatcgtcggcatgcggctcatcatcgtc
 601  gccttcgtcatctggctgtggtaccgcacggcgaaggatcgctggatcgc
 651  ccatctcggttatgcgctgatcatcgccggtgcgatcggcaatctcgtcg
 701  accgcttcgcctatgggcatgtgatcgattacattcttttctacaccgag
 751  agctggtccttcgcggtcttcaatctcgccgacagtttcatcaccatcgg
 801  ggccggctgcgtcattctcgacgagctgcttttgccgaaaaaggccggcc
 851  gctaaaatcttatcgaaatcctgaaggcattcggaacgggcctcatgtta
 901  gctagatagcatgcagaacctgggacagttgcaggaaagattgtccgccg
 951  ccttcggtggacccgccgccaagccgcttgagaagcggatggaggtcttt
1001  tccccacggcccaacgcagcggcaccccggggcgccgatgaccgccaggg
1051  cgggcctgcaccgacacggcgcctgctgttctactggctgggcggcgccg
1101  gccttctggccgcaacaatcctgatgctgctagcccatgccggcgatccg
1151  ctgctgctttctggcggtctcgtcgttcttgggcttgccgtcatcgccag
1201  ttatgccctgctgatggttcgctcgcgcagggccgggcagcgccccggtc
1251  aaaccatgccggatgggaacggcggcgcaaagctgttcgccgacgtgcac
1301  gacgtgctcggcgacatcacggtcagccgcaccatggaccggcgcatcat
1351  ctccgccaacgacacatttcgccgcctgaccgggcggctgcgcccggaag
1401  gaaaaacctgcgaggaaatcgggcttgctttccgccccggtccgataccg
1451  cattgctacgatgtcgagatctcgacaccggagggccagcgtatcttcct
1501  ctggcgcgatgtcgtcacccgcgacccggcgaacggccgtctgctgctgc
1551  agagcgttgcccgcgacgtgaccgacgagcggctgatcgcgcagggccgg
1601  gaggaggcccgccagaaggccgagtataacagcgccgccaaatcgaggct
1651  gcttgccaccgtcagccacgaggtgcgcacgccgctttccggcatccttg
1701  gcatgacacatctgatcgccgagacgcggctgacgcaggaacagcagaac
1751  tatctcgcgagcatccgccagtccggccatgcattgacgcagctcgtcga
1801  ggatctgctcgatttctccaccatcgaggtcggccgcttcgcgctgcacc
1851  cgcgctcggaatcgctgcgcaagcttctggaaagcgtcgtcgagatgctt
1901  gcccaccgggcgcatgaaaagggtatcgagatcggcgccaccgtgtcatc
1951  cgacgtgcccgaaaatatgagcttcgacccggcgcggctgcgccaggttc
2001  tgttcaacgttatcggcaacgccgtgaaattcacccaggtcggcggcgtc
2051  ttcatccgcgtgtcgctcgacggcgacgacctgtcgatc
```

pRU870

```
   1  tctctacggctggacgaaggacggctccttccccaaggaaatgtggggtg
  51  ctgccggcggaacccagtacaagaacatgggcgacgagttcgtcaacggc
 101  aacgtcgtgacctatctcgccggcaactggatggtcaatccgttccagaa
 151  aaagatcggcgacgccttcgactggacggcgatcagcgcgccgtgcggcg
 201  atgccggctgctatgcgatgcccggtggcaccgcgatcgtcggcttcaag
 251  cgcaccaagtctccggaggccgtggccgcgttcatcgagttcctcggctc
 301  tgaaaaggtccagcgtgaaatcgctgaaaactacgtcgtcctgacgggcg
 351  ctgacatcaaggacccgcaatacaagctcgagagcaagaatgccaaggat
 401  gccatggccgtcttcctcggaagccgcaatagcgtgccgcaggcagcccg
 451  ggacctggaacgtctcaagggatcctccgccatctatcagctcatcgtcc
 501  agaggatgagccagctgatcgtcggcgagctttccctcgacgagaccttc
 551  aaggcaatgaatgccgacgtcgacaaggtcaacgaggcgctcgccgccaa
 601  gtaaccggtcacgcgttgtcccgcatcatgcgcgcatgcagcagccacgg
 651  cctcgcaacagggccgtggcttttccttccccatggacacttcgaccag
 701  tttcgaattgtcgatggcaacgccggagatctaccgatgcggatgcgcgg
 751  tctatcgcgggtcatctgtggcgaagcggtcatggagccacgcgcggccg
 801  agcgaatgccggccttcgcaccgatggcgaaggcatggtgacatgcggtt
 851  cgggagaggctataggcgttgtcggccttcgccggctcgcaggcgccatg
 901  gtcaacttacaaatgccggcaacccgattacatcaggcgacccagcgact
 951  gtctaccgcatagaggtgcatttccatgccgatgggctccgagcatccgc
1001  tgcgaagggaacttcacaacgagctccacgcgcggccgtcgctttacttc
1051  gatggcgacacggatgtgtggcacgtcgctatcgtcggagaaaatgcccc
1101  tcctcagataccaaattccctgccgggactggaggatgtctccacgaccc
1151  gcgaaggcaaccacggcatcggccgcatcggcgacggccggctgaagtgg
1201  gaggcccataccgagttcctcaccctgaccttcgtcgttccggcatcagc
```

```
1251  cgaccccggcagcaatccgcctgaagcctttcaggcctgctgccgccaga
1301  tcgacgggaaggtcatcgcagccgtccgcgtgctggtgcgcgacgagaag
1351  gacgggcagcgtccagaaaaaccgaagttcgactatgtcgcctctcaggt
1401  cggcggcggcgacgcggaagtgcattcgaacttccggctgaccgacagcg
1451  gctttgtggaattcctgtttttcaaccgcaacctcaatgcctatcgcacc
1501  ggccgcatggtcaggcgttttctggagatcgagacataccggatgatggc
1551  gctgctggccttgccgatggcgcgcgagacggtgtcgaagctttccgtct
1601  tcgaccggcgcctcgacctgctgatcgcgcatatgcagagtgcggtcaag
1651  gtcgacaaggccctgctttccgaggtcaccaagctctcgtcggatgtgct
1701  caatttctccgcgctcgcccgccaccgcttcggcgcgacgaaagcctatg
1751  ccgagatcgtcgcaagcagatcgtcggagctgcgcgaggaacgcgtcgag
1801  cagcgccagcggatcggcaccttcatcgaccggcgctttcaaccggccgt
1851  ccgcgcggttcacgccgccgaacgccgcctcgatgaactggccgaacggg
1901  tgagcctggcgggagacctgctcagaaccaccgtgcaggttcagctcgaa
1951  gatcagaacgcgtcgctgctaacgtcgatggaagagcgggcgcgcatcca
2001  ggtgcatatccagcaggcggtcgaaggcttctccgtcatcgccatcacct
2051  attacaccgtcggcctcgcaaagatc
```

pRU871

```
   1  tcttcaccaccggcaactattccaaagtcgtcggcgtcacccgcctcgac
  51  gaccgcaatctccaggaaggcccggtcacccgcaaggcgctggaacttta
 101  catggactgggccttcggccgcagcgagagcgaggagtgaggagaggtcc
 151  cgcgaagcgggagcaatcgatccagtgaatcgattgcaacgacgaacgcc
 201  cgaagcgcaagcgcagggctgcagcggttcggcggatcagagcccagccc
 251  gcccctcatccgcctgccggcaccttctccccgcttgcggggcgaagggg
 301  gatagccgcagcctctcagttcacgcaagggtcgtcaaaagtactcatct
 351  cgcctgtagaaattgccgtagaggtgcgtcagcaccgaattggctgccaa
 401  tttttcttcatcgctgtcgccataaaccagtttgggcggatcgctttcgc
 451  taaagatgagagcgacaaaatgattgtaacccgcgccggcgatccctgtc
 501  gaaacgtcgcgcccctcgatgatcgtaccgttctgtgcagagccactcgc
 551  ggctgccgacgcttcgaaacgcgatttattgaaaatgacgtaccatagcc
 601  aaggatcgaacaaaggcacggtgatcatcgtcaaaccggaacaatgatga
 651  tagcaataagacgactcaccattccacggatacgacgacgaaataacgcg
 701  actaatgcgccgataaaccaaacgaacagtatgcccgcgaccgagagaaa
 751  gagcagcggccctcccaatagatccaagagtcctcgtagcccattggccg
 801  aatcaagcggccaaaaatccaaggcgcgatggtaagctgcacacgcgatg
 851  ccaagaaaagcggcgatgaccattcgccgaacattccaaatatctatggg
 901  cacagccgacacaccatcctgttgtttttcatagtcgaatggtgtccagc
 951  gaagtcgaggcatttttggggccgaacggcattgctgccgaattccgctg
1001  tgttgcagtgaagggtgacctccaatgtaaaattacccctcacgtcatcc
1051  acttccgatgctcagcaaatctctccgccaagagatcgatgaacagccgc
1101  accttcgtcggcagatggctgcggttggggtagacggcgttgatgttgaa
1151  ttccaccggccggtagtccggcatgatcttcaccagccgcccttcggcga
1201  tgtcgtcgaacaccacgaaactcggcgccaggaagatgccccgcccggtc
1251  agcgtcaggaaacgcagcatctcggcgctgttggagacgacgttgccgct
1301  gatcttgacgctctccttgttcccctcgccatcctcgaagcgccattcgt
1351  cgccataggggtaataggcatattgcaggcaattgtgatcggcgacctcg
1401  gccggcgtcttcggcatcggatggctctggaaataggccggcgagcagac
1451  gagcatatggcgccagggtgtcagcttgcgggcgacgagtgatgaatccg
1501  gcggcggtacggtgcgcatcaccaggtcgtagccatcctcgatcatgtcg
1551  accatccgttcgccgacgctgaaatcgagtgagatcgacggataaagctc
1601  cataaattcgctgacgacgggcagcaggaaacgcacgatggcactgctgg
1651  tatagaccttcagcgtgccccgcggcgtcgtgctcagcgcgccggccgtc
1701  cggtccgcctcatcgagttcggccaggatctgcgacgagcgctcataata
1751  atatttgccggtctccgtcaggctgaccttgcgcgtcgtgcggttaagca
1801  gccgcacgccgagccggtcctccagcgattgcacgtgattgccgaccatg
1851  gtgacggacatgttgagccggcgcgcggcggcggaaaaaccgccgcattc
1901  caccacgcggccaaacaccatgaggctggttagtcgatccatattgccct
1951  cggattatccgctgagagttgatgatccttcccgatttaagcagattatc
2001  aaaatgaatgtcagggtgcattttccttcgcatcgaacagggcctctcgc
2051  agaggaggccaacggtgaaggagaatgacgatggtcgagttgccccgcaa
2101  agacgttttcgaaagtgcgagggaggctgagcagatcctggccgaggaag
2151  ccgccagggcacccgccgtcgaggcgcgccccatgcctgtctccgaagcg
```

```
2201  ccggttgcggacgcaccggtcgcggatgctcccaaaaagactggccgtcg
2251  catcgtcaagcgcgccgtcatcgccgccgtcctgcttgccggcgtcgctt
2301  tcgctggcgacttcggctaccgctactggacggtcggccgcttcatcgaa
2351  tccaccgacgatgcctatgtgaaagccgattacaccaccgtcgccccgaa
2401  ggtcgccggctatatcaaggcagtgctcgtcaacgacaatgacgcggtca
2451  aggccggccaggttctcgcccgtatcgacgaccgcgactttcaggccgca
2501  ttgtcgcaggcgaaggccgatgtgaaggcggccgaggccgccatcaccaa
2551  tatcgacgcccagatctctctgcagcaatcggtgatcga
```

pRU872

```
   1  acggcaggaacgcccgcaaggcctcggtcgcggaagttgccgaacgcacc
  51  gtcaaggtgctgaaggcgaccgttccgcctgccgttcccggcatcgcctt
 101  cctctccggcggccagacgaccgaagaagcgacagcccacctttcggcga
 151  tcaatgccatcggccacctgccctggttcgtcaccttctcctacggtcgc
 201  gccctgcaggacagctcgctcaaggcctggaacggcaagcaggaaaatgt
 251  cgccgccggccagcgcgaattcacccaccgcgccgagatgaacagcctcg
 301  ccgccaagggcaactggaagaaggacctggaaaaggccgcctgagtatta
 351  acgcttaggacttgggcagggatggcaagcgcagcgtagccgcccctgcc
 401  atccctctctcctgcctctcattgtcacggcgacaagaaaccgcttcgtt
 451  ccgcgaaaaaccgtgattactcatagcgccctgtcacaacaggagctgcc
 501  atgaacaccctctcctacgtcaccgtcgatgtcttcacctccacccgctt
 551  cgagggcaatccgcttggcgtcatctccgatgcgcgcggcctgagcgatg
 601  cggcgatgcaaaagatcgcaaccgagttcaattattccgaagtcaccttc
 651  gtcctgcccccggaagaccctcaaaattccgcccgcgtgcgcatcttcac
 701  cccgacgatggaaataccctttgccggccatccgaatgtcggcaccgcct
 751  atgtgctcggccagcaggcggagattttcggcaagccggtcggctataag
 801  ctgcgtttcgaggaaaaggccggcatcgtcgaagtcagcctgaaacgcag
 851  cggcggaaaggtttctgctgccgccatccgcgcgccgcagccgctgacga
 901  tcggcgacaccatcgccgcgcaaaccatcgccggctgcgtctcgctcgac
 951  ccaggcgtcatcgtcgacaccagccatgccccggtctttgcctcggtcgg
1001  gctgaacttcgccatcgcagagctgaacgggctcgaagcgctggccgccg
1051  cccgcccgaaccttgccggattccaggcagcggccggccgccagacgacg
1101  agcggccacgacttctcgctcttcctctatgtgaggaccgccgaaagacc
1151  atggaatattcgcgcccgcatgttcgcgcccctcgacaatgtgcccgagg
1201  atccggcaacgggcagcgcttcagccgcgcttggcgcctatctcgtctcg
1251  cttgcgccggaggccgatatgaacgcccgcatcaccatcgaacagggcgt
1301  cgaaatgggccgccgcagcgtcatcacccttgatgtcgtgaaatccaacg
1351  gcatcgttaccgatgtcgtcatctcgggagactgcgtttccgtcatgcgc
1401  ggagaaattagcttgcaagactgacggtcgcagctcgaccgtacccctgc
1451  gttacacgtaatccggcgctacatcaggaaatcgcgcgaaagcagaacca
1501  ctgcccagacggcgaaagcgatcagcgcgagcttccagaaatccatccgc
1551  cgcctgaggccgggaaacccaagcggcttgatcacctgtatcgccatggc
1601  aaggatgagcagcaaggctatcagctttgtcatgctttatttttccgtt
1651  tttcggaatgtcacaggacggacatttttccgcgcgaacacaggatctcg
1701  acacgccgtaaagacattccggggcacaatcaatcatcttgaggctgggc
1751  tgaagcaggtcctgccttatctcattgcaaatcgctaagaccgagtctgg
1801  ggaaatgaagagaaatctgctgtccgtcgctgcgctgctctttggcacgc
1851  tcttccttttcatgggcaacggcctgcagggcatcctgcttcccgtgcgc
1901  ggcaatctcgaaggctacgcaacgacgacgctcggcctgctcggcacttc
1951  gtgggcgggggggcttcgtcatcggctgcctgattgcgccgaagattgtgc
2001  gccgcgtcggccacgtgcgtgccttttcggggttcatctcgatcatcgcc
2051  atcattgcgctggtcagcggcatcatcatcgatccggtctggtgggtggt
2101  cctgcgcgccgtcaccggcttctccaccgccggtacgtcgatgatc
```