EXPRESSION ANALYSIS OF THE TRANSPORTERS OF *SINORHIZOBIUM MELILOTI*

EXPRESSION ANALYSIS OF THE TRANSPORTERS OF *SINORHIZOBIUM MELILOTI*

By

Andrea L. Sartor

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requitements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE
(Biology)

MCMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Expression Analysis of The Transporters of *Sinorhizobium meliloti*

AUTHOR: Andrea Leigh Sartor, B. Sc.

SUPERVISOR: Professor T.M. Finan

Number of Pages: 153

ABSTRACT

Sinorhizobium meliloti is an alpha-proteobacterium that forms symbiotic nodules on the roots of *Medicago sativa* (alfalfa). The ability to catabolize specific compounds available in the soil is one of the best-characterized factors to increase competition for nodulation. In order to successfully attain symbiosis *S. meliloti* must compete for nutrients in the rhizosphere, which can be done by having a large number of transport systems encoded in its genome. Genes encoding proteins involved in transport constitute the largest (12%) class of genes in the *S. meliloti* genome. Great interest now lies in determining substrates for the transport systems and their role in the survival and fitness of *S. meliloti*.

An estimated 824 transport genes in the genome of the soil bacterium *Sinorhizobium meliloti* are predicted to encode 382 transport systems. All of the *S. meliloti* transporters had been studied under 120 different conditions, including growth on various carbon and nitrogen sources, seed and root exudates and starvation conditions.

From this screen of every transport system in *S. meliloti*, the substrates that induce expression of over 50 transport systems have been identified. We have found putative transporters for amino acids, sugars, sugar alcohols, amino sugars, betaines and other compounds that might be found in the soil. This large scale expression analysis gives insight into the natural environment of *S. meliloti* by studying those genes that are induced by compounds that would be found in the soil.

ACKNOWLEDGEMENTS

I would like to thank Dr. Finan for giving me the opportunity to work in his lab. I have learned a lot not only about science and research but also about myself.

I thank all the members of the Finan lab for being supportive, helpful, and easy and fun to live with for the last three and a half years. I especially want to thank those that helped guide me in my research and encouraged me to keep going forward. I have had so many great times with all of you and a lot of good laughs.

I also thank my family, friends, and boyfriend who have given me love and support throughout this experience. My parents have helped me so much through this journey and always supported and loved me. I appreciate all your love and affection and could never have made it remotely this far without you.

I want to specifically thank Jane Fowler for all her help, hard work, and support throughout this experience. Jane not only worked very hard helping with experiments and data analysis but was and still is a great friend. All of your advice for inside and outside of the lab was invaluable to me and I could not have asked for a better friend. Thank-you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Km	kanamycin
Cm	chloramphenicol
Sp	spectinomycin dihydrochloride
Tc	tetracyclin
Gm	gentamicin sulphate
Rm	rifampicin
Nm	neomycin
Amp	ampicillin
PNPG	p-nitrophenyl β-D-glucuronide
ONPG	p-nitrophenyl β-D-galactopyranoside
pNPβG	p-nitrophenyl β-D-glucopyranoside
DTT	DL-dithiolthreitol
SDS	sodium dodecyl sulphate
EDTA	(ethylenedinitrilo)-tetraacetic acid
bp	base pair
kb	kilobase
OD	optical density
AS	alfalfa seed exudates
AR	alfalfa root exudates
PS	pea seed exudates
PR	pea root exudates
SRS	scarlet runner seed exudates
SRR	scarlet runner root exudates
KWS	kentucky wonder seed exudates
KWR	kentucky wonder root exudates
RCS	red clover seed exudates
RCR	red clover root exudates
WCS	white clover seed exudates
WCR	white clover root exudates
LS	lima seed exudates
LR	lima root exudates
YLS	yard long seed exudates
YLR	yard long root exudates
MMS	mono-methyl succinate

CHAPTER 1. INTRODUCTION

Sinorhizobium meliloti

Sinorhizobium meliloti is a gram negative alpha proteobacterium that forms a symbiotic relationship with the legume alfalfa (*Medicago sativa*). In this relationship *S. meliloti* reduces dinitrogen (N_2) to ammonium (NH_4^+) for the plant and in return the bacteria are supplied with carbon sources, mainly in the form of C_4-dicarboxylates. The study of this organism is of great ecological and economical importance due to its role in agriculture with alfalfa being a major food crop. *S. meliloti* has a tripartite genome, comprised of a 3.65-Mb chromosome, and 1.35-Mb pSymA and 1.68-Mb pSymB megaplasmids.

The genome consists of 6204 predicted protein-coding regions and about 3000 of these encode Proteins of Unknown Function (PUFs). Of the known genes on pSymA, many are involved in nodulation and nitrogen fixation. Less known are the functions of the genes on pSymB. In an attempt to understand the role of this megaplasmid, most of pSymB has been deleted and there are only a few genes that have been identified as being essential; $tRNA^{Arg}$, *engA* (Charles and Finan, 1991, Paduska, B., unpublished data). There are an unusually large number of transport genes found on pSymA and pSymB and it has been suggested that these megaplasmids play a key role in the competitive nature of *S. meliloti* as a free-living organism (Barnett et al., 2001, Finan et al., 2001).

Within the rhizosphere a competition exists between strains of Rhizobium and a large number of factors influence the competition of these strains (Dowling and Broughton, 1986). In order to successfully attain symbiosis *S. meliloti* must maintain

their populations in the soil and establish themselves competitively in the rhizosphere. The ability to catabolize specific compounds available in the soil is one of the best-characterized factors to increase competition. Genes encoding proteins involved in transport constitute the largest (12%) class of genes in the *S. meliloti* genome (Galibert et al, 2001). With the anotation of the *S. meliloti* genome complete, interest now lies in gene function and the role that these unknown genes play in the survival and fitness of *S. meliloti*.

Transport Families

Transport DB is a database designed for describing and predicting cellular membrane transport proteins of organisms whose genome sequence is available (Ren et al., 2004). Ren et al. (2004) identified the complete set of *S. meliloti* membrane transport systems and classified them into different types and families according to putative membrane topology, protein family, bioenergetics and substrate specificities (see Table 1-1). According to this database and the analyses performed by Galibert et al. (Galibert et. al., 2001), more than half of the transport genes found in *S. meliloti* belong to ABC (ATP binding cassette) type transport systems.

Table 1-1: Transport DB classification of *S. meliloti* transport genes and number of unique clones used in this study.

	Number of Transporters	Number of Clones
ATP-Dependent		
The ATP-binding Cassette (ABC) Superfamily	200(201*)	182(187)
The H+- or Na+-translocating F-type, V-type and A-type ATPase (F-ATPase) Superfamily	2	2
The P-type ATPase (P-ATPase) Superfamily	9	8
Ion Channels		
The Major Intrinsic Protein (MIP) Family	3	2(3)

The CorA Metal Ion Transporter (MIT) Family	3	2
The Large Conductance Mechanosensitive Ion Channel (MscL) Family	1	0
The Small Conductance Mechanosensitive Ion Channel (MscS) Family	7	5
The Voltage-gated Ion Channel (VIC) Superfamily	1	1
Phosphotransferase System (PTS)		
General PTS	2	0
Sugar Specific PTS	2	0
Secondary Transporter		
The Auxin Efflux Carrier (AEC) Family	3	2
The Alanine or Glycine:Cation Symporter (AGCS) Family	2	2
The Ammonium Transporter (Amt) Family	1	1
The Amino Acid-Polyamine-Organocation (APC) Family	7	7
The Betaine/Carnitine/Choline Transporter (BCCT) Family	1	1
The Benzoate:H+ Symporter (BenE) Family	1	1
The Ca2+:Cation Antiporter (CaCA) Family	2	1
The Cation Diffusion Facilitator (CDF) Family	2	2
The Chromate Ion Transporter (CHR) Family	1	1
The Monovalent Cation:Proton Antiporter-2 (CPA2) Family	3	2
The Monovalent Cation (K+ or Na+):Proton Antiporter-3 (CPA3) Family	3	3
The Dicarboxylate/Amino Acid:Cation (Na+ or H+) Symporter (DAACS) Family	1	1
The Divalent Anion:Na+ Symporter (DASS) Family	2	1(2)
The Drug/Metabolite Transporter (DMT) Superfamily	11	9
The K+ Uptake Permease (KUP) Family	2	2
The Major Facilitator Superfamily (MFS)	44	26
The Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) Flippase Superfamily	4	2
The Nucleobase:Cation Symporter-1 (NCS1) Family	1	1
The Nucleobase:Cation Symporter-2 (NCS2) Family	5	4
The NhaA Na+:H+ Antiporter (NhaA) Family	1	1
The Metal Ion (Mn2+-iron) Transporter (Nramp) Family	1	1
The Cytochrome Oxidase Biogenesis (Oxa1) Family	1	1
The Inorganic Phosphate Transporter (PiT) Family	2	2
The Phosphate:Na+ Symporter (PNaS) Family	3	2
The Resistance to Homoserine/Threonine (RhtB) Family	11	7
The Resistance-Nodulation-Cell Division (RND) Superfamily	12	10
The Sulfate Permease (SulP) Family	2	1
The Twin Arginine Targeting (Tat) Family	1	1
The Tripartite ATP-independent Periplasmic Transporter (TRAP-T) Family	13	12
The K+ Transporter (Trk) Family	2	2
The Tricarboxylate Transporter (TTT) Family	4	4
Unclassified		
The MerTP Mercuric Ion (Hg2+) Permease (MerTP) Family	1	0
The Mg2+ Transporter-E (MgtE) Family	1	1
The Peptide Uptake Permease (PUP) Family	1	0

*There is one transporter that db annotates as 1 but is probably 2 separate transporters (Ren et al, 2004, www.membranetransport.org)

All ABC tranporter systems are composed of four protein domains; two transmembrane permeases and two ATPases. These domains can be expressed as four different proteins or in various combinations of protein fusions (Linton and Higgins, 1998). The genes for the components of these transport systems are often found in one operon and in many cases the metabolism genes associated with the imported substrate are also included in the operon (6+ another). In gram negative bacteria ABC uptake systems contain an additional protein, the periplasmic solute binding protein. In gram positive bacteria a membrane bound lipoprotein mediates solute uptake whereas eukaryotic cells have no homologues to these two proteins. The two transmembrane domain (TMDs) permeases contain membrane-spanning α-helices (typically six each) (Higgins, 1992). The two ATP or nucleotide –binding domains (NBDs) are hydrophilic and are found on the cytoplasmic side of the membrane. These domains contain the signature ABC domain by which this class of transporters is characterized. A typical ATP binding domain consists of approximately 215 amino acid residues containing Walker A and B motifs and also a C motif just upstream of the Walker B sequence and is responsible for the hydrolysis of ATP for the energy needed to move the solute across the membrane (Jones and George, 1999). A multitude of solutes are imported by ABC transporters including sugars, amino acids, peptides, opines, phosphate, sulphate and metals. ABC export systems transport various drugs, toxins, and antibiotics (Higgins, 1992).

The other major group of transport systems represented in the *S. meliloti* genome is the secondary transporters (40 % of transport systems). Transporters from this superfamily couple solute transport with cation transport to actively move nutrients into

or waste out of the cell (Leblanc et al., 1989). Within this class of transporters, the major facilitator superfamily transporters (MFS) are the most common (44 out of 150) (Ren et al., 2004). The major facilitator transport family includes the antiport, symport, and uniport tranporters. They can function by uniport, solute:solute antiport, and/or solute:cation symport, depending on the system and/or conditions. These systems are typically composed of one protein with 12 transmembrane spanning helices (TMSs). This family also includes the drug efflux systems that have 12 and 14 TMSs. These permeases catalyze drug:H^+ antiport (Leblanc et al., 1989).

Another common secondary transporter group, Trap-T, shares some characteristics with the ABC type transporters. Trap-T transporters (Tripartite ATP-independent periplasmic transporters) are characterized by a periplasmic binding protein, a small integral binding protein (with four putative TMSs) and a large integral membrane protein with 12 putative TMSs. The activity of these transporters is dependent on the proton motive force. In *Rhodobacter capsulatus* these systems have been shown to be specific for C4-dicarboxylates. Likewise, the TTT transport system (tripartite tricarboxylate transporter) is very similar to the Trap-T transport family except the TTT system transports tricarboxylates. However there seems to be no discernable sequence similarity between the proteins of these sytems and therefore these are categorized as parts of different families (Winnen et al., 2003).

The RND superfamily (Resistance-Nodulation-Cell Division Superfamily) of secondary transport systems are known to transport cations and export drugs, antibiotics and toxins. This system is characterized by proteins that have 12 TMSs, which are typically associated with at least two other proteins: the periplasmic membrane fusion

proteins and the outer membrane channels. These accessory proteins, together with a unique RND transporter, form complexes that span both the inner and the outer membranes (Tikhonova et al., 2002).

Finally, the APC (amino acid, polyamine, organocation) family of transporters consists of one protein that has 12, 13, or 14 (rarely) TMSs. This superfamily is the largest superfamily of amino acid transporters with ten well defined families that exist in both prokaryotes and eukaryotes (Jack et al., 2000). Transport DB has identified seven such transport systems in *S. meliloti* (Ren et al., 2004).

The Transport DB classification has also identified 15 ion channels, four phosphotransferase systems (PTS), and three unclassified transport systems (Ren et al., 2004) whereas, previous analysis of the genome sequence did not yield findings of any PTS system in *S. meliloti* (Galibert et al., 2001).

Transport in *S. meliloti*

There is an ongoing interest in the transport systems of *S. meliloti*. Over the years several *S. meliloti* transporters have been identified and characterized. The main focus, however, has been on the ABC type transporters. Transposon insertion mutants in the *frcBCA* genes failed to grow on media with fructose as a sole carbon source (Lambert et al., 2001). Transport assays further proved that the *frcBCA* was transporting fructose. Knockout analysis of the transporter showed that this was the sole transport system for this compound. This system was also shown to be involved in ribose and mannose transport. As mutants of this transporter could still grow on ribose and mannose as a sole carbon source, other systems can transport these compounds (Lambert et al., 2001).

Furthermore, an α-glucoside ABC transporter was discovered by introducing a cosmid library of *S. meliloti* DNA into a heterologous host, *Ralstonia eutropha*, unable to utilize sucrose and selecting for derivatives that could grow on sucrose. Tn5 insertion mutants with a disruption in the *agl* transporter were still capable of growth on α-glucosides. This strongly suggests that there is at least one more α-glucoside transport system (Willis and Walker, 1999).

One important set of systems in *S. meliloti* transport inorganic phosphate. There are two ABC-type transport systems with high affinity for inorganic phosphate (P_i); *phoCDET*, which transports phosphonates as well as P_i (Bardin et.al., 1996), and *pstSCAB*, which is P_i specific (Yuan et al., 2006). An additional transport system exists, OrfA-pit, which is a low affinity transport system (Voegele et al., 1997).

There has also been several transport systems found in *S. meliloti* that are not of the ABC type. One that is fairly well studied is the Dct transporter. This transporter was discovered by creating Tn5 mutants and selecting for those that could not grow on C4-dicarboxylates (Dct) such as succinate, fumarate, and malate (Watson et al., 1998). The transporter was further characterized through transposon mutagenesis and complementation analysis. This transport system was named *dctA* and is located on the pSymB megaplasmid. Mutants of this transporter have also been shown to have Fix^- phenotypes (*dctA*) or reduced nitrogen fixing activities (*dctB* and *dctD*) (Yarosh et al., 1989).

CHAPTER 2. MATERIALS AND METHODS

Bacterial Strains and Growth Conditions

S. meliloti was grown at 30°C and *E. coli* was grown at 37°C. Cultures were inoculated with single colonies that had been streak purified three times on selective media. Small scale cultures (2 mL) used for genetic experiments and for plasmid DNA preparations were grown in test tubes on a rotary mixer overnight. Larger cultures were grown in Erlenmeyer flasks in a rotary shaker. *S. meliltoi* and *E. coli* were routinely grown in Luria-Bertani broth (LB), which contains 10 g tryptone (Difco), 5 g yest extract (Difco), and 5 g NaCl per litre of double distilled water. For growth of *S. meliloti* LB broth was supplemented with $MgSO_4$ (2.5 mM) and $CaCl_2$ (2.5 mM). Solid media was prepared by the addition of 15 g agar (Difco) to 1L of LB before sterilization. Defined growth medium was M9 minimal media. This media contains 5 X M9 salts (Difco), which consists of Na_2HPO_4 (33.9 g/L), KH_2PO_4 (15 g/L), NH_4Cl (2.5 g/L) as a nitrogen source unless otherwise stated, and a carbon source. For *S. meliloti* the medium was supplemented with biotin (0.3 mg/ml), $CoCl_2$ (10 ng/ml), $MgSO_4$ (1.0 mM), and $CaCl_2$ (0.25 mM). For growth of *E. coli* the medium was supplemented with L-argenine (1 mM), thiamine (5 μM), and trace elements (1000X trace minerals contained per litre: 0.1 g H_3BO_3, 0.1 g $ZnSO_4 \cdot 7H_2O$, 0.05 g $CuSO_4 \cdot 5H_2O$, 0.05 g $MnCl_2 \cdot 4H_2O$, 0.1 g $Na_2MoO_4 \cdot 2H_2O$, 1 g Na_2EDTA, 0.2 g FeEDTA). Media was sterilized at 15 pounds/square inch at 121°C for 30 minutes. Temperature labile compounds were filter sterilized through a 0.45 or a 0.20 μm filter.

Antibiotics were obtained from Sigma or Boehringer Mannhiem and were stored at -20°C as stock solutions in ethanol (tetracycline, chloramphenicol) and the remaining

in water. They were filter sterilized and used at the following concentrations for the growth of *E. coli* on solid agar media (μg/ml): kanamycin sulphate (Km), 20; chloramphenicol (Cm), 10; spectinomycin dihydrochloride (Sp), 100; tetracycline (Tc), 10; gentamicin sulphate (Gm), 10. For *E. coli* growth in liquid media the indicated concentrations were halved. For growth of *S. meliloti* on solid agar medium the following antibiotic concentrations were used (μg/ml): streptomycin sulphate (Sm), 200; neomycin sulphate (Nm), 200; spectinomycin dihydrochloride (Sp), 200; tetracycline (Tc), 5; gentamicin sulphate (Gm), 60; ampicillin (Amp), 100. For *S. meliloti* in broth the antibiotic concentrations used were half those used in solid media. All plasmids, strains, and primers used in this study are listed in Tables 2-1 through 2-4.

Table 2-1. Plasmids used in this study

	Relevant Characteristics	**Reference**	**Primer Sets (5'-3')**	**Stock Number**
pRK600	pRK2013 *npt*::Tn9, Cm^R	(Finan et al., 1986)		
pTH1582	GusA from pFus1 (*pstI* sites) into pTH1581 (modified pJP2), Tc^R	Finan lab (Prell et al., 2002)		M462
pTH1522	library vector containing MCS with reporter proteins *gfp/lacZ* in one orientation and *gusA/rfp* in the other, Gm^R	(Cowie et al., 2006)		M411
pTH1703	MCS-gfp in pTH1591 (resembles pTH1522 but with MCS)	(Cowie et al., 2006)		M589
pTH1508	pTR102 with *attB* (used to create replicating plasmids in *S. meliloti*)	Finan Lab		M395
pTH1360	pV0155 with *gusA* cassette from pFus1	Finan lab		M216
pFL2765	Library plasmid; pTH1522 containing 1746 bp from smc04259 to smc04260.	(Cowie et al., 2006)		
pTH1937	pACYC 177 (MCS; oriT from RK2), Nm^R	B. Paduska, Finan Lab		M835
pTH2310	pTH1703 with 5'end and upstream promoter region of smc02619 via *ApaI/XhoI*; gfp/lacZ	This work	smc02619F smc02619R	M1220
pTH2311	pTH1703 with 3'end of smc02618 via *ApaI/XhoI*; gfp/lacZ	This work	smc02618F smc02618R	M1221
pTH2312	pTH1703 with 5'end of smc02615 via *ApaI/XhoI*; gfp/lacZ	This work	smc02615F smc02615R	M1222
pTH2313	pTH1937 with smc04259 to smc04251 using flp recombinase system	This work B. Paduska, Finan lab		M1223
pTH2327	pTH1703 with promoter region of smc04247 via *ApaI-BglII*	This work	4247F 4247R	

pTH2328	pTH1703 with promoter region of smc04248 via *ApaI-BglII*	This work	4248F 4248R	
pTH2329	pTH1703 with promoter region of smc04251via *ApaI-BglII*	This work	4251F 4251R	
pTH2330	pTH1703 with promoter region of smc04253 via *ApaI-BglII*	This work	4253F 4253R	
pTH2331	pTH1703 with promoter region of smc04259 via *ApaI-BglII*	This work	4259F 4259R	
pTH2332	pTH1703 with promoter region of smc04260 via *ApaI-BglII*	This work	4260F 4260R	
pTH2333	pTH1703 with promoter region of smc04258 via *ApaI-BglII*	This work	4258F 4258R	
pTH2334	pTH1703 with promoter region of *gnd* via *ApaI-BglII*	This work	gndF gndR	

Table 2-2. *E. coli* strains used in this study

Strain	**Relevant Characteristics**	**Reference**
DH5α	F^-, *endA1, hsdR17* (r_K^-, m_K^-), *supE44, thi-1, recA1, gyrA96, relA1,* Δ(*argF-lacZYA*), U169, Φ80d*lacZ*, ΔM15	B.R.L. Inc
MT616	MT607/pRK2013 *npt*::Tn9	(Finan et al., 1986)
M462	DH5α (pTH1582)	Finan Lab
M411	DH5α (pTH1522), Gm^R	(Cowie et al., 2006)
EcFL1580	DH5α (pFL1580), Gm^R	Finan lab
M216	DH5α(pTH1360), Km^R/Amp^R	Finan Lab
M842	pTH1944, carrying *flp* recombinase, Tc^R	B. Paduska, Finan Lab
M928	DH5α, Rf^R	Finan Lab
M592 (MT620)	MT620 expressing integrase from ΦC31, Rf^R	Finan Lab
M395	DH5α (pTH1508 (pTR102 with *attB*)), Tc^R	Finan Lab
M835	DH5α(pTH1937), Km^R	B. Paduska, Finan Lab
M1220	DH5α (pTH2310), Gm^R	This work

M1221	DH5α (pTH2311), Gm^R	This work
M1222	DH5α (pTH2312), Gm^R	This work
M1223	DH5α (pTH2313), Km^RGm^R	This work
M1237	DH5α (pTH2328), Gm^R	This work
M1238	DH5α (pTH2329), Gm^R	This work
M1239	DH5α (pTH2330), Gm^R	This work
M1240	DH5α (pTH2331), Gm^R	This work
M1241	DH5α (pTH2332), Gm^R	This work
M1242	DH5α (pTH2333), Gm^R	This work
M1243	DH5α (pTH2334), Gm^R	This work

Table 2-3. *S. meliloti* strains used in this study not involved in the screen

Strain	**Relevant Characteristics**	**Reference**
Rm1021	SU47, *str*-21	(Meade et al., 1982)
SmP110	Rm1021 with corrected *pstC*	(Cowie et al., 2006)
RmF250	Rm1021(*ntrA74*::Tn5-233), Sp^RGm^R	Finan lab
RmP1489	SmFL1790(*ntrA*$^-$), $Sp^RGm^RSm^R$	This work
RmP1490	SmFL1790(*ntrC*$^-$), $Sp^RGm^RSm^R$	This work
RmP1491	SmFL4232(*ntrA*$^-$), $Sp^RGm^RSm^R$	This work
RmP1492	SmFL4232(*ntrC*$^-$), $Sp^RGm^RSm^R$	This work
RmP1493	SmFL3396(*ntrA*$^-$), $Sp^RGm^RSm^R$	This work
RmP1494	SmFL3396(*ntrC*$^-$), $Sp^RGm^RSm^R$	This work
RmP1495	SmP110 (pTH2328+pTH1508 via *attP*/*attB* integrase system), $Sm^RGm^RTc^R$	This work
RmP1496	SmP110 (pTH2329+pTH1508 via *attP*/*attB* integrase system), $Sm^RGm^RTc^R$	This work
RmP1497	SmP110 (pTH2330+pTH1508 via *attP*/*attB* integrase system), $Sm^RGm^RTc^R$	This work
RmP1498	SmP110 (pTH2331+pTH1508 via *attP*/*attB* integrase system), $Sm^RGm^RTc^R$	This work
RmP1499	SmP110 (pTH2332+pTH1508 via *attP*/*attB* integrase system), $Sm^RGm^RTc^R$	This work
RmP1500	SmP110 (pTH2333+pTH1508 via *attP*/*attB* integrase system), $Sm^RGm^RTc^R$	This work

RmP1501	SmP110 (pTH2334+pTH1508 via *attP/attB* integrase system), $Sm^RGm^RTc^R$	This work
RmP1517	SmFL5992 (pTH2313), $Sm^RGm^RNm^R$	This work

Table 2-4. *Sinorhizobium meliloti* library strains used in this study from the random fusion library

SMa0036	SmFL686	Library clone 686 containing 859 bp from sma0034 to sma0036;gusA/rfp
SMa0081	SmFL417	Library clone 417 containing 1418 bp from SMa0078 to SMa0081;gusA/rfp
SMa0101	SmFL2364	Library clone 2364 containing 1082 bp from SMa0097 to SMa0101gusA/rfp
SMa0110	SmFL2812	Library clone 2812 containing 1140 bp from SMa0110 to SMa0112;lacZ/gfp
SMa0151	SmFL2485	Library clone 2485 containing 665 bp from SMa0151 to SMa0155;lacZ/gfp
SMa0185	SmFL7022	Library clone 7022 containing 514 bp of the upstream and 5' region of SMa0185; lacZ/gfp
SMa0198	SmFL1344	Library clone 1344 containing 2182 bp from SMa0198 to SMa0203;lacZ/gfp
SMa0217	SmFL3389	Library clone 3389 containing 1493 bp from SMa0217 to SMa0218;lacZ/gfp
SMa0224	SmFL1084	Library clone 1084 containing 1340 bp of SMa0224;lacZ/gfp
SMa0252	SmFL630	Library clone 630 containing 1494 bp from SMa0250 to SMa0252;gusA/rfp
SMa0270	SmFL2763	Library clone 2763 containing 1804 bp from SMa0265 to SMa0270lacZ/gfp
SMa0300	SmFL2623	Library clone 2623 containing 1590 bp from SMa0300 to SMa0301;lacZ/gfp
SMa0383	SmFL7062	Library clone 7062 containing 432 bp of the upstream and 5' region of sma0383, lacZ/gfp
SMa0396	SmFL3319	Library clone 3319 containing 1614 bp from SMa0396 to SMa0400;gusA/rfp
SMa0469	SmFL4058	Library clone 4058 containing 826 bp from SMa0467 to SMa0469;lacZ/gfp
SMa0495	SmFL7023	Library clone 7023 containing 465 bp of the upstream and 5' region of SMa0495, lacZ/gfp
SMa0501	SmFL5380	Library clone 5380 containing 1723 bp from SMa0498 to SMa0501;lacZ/gfp
SMa0527	SmFL7045	Library clone 7045 containing 305 bp of the upstream and 5' region of sma0527, lacZ/gfp
SMa0579	SmFL341	Library clone 341 containing 2430 bp from SMa0575 to SMa0579;gusA/rfp
SMa0583	SmFL4232	Library clone 4232 containing 1777 bp from SMa0583 to SMa0585;gusA/rfp
SMa0627	SmFL4612	Library clone 4612 containing 1132 bp from SMa0626 to SMa0627;lacZ/gfp
SMa0630	SmFL4547	Library clone 4547 containing 1994 bp from SMa0630 to SMa0633;lacZ/gfp
SMa0675	SmFL7001	Library clone 7001 containing 651 bp of 3' end of sma0675;lacZ/gfp
SMa0675	SmFL7063	Library clone 7063 containing 352 bp of sma0675;lacZ/gfp
SMa0677	SmFL1689	Library clone 1689 containing 1588 bp from SMa0677 to SMa0678;gusA/rfp
SMa0677	SmFL1689	Library clone 1689 containing 1588 bp from SMa0677 to SMa0678;gusA/rfp

SMa0682	SmFL4903	Library clone 4903 containg 1777 bp from SMa0682 to SMa0683;lacZ/gfp
SMa0683	SmFL54	Library clone 54 containing 1858 bp from SMa0683 to SMa0684;gusA/rfp
SMa0684	SmFL535	Library clone 535 containing 1523 bp from SMa0684 to SMa0689;lacZ/gfp
SMa0830	SmFL3318	Library clone 3318 containing 2030 bp from SMa0287 (nifD) to SMa0830 (nifE);lacZ/gfp
SMa0875	SmFL97	Library clone 97 containing 1229 bp of SMa0875 (nolG);lacZ/gfp
SMa0937	SmFL2040	Library clone 2040 containing 1757 bp of SMa0937;lacZ/gfp
SMa0951	SmFL5235	Library clone 5235 containing 2311 bp from SMa952 to SMa0955;lacZ/gfp
SMa1008	SmFL7024	Library clone 7024 containing 474 bp of the upstream and 5' region of SMa1008, lacZ/gfp
SMa1153	SmFL6159	Library clone 6159 containing 1094 bp from SMa1149 to SMa1153;gusA/rfp
SMa1155	SmFL7002	Library clone 7002 containing 335 bp of the 3' end of sma1155, lacZ/gfp
SMa1328	SmFL7025	Library clone 7025 containing 519 bp of the upstream and 5' region of SMa1328; lacZ/gfp
SMa1337	SmFL5184	Library clone 5184 containing 1414 bp from SMa1335 to SMa1337;lacZ/gfp
SMa1364	SmFL5353	Library clone 5353 containing 1492 bp of SMa1364;gusA/rfp
SMa1365	SmFL3528	Library clone 3528 containing 800 bp from SMa1365 to SMa1367;lacZ/gfp
SMa1365	SmFL3048	Library clone 3048 containing 1181 bp from SMa1365 to SMa1368;lacZ/gfp
SMa1371	SmFL431	Library clone 431 containing 1562 bp from SMa1371 to SMa1373;gusA/rfp
SMa1447	SmFL4563	Library clone 4563 containing 1119 bp from SMa1447 to SMa1450;gusA/rfp
SMa1466	SmFL501	Library clone 501 containing 1434 bp from SMa1466 to SMa1471;lacZ/gfp
SMa1538	SmFL781	Library clone 781 containing 1137 bp of SMa1538;lacZ/gfp
SMa1541	SmFL974	Library clone 974 containing 1115 bp from SMa1541 to SMa1544;gusA/rfp
SMa1600	SmFL4016	Library clone 4016 containing 933 bp of SMa1600;gusA/rfp
SMa1641	SmFL1190	Library clone 1190 containing 1543 bp from SMa1641 to SMa1644;gusA/rfp
SMa1647	SmFL2743	Library clone 2743 containing 1887 bp from SMa1647 to SMa1651;gusA/rfp
SMa1662	SmFL1751	Library clone 1751 containing 1933 bp of SMa1662;lacZ/gfp
SMa1667	SmFL274	Library clone 274 containing 2234 bp from SMa1664 to SMa1667(arcD1);lacZ/gfp
SMa1668	SmFL3335	Library clone 3335 containing 2168 from SMa1667 to SMa1668(arcD2);gusA/rfp
SMa1675	SmFL4876	Library clone 4876 containing 1634 bp from SMa1675 to SMa1677;lacZ/gfp
SMa1691	SmFL1505	Library clone 1505 containing 1747 bp from SMa1688 to SMa1691;gusA/rfp
SMa1697	SmFL5509	Library clone 5509 containing 2356 bp from SMa1697 to SMa1792;gusA/rfp
SMa1742	SmFL3361	Library clone 3361 containing 1636 bp from SMa1742 to SMa1745;gusA/rfp

SMa1753	SmFL819	Library clone 819 containing 1462 bp from SMa1753 to SMa1754;gusA/rfp
SMa1757	SmFL5070	Library clone 5070 containing 1715 bp from SMa1755 to SMa1757;lacZ/gfp
SMa1789	SmFL2973	Library clone 2973 containing 835 bp of SMa1798(kup2);gusA/rfp
SMa1814	SmFL3514	Library clone 3514 containing 771 bp of SMa1814;gusA/rfp
SMa1862	SmFL2427	Library clone 2427 containing 1393 bp from SMa1860 to SMa1862;gusA/rfp
SMa1913	SmFL4961	Library clone 4961 containing 1494 bp from SMa1910 to SMa1913;lacZ/gfp
SMa1916	SmFL2853	Library clone 2853 containing 1481 bp of SMa1916;lacZ/gfp
SMa1937	SmFL2344	Library clone 2344 containing 1658 bp from SMa1935 to SMa1937;lacZ/gfp
SMa1959	SmFL807	Library clone 807 containing 1343 bp from SMa1959 to SMa1961;gusA/rfp
SMa1998	SmFL3574	Library clone 3574 containing 799 bp from SMa1998 to SMa2000;gusA/rfp
SMa2000	SmFL2524	Library clone 2524 containing 1483 bp from SMa2000 to SMa2004;gusA/rfp
SMa2075	SmFL4691	Library clone 4691 containing 2273 bp from SMa2075 to SMa2077;lacZ/gfp
SMa2085	SmFL1109	Library clone 1109 containing 1931 bp from SMa2085 to SMa2087;gusA/rfp
SMa2123	SmFL1501	Library clone 1501 containing 1363 bp from SMa2123 to SMa2125;lacZ/gfp
SMa2125	SmFL4493	Library clone 4493 containing 1444 bp from SMa2125 to SMa2127;gusA/rfp
SMa2199	SmFL4094	Library clone 4094 containing 918 bp from SMa2199 to SMa2201;gusA/rfp
SMa2205	SmFL732	Library clone 732 containing 1891 bp from SMa2205 to SMa2209;gusA/rfp
SMa2305	SmFL1860	Library clone 1860 containing 1669 bp from SMa2301 to SMa2305;lacZ/gfp
SMa2337	SmFL5272	Library clone 5272 containing 1464 bp of SMa2337;lacZ/gfp
SMa2367	SmFL1528	Library clone 1528 containing 2739 bp from SMa2363 to SMa2367;gusA/rfp
SMa2377	SmFL4088	Library clone 4088 containing 1168 bp from SMa2373 to SMa2377;lacZ/gfp
SMa2385	SmFL3836	Library clone 3836 containing 620 bp from SMa2385 to SMa2387;gusA/rfp
SMb20002	SmFL5958	Library clone 5958 containing 1516 bp from SMb21655 (lacZ/gfp1) to SMb20002 (lacK1);gusA/rfp
SMb20015	SmFL1845	Library clone 1845 containing 1264 bp from SMb20015 to SMb20016;gusA/rfp
SMb20018	RmP211	Rm1021 ΦpTH1664, gusA
SMb20025	SmFL4077	Library clone 4077 containing 918 bp from SMb20024 to SMb20025;gusA/rfp
SMb20027	SmFL3291	Library clone 3291 containing 1418 bp of SMb20027;gusA/rfp
SMb20027	RmP230	Rm1021 ΦpTH1683, gusA
SMb20030	RmP190	Rm1021 ΦpTH1643; gusA
SMb20035	SmFL567	Library clone 567 containing 1598 bp from SMb20034 to SMb20035;lacZ/gfp

SMb20036	RmP218	Rm1021 ΦpTH1671;gusA
SMb20069	SmFL262	Library clone 262 containing 1558 bp from SMb20067 to SMb20069;gusA/rfp
SMb20070	SmFL7046	Library clone 7046 containing 347 bp of the upstream and 5' region of smb20070;lacZ/gfp
SMb20071	SmFL1059	Library clone 1059 containing 1520 bp from SMb20071 to SMb20072;lacZ/gfp
SMb20108	SmFL3241	Library clone 3241 containing 1811 bp from SMb20107 to SMb20108;gusA/rfp
SMb20112	SmFL265	Library clone 265 containing 1011 bp from SMb20111 to SMb20112;lacZ/gfp
SMb20112	RmP212	Rm1021 ΦpTH1665;gusA
SMb20124	SmFL5164	Library clone 5164 containing 2909 bp from SMb20124 to SMb20126;gusA/rfp
SMb20124	RmP191	Rm1021 ΦpTH1644;gusA
SMb20128	SmFL2321	Library clone 2321 containing 1633 bp from SMb20128 to SMb20130;lacZ/gfp
SMb20134	SmFL458	Library clone 458 containing 832 bp from SMb20134 to SMb20135;lacZ/gfp
SMb20141	SmFL2228	Library clone 2228 containing 1597 bp from SMb20141 to SMb20142;lacZ/gfp
SMb20153	SmFL2589	Library clone 2589 containing 1772 bp from SMb20152 to SMb20153;lacZ/gfp
SMb20155	SmFL7064	Library clone 7064 containing 449 bp of SMb20155; lacZ/gfp
SMb20158	RmP224	Rm1021 ΦpTH1677;gusA
SMb20181	SmFL1540	Library clone 1540 containing 1117 bp from SMb20179 to SMb20181;lacZ/gfp
SMb20231	SmFL1120	Library clone 1120 containing 1564 bp from SMb20230 to SMb20231;gusA/rfp
SMb20235	RmP196	Rm1021 ΦpTH1649;gusA
SMb20263	RmP215	Rm1021 ΦpTH1668;gusA
SMb20263	SmFL7003	Library clone 7003 containing 472 bp of smb20263;lacZ/gfp
SMb20268	SmFL6491	Library clone 6491 containing 1564 bp from SMb20266 to SMb20268;gusA/rfp
SMb20272	SmFL154	Library clone 154 containing 1401 bp from SMb20272 to SMb20275;lacZ/gfp
SMb20272	SmFL4010	Library clone 4010 containing 849 bp of SMb20272;lacZ/gfp
SMb20282	SmFL5672	Library clone 5672 containing 1955 bp from SMb20282 to SMb20284;gusA/rfp
SMb20289	SmFL7027	Library clone 7027 containing 346 bp of the upstream and 5' region of SMb20289;lacZ/gfp
SMb20299	SmFL661	Library clone 661 containing 620 bp of the 3' end of smb20299;lacZ/gfp
SMb20300	SmFL5122	Library clone 515 containing 1553 bp from SMb20299 (nanA) to SMb20300 (cyaF7) ;gusA/rfp
SMb20315	SmFL1826	Library clone 1826 containing 1688 bp from SMb20314 to SMb20315;lacZ/gfp
SMb20318	RmP199	Rm1021 ΦpTH1652;gusA
SMb20320	RmP217	Rm1021 ΦpTH1670;gusA

SMb20321	SmFL4594	Library clone 4594 containing 901 bp from SMb20321 to SMb20322;gusA/rfp
SMb20328	SmFL2265	Library clone 2265 contining 2615 bp from SMb20325 (thuE) to SMb20328 (thuK);lacZ/gfp
SMb20333	SmFL2487	Library clone 2487 containing 1567 bp of SMb20333;lacZ/gfp
SMb20345	SmFL631	Library clone 631 containing 1314 bp of SMb20345;gusA/rfp
SMb20349	RmP220	Rm1021 ΦpTH1673;gusA
SMb20351	SmFL3286	Library clone 3286 containing 1493 bp of SMb20350 to SMb20351;gusA/rfp
SMb20354	RmP219	Rm1021 ΦpTH1672;gusA
SMb20361	SmFL2301	Library clone 2301 containing 1144 bp from SMb20361 to SMb20363;gusA/rfp
SMb20365	SmFL519	Library clone 519 containing 1923 bp from SMb20365 to SMb20366;lacZ/gfp
SMb20369	SmFL4857	Library clone 4857 containing 1800 bp from SMb20367 to SMb20369;gusA/rfp
SMb20380	SmFL2526	Library clone 2526 containing 1136 bp from SMb20380 to SMb20381;gusA/rfp
SMb20402	SmFL3133	Library clone 3133 containing 1445 bp from SMb20400 to SMb20402;gusA/rfp
SMb20416	SmFL7028	Library clone 7028 containing 923 bp of the upstream and 5' region of SMb20416;lacZ/gfp
SMb20428	SmFL2253	Library clone 2253 containing 1221 bp from SMb20426 to SMb20427;gusA/rfp
SMb20428	RmP226	Rm1021 ΦpTH1679;gusA
SMb20433	RmP225	Rm1021 ΦpTH1678;gusA
SMb20436	SmFL602	Library clone 602 containing 1364 bp from SMb20435 to SMb20436;lacZ/gfp
SMb20442	RmP227	Rm1021 ΦpTH1680;gusA
SMb20444	SmFL1336	Library clone 1336 containing 2307 bp from SMb20442 to SMb20444;lacZ/gfp
SMb20476	SmFL1557	Library clone 1557 containing 1330 bp of SMb20476;lacZ/gfp
SMb20476	RmP231	Rm1021 ΦpTH1684;gusA
SMb20484	SmFL4183	Library clone 4183 containing 1235 bp from SMb20483 to SMb20484;gusA/rfp
SMb20488	RmP200	Rm1021 ΦpTH1653;gusA
SMb20502	RmP229	Rm1021 ΦpTH1682;gusA
SMb20506	RmP201	Rm1021 ΦpTH1654;gusA
SMb20508	SmFL4550	Library clone 4550 containing 1278 bp from SMb20508 to SMb20509;lacZ/gfp
SMb20568	SmFL1131	Library clone 1131 containing 2619 bp from SMb20568 to SMb20571;lacZ/gfp
SMb20568	RmP213	Rm1021 ΦpTH1666;gusA
SMb20571	RmP214	Rm1021 ΦpTH1667;gusA
SMb20604	SmFL1790	Library clone 1790 containing 1933 bp from SMb20604 to SMb20605;gusA/rfp

SMb20625	SmFL2493	Library clone 2493 containing 1434 bp from SMb20625 to SMb20626;gusA/rfp
SMb20634	SmFL2393	Library clone 2393 containing 2055 bp from SMb20634 to SMb21706;lacZ/gfp
SMb20661	SmFL3905	Library clone 3905 containing 641 bp from SMb20661 to SMb20662;gusA/rfp
SMb20671	SmFL4523	Library clone 4523 containing 1427 bp from SMb20671 to SMb20672;gusA/rfp
SMb20697	RmP210	Rm1021 ΦpTH1663;gusA
SMb20701	SmFL7029	Library clone 7029 containing 389 bp of the upstream and 5' region of SMb20701;lacZ/gfp
SMb20705	SmFL7004	Library clone 7004 containing 316 bp of smb20705;lacZ/gfp
SMb20706	SmFL1712	Library clone 1712 containing SMb20705 and SMb20706;gusA/rfp
SMb20713	SmFL7047	Library clone 7047 containing 270 bp of the 3' end of smb20713;lacZ/gfp
SMb20716	SmFL1274	Library clone 1274 containing 1063 bp from SMb20715 to SMb20716;lacZ/gfp
SMb20718	SmFL4607	Library clone 4607 containing 1732 bp from SMb20717 to SMb20718;lacZ/gfp
SMb20720	RmP209	Rm1021 ΦpTH1662;gusA
SMb20724	SmFL3347	Library clone 3347 containing 1561 bp from SMb20723 to SMb20724;lacZ/gfp
SMb20771	SmFL3481	Library clone 3481 containing 775 bp from SMb20771 to SMb20772;lacZ/gfp
SMb20784	SmFL1872	Library clone 1872 containing 2519 bp from SMb20784 to SMb20787;gusA/rfp
SMb20784	RmP188	Rm1021 ΦpTH1641;gusA
SMb20813	SmFL7030	Library clone 7030 containing 516 bp of the upstream and 5' region of SMb20813;lacZ/gfp
SMb20833	SmFL4958	Library clone 4958 containing 1596 bp from SMb20833 (*rkpT1*) to SMb20834 (*rkpZ1*);gusA/rfp
SMb20854	RmP197	Rm1021 ΦpTH1650;gusA
SMb20856	SmFL4881	Library clone 4881 containing 983 bp from SMb20856 to SMb20859;lacZ/gfp
SMb20863	SmFL7005	Library clone 7005 containing 334 bp of the 3' end of smb20863;lacZ/gfp
SMb20895	SmFL7031	Library clone 7031 containing 445 bp of the upstream and 5' region of chvE;lacZ/gfp
SMb20902	SmFL7032	Library clone 7032 containing 425 bp of the upstream and 5' region of SMb20902;lacZ/gfp
SMb20904	RmP206	Rm1021 ΦpTH1659;gusA
SMb20929	SmFL315	Library clone 315 containing 1954 bp from SMb20929 to SMb20931;lacZ/gfp
SMb20931	SmFL315	Library clone 315 containing 1954 bp from SMb20929 to SMb20932;gusA/rfp
SMb20931	RmP223	Rm1021 ΦpTH1676;gusA
SMb20972	SmFL2779	Library clone 2779 containing 1496 bp from SMb20972 to SMb20973;lacZ/gfp
SMb20975	SmFL3843	Library clone 3843 containing 1285 bp from SMb20975 to SMb20976;lacZ/gfp
SMb20979	RmP198	Rm1021 ΦpTH1651;gusA

SMb20981	SmFL7048	Library clone 7048 containing 319 bp of the 5' end of smb20981;lacZ/gfp
SMb20999	SmFL7033	Library clone 7033 containing 350 bp of the upstream and 5' region of bacA;lacZ/gfp
SMb21016	RmP202	Rm1021 ΦpTH1655;gusA
SMb21019	SmFL2412	Library clone 2412 containing 1565 bp of SMb21019;lacZ/gfp
SMb21037	SmFL56	Library clone 56 containing 2079 bp from SMb21038 to SMb21040;gusA/rfp
SMb21050	SmFL7006	Library clone 7006 containing 341 bp of smb21050;lacZ/gfp
SMb21097	SmFL5733	Library clone 5733 containing 1729 bp from SMb21095 to SMb21097;gusA/rfp
SMb21097	RmP189	Rm1021 ΦpTH1642;gusA
SMb21103	SmFL2282	Library clone 2282 containing 776 bp from SMb21102 to SMb21103;lacZ/gfp
SMb21130	SmFL1226	Library clone 1226 containing 1051 bp from SMb21130 to SMb21131;lacZ/gfp
SMb21130	RmP232	Rm1021 ΦpTH1685;gusA
SMb21138	RmP193	Rm1021 ΦpTH1646;gusA
SMb21139	SmFL1693	Library clone 1693 containing 1070 bp from SMb21137 to SMb21139;lacZ/gfp
SMb21145	SmFL2094	Library clone 2094 containing 1736 bp from SMb21144 to SMb21145;lacZ/gfp
SMb21151	SmFL32	Library clone 32 containing 1066 bp from SMb21149 to SMb21151;lacZ/gfp
SMb21151	RmP221	Rm1021 ΦpTH1674;gusA
SMb21162	SmFL7034	Library clone 7034 containing 355 bp of the upstream and 5' region of smb21162;lacZ/gfp
SMb21169	SmFL318	Library clone 318 containing 1072 bp from SMb21168 to SMb21169;lacZ/gfp
SMb21191	SmFL112	Library clone 112 containing 2085 bp from SMb21190 to SMb21191 (*msbA2*);gusA/rfp
SMb21196	SmFL3952	Library clone 3952 containing 1033 bp from SMb21192 (cbbA2) to SMb21196 (oppA);lacZ/gfp
SMb21205	SmFL5338	Library clone 5338 containing 2718 bp from SMb21205 to SMb21207;lacZ/gfp
SMb21216	SmFL4061	Library clone 4061 containing 1270 bp from SMb21216 to SMb21217;gusA/rfp
SMb21216	RmP192	Rm1021 ΦpTH1645;gusA
SMb21251	SmFL4275	Library clone 4275 containing 1543 bp from SMb21250 to SMb21251;gusA/rfp
SMb21260	SmFL5050	Library clone 5050 containing 933 bp from SMb21259 to SMb21260;gusA/rfp
SMb21273	SmFL4241	Library clone 4241 containing 1146 bp from SMb21272 to SMb21273 (potD);lacZ/gfp
SMb21281	SmFL7035	Library clone 7035 containing 328 bp of the upstream and 5' region of smb21281;lacZ/gfp
SMb21316	SmFL7036	Library clone 7036 containing 340 bp of the upstream and 5' region of expD1;lacZ/gfp
SMb21342	RmP203	Rm1021 ΦpTH1656;gusA
SMb21343	SmFL663	Library clone 663 containing 1791 bp from SMb21343 to SMb21344;lacZ/gfp

SMb21351	SmFL1717	Library clone 1717 containing 2472 bp from SMb21351 (*dctM*) to SMb21353 (*dctP*);gusA/rfp
SMb21352	SmFL1582	Library clone 1582 containing 1665 bp from SMb21353 to SMb21354 (uxaC);lacZ/gfp
SMb21375	RmP195	Rm1021 ΦpTH1648;gusA
SMb21376	SmFL738	Library clone 738 containing 2115 bp from SMb21376 to SMb21378;gusA/rfp
SMb21424	RmP207	Rm1021 ΦpTH1660;gusA
SMb21424	SmFL7065	Library clone 7065 containing 257 bp of the 3' end of smb21424;lacZ/gfp
SMb21430	SmFL977	Library clone 977 containing 2636 bp from SMb21430 to SMb21433;lacZ/gfp
SMb21438	SmFL908	Library clone 908 containing 2045 bp from SMb21438 to SMb21441;gusA/rfp
SMb21458	SmFL3971	Library clone 3971 containing 836 bp from SMb21458 to SMb21459;gusA/rfp
SMb21458	RmP208	Rm1021 ΦpTH1661;gusA
SMb21465	SmFL973	Library clone 973 containing 1504 bp from SMb21465 (prsE) SMb21466 (prsD);gusA/rfp
SMb21486	SmFL3579	Library clone 3579 containing 731 bp of SMb21486;lacZ/gfp
SMb21489	SmFL3146	Library clone 3146 containing 1264 bp from SMb21498 (acrF);gusA/rfp
SMb21507	SmFL7061	Library clone 7061 containing 357 bp of the upstream and 5' region of smb21507;lacZ/gfp
SMb21512	SmFL105	Library clone 105 containing 1394 bp from SMb21512 to SMb21513 (wzx2);lacZ/gfp
SMb21528	SmFL627	Library clone 627 containing 2768 bp from SMb21525 to SMb21526 (tauA);gusA/rfp
SMb21536	SmFL7007	Library clone 7007 containing 513 bp of smb21536;lacZ/gfp
SMb21555	SmFL5693	Library clone 5693 containing 1644 bp from SMb21555 (kefB2);gusA/rfp
SMb21575	SmFL7060	Library clone 7060 containing 289 bp of the upstream and 5' region of smb21575;lacZ/gfp
SMb21578	SmFL3376	Library clone 3376 containing 1468 bp from SMb21577 to SMb21578 (atcU2);gusA/rfp
SMb21587	SmFL3916	Library clone 3916 containing 957 bp from SMb21586 (gshB2) to SMb21587 ;lacZ/gfp
SMb21587	RmP222	Rm1021 ΦpTH1675;gusA
SMb21592	SmFL4954	Library clone 4954 containing 2459 bp from SMb21592 to SMb21594;lacZ/gfp
SMb21592	RmP204	Rm1021 ΦpTH1657;gusA
SMb21602	RmP205	Rm1021 ΦpTH1658;gusA
SMb21603	SmFL1615	Library clone 1615 containing 2250 bp from SMb21603 to SMb21605;gusA/rfp
SMb21644	SmFL1811	Library clone 1811 containing 1779 bp from SMb21644 to SMb21645;gusA/rfp
SMb21707	RmP194	Rm1021 ΦpTH1647;gusA
SMc00028	SmFL4438	Library clone 4438 containing 1560 bp from SMc00027 to SMc00028;gusA/rfp/rfp
SMc00044	SmFL3814	Library clone 3814 containing 706 bp from SMc00107 to SMc00044;gusA/rfp

SMc00172	SmFL2136	Library clone 2136 containing 2482 bp from SMc00172 to SMc00169 (dme);gusA/rfp
SMc00174	SmFL1312	Library clone 1312 containing 2351 bp from SMc00174 to SMc00172;gusA/rfp
SMc00185	SmFL1326	Library clone 1326 containing 1247 bp from SMc00187 to SMc00184;gusA/rfp
SMc00196	SmFL2030	Library clone 2030 containing 834 bp from SMc00196 to SMc00181;gusA/rfp
SMc00233	SmFL135	Library clone 135 containing 968 bp from SMc00233 to SMc00234 (ppiD);gusA/rfp
SMc00243	SmFL534	Library clone 534 containing 1711 bp from SMc00243 to SMc00245;lacZ/gfp
SMc00265	SmFL7069	Library clone 7069 containing 492 bp of the 3' end of smc00265;lacZ/gfp
SMc00273	SmFL4553	Library clone 4553 containing 1357 bp from SMc00271 to SMc00273;lacZ/gfp
SMc00317	SmFL7008	Library clone 7008 containing 309 bp of the 5' end of smc00317;lacZ/gfp
SMc00350	SmFL7058	Library clone 7058 containing 487 bp of the upstream and 5' region of smc00350;lacZ/gfp
SMc00381	SmFL7049	Library clone 7049 containing 328 bp of the upstream and 5' region of smc00381;lacZ/gfp
SMc00422	SmFL7009	Library clone 7009 containing smc00422;lacZ/gfp
SMc00422	SmFL7066	Library clone 7066 containing 731 bp from smc00422 to smc00423;gusA/rfp
SMc00423	SmFL7037	Library clone 7037 containing containing 731 bp from smc00422 to smc00423;lacZ/gfp
SMc00428	SmFL186	Library clone 186 containing SMc00428 and SMc00429;lacZ/gfp
SMc00476	SmFL4182	Library clone 4182 containing 636 bp from SMc00477 to SMc00476;gusA/rfp
SMc00498	SmFL6471	Library clone 6471 containing 1585 bp from SMc00500 to SMc00498;gusA/rfp
SMc00537	SmFL536	Library clone 536 containing 1306 bp from SMc00537 to SMc00536;gusA/rfp
SMc00537	SmFL7010	Library clone 7010 containing 348 bp of the 3' end of smc00537;lacZ/gfp
SMc00550	SmFL7050	Library clone 7050 containing 498 bp of the upstream and 5' region of smc00550;lacZ/gfp
SMc00564	SmFL7011	Library clone 7011 containing 341 bp of the 3' end of smc00564;lacZ/gfp
SMc00590	SmFL2648	Library clone 2648 containing 1821 bp from SMc00590 to SMc00592;gusA/rfp
SMc00642	SmFL4001	Library clone 4001 containing 392 bp from SMc00643 (purA) to SMc00642;lacZ/gfp
SMc00717	SmFL5016	Library clone 5016 containing 1621 bp from SMc00719 (hpt) to SMc00717;lacZ/gfp
SMc00744	SmFL834	Library clone 834 containing 1676 bp from SMc04458 (secA) to SMc00744;gusA/rfp
SMc00773	SmFL7038	Library clone 7038 containing 357 bp of the 5' end of potI;lacZ/gfp
SMc00787	SmFL3438	Library clone 3438 containing 1020 bp from SMc00786 (dppA1) to SMc00787 (dppB1);gusA/rfp
SMc00808	SmFL3545	Library clone 3545 containing 907 bp from SMc00807 to SMc00808 (chrA);gusA/rfp
SMc00813	SmFL2	Library clone 2 containing 1269 bp of SMc00813;lacZ/gfp
SMc00827	SmFL4136	Library clone 4136 containing 531 bp of SMc00827 and SMc00828;gusA/rfp

SMc00868	SmFL5319	Library clone 5319 containing 1263 bp from SMc00871 (atpB) to SMc00868 (atpF);gusA/rfp
SMc00873	SmFL2325	Library clone 2325 containing 1682 bp from SMc00874 (corA2) to SMc00873 (kup1);gusA/rfp
SMc00874	SmFL824	Library clone 824 containing 2428 bp from SMc00876 to SMc00874 (corA2);gusA/rfp
SMc00898	SmFL7051	Library clone 7051 containing 442 bp of the upstream and 5' region of kefB1;lacZ/gfp
SMc00922	SmFL580	Library clone 580 containing 1790 bp from SMc00922 to SMc00931;gusA/rfp
SMc00937	SmFL2813	Library clone 2813 containing 1834 bp from SMc00936 (ilvA) to SMc00937 (trkH);gusA/rfp
SMc00954	RmP228	Rm1021 ΦpTH1681;gusA
SMc00963	SmFL2409	Library clone 2409 containing 2068 bp from SMc00963 to SMc00961;gusA/rfp
SMc00978	SmFL7057	Library clone 7057 containing 314 bp of smc00978;lacZ/gfp
SMc01136	SmFL4858	Library clone 4858 containing 1637 bp from SMc01136 to SMc01134 (ihfB);gusA/rfp
SMc01141	SmFL7012	Library clone 7012 containing 294 bp of the 3' end of smc01141;lacZ/gfp
SMc01211	SmFL7052	Library clone 7052 containing 348 bp of the upstream and 5' region of smc01211;lacZ/gfp
SMc01212	SmFL5381	Library clone 5381 containing 2388 bp from SMc01214 to SMc01212;gusA/rfp
SMc01217	SmFL2613	Library clone 2613 containing 1924 bp from SMc01218 (greA) to SMc01217;gusA/rfp
SMc01261	SmFL3149	Library clone 3149 containing 2643 bp from SMc01264 to SMc01261;gusA/rfp
SMc01361	SmFL2085	Library clone 2085 containing 2010 bp from SMc01361 to SMc01359 (aidB);lacZ/gfp
SMc01368	SmFL5379	Library clone 5379 containing 1859 bp from SMc01366 to SMc01368;lacZ/gfp
SMc01376	SmFL7053	Library clone 7053 containing 296 bp of the upstream and 5' region of smc01376;lacZ/gfp
SMc01457	SmFL2957	Library clone 2957 containing 1767 bp from SMc01457 to SMc01459;lacZ/gfp
SMc01496	SmFL2002	Library clone 2002 containing 2486 bp from SMc01496 (smoE) to SMc01819;gusA/rfp
SMc01512	SmFL4480	Library clone 4480 containing 1618 bp from SMc01747 to SMc01512 (hmuT);lacZ/gfp
SMc01525	SmFL2542	Library clone 2542 containing 2466 bp from SMc01525 (dppA2) to SMc01524;gusA/rfp
SMc01584	SmFL3595	Library clone 3595 containing 714 bp from SMc01583 to SMc01584;lacZ/gfp
SMc01597	SmFL1629	Library clone 1629 containing 1738 bp from SMc01597 to SMc01602;gusA/rfp
SMc01606	SmFL726	Library clone 726 containing 1819 bp from SMc01602 to SMc01606;gusA/rfp
SMc01624	SmFL333	Library clone 333 containing 1470 bp from SMc01624 to SMc01625;gusA/rfp
SMc01633	SmFL124	Library clone 124 containing 977 bp from SMc01632 to SMc01633;lacZ/gfp
SMc01654	SmFL1889	Library clone 1889 containing 2311 bp from SMc01652 to SMc01654;gusA/rfp
SMc01729	SmFL4154	Library clone 4154 containing 1006 bp from SMc01730 to SMc01729;gusA/rfp
SMc01823	SmFL3856	Library clone 3856 containing 1584 bp from SMc01821 (dht) to SMc01823;lacZ/gfp

SMc01829	SmFL7013	Library clone 7013 containing 381 bp of the 3' end of smc01829;lacZ/gfp
SMc01869	SmFL622	Library clone 622 containing 612 bp of SMc01870 (agpZ1) to SMc01869;gusA/rfp
SMc01870	SmFL622	Library clone 622 containing 411 bp of SMc01870 (agpZ1) to SMc01869;lacZ/gfp
SMc01965	SmFL297	Library clone 297 containing 1584 bp from SMc01965 to SMc01967 (speB2);gusA/rfp
SMc01970	SmFL4977	Library clone 4977 containing 1616 bp from SMc01970 to SMc01971;lacZ/gfp
SMc01980	SmFL130	Library clone 130 containing 961 bp from SMc01978 to SMc01980;lacZ/gfp
SMc02020	SmFL3834	Library clone 3834 containing 845 bp from SMc02020 to SMc02021;lacZ/gfp
SMc02027	SmFL1276	Library clone 1276 containing 2355 bp from SMc02027 to SMc02029;gusA/rfp
SMc02033	SmFL116	Library clone 116 containing 1055 bp of SMc02033;lacZ/gfp
SMc02057	SmFL2733	Library clone 2733 containing 1957 bp from SMc02057 (secD1) to SMc02056;gusA/rfp
SMc02065	SmFL1371	Library clone 1371 containing 2026 bp from SMc02068 to SMc02065 (tatC);gusA/rfp
SMc02066	SmFL2322	Library clone 2322 containing 1516 bp from SMc02066 (tatB) to SMc02064 (serS);;lacZ/gfp
SMc02067	SmFL4354	Library clone 4354 containing 1644 bp from SMc02069 to SMc02067;gusA/rfp
SMc02068	SmFL1771	Library clone 1771 containing 2408 bp from SMc02071 to SMc02068;lacZ/gfp
SMc02118	SmFL1644	Library clone 1644 containing 1917 bp from SMc02118 (aapJ) to SMc02117 (metC);lacZ/gfp
SMc02141	SmFL2307	Library clone 2307 containing 2034 bp from SMc02143 (pstA) to SMc02141 (phoU);gusA/rfp
SMc02161	SmFL1665	Library clone 1665 containing 2332 bp from SMc02162 (fadD) to SMc02161;gusA/rfp
SMc02169	SmFL481	Library clone 481 containing SMc02170 and SMc02169;lacZ/gfp
SMc02170	SmFL624	Library clone 624 containing 2134 bp from SMc02170 to SMc02167;gusA/rfp
SMc02224	SmFL1578	Library clone 1578 containing 826 bp from SMc02223 to SMc02224 (chaA);lacZ/gfp
SMc02250	SmFL7014	Library clone 7014 containing 344 bp of the upstream and 5' region of mscL;lacZ/gfp
SMc02260	SmFL5345	Library clone 5345 containing 1250 bp from SMc02260 to SMc02262;gusA/rfp
SMc02265	SmFL1516	Library clone 1516 containing 2158 bp from SMc02263 (ilvB1) to SMc02265 (secD2);lacZ/gfp
SMc02325	SmFL58	Library clone 58 containing 1466 bp from SMc02324 to SMc02325;gusA/rfp
SMc02343	SmFL7054	Library clone 7054 containing 318 bp of the 5' end of smc02343;lacZ/gfp
SMc02344	SmFL2637	Library clone 2637 containing 1333 bp from SMc02344 to SMc02345;lacZ/gfp
SMc02344	SmFL2829	Library clone 2829 containing 1888 bp from SMc02344 to SMc02347 (asfB);gusA/rfp
SMc02357	SmFL446	Library clone 446 containing 1732 bp from SMc02355 to SMc02357;lacZ/gfp
SMc02407	SmFL2564	Library clone 2564 containing 1563 bp of SMc04207;lacZ/gfp
SMc02418	SmFL3256	Library clone 3256 containing 1857 bp from SMc02417 to SMc02418;lacZ/gfp

SMc02419	SmFL167	Library clone 167 containing 1398 bp from SMc02418 to SMc02419;lacZ/gfp
SMc02424	SmFL2819	Library clone 2819 containing 2340 bp from SMc02424 to SMc02427;lacZ/gfp
SMc02437	SmFL7059	Library clone 7059 containing 336 bp of the 5' end of ptsP;lacZ/gfp
SMc02472	SmFL3208	Library clone 3208 containing 1268 bp from SMc02472 to SMc02471;gusA/rfp
SMc02484	SmFL7039	Library clone 7039 containing 414 bp of the upstream and 5' region of smc02484;lacZ/gfp
SMc02516	SmFL4050	Library clone 4050 containing 647 bp from SMc02517 to SMc02516;gusA/rfp
SMc02571	SmFL3495	Library clone 3495 containing 1369 bp from SMc02661 to SMc02571;lacZ/gfp
SMc02589	SmFL5160	Library clone 5160 containing 1783 bp from SMc02589 SMc02676 (16S);lacZ/gfp
SMc02603	SmFL1239	Library clone 1239 containing 1321 bp of SMc02603;gusA/rfp
SMc02616	SmFL5242	Library clone 5242 containing 2065 bp from SMc02616 to SMc02620;gusA/rfp
SMc02648	SmFL2355	Library clone 2355 containing 1103 bp from SMc02650 (arsH) to SMc02648;lacZ/gfp
SMc02648	SmFL7017	Library clone 7017 containing smc02648;lacZ/gfp
SMc02724	SmFL3225	Library clone 3225 containing 1430 bp from SMc02724 to SMc02725 (trpE);gusA/rfp
SMc02737	SmFL4177	Library clone 4177 containing 653 bp of SMc02737 (opuC);gusA/rfp
SMc02753	SmFL7040	Library clone 7040 containing 380 bp of the upstream and 5' region of smc02753;lacZ/gfp
SMc02773	SmFL2443	Library clone 2443 containing 1557 bp from SMc02774 to SMc02773 ;lacZ/gfp
SMc02793	RmP216	Rm1021 ΦpTH1669;gusA
SMc02814	SmFL7055	Library clone 7055 containing 407 bp of the upstream and 5' region of smc02814;lacZ/gfp
SMc02836	SmFL5872	Library clone 5872 containing 1638 bp from SMc02835 (glk) to SMc02836;gusA/rfp
SMc02855	SmFL556	Library clone 556 containing 1636 bp from SMc02855 to SMc02856;lacZ/gfp
SMc02861	SmFL4512	Library clone 4512 containing 993 bp from SMc02861 (pit) to SMc02862;gusA/rfp
SMc02867	SmFL7015	Library clone 7015 containing 333 bp of smc02867;lacZ/gfp
SMc02872	SmFL5374	Library clone 5374 containing 1629 bp from SMc02872 to SMc02873;gusA/rfp
SMc02888	SmFL7071	Library clone 7071 containing 374 bp of the 3' end of smc02888;lacZ/gfp
SMc02890	SmFL4616	Library clone 4616 containing 1515 bp from SMc02889 to SMc02890;gusA/rfp
SMc02892	SmFL7041	Library clone 7041 containing 433 bp of the upstream and 5' region of smc02892;lacZ/gfp
SMc02895	SmFL3779	Library clone 3779 containing 1108 bp from SMc02893 to SMc02895;gusA/rfp
SMc02907	SmFL2512	Library clone 2512 containing 1899 bp from SMc02905 (dnaX) to SMc02907;lacZ/gfp
SMc02910	SmFL1043	Library clone 1043 containing 1346 bp from SMc02910 to SMc02912 (nusA);gusA/rfp
SMc02981	SmFL7070	Library clone 7070 containing 385 bp of the upstream and 5' region of smc02981;lacZ/gfp

SMc03000	SmFL2481	Library clone 2481 containing 974 bp from SMc02325 to SMc02300;gusA/rfp
SMc03061	SmFL5246	Library clone 5246 containing 1586 bp from SMc03060 (aglR) to SMc03061 (aglE);lacZ/gfp
SMc03121	SmFL3349	Library clone 3349 containing 1090 bp from SMc03122 to SMc03121;gusA/rfp
SMc03127	SmFL983	Library clone 983 containing 1176 bp from SMc03127 to SMc03126;gusA/rfp
SMc03146	SmFL1155	Library clone 1155 containing 1939 bp from SMc03147 to SMc03146;gusA/rfp
SMc03157	SmFL5122	Library clone 5122 containing 1901 bp from SMc03157 to SMc03159;gusA/rfp
SMc03168	SmFL7018	Library clone 7018 containing 635 bp of the upstream and 5' region of smc03168;lacZ/gfp
SMc03179	SmFL1178	Library clone 1178 containing 1678 bp from SMc03179 (phaA1) to SMc03181 (phaD1);gusA/rfp
SMc03237	SmFL7042	Library clone 7042 containing 500 bp of the upstream and 5' region of smc03237;lacZ/gfp
SMc03269	SmFL5005	Library clone 5005 containing 1896 bp from SMc03268 to SMc03269;lacZ/gfp
SMc03277	SmFL7043	Library clone 7043 containing 402 bp of the upstream and 5' region of smc03277;lacZ/gfp
SMc03807	SmFL3396	Library clone 3396 containing 1656 bp from SMc03805 (tesB) to SMc03805 (amtB);lacZ/gfp
SMc03815	SmFL1733	Library clone 1733 containing 804 bp from SMc03815 to SMc03816;lacZ/gfp
SMc03824	SmFL7044	Library clone 7044 containing 495 bp of the upstream and 5' region of smc03824;lacZ/gfp
SMc03827	SmFL1090	Library clone 1090 containing 2188 bp from SMc03825 to SMc03827;lacZ/gfp
SMc03829	SmFL1186	Library clone 1186 containing 2022 bp from SMc03829 to SMc03831;lacZ/gfp
SMc03838	SmFL4340	Library clone 4340 containing 1666 bp from SMc03837 to SMc03838;gusA/rfp
SMc03865	SmFL144	Library clone 144 containing 1018 bp from SMc03864 to SMc03865;lacZ/gfp
SMc03869	SmFL2238	Library clone 2238 containing 1278 bp from SMc03868 to SMc03869;gusA/rfp
SMc03900	SmFL2020	Library clone 2020 containing 1920 bp from SMc03948 (TRm1b) to SMc03900 (ndvA);gusA/rfp
SMc03971	SmFL746	Library clone 746 containing 1480 bp from SMc03971 (mexF2) to SMc03972 (mexE2);gusA/rfp
SMc03991	SmFL1220	Library clone 1220 containing 2809 bp from SMc03989 to SMc03992;gusA/rfp
SMc04037	SmFL1593	Library clone 1593 containing 1812 bp from SMc04036 to SMc04037;lacZ/gfp
SMc04126	SmFL2004	Library clone 2004 containing 1205 bp from SMc04127 to SMc04126;gusA/rfp
SMc04128	SmFL5864	Library clone 5864 containing 1620 bp from SMc04129 to SMc04128;gusA/rfp
SMc04137	SmFL836	Library clone 836 containing 833 bp from SMc04137 to SMc04136;gusA/rfp
SMc04140	SmFL64	Library clone 64 containing 1910 bp from SMc04140 to SMc04138;lacZ/gfp
SMc04147	SmFL4572	Library clone 4572 containing 2114 bp from SMc04148 to SMc04147;lacZ/gfp
SMc04167	SmFL214	Library clone 214 containing 1500 bp from SMc04167 to SMc04169;gusA/rfp
SMc04167	SmFL7021	Library clone 7021 containing 419 bp of the 3' end of smc04167;lacZ/gfp

SMc04179	SmFL5514	Library clone 5514 containing 1237 bp from SMc04178 to SMc04179;gusA/rfp
SMc04179	SmFL7019	Library clone 7019 containing 297 bp of smc04179;lacZ/gfp
SMc04218	SmFL321	Library clone 321 containing 1498 bp from SMc04218 to SMc04219;lacZ/gfp
SMc04242	SmFL394	Library clone 394 containing 1478 bp from SMc04242 (zur) to SMc04244 (znuC);lacZ/gfp
SMc04250	SmFL5992	Library clone 5992 containing 1302 bp of the 3' end of smc04250;lacZ/gfp
SMc04259	SmFL1580	Library clone 1580 containing 1781 bp from SMc04259 to SMc04260;gusA/rfp
SMc04283	SmFL2089	Library clone 2089 containing 1323 bp from SMc04283 to SMc04289;lacZ/gfp
SMc04287	SmFL1077	Library clone 1077 containing 2452 bp from SMc04287 to SMc04291;lacZ/gfp
SMc04287	SmFL1077	Library clone 1077 containing 2452 bp from SMc04287 to SMc04291;lacZ/gfp
SMc04289	SmFL3818	Library clone 3818 containing 911 bp from SMc04289 to SMc04291;lacZ/gfp
SMc04317	SmFL7020	Library clone 7020 containing 607 bp of the 3' end of smc04317;lacZ/gfp
SMc04351	SmFL49	Library clone 49 containing 1451 bp from SMc04350 to SMc04351;lacZ/gfp
SMc04362	SmFL7067	Library clone 7067 containing 587 bp of the upstream and 5' region of smc04362;lacZ/gfp
SMc04393	SmFL39	Library clone 39 containing 1014 bp from SMc04393 to SMc04395;gusA/rfp
SMc04404	SmFL2556	Library clone 2556 containing 1551 bp from SMc04404 to SMc04405 (leuB);gusA/rfp
SMc04407	SmFL1286	Library clone 1286 containing 1591 bp from SMc04407 to SMc04408;gusA/rfp
SMc04439	SmFL3139	Library clone 3139 containing 1243 bp from SMc04311 to SMc04439;gusA/rfp
SMc04454	SmFL7056	Library clone 7056 containing 422 bp of the upstream and 5' region of smc04454;lacZ/gfp

Table 1-6. Primers used in this study. The first 70 primer sets were used to create fusions to the transporters not represented in the library. The gene to which the fusion was built is in the primer name.

5' Primer/Sense (5'-3')	3' Primer/Antisense (5'-3')	Restriction Sites	Notes
sma0489F CAGATCTTAATTAAGTTGTCGGTAATCGACATCGAGG	sma0489R ATCACCGGATCTCGACAGGCGAAATGTAGCTCAAGTGC	ApaI-XhoI	
smb20156F CAGATCTTAATTAAGCGGTCTGAATCGATCATACGTTCC	smb20156R TCACCGGATCTCGAGGTCGAGCACGATGGAAATGATCG	ApaI-XhoI	

smb20419F TCACCGGATCTCGAGGTCCTCACAATCTTTCCGGACG	smb20419R CAGATCTTAATTAAGGGCCCTTGTAGGTGCTCTGGCTCTCG	XhoI-ApaI	
smb20813F TCACCGGATCTCGAGGGTAATGGCTCCTGTATAAGTGC	smb20813R CAGATCTTAATTAAGGGCCCGCGAAAGCTTCGTTGATCACCG	XhoI-ApaI	
smb20894F CAGATCTTAATTAAGGAACAGCTTCTCCATGGTCTCG	smb20894R TCACCGGATCTCGATGCCATCGAGCATCGCATTCG	ApaI-XhoI	
smb20902F TCACCGGATCTCGAGGCTGCAGCTCTTCAAGGTGC	smb20902R CAGATCTTAATTAAGGGCCCGCCGATGACCAGATCCTTGG	XhoI-ApaI	
smb21316F TCACCGGATCTCGAGGGTTCATCCCTCTCTCCATTGG	smb21316R CAGATCTTAATTAAGGGCCCAAGCGTCACGACGAACACACG	XhoI-ApaI	
sma0185F TCACCGGATCTCGAGGATCTTGACCGACGTCGTAGC	sma0185R CAGATCTTAATTAAGGGCCCACAAGTCCATCGAGTAGATCAGGC	XhoI-ApaI	
sma1328F CAGATCTTAATTAAGATCCAGCCGAAGAGGAAAGCG	sma1328R ATCACCGGATCTCGAGGCTTTACTCCAGCGATACAGG	ApaI-XhoI	
sma1009F TCACCGGATCTCGAGACAGCTTCCTCGGTAGAGACC	sma1009R CAGATCTTAATTAAGGGCCCTTATAGCCAGCGTCTTCAATCGCG	XhoI-ApaI	
smb20698F CAGATCTTAATTAAGTTCAGCAGCCAGCAGGTTGC	smb20698R ATCACCGGATCTCGAGAACAACCGCTCCTTTGGACG	ApaI-XhoI	
smb20289F CAGATCTTAATTAAGGAAACTCTGCAGGATGGTGCC	smb20289R ATCACCGGATCTCGACGGATGATGGCATTGCGACG	ApaI-XhoI	
smb21507F TCACCGGATCTCGAGGAGCCACCAAGGTGCTTATTGCC	smb21507R CAGATCTTAATTAAGGGCCCAGCAAGGGTACCTGGACAATGC	XhoI-ApaI	
smc02250F TCACCGGATCTCGAGATGACCGAACCCGAAAGCATGC	smc02250R CAGATCTTAATTAAGGGCCCTGACGACCGAATCGACAATCTTGC	XhoI-ApaI	
smc00350F CAGATCTTAATTAAGGGCCCAATGACCAGCGCCATCGAGC	smc00350R TCACCGGATCTCGAGGCATGGCCATGGTCTCTTCG	ApaI-XhoI	
smc03167F CAGATCTTAATTAAGGGCCCGCAATGCGGTAATCTCC	smc03167R TCACCGGATCTCGAGGGACGTATCCACTCCTTTGATCG	ApaI-	

GTCG		XhoI	
smc02814F CAGATCTTAATTAAGGGCCCGTCTCGCTGATCACGTACAGC	smc02184R TCACCGGATCTCGAGCTGGCAAGTATAAGGCAAGTCTGC	ApaI-XhoI	
smc02892F CAGATCTTAATTAAGGGCCCACCGTGAACAGCAGGATTCCG	smc02894R TCACCGGATCTCGAGCAGTAGCACAGGTTTTATCCAGGC	ApaI-XhoI	
smc03237F CAGATCTTAATTAAGGGCCCGATGAGACCGAGCGAGAAAGG	smc03237R TCACCGGATCTCGAGACGTTGAACGCATAGAGCCAGG	ApaI-XhoI	
smc03277F TCACCGGATCTCGAGTTCTTGAGGTGCAGTCCGAATCG	smc03277R CAGATCTTAATTAAGGGCCCACCAGTGCAACCATCAGATCGG	XhoI-ApaI	
smc00423F TCACCGGATCTCGAGTCGATGTTCTGCGACAGGAGC	smc00423R CAGATCTTAATTAAGGGCCCTGCCAGGCGAGATAAAGCAGG	XhoI-ApaI	
smc00422F AAGGGCCCTCGATGTTCTGCGACAGGAGC	smc00422R AACTCGAGTGCCAGGCGAGATAAAGCAGG	ApaI-XhoI	
smc02484F CAGATCTTAATTAAGGGCCCATGCCGAAGGAGGTGATGACC	smc02484R TCACCGGATCTCGAGTGGTGCTGGATCTCTAGATCACG	ApaI-XhoI	
smc02981F TCACCGGATCTCGAGATGCGTCATGACAGGGTCTCC	smc02981R CAGATCTTAATTAAGGGCCCAACGCGAACCATTGGTGAGCG	XhoI-ApaI	
smc02753F CAGATCTTAATTAAGGGCCCATATCCTGACGCCGCTGATCC	smc02753R TCACCGGATCTCGAGCGCGTCTCAACTTCTAGGAATCG	ApaI-XhoI	
smc00898F CAGATCTTAATTAAGGGCCCTTCGAGAAGACCGCTGAACGC	smc00898R TCACCGGATCTCGAGCATCAAGGGCATCATCGTGATTCC	ApaI-XhoI	
smc00381F TCACCGGATCTCGAGACCTCACGAGCGAGATGAGC	smc00381R CAGATCTTAATTAAGGGCCCGGAATTCCGTTGTGACCAGATAGG	XhoI-ApaI	
smc04362F CAGATCTTAATTAAGGGCCCCGTGTTCTGTTCGAAGCTGTCG	smc04362R TCACCGGATCTCGAGGCTTATCTCGCCGTCGATAGG	ApaI-XhoI	
sma0526F	sma0526R	XhoI-	

TCACCGGATCTCGAGAGGATGCCTATCTTGGCTGGC	CAGATCTTAATTAAGGGCCCAATCCGGAGCGTCGAACACG	ApaI	
20713F TCACCGGATCTCGAGCAACCAGCAGAAGGTACTGATCG	smb20713 TCAGATCTTAATTAAGGGCCCCTCCATGACCTTGATTTGGGTGG	ApaI-XhoI	
00550F TCACCGGATCTCGAGACACGGTGCTCTTCAACGATACG	00550R CAGATCTTAATTAAGGGCCCACGCGCTTCAACTGTTCCTCG	XhoI-ApaI	
00771F TCACCGGATCTCGAGCACCTTCACCATGTGTTTCGTCG	00771R CAGATCTTAATTAAGGGCCCACGACGCTTCGTGACGATCG	XhoI-ApaI	
01376F TCACCGGATCTCGAGCTGGACTATATGCGCATCGGC	01376R CAGATCTTAATTAAGGGCCCCTATTTCAGTTCGACCACCTTGCC	XhoI-ApaI	
04454F TCACCGGATCTCGAGATCATCAAGCTCGGCAAGCACG	04454R CAGATCTTAATTAAGGGCCCCGTCAGACGCTTGTAGGTTACC	XhoI-ApaI	
0383F TCACCGGATCTCGAGCAGTCCACCTTGATCATTGGAGC	0383R CAGATCTTAATTAAGGGCCCTGAAGTGACGGACGACAGGC	XhoI-ApaI	
21162F TCACCGGATCTCGAGATCGGCATCATCGTCGGTATCG	21162R CAGATCTTAATTAAGGGCCCGTTCCCGAATAGTCGCTGATGC	XhoI-ApaI	
21575F TCACCGGATCTCGAGTGGTCTTCGCGATCTGTCTCG	21575R CAGATCTTAATTAAGGGCCCTCAAAGCGTGAGAGCGTCCG	XhoI-ApaI	
21281F TCACCGGATCTCGAGGATCTCGGTGCTGTTCCTCG	21281R CAGATCTTAATTAAGGGCCCACTCGGCGAAGAATGCGAAGC	XhoI-ApaI	
20070F TCACCGGATCTCGAGTGTTCTTCGCCGGAAAGGTCG	20070R CAGATCTTAATTAAGGGCCCAATGTGTCACGGTGAGCGACG	XhoI-ApaI	
20981F TCACCGGATCTCGAGGAGACCCGCAAGGTCATTTCG	20981R CAGATCTTAATTAAGGGCCCTACTCTGCCATCGCCTTCTGC	XhoI-ApaI	
20999F TCACCGGATCTCGAGAGCTCTTCGACAATGTCAGGCG	20999R CAGATCTTAATTAAGGGCCCACAGAGCCAGCTCTTCCTTACC	XhoI-ApaI	
02437F TCACCGGATCTCGAGGCAGCTCGACGAACTGATGG	02437R CAGATCTTAATTAAGGGCCCGAGGAGTTCGTTGAGCCTGC	XhoI-ApaI	
00317F	00317R	XhoI-	

TCACCGGATCTCGAGCAGCGGCTTATCGACTATCTGG	CAGATCTTAATTAAGGGCCCGCAACGTCGTGATGAGGATCG	ApaI	
02343F TCACCGGATCTCGAGCATTTGTCGACGGCGAGACC	02343R CAGATCTTAATTAAGGGCCCAGAAAGCGCAGTGCGGAAGC	XhoI-ApaI	
03825F TCACCGGATCTCGAGGCTTGTGATCCTGCAGCACG	03825R CAGATCTTAATTAAGGGCCCTCGGTCACGATGACGTTCTTGG	XhoI-ApaI	
01211F TCACCGGATCTCGAGACCATCTTCGCTCTGTTCCCG	01211R CAGATCTTAATTAAGGGCCCTACGGAGCAGACGGAAGAAGC	XhoI-ApaI	
ML8388 GTACTAGTTGCTCAATCAATCACCGG	ML8389 CCGAATTCGCTAGCCATTATTAATCTCC	SpeI-EcoRI	amplify insert of pFL2765 for the construction of pTH2313
smc02619F TCACCGGATCTCGAGGCTTTGCCTGACATTCCTGTTAGG	smc02619R CAGATCTTAATTAAGGGCCCGTCACGTCTGTGCATTGCATCG	ApaI-XhoI	amplify 5' and promoter region of smc02619
smc02618F TCACCGGATCTCGAG CTCGTGCTGAACATCGGAGC	smc02618R CAGATCTTAATTAAGGGCCCGCTCTTCGATCACTTCGCGG	ApaI-XhoI	amplify 3' end of smc02618
smc02615F TCACCGGATCTCGAGAACTGTGGCTGAAGCTCACTTTGC	smc02615R CAGATCTTAATTAAGGGCCCAGCGAATAAGCGGGCTCATCG	ApaI-XhoI	amplify 5' end of smc02615
4247F ATGGGCCCCATTTCAAGCCTGCTTTCAAGTGC	4247R ATAGATCTGTGTAGACGAGCCAGAACAGC	ApaI-BglII	amplify promoter region of smc04247
4248F ATGGGCCCTTCAAGACCTTTCACGCAGGCG	4248R ATAGATCTGTTCTGGCTCAGCGTGTAGG	ApaI-BglII	amplify promoter region of smc04248
4251F ATGGGCCCGGTCGCTGGACATGCATTATCG	4251R ATAGATCTCGTTGAAGAGGTCGATGCCG	ApaI-BglII	amplify promoter region of smc04251
4253F ATGGGCCCAGCCTGCTTCTGCGTTGACC	4253R ATAGATCTCCTTGGACGAGATGATGAGTTCG	ApaI-BglII	amplify promoter region of smc04253

4259F ATGGGCCCCAGCGTTACTAACAGCTTACCTCG	4259R ATAGATCTGGTGAGATCGCGCATCAATCC	ApaI-BglII	amplify promoter region of smc04259
4260F ATGGGCCCGCCTTTAATGTCTTCCCATATGCG	4260R ATAGATCTGACCAGGTGGTATTGCGTGC	ApaI-BglII	amplify promoter region of smc04260
4258F ATGGGCCCCAATGTGATCCAGGGCACGG	4258R ATAGATCTCCACGAAGATGACCATGGCC	ApaI-BglII	amplify promoter region of smc04258
gndF ATGGGCCCCATGAGGCGATATATCCCGTGG	gndR ATAGATCTCATGCCGATGAAGGTGAGACC	ApaI-BglII	amplify promoter region of *gnd*

Library Plasmid pTH1522

The plasmid pTH1522, designed for the construction of the random *S. meliloti* fusion library is diagramed below. The *XhoI* restriction site is flanked by two sets of reporter genes. In one direction *gusA* and *rfp* (red fluorescent protein) are the enzyme and fluorescent reporter genes and in the opposite direction gfp^{+} (green fluorescent protein plus) and *lacZ* are the fluorescent and enzyme reporter genes. With this system, you can have a DNA fragment cloned in either direction and can generate a transcriptional reporter fusion. Another important feature of this construct is that it has a pMB1 origin of replication and therefore cannot replicate in *S. meliloti*. Thus any SmP110 transconjugants with this plasmid must have an integrated plasmid (Cowie et al., 2006).

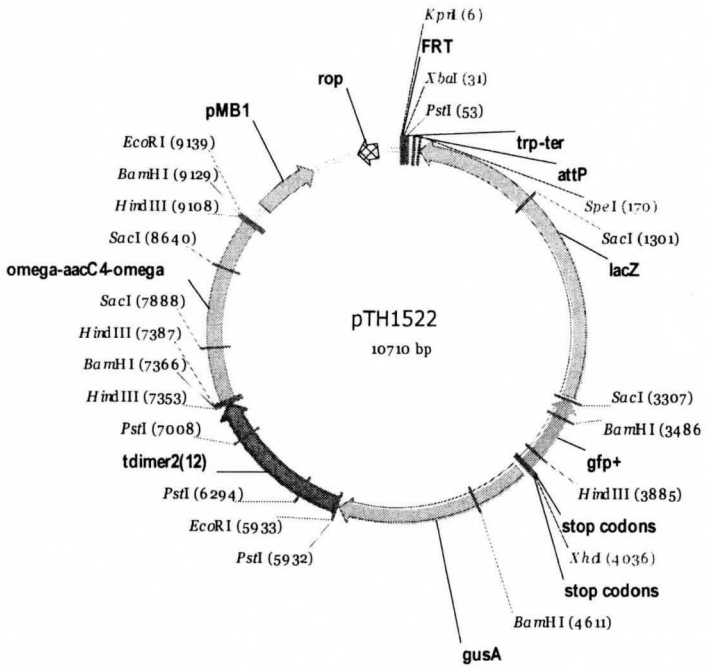

Figure 2-1: Plasmid map pTH1522 used in the construction of the library (Cowie, et al., 2006).

Recombination Event

Figure 2-2 illustrates the recombination event that occurs between the cloned DNA fragment in pTH1522 and the SmP110 genomic DNA. In this diagram the pTH1522 clone carries the promoter region and the 5' end of smc00157 along with the 3' end of smc00156. When mated into SmP110 a single crossover between the cloned region in pTH1522 and the homologous sequences in SmP110 genome will result in a fusion strain. In this particular example, the smc00157 promoter will be driving the expression of the *gfp* and *lacZ* genes. This fusion does not result in a smc00157 mutation because the recombination results in a duplication of the promoter and 5' region of smc00157.

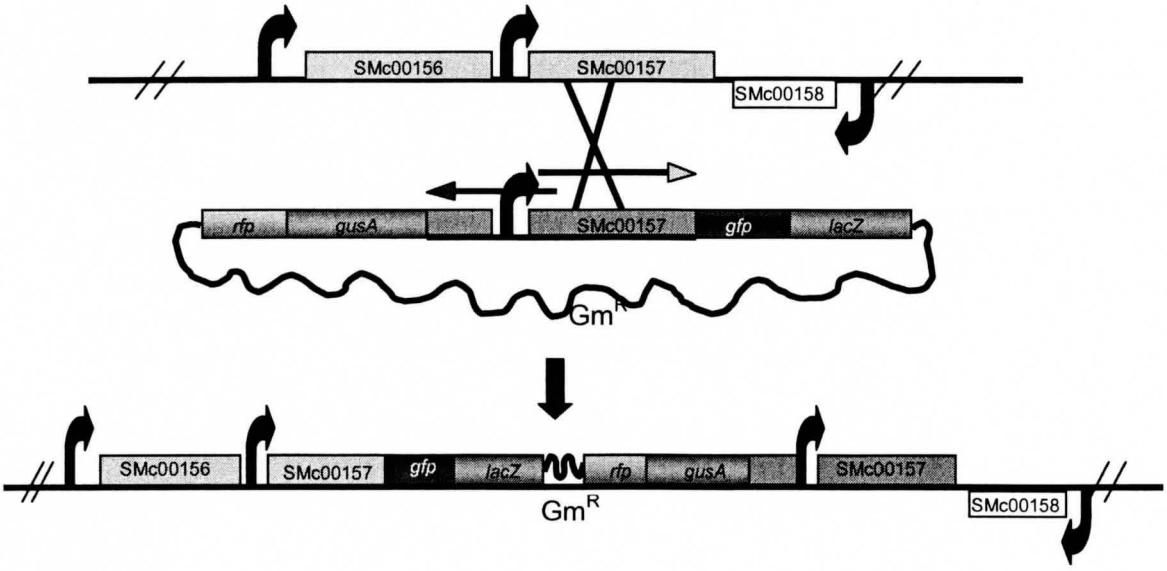

Figure 2-2: Schematic of the recombination event that occurs between library plasmid pTH1522 with cloned DNA and SmP110 genomic DNA (Chris Sibley).

In previous work, Jane Fowler and Rahat Zaheer had constructed 45 fusion strains in Rm1021 background using pTH1360 as a vector. This vector has a ß-glucuronidase

reporter gene and, like pTH1522, cannot replicate in *S. meliloti*. The transcriptional fusion strains were built at the 3' ends of putative operons in an attempt to keep the proteins functional. Below is a diagram illustrating the single cross over event that occurs when pTH1360 carrying a cloned Rm1021 DNA fragment is mated into Rm1021.

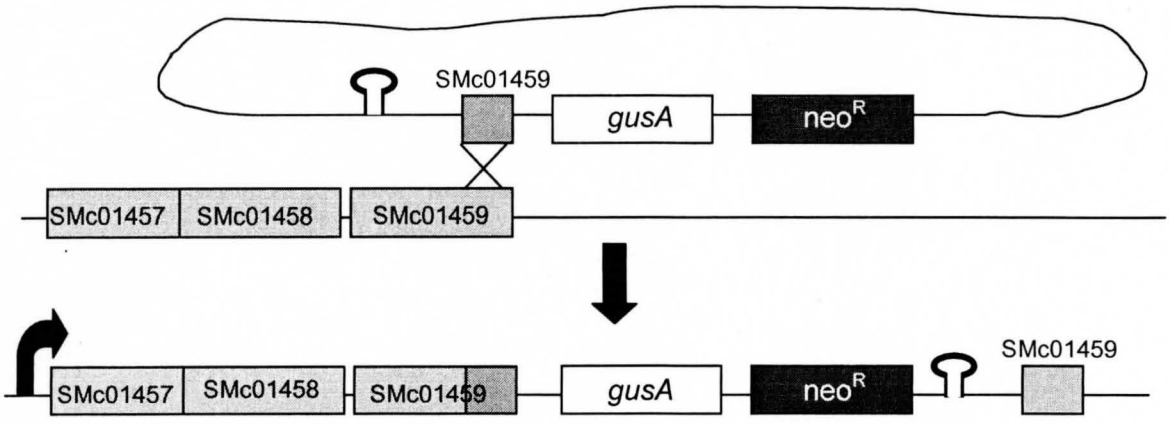

Figure 2-3: Recombination event between pTH1360 with cloned DNA and Rm1021 genomic DNA (adapted from Fowler, 2005).

Screening the Transportome

A list of all the fusions from the pTH1522 fusion library spanning the transport genes was generated using the Transport DB database for *S. meliloti*. The list was manually analyzed to identify each fusion that was ultimately included. At the same time, I attempted to choose those fusions that did not generate a knockout of the transport system. The transporters of *S. meliloti* have been termed 'The Transportome of *S. meliloti*' and the fusions to these transport genes included in this study are listed in Table 2-4. At this stage of the construction of the library not all of the fusions had been mated

into *S. meliloti* strain SmP110 so those fusions of interest, about 50 fusions, were mated into SmP110.

Each test strain was streaked out onto LB $Sm_{200}Gm_{60}$ agar medium and then inoculated into broth. Strains were frozen down into six separate 96 well microtiter plates divided by the nature of the fusion (i.e. whether it was a gusA/rfp or a gfp/lacZ fusion). The frozen stocks were prepared by combining 60 µl of overnight culture and 60 µl of 14% DMSO in LB, followed by freezing in -80°C.

In this study a high-throughput screening method was used to measure ß-glucuronidase and ß-galactosidase activity in a 96 well microtitre format to quantify expression of many reporter gene fusions under various conditions. The fusion strains and controls were grown in 1.6 ml of $LB_{mc}Sm_{100}Gm_{30}$ in deep well microtiter plates for 24 hours at 30°C (to an OD_{600} of approximately 1.0). The strains were then subcultured into microtitre plates containing 215 µl of various test media using a replicating tool which inoculates approximately 5 µl of culture. Growth in test media was approximately 48 hours at 30°C. The MultiProbeII was used to dispense reagents into the 96 well plates in order to achieve a more efficient screen.

As the screening was on a large scale, the buffers were made as a master mix, where 80 µl of either Gus buffer or LacZ buffer was added to 20 µl of cell culture. The reactions were allowed to develop for one hour then stopped by the addition of 100 µL of 1 M Na_2CO_3. The 80 µL GusA developing mixture contained: 78 µl GUS buffer, 1 µL p-nitrophenyl-ß-D-glucuronide (PNPG) (35 mg/ml) and 1 µl of 1% SDS (final concentration of 0.01%). Similarly, the 80 µl LacZ developing mixture contained: 72 µL Z Buffer, 0.216 µL 2-mercaptoethanol, 0.1 µl of 10 % SDS (final concentration of 0.01

%), and 8 µL 2-nitrophenyl β-D-galacto pyranoside (ONPG) (64 mg/ml) dissolved in 0.8 mL Z Buffer.

A microtitre plate containing 100 µl of cell culture was used to generate an OD_{600} reading for cell density and also a Gfp or Rfp value for each culture. The Gfp was read at an excitation wavelength of 485 nm and an emission wavelength of 510 nm. The Rfp was read at an excitation wavelength of 552 nm and an emission wavelength of 579 nm. Fluorescent readings were measured using the Tecan Safire microtiter plate reader to calculate the relative Gfp and Rfp fluorescence. The fluorescent readings were measured during the same time period as are the enzyme activities. The enzyme reaction plate was read at 405 nm (GusA) or 420 nm (LacZ) and using the Tecan Safire microtitre plate reader so that Miller Units could be calculated using the following formula:

$(1000 \times A_{405}/A_{420}) / (A_{600} \times$ reaction time in minutes x volume of culture used in ml).

This optimized assay has eliminated the separate steps of permeabilizing agent addition and substrate addition by using one pre-mixed reaction mixture. This method has been demonstrated to be a rapid and reproducible assay (Cowie, et al., 2006). Any positive results obtained were repeated at least once but often twice using the high-throughput method. 320 different fusion strains were tested for each substrate in addition to two positive controls, RmK990 (*pckA*::*gusA/Rfp*) and RmK991 (*pckA*::*gfp/lacZ*), which have been shown to be constitutively expressed under various conditions. SmP110 was used as a negative control and a blank media control was used to ensure no contamination and to subtract the background activity that may be recorded from the test media.

Preparation of Seed and Root Exudates

Alfalfa, sweet clover, lentil and beans were surface sterilized in 95% ethanol (1 minute) and 1.25% (w/v) hypochlorite (15 minutes) and rinsed five times with sterile water. For preparation of seed exudates, seeds were imbibed in sterile double distilled water in the dark for 6 hours. Two seeds per ml of water were used for large seeds (pea, lentil and bean) and for small seeds (alfalfa and sweet clover) the volume of water was four times the volume of seeds used. Alfalfa, sweet clover and pea root exudates were prepared by germinating seeds on water agar for 2 days and then imbibing seedlings in sterile water in the dark for 5 days. This was done by placing 30 seedlings in a 50 mL falcon tube and then filling the remaining volume with water. All exudates were filtered through Whatman filter paper (No.7) to remove plant debris, then through a 0.45 μm syringe filter to ensure sterility. If not used immediately exudates were stored at –20°C.

Bacterial Matings

Plasmids were transferred to *S. meliloti* from *E. coli* by triparental mating using overnight cultures of the recipient strain, donor strain, and helper strain MT616 (which carries the self-transmissible plasmid pRK600 that provides the transfer functions in trans). These strains were centrifuged at 13 000 RPM for one minute in a 1.5 mL eppendorf tube using a table top centrifuge, washed twice and resuspended in 0.5 ml of sterile 0.85% NaCl. 20 μl of each culture was then spotted onto an LB agar plate and incubated overnight at 30°C. The mating spot was resuspended in 1 ml sterile 0.85%

NaCl and dilutions were plated onto appropriate media for selection of the *S. meliloti* recipient carrying the plasmid.

β-glucosidase Assay

Two litres of *S. meliloti* P110 were grown in M9 minimal media supplemented with cellobiose, salicin, or glycerol as the sole source of carbon to an OD_{600} of 0.8 to 1.0. Cells were pelleted using Beckman centrifuge and spinning for 30 minutes at 10 000 RPM at 4°C. The pellet was resuspended in 10 ml sterile 0.85% NaCl. Pellets were stored at -20°C until needed.

Pellets were resuspended in 15 ml of 100 mM Tris (pH 7) and 0.5 mM DL-dithiothrietol (DTT) (Bioshop). Each sample was passed through the French Press three times. The soluble fractions were isolated from the insoluble fractions by centrifuging samples in the Sorvall centrifuge at 4°C for 30 minutes at 15 000 RPM. The supernatant was removed and saved and the pellet was resuspended in 5 ml of Tris + DTT buffer. Both fractions were frozen down in 1 ml samples and stored at -80°C.

For a 1 ml reaction 1 µl of crude cell extract was added to 50 mM phosphate buffer (pH 7.0), 50 µl of 100 mM p-nitrophenyl β-D-glucopyranoside (pNPβG) (Sigma). Reactions were placed at 30°C for 30 minutes and to measure the activity OD401 readings were taken using the Tecan Safire microtitre plate reader. Protein content of samples was determined according to Bradford (Bradford 1976). The following equation was used to calculate the specific activity of the β-glucosidase being measured in the assays:

Activity = (Absorbance $O.D._{400}$*1000)/(time in minutes*protein in mg)

Construction of *E. coli* M1223

To determine if the suspected *S. meliloti* β-glucoside transporter and associated metabolism genes were sufficient to support growth on β-glucosides, the genes were captured from *S. meliloti* and transferred into DH5α using the Flp recombinase system. Before advancing with the cloning, the system was tested for expression in *E. coli*. This was done by simply growing the *E. coli* library strain EcFL1580 and M411 (DH5α (pTH1522)) in LB to see if the fusion was turned on. The EcFL1580 strain showed a 6.3-fold induction over M411 (data not shown).

The library plasmid pFL2765 was used as a template for PCR amplification where the cloned upstream region of the transporter was amplified using primers ML8388 and ML8389 with engineered *SpeI* and *EcoRI* sites. This was cloned into plasmid pTH1937 using these engineered restriction sites. The resulting Nm^R plasmid carrying a FRT site was mated into library strain SmFL5992. The resulting *S. meliloti* strain, RmP1517, now has FRT sites flanking the entire transporter and associated metabolism genes (see Figure A-1 in the Appendix). Using a quadra-parental mating with the newly constructed *S. meliloti* strain carrying the two FRT sites, MT616 as the helper strain, M842 (carrying pTH1944 (Tc^R) to provide the flp recombinase), and M928, which is a Rf^R variant of DH5α. This *E. coli* strain needs to be used so it can be selected for after the quadra-parental mating.

DNA Manipulations and Transformations

Plasmid DNA was isolated using the QIAquick miniprep kit (Quiagen) and the Gene Elute miniprep kit (Sigma) by following the manufacturer's directions.

Restriction enzyme digests were carried out by following the manufacturer's directions (Roche, NEB, Invitrogen).

For ligation reactions, an excess of purified PCR product and the plasmid DNA were passed through a QIAQuick PCR purification kit. The ligation reaction was carried out in a 10 or 20 µl reaction containing ligase and ligation buffer as suggested by the manufacturer (NEB) and incubated overnight at 16°C.

All competent cells used in this work were prepared by myself or purchased from Invitrogen. 50 µl of competent cells and 5 µl of ligation mixture were kept on ice until the former had thawed. The competent cells were added to the ligation sample, and the resulting mixture was incubated on ice for 30 minutes. The cells were then heat shocked at 42°C for one minute, followed by a two minute incubation on ice. 950 µl of LB broth containing no antibiotics was added, and the cells were incubated at 37°C for a minimum of two hours. 100 µl aliquots were plated onto LB agar with selective antibiotics and incubated overnight at 37°C.

Transduction of ntrA- and ntrC- into SmFL1790, SmFL3396, and SmFL4232

An overnight culture of the donor was subcultured into 5 ml of LBmc. When the optical density reached 0.4, 50 µl of undiluted ΦM12 was added to the culture. This was incubated overnight at 30°C until clear. Two drops of chloroform were added to the lysate and this sample was then diluted 1:20 in LBmc. An equal volume (500 µl) of the

diluted lysate and overnight culture of recipient culture (OD of 0.8) were mixed and incubated for 20 minutes at 30°C, not shaking. 3 ml of sterile 0.85% NaCl was added and centrifuged for five minutes at 5000 RPM, repeated twice. The final pellet was resuspended in 250 µl sterile 0.85% NaCl and 100 µl was spread plated onto an appropriate selective media.

PCR

Primers were synthesized (Sigma Genosys) and resuspended in distilled water to a concentration of 100 pmol/µl. PCR reactions were carried out in an Eppendorf Mastercycler epgradient S. PCR amplifications were carried out in 100 µl volumes containing: 10 µl 10x PCR buffer, 16 µl dNTPs (1.25 mM stock), 1 µl of each primer, 1.0 to 2.5 mM $MgCl_2$, 0.2 µl Platinum Taq polymerase (invitrogen) and brought up to 95 µl with ddH_20 and mixed by slight aspiration with the pippett. This mixture was added to 5 µl of the template DNA (2 ng/ul genomic DNA). Each reaction began with an initial melting for two minutes at 95°C followed by 30 cycles of amplification with 30 seconds of melting (95°C), 40 seconds of annealing ranging from 55 to 62°C depending on the melting temperatures of the primers, an extension at 72°C for 1 minute per expected kilobase of product. The final step was an extension for seven minutes at 72°C. Gel electrophoresis was performed to confirm the presence of PCR product. PCR purification using a QIAgen PCR purification kit was performed on all inserts prior to cloning.

DNA Sequencing and Analysis

DNA sequencing was carried out using dye terminator chemistry and cycle sequencing on an ABI 373 Stretch automatic sequencer (Mobix).

CHAPTER 3. RESULTS FROM SCREENING

In this chapter I have outlined the process of the screening and analysis that took place during my research. To illustrate this process, I have used several unrelated examples of results. In some of the cases the inducing condition was very evident, however there were a multitude of fusion strains that showed potential induction in several conditions and various different approaches were taken to identify the actual inducing compound(s).

Chapter 3-1. Screening the Transportome Library

The TransportDB database was used for identifying all of the transport systems in the *S. meliloti* genome (www.membranetransport.org). Reporter gene fusions to the majority of these systems were available in the *S. meliloti* random library. These strains were made by a single homologous recombination event between the suicide vector pTH1522 containing the randomly inserted DNA and the genomic DNA of *S. meliloti* wildtype strain SmP110 (Cowie et al, 2006). These 359 library fusion strains were combined with 46 of Jane Fowler's strains (see Materials and Methods for further explanation) for a total of 405 integrated fusions. These gene fusion strains were screened for induction following growth in various conditions. Table 3-1 lists the conditions tested for possible induction of the fusion strains, including seed and root exudates. As a reminder, those compounds that were used as sole carbon sources were

added to the media to concentration of 10 mM while those used as nitrogen or carbon and nitrogen sources were used at 5 mM. The exception is nucleosides, which were used as sole nitrogen sources at a concentration of 2.5 mM. Also, those compounds that were only added to the media as inducers, with 0.5% glycerol and ammonium chloride as the carbon and nitrogen sources, respectively, were added at a concentration of 5 mM.

Table 3-1. Substrates, exudates, and conditions used in the high-throughput screen to investigate expression of fusion strains.

2-Deoxy-D-Ribose (C)	L-Glutamate (N)	L-Proline (CN)
5-Hydroxy-L-Tryptophan (I)*	L-Glutamine (CN)	Propionic Acid (C)
Adenine Sulphate (N)	Glutarate (C)	Protochatechuic Acid (C)
Adenosine (N)	Gly-Asp (N)	D-Psicose (C)
Adonitol (C)	Glycerol (C)	Putrescine (N)
Agmatine sulfate (N)	L-Glycine (N)	Pyruvate (C)
L-Alanine (N)	Glycolate (C)	(-)Quinic Acid (I)
Alanineamide hydrochloride (N)	GlycylGlycin (N)	D(+)Raffinose (C)
Alfalfa root (CN)	Gly-Glu (N)	Red Clover root
Alfalfa seed (CN)	L-Histidine (CN)	Red Clover seed
Allantoin (N)	p-Hydroxybenzoic Acid (I)	L-Rhamnose (C)
alpha-ketoglutarate (C)	β-Hydroxybutyric Acid (I)	D-Ribose (C)
Aly-Gly (N)	Hydroxylamine HCl (N)	D-Salicin (C)
D(-)Arabinose (C)	L-HydroxyProline (CN)	L-Serine (N)
D(+)Arabitol (C)	Inosine (N)	D-Sorbitol (C)
L-Arginine (N)	L-Isoleucine (N)	L-Sorbose (C)
L-Asparagine (N)	Iron starvation	Spermidine (I)
L-Aspartate (N)	Lactate (C)	Stachyose (I)
Glyine Betaine (N)	α-D-Lactose (C)	Succinic acid (C)
Caffeine (N)	Lactulose (C)	Sucrose (C)
Calcium starvation	L(+)Leucine (N)	Sulphur starvation
L-Canavanine (C)	L-Lysine (CN)	D-Tagatose (C)
(+/-)Carnitine HCl (C)	L-Lyxose (C)	D(+)Talose (C)
D(+)Cellobiose (C)	D,L Malic acid (C)	Taurine (CN)
Choline (N)	Maltitol (C)	Theobromine (N)
L-Citrulline (N)	Maltose (C)	L-Threonine (N)
L-Cystine (N)	Maltotriose (C)	Thymidine (N)
Cytosine (N)	D-Mannitol (C)	Thymine (N)
Deoxyadenosine (N)	Mannose (C)	D(+)Trehalose (C)
Dextran (C)	α-D-Melibiose (C)	Trigonelline HCl (CN)
Dextrin (C)	Meso-Erythritol (C)	D(+)Turanose (C)
Dulcitol (C)	L-Methionine (N)	L-Tyrosine (N)
D(-)Fructose (C)	Methyl-pyruvate (C)	Uracil (N)
D(+)Fucose (C)	Mono-methyl-succinate (C)	Urea (N)
Fumarate (C)	Myo-Inositol (C)	Uridine (N)

D-Galactosamine (C)	Nitrogen starvation	L-Valine (N)
D(+)Galactose (C)	L-Ornithine (CN)	Various Bean seed and root
α-D-Galacturonic Acid (C)	Palatinose (C)	White Clover root
Gamma-Amino-n-Buteric Acid (N)	Parabanic Acid (C)	White Clover seed
β-Gentiobiose (C)	Pea root	Xanthine (N)
D-Gluconic Acid (C)	Pea seed	Xanthosine (N)
D(+)Glucose (C)	L-Phenylalanine (N)	Xylitol (C)
D-Glucosamine (C)	Phosphate starvation	D-Xylose (C)

*Compounds that could not be used as a carbon or nitrogen source were tested as inducers (I) at a final concentration of 5 mM in the presence of ammonium chloride as the nitrogen source and 0.5% glycerol as the carbon source. Any compound used as a carbon source (C) was used at 10 mM and any compound used as a sole nitrogen source (N) or a sole nitrogen and carbon source (CN) was used at 5 mM except nucleosides which were used at a 2.5 mM concentration.

Chapter 3-2. Results of high-throughput screening

. This section addresses the various challenges that had to be tackled in order to fully gain insight into the data that were accumulated. The examples used here were selected in order to give the reader an understanding and appreciation for the process of determining whether a fusion was considered to be induced.

The following diagrams depict typical results of the high-throughput screening assays. The assays were carried out with one replicate in a 96-well format and any putative positive results were retested manually in triplicate. Of all the fusions tested, in over 120 different test conditions, 52 were found to be specifically induced by one or more of the conditions. Fusions were considered induced by a cut-off where if the fusion was expressed at least 3-fold in one compound(s) over that when grown in glycerol.

Because of the large number of strains used in this project, the various reporter fusions are referred to by their strain name (e.g. SmFL2282) and also by the gene name (to which the fusion is) as given by the *Sinorhizobium meliloti strain 1021* Genome

Project website (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/) (e.g. SMc04259).

Figure 3-1 shows the β-galactosidase activity measured when strain SmFL2282 (SMb21103::*lacZ*) was cultured under the various test conditions. This fusion strain was specifically induced by D(+)fucose. For clarity only every third condition (tick) on the X-axis is labeled and this is carried throughout this report. The conditions are always plotted in the same order along the X-axis and this order can be found in Table A-1 of the Appendix. For further clarification, those peaks on the graphs that show specific induction have been labeled.

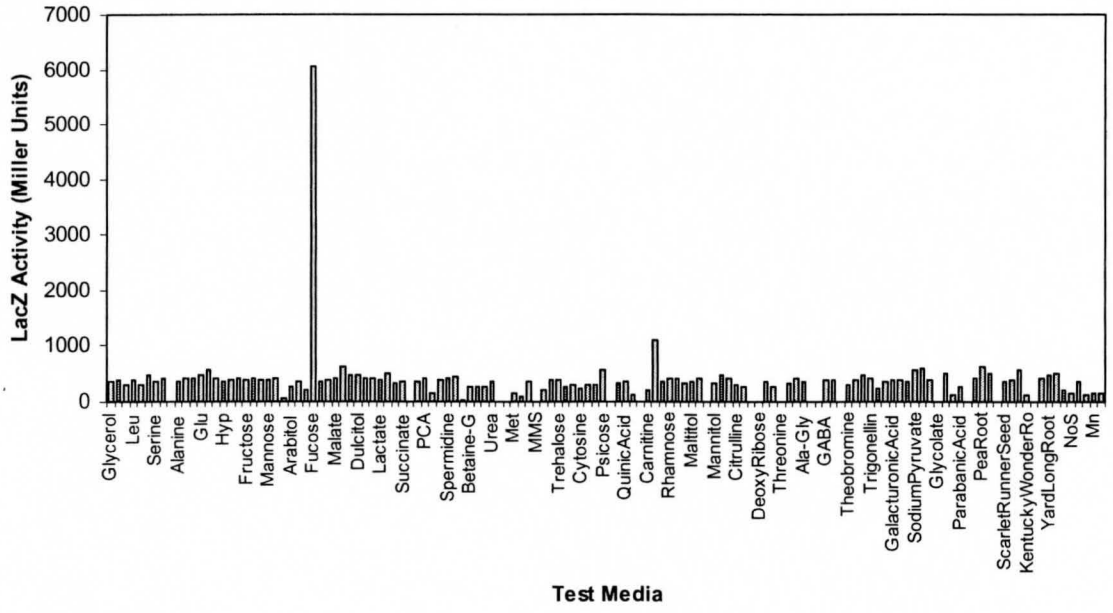

Figure 3-1. β-galactosidase activity of SmLF2282 (SMb21103::*lacZ*) when grown in the different test media.

Figure 3-2 shows the corresponding green fluorescent protein (Gfp) graph from the same strain grown under the same conditions.

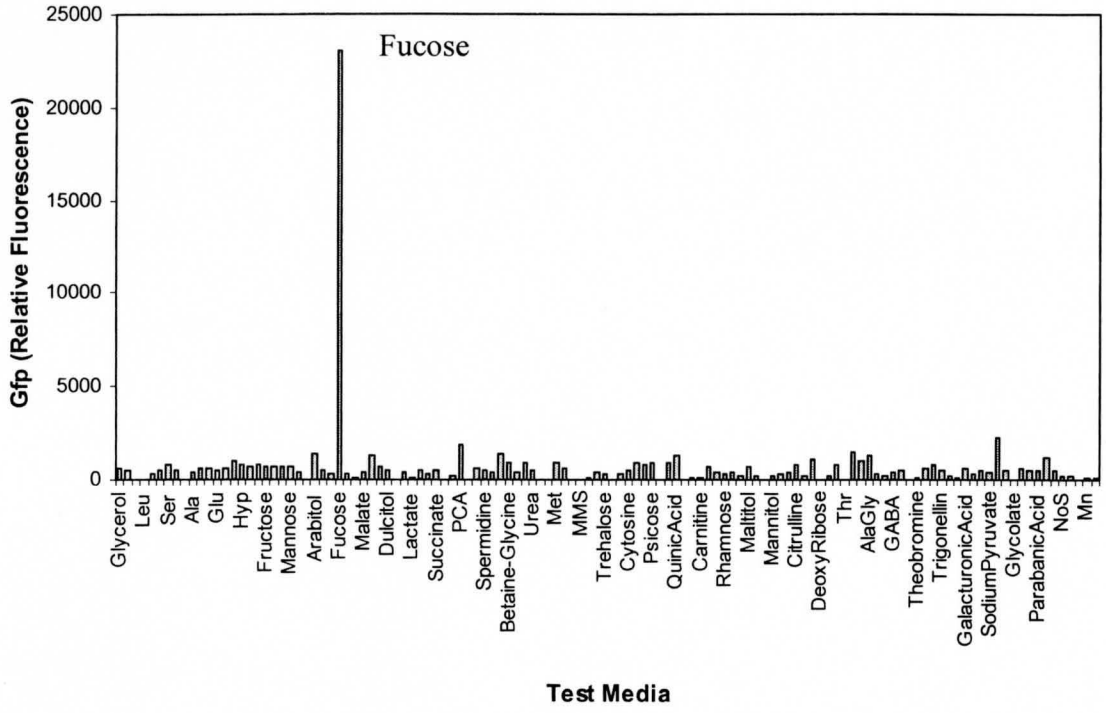

Figure 3-2. Gfp specific activity of fusion strain SmFL2282 (SMb21103::*lacZ*) when grown in the different test media

Data represented by the LacZ and the Gfp graphs show that SmFL2282 (SMb21103::*lacZ*) is indeed induced by D(+)-fucose. This is an example of where the identification of the inducing compound of a fusion is not difficult.

The following example depicts the use of an additional fusion in the transportome library to verify the results of another fusion strain. Figure 3-3 shows the β-glucuronidase activity when SmLF4493 (SMa2125::*gusA*) was grown and tested for induction in the different media. The corresponding red fluorescent protein (Rfp)

readings were measured and recorded but were not used in the analysis of the data due the low sensitivity of the Rfp readings (Cowie, et.al, 2006). This fusion strain was clearly induced specifically by caffeine and theobromine as nitrogen sources.

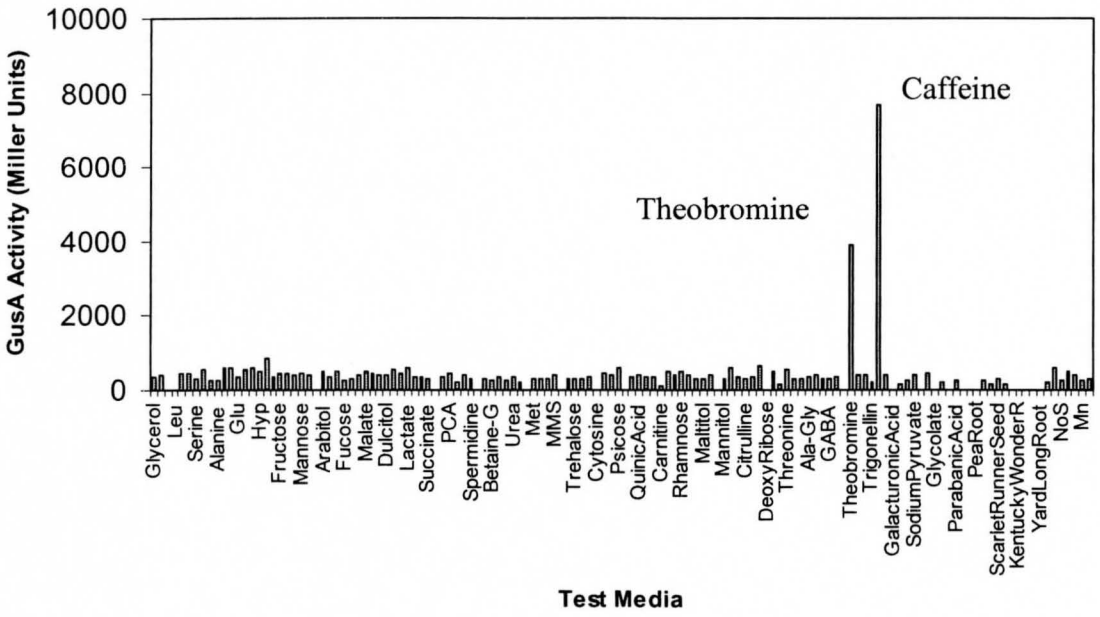

Figure 3-3. β-glucuronidase activity of SmFL4493 (SMa2125::*gusA*) when grown in the different test media.

This fusion is clearly induced by caffeine and theobromine. However, the use of another fusion to the same operon is useful for verification of such results. This is especially useful as the compounds caffeine and theobromine do not support good growth of *S. meliloti* ($O.D._{600}$ below 0.1 as measured by the Tecan Safire) and such high induction could possibly artifactual. Thus the fusion strain SmFL1501 (Sma2123::*lacZ*) was referred to and found to also be induced, though the induction pattern was much less obvious.

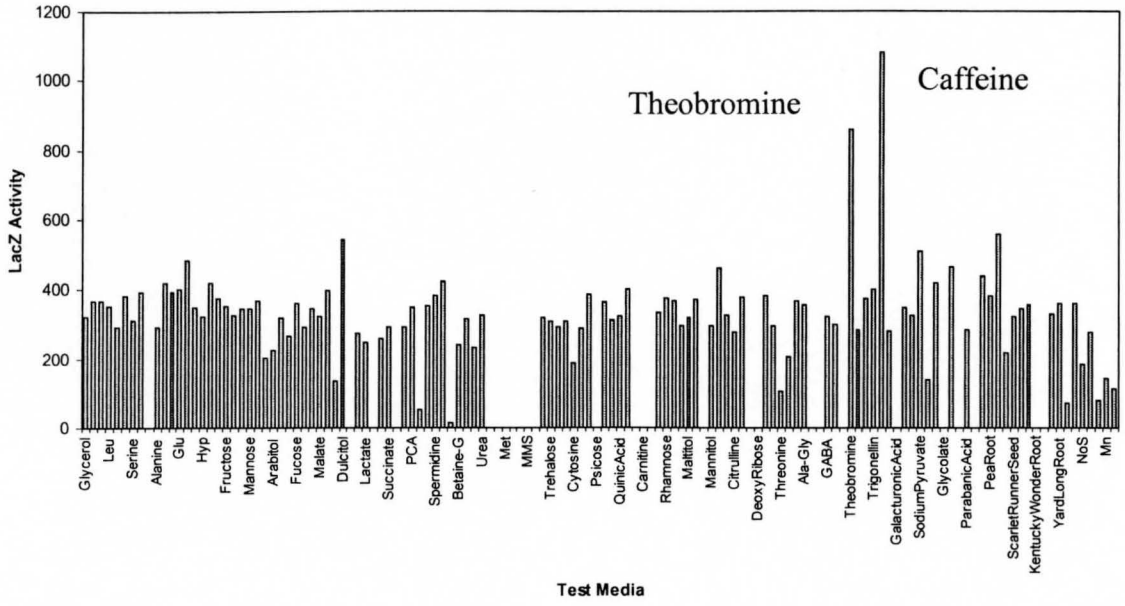

Figure 3-4. β-galactosidase activity of SmFL1501 (SMa2123::*lacZ*) when tested in all the test media showing, to a lesser extent, specific induction by theobromine and caffeine.

In some cases the inducer of a fusion strain was not clear and additional screening had to be applied. This involved performing retest experiments several times with the various media in question. Examples of this are shown in Figures 3-5 and 3-6. Fusion strain SmFL333 (SMc01624::*gusA*) appears to be induced by arabinose, arabitol, psicose, rhamnose, sorbose, deoxyribose, and red clover root and kentucky wonder root exudates. However, after retesting in the putative inducers along with other compounds (that were putative inducers for other strains) it was concluded that erythritol, adonitol, sorbitol, and xylitol were the inducers for this strain.

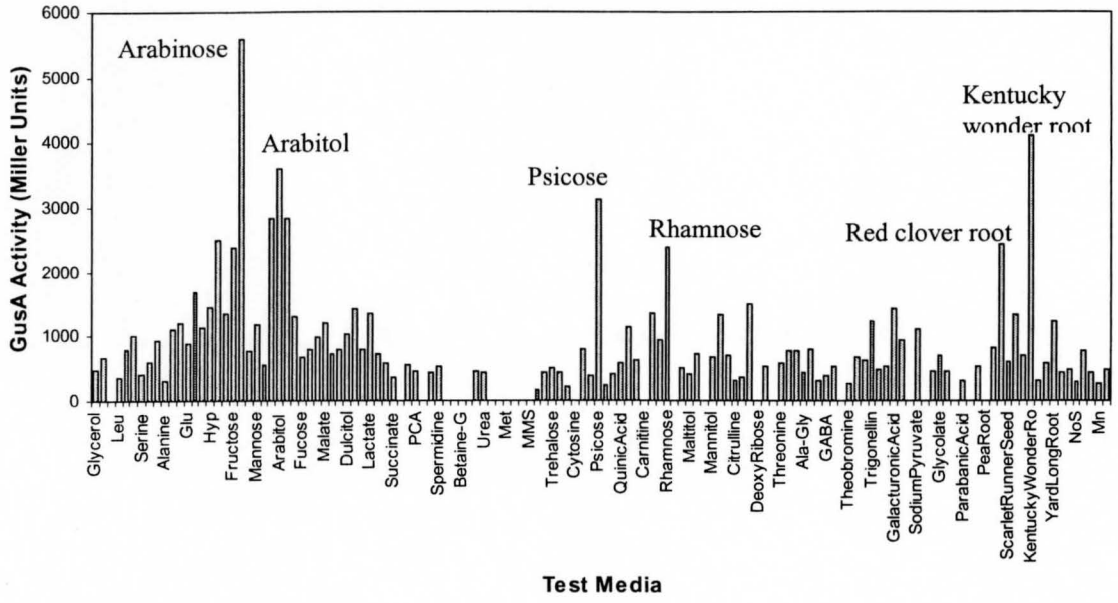

Figure 3-5. β-glucuronidase activity of SmFL333 (SMc01624::*gusA*) when grown in the different test media.

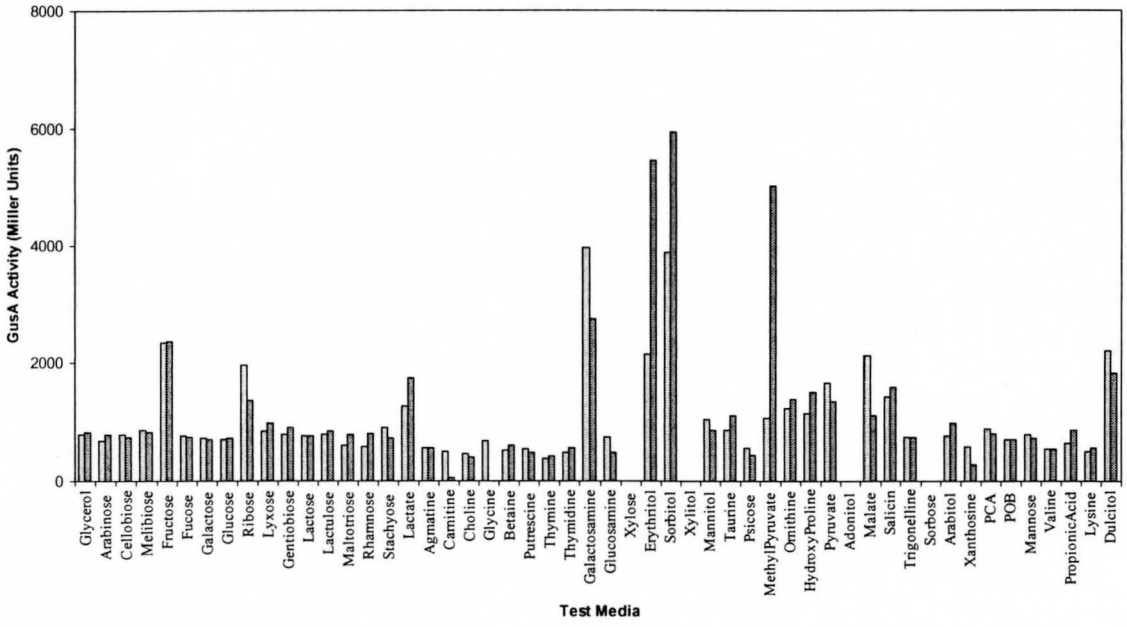

Figure 3-6. β-glucuronidase activity of SmFL333 (SMc01624::*gusA*) when grown in a subset of the test media, including those that are suspect inducers (arabinose, arabitol, psicose, rhamnose, sorbose, deoxyribose, and red clover root and kentucky wonder root exudates). The white and black bars are representative of duplicate testing, with the light grey being one replicate and the dark grey being another replicate.

In some instances there were multiple fusions to one operon represented in the transporter library. This served as a very useful tool in quality control. Fusion strains SmFL4594 (SMb20321::*gusA*) and RmP227 (SMb20320::*gusA*) are fusions to the same ABC transport system and both separately show specific induction when grown in hydroxyproline as a sole carbon and nitrogen source, as shown in Figures 3-7 and 3-8.

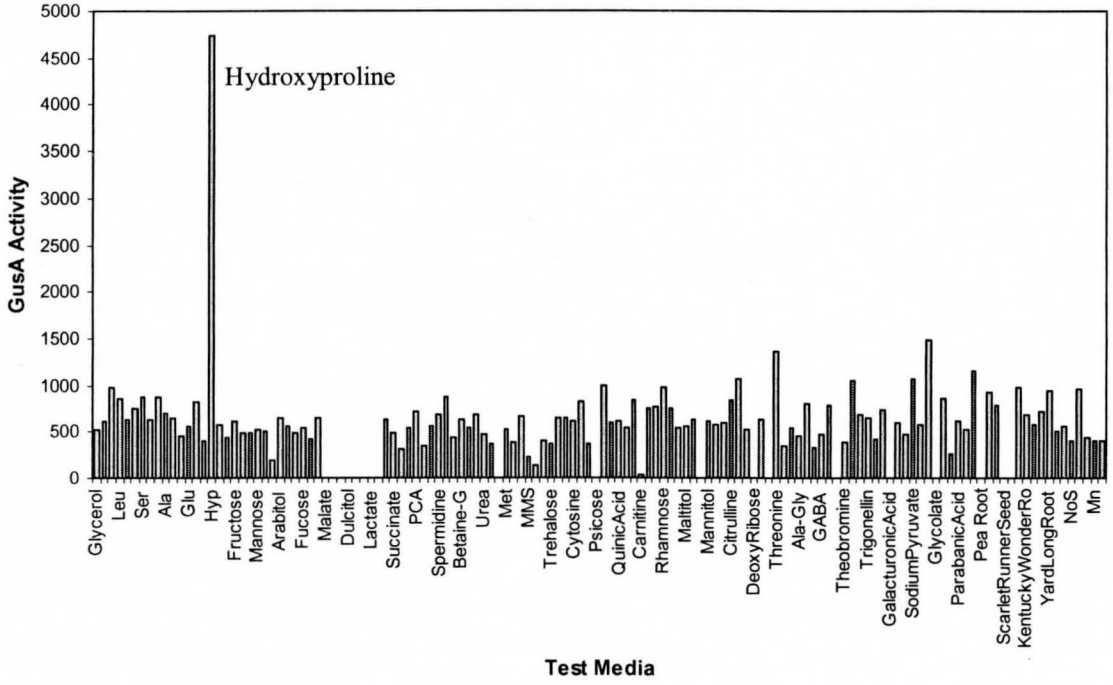

Figure 3-7. β-glucuronidase activity of SmFL4594 (Smb20321::*gusA*) grown in the different test media, showing specific induction in hydroxyproline. The missing bars are due to the very low values obtained from the screen. Sometimes these compounds did not support good growth of *S. meliloti* and in some strains do not give usable data.

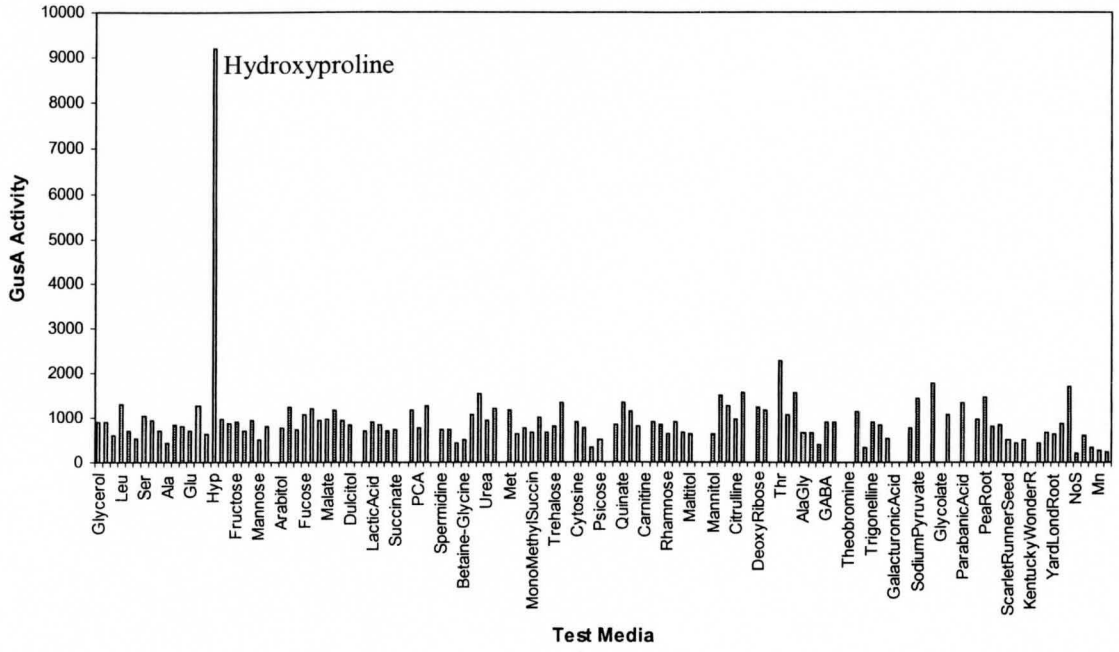

Figure 3-8. β-glucuronidase activity of RmP227 (SMb20320::*gusA*) showing specific induction in hydroxyproline when grown in all the different test media.

In some other cases the results from a fusion to a transporter gene were complimented by a reporter fusion to a gene that appeared to be a metabolism gene in the same operon. For example figure 3-9 shows that RmP193 (SMb21138::*gusA*) was found to be induced by galactosamine and glucosamine in this study (also found to be induced by pea seed exudate in Jane Fowler's M.Sc study). The library fusion SmFL1693 is a LacZ/Gfp fusion to the metabolism gene SMb21139, which is associated with SMb21138, the ATP-binding protein of an ABC transporter. In figure 3-10 it is shown that SmFL1693 (SMb21139::*lacZ*) is induced by galactosamine and glucosamine as well. The Gfp data from this fusion also shows induction by galactosamine and glucosamine (Figure 3-12).

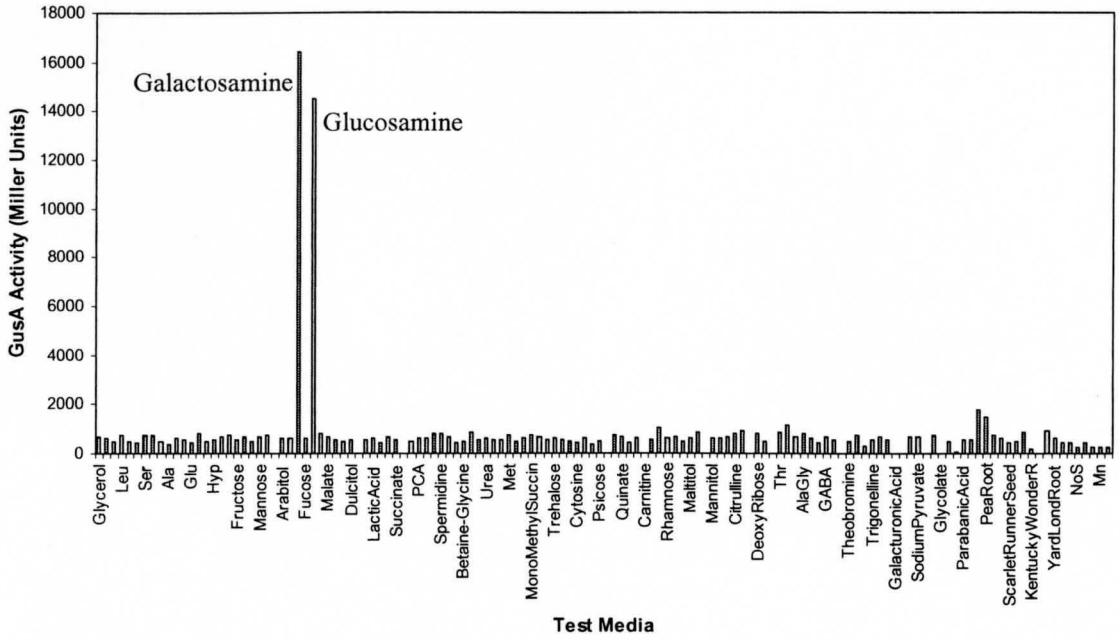

Figure 3-9. β-glucuronidase activity of RmP193 (SMb21138::*gusA*) tested for induction in all the test media. This fusion was not retested because it was extensively studied by Jane Fowler in her M.Sc.

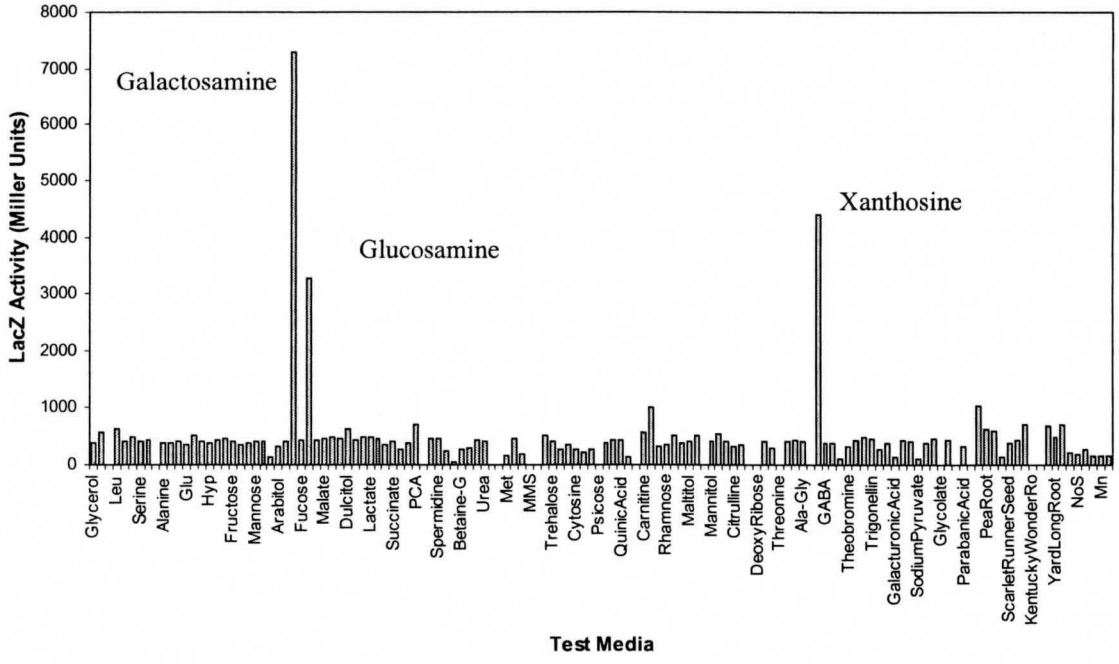

Figure 3-10. β-galactosidase activity of SmFL1693 (SMb21139::*lacZ*) when tested under the test media. This fusion strain shows induction in glucosamine and galactosamine like RmP193, but also seems to be induced by xanthosine.

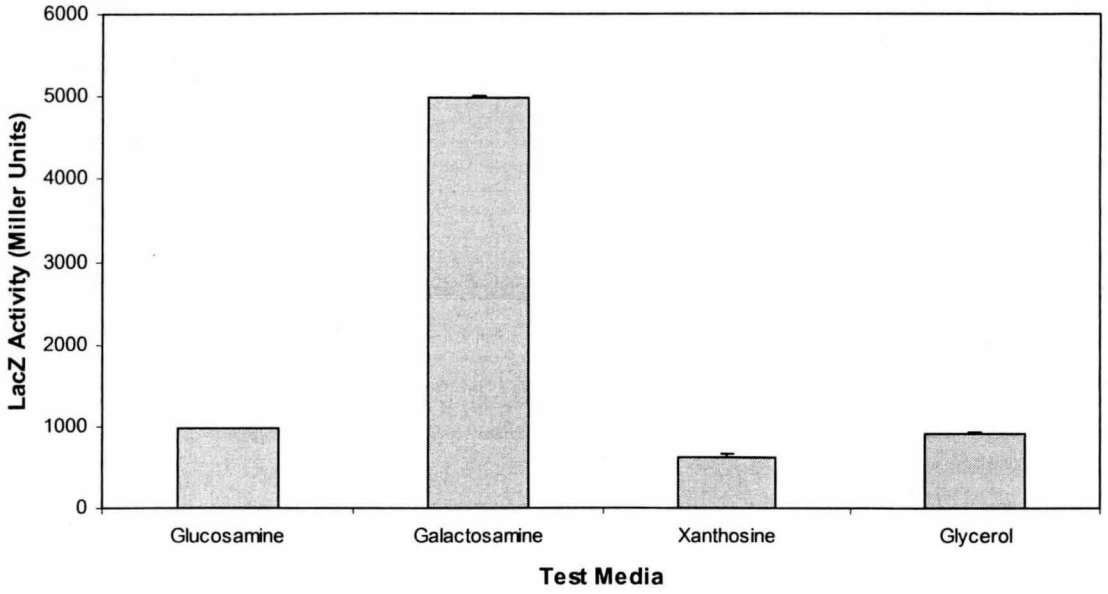

Figure 3-11. β-galactosidase activity of SmFL1693 (SMb21139::*lacZ*) when retested in triplicate in the putative inducing media, showing only real induction by galactosamine.

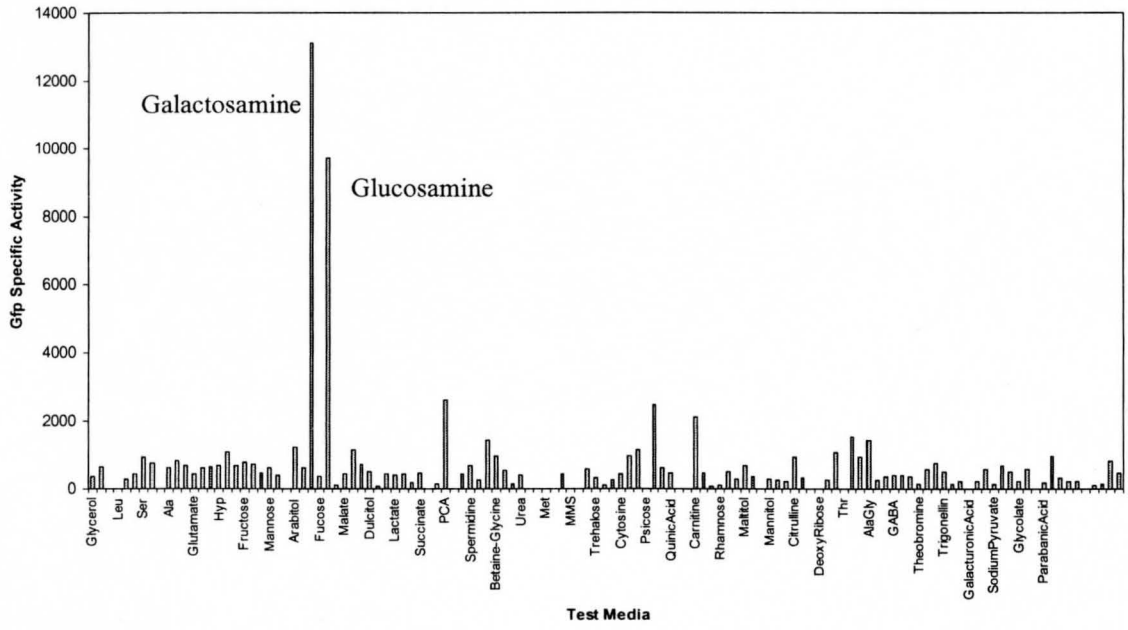

Figure 3-12. Gfp specific activity of SmFL1693 (SMb21139::*gfp*) when tested for induction in all the test media. This fusion shows specific induction in glucosamine and galactosamine but not xanthosine.

Comparing the data from these two fusion strains we can conclude that the ABC transporter containing SMb21138 is induced by glucosamine and galactosamine. The

metabolism gene SMb21139 may only be involved with the metabolism of galactosamine and thus is not induced by the presence of glucosamine in the media.

As mentioned earlier, more than one fusion to the same gene or operon can be helpful in resolving ambiguous situations. As an example, Figure 3-13 shows SmFL1336 (SMb20444::*lacZ*) with probable induction in mannose and fucose. There is also potential induction in lyxose, talose and by calcium starvation. The Gfp data from this fusion did not show specific induction in any compound and thus offered no help in resolving the inducing conditions (attched CD for raw data). However, there is another fusion, RmP227 (built by Jane Fowler), to a different gene in the same operon tested, it did not show induction in any compound but only in pea seed exudates (Figure 3-14).

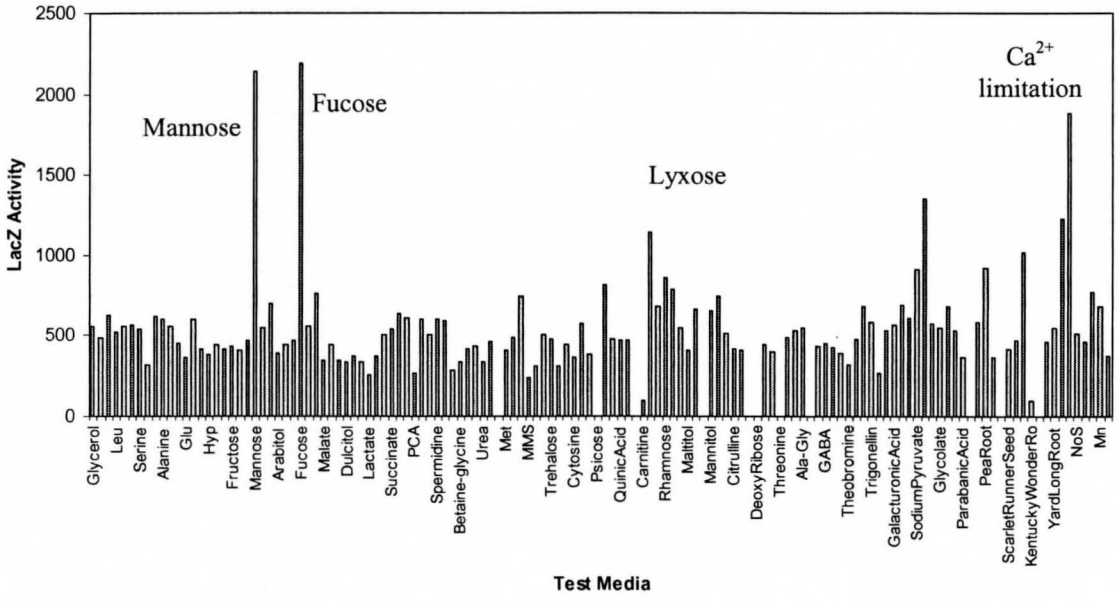

Figure 3-13. β-galactosidase activity of SmFL1336 (SMb20444::*lacZ*) when tested in all the test media, showing probable induction in mannose and fucose and potential induction in lyxose, and calcium starvation.

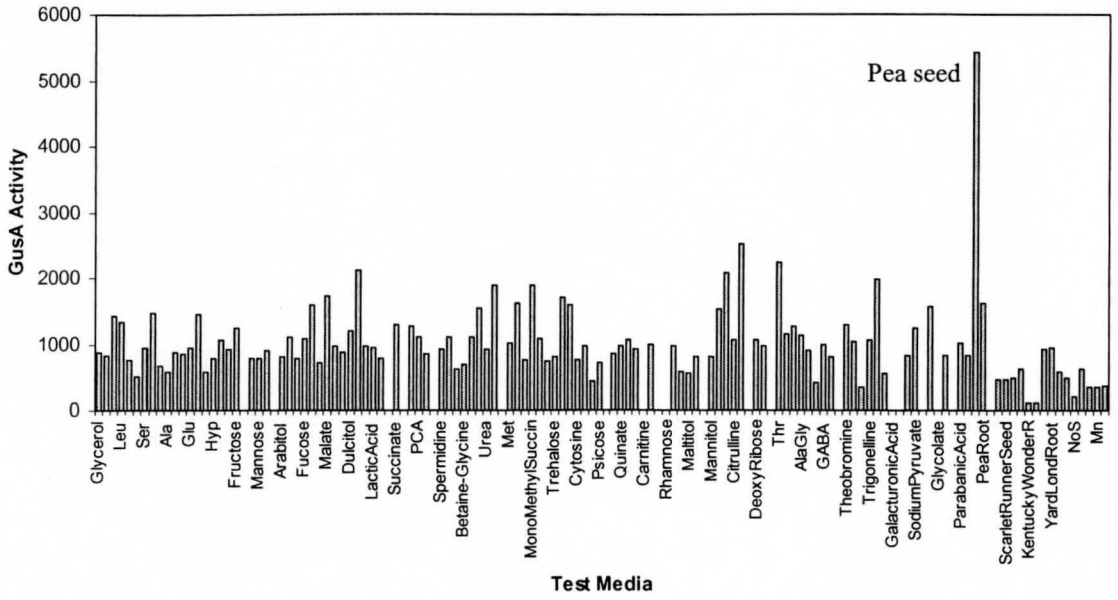

Figure 3-14. β-glucuronidase activity of RmP227 (SMb20442::*gusA*) when grown in the test media, showing specific induction in pea seed exudates.

To determine whether the operon containing SMb20442 and SMc20444 is induced by the above mentioned sugars, another fusion from the library, SmFL1446, which is a non-knockout of the operon in question, was used in the analysis. This strain is particularly useful because SmFL1336 (SMb20444::*lacZ*) is a knockout fusion and RmP227 (SMb20442::*gusA*) is a fusion to 5'end of the last gene in the operon and thus disrupts the gene and may have a knockout phenotype. According to this retest, mannose is a definite inducer and fucose is a minor inducer as there appears to be only a 2.5 fold induction over glycerol. As Figure 3-15 indicates that SMb20444 is also induced under calcium limitation, however this condition was later retested for SmFL1336 (SMb20444::*lacZ*) and no induction was found (data not shown).

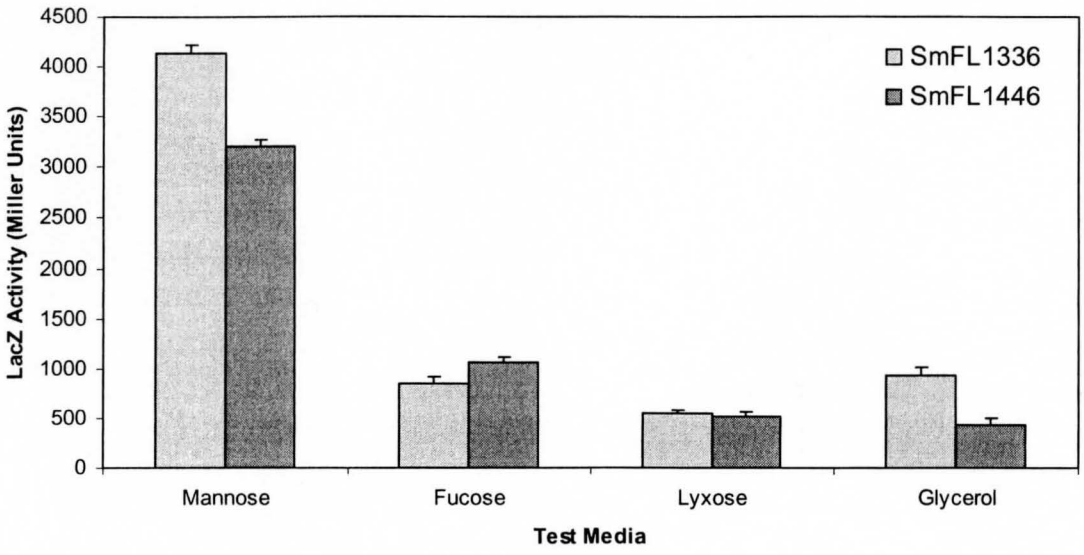

Figure 3-15. β-galactosidase activities of library fusions SmFL1336 and SmFL1446 when grown in the putative inducing conditions.

Another example of where an additional fusion from the library was used to help determine whether or not a compound was an inducer is in the case of SmFL2443 (SMc02773::*lacZ*) and SmFL3038 (SMc02776::*lacZ*). In the initial screening, SmFL2443 (SMc02773::*lacZ*) appeared to be induced by fucose and potentially by pyruvate (Figure 3-16 and 3-17). To verify these results, an additional fusion, SmFL3038 (SMc02776::*lacZ*), was included in an additional screen (Figures 3-18 and 3-19). In contrast to SmFL2443 (SMc02773::*lacZ*), SmFL3038 (SMc02776::*lacZ*) does not create a knockout of the operon and thus is more representative of a wildtype situation. In this screen the fusions were tested in duplicate in the putative inducing test media along with other test media.

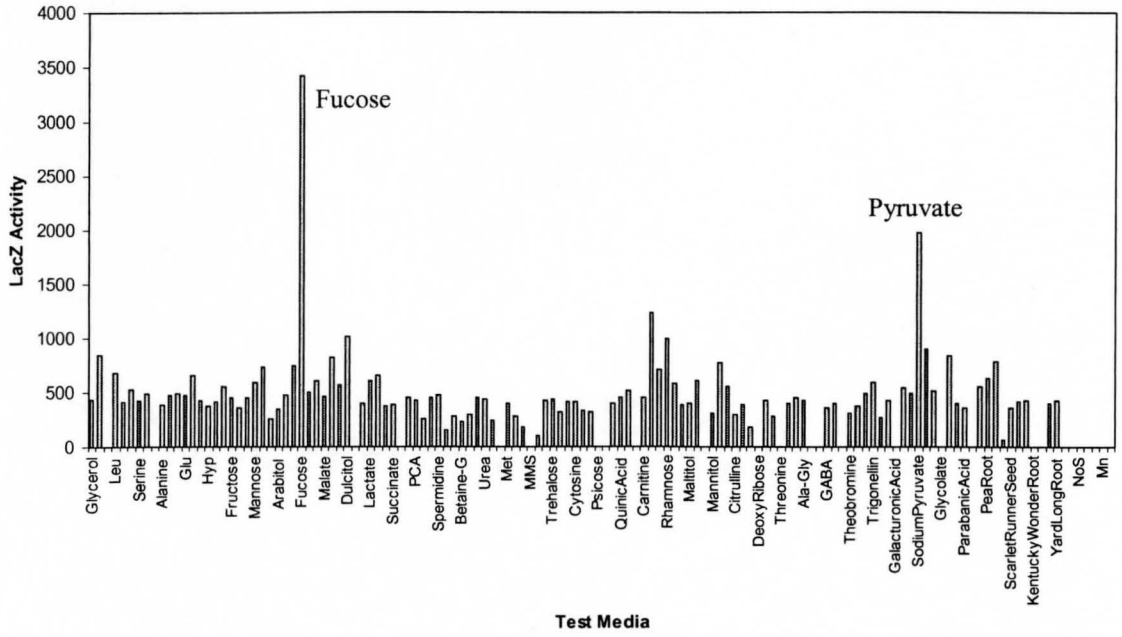

Figure 3-16. β-galactosidase activity of fusion strain SmFL2443 (SMc02773::*lacZ*) tested in all the test media showing induction in fucose and partial induction by pyruvate.

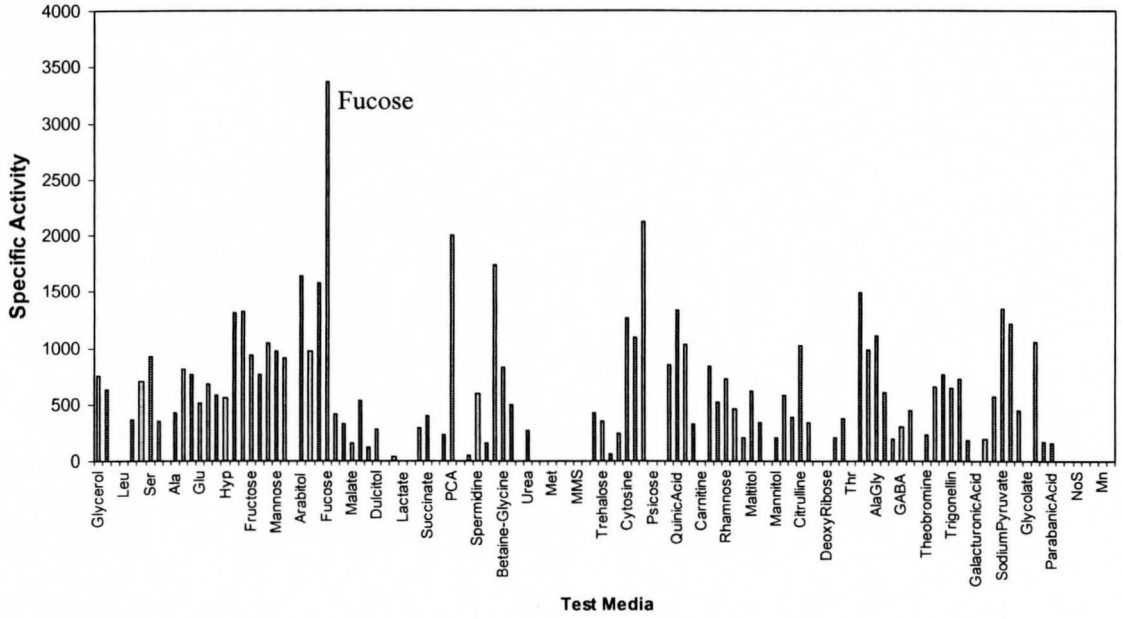

Figure 3-17. Gfp activity of SmFL2443 (SMc02773::*lacZ*) tested under all conditions. The highest induction is in fucose, though the background Gfp activity is random.

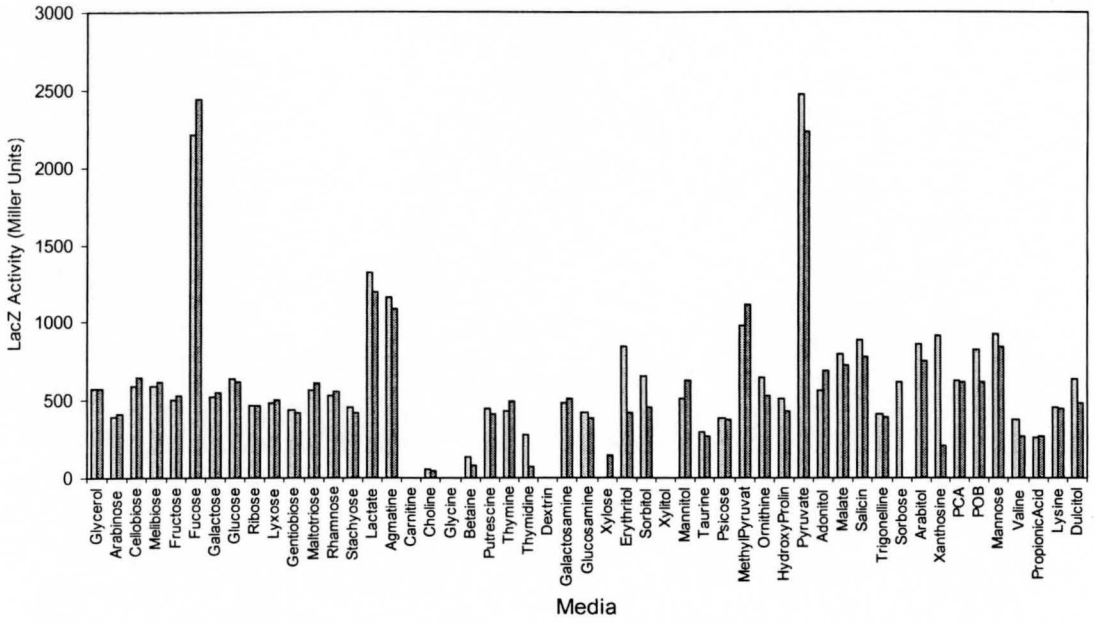

Figure 3-18. β-galactosidase activity of fusion strain SmFL2443 (SMc02773::*lacZ*) retested in select media including fucose and pyruvate and showing induction in both the media. The light grey is one replicate and the dark grey is another replicate.

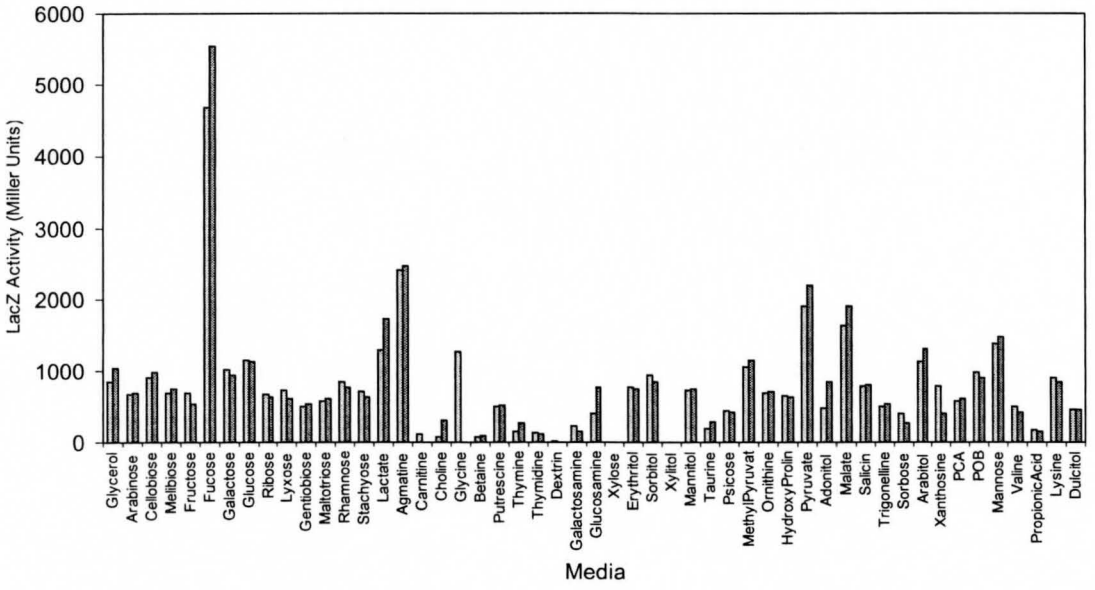

Figure 3-19. β-galactosidase activity of SmFL3038 (SMc02776::*lacZ*) tested in select media and showing induction in fucose and slight induction in pyruvate. The light grey is one replicate and the dark grey is another replicate.

Though fusion SmFL3038 (SMc02776::*lacZ*) showed only slight induction in pyruvate, it reinforced that induction in SmFL2443 (SMc02773::*lacZ*) exists and it was therefore concluded that fucose was the primary inducer and pyruvate was a slight inducer of this operon.

Since the high-throughput screen produces crude data with only one replicate, it was necessary to retest all of the potentially induced fusions in the test media in triplicate assays. As there were so many fusions and compounds to be retested, the retest method was the same as that of the high-throughput screen but was done by hand in triplicate. The following examples are representative of the process taken for all of those fusions that were eventually considered to have positive results. In both cases, there appear to be inducers and some of these inducing conditions are verified through retesting, whereas others are found to be false positives.

The results obtained from the initial screen for SmFL1889 (SMc01654::*gusA*) showed potential induction in rhamnose, agmatine, and putrescine (Figure 3-20). These results were verified by testing the fusion in triplicate in the potential inducers. The first test included just rhamnose as the test media and was found to be negative as shown in Figure 3-21. The second retest included the three polyamines, agmatine, putrescine, and spermidine as test inducers (Figure 3-22). Even though it did not show induction in the initial screen, spermidine was included in the retest due to its very close structure to putrescine and agmatine. As shown in Figure 2-22, only spermidine and agmatine showed specific induction of this operon.

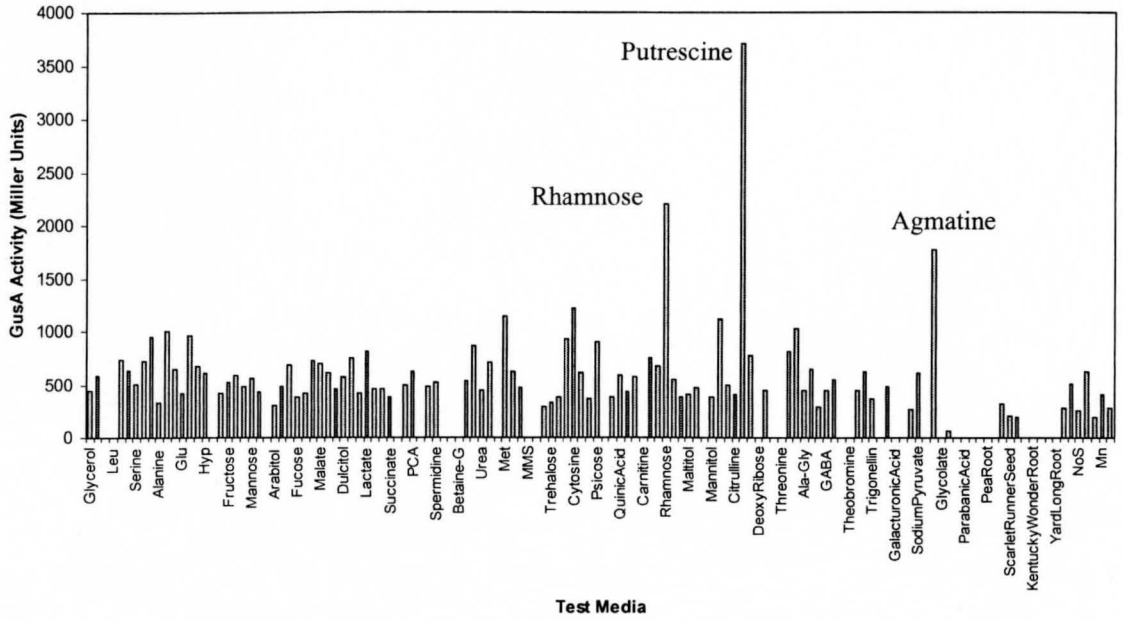

Figure 3-20. β-glucuronidase activity of SmFL1889 (SMc01654::*gusA*) grown and tested in all the test media showing induction in rhamnose, putrescine, and agmatine.

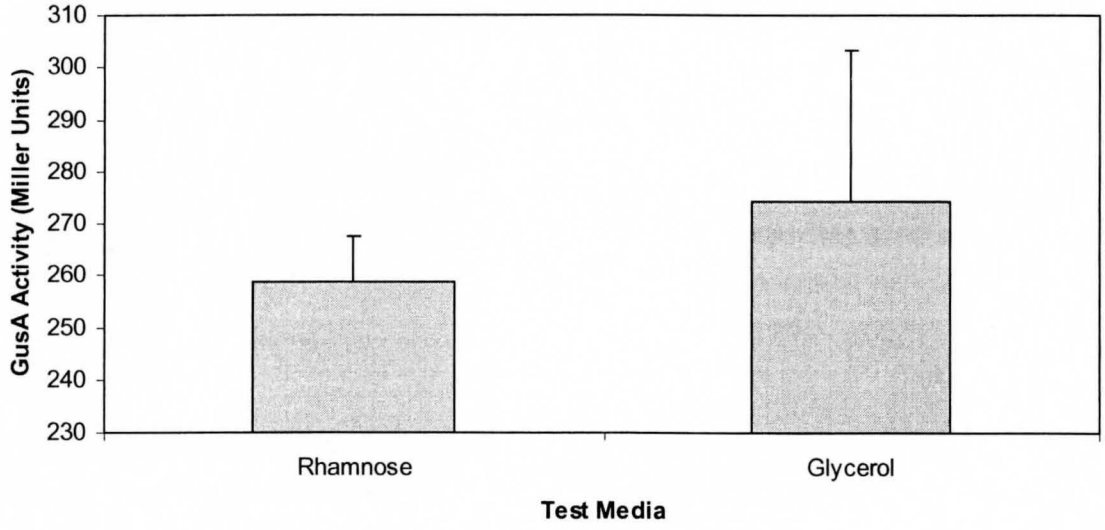

Figure 3-21. β-glucuronidase activity of SmFL1889 (SMc01654::*gusA*) showing no induction when grown in rhamnose and compared with glycerol.

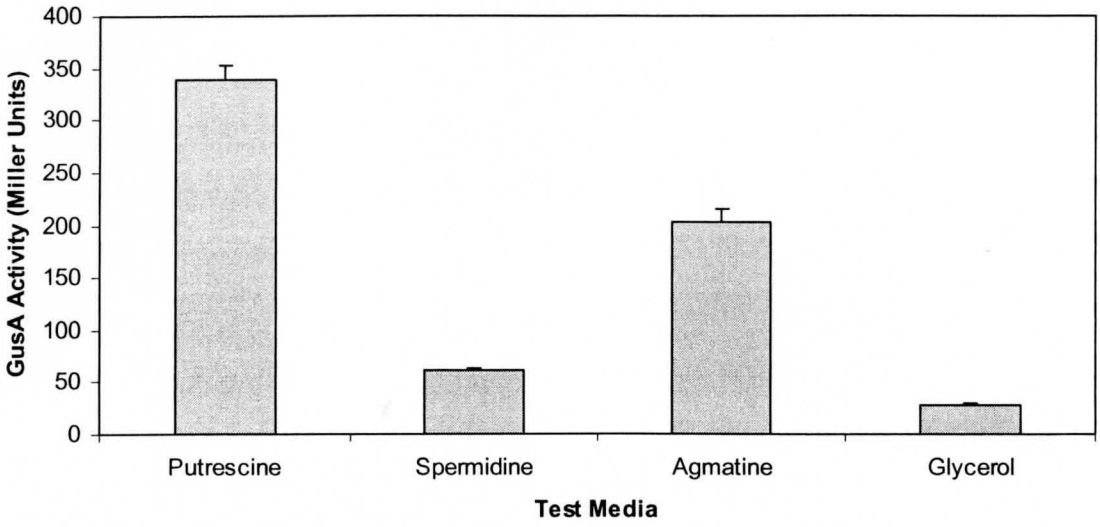

Figure 3-22. β-glucuronidase activity of SmFL1889 (SMc01654::*gusA*) showing specific induction by putrescine and agmatine but not spermidine.

Another example of the importance of retesting is with fusions SmFL3856 (SMc1823::*lacZ*). In this case, the enzyme activity, the Gfp specific activity and retest data were all used in the analysis of this strain. Figures 3-23 shows library fusion SmFL3856 (SMc1823::*lacZ*) with potential induction in uracil, uridine, and magnesium starvation. The retest data, shown in Figure 3-24, however demonstrates that this operon is only induced by the two pyrimidines when used as a nitrogen source (with 0.5% glycerol as the carbon source). The Gfp data for this fusion is scattered and does not give insight into the compound(s) that may cause specific induction (see attached CD for raw Gfp data).

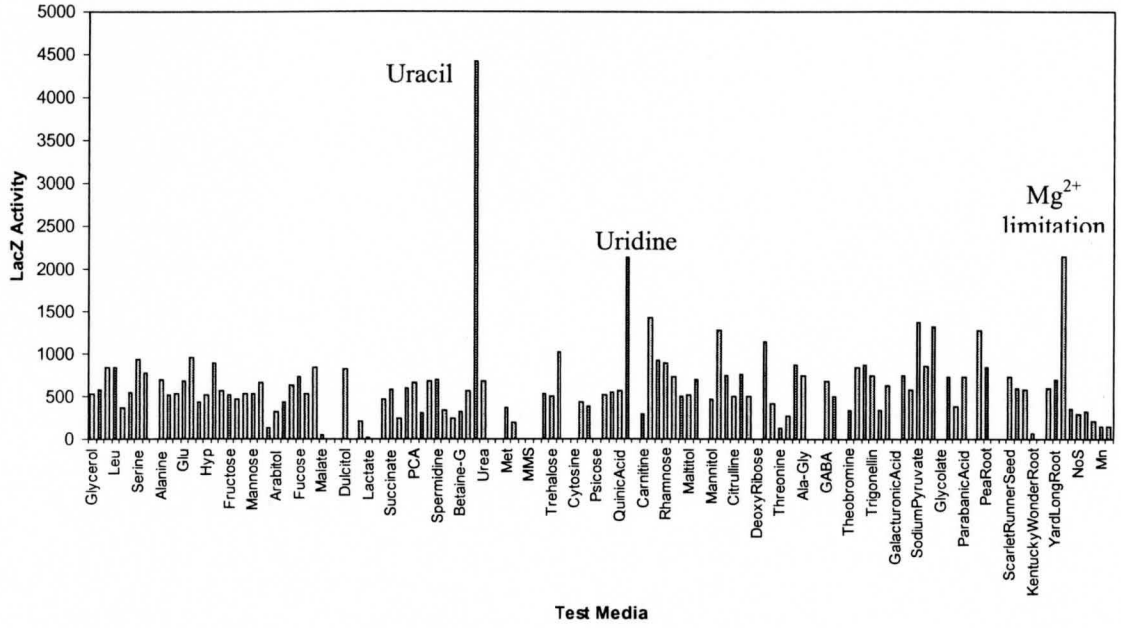

Figure 3-23. β-galactosidase activity of SmFL3856 (SMc1823::*lacZ*) tested for induction in all the test media, showing potential induction in uracil, uridine, and magnesium starvation.

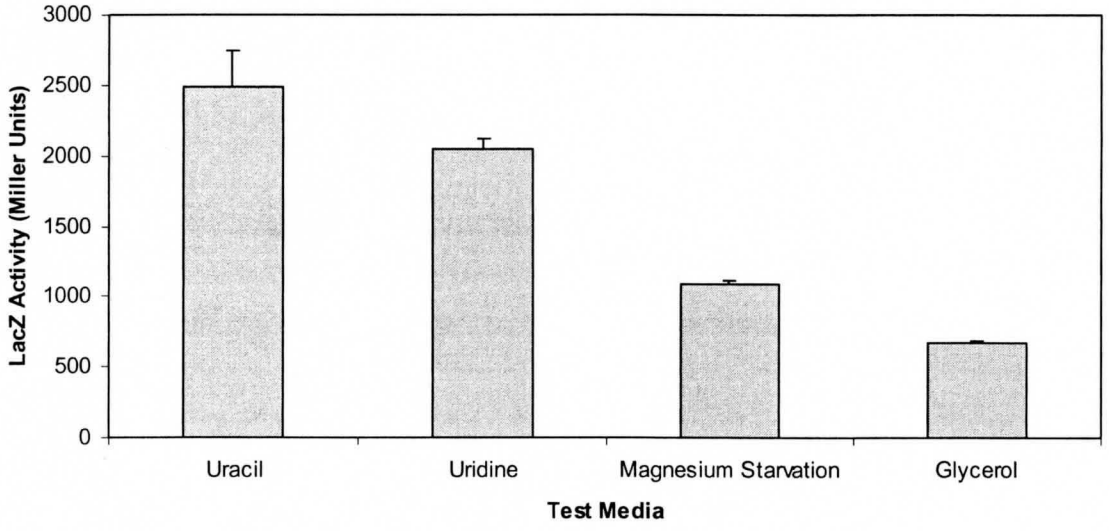

Figure 3-24. β-galactosidase activity of SmFL3856 (SMc01823::*lacZ*) retested in triplicate for induction in the potential inducing compounds. This operon was found to be induced only by uracil and uridine when used as the nitrogen source with 0.5% glycerol as the carbon source.

Through analyzing all three data sets for this fusion, LacZ, Gfp and retesting, it was concluded that SMc01823 is induced by uracil and uridine when used as nitrogen sources with 0.5% glycerol as the carbon source.

Positive Results

The following table includes the gene fusions for which positive inducers were found. The fold increase of LacZ or GusA enzyme activity in the presence of an inducing compound over the enzyme activity when that fusion was grown in M9 minimal media with 0.5% glycerol as the carbon source and 5 mM NH_4Cl as the nitrogen sources unless otherwise noted.

Table 3-2. Summary of all positive inducers for transport gene fusions

Gene	Transport family	Inducing compounds	Fold increase
SMa0151	Trap-T	malonate	4.4
SMa0198	ABC	fucose	11.5
		galactose	6.9
SMa0583	ABC	0.10% glutamine	29.7
		0.5 mM KNO_3	29.6
SMa1153	hypothetical protein	$CaCl_2$ starvation	6.7
SMa1447	MFS	isoleucine	25.7
		leucine	11.9
SMa2125*	ABC	caffeine	22.5
		theobromine	11.4
SMb20002 (*lacK1*)	ABC	lactulose	10.8
		lactose	21.7
		melibiose	4.6
SMb20036*	Trap-T	quinate	8.7
SMb20124	ABC	xanthosine	2.3
		xanthine	2.8
SMb20272	MFS	glycerol**	6.3
SMb20315	ABC	erthritol	3.7
SMb20321*	Trap-T	hydroxyproline	9.2
SMb20328	ABC	maltose	4.0

		trehalose	10.0
		turanose	3.0
		maltitol	3.6
SMb20345*	RND	glycerol**	3.9
SMb20444*	Trap-T	mannose	3.8
		fucose	3.9
SMb20571	ABC	choline	2.5
		glycine betaine	4.8
		sulphur starvation	5.1
SMb20604	ABC	0.10% glutamine	24.4
		0.5 mM KNO_3	22.9
SMb20784*	ABC	protocatechuate	6.8
		p-hydroxy benzoate	3.3
		quinate	2.0
SMb20854*	ABC	deoxyribose	5.1
SMb20904*	ABC	mannose	4.8
		sorbose	3.2
		glucose	6.6
		lyxose	6.9
		D-xylose	7.3
SMb20931	ABC	lactose	5.3
		lactulose	3.4
		melibiose	3.0
		raffinose	2.8
SMb20979	Trap-T	fucose	13.4
SMb21097*	ABC	citrulline	24.7
		red clover seed	4.8
SMb21103	ABC	fucose	17.1
SMb21138*	ABC	galactosamine	19.7
		glucosmaine	22.0
		pea seed exudates	2.7
SMb21151*	ABC	galactosamine	2.3
SMb21196	ABC	methionine	17.9
SMb21216*	ABC	galactosamine	11.9
		glucosamine	20.0
		N-acetyl-glucosamine	2.0
		pea root exudates	2.2
SMb21342*	ABC	sorbose	2.5
		galactose	4.2
SMb21353*	Trap-T	pyruvate	15.4
		methyl pyruvate	4.7
		gentiobiose	4.3
		talose	5.5
		lyxose	7.3
		galactose	4.3
		galactosamine	3.0
SMb21375*	ABC	galactose	9.0
		galactitol	29.4

		tagatose	16.6
		sorbose	11.3
		lyxose	8.5
SMb21486	MFS	$CaCl_2$ starvation	4.0
SMb21528 (*tauC*)	ABC	taurine (C and N or S)	49.4
SMb21587	ABC	arabinose	22.7
		talose	16.0
		stachyose	2.7
		galactose	6.4
SMb21644	ABC	galactosmaine	49.4
		galactitol	41.6
		raffinose	92.2
		lactose	16.5
		stachyose	33.6
		galactose	34.9
SMc01457	RND	glycine	2.1
SMc01496	ABC	glucosamine	14.9
		sorbitol	31.3
		galactitol	46.6
		maltitol	21.9
		mannitol	23.9
		sorbose	9.8
SMc01624*	ABC	erythritol	6.6
		sorbitol	7.2
		adonitol	14.4
		xylitol	17.3
SMc01654	ABC	putrescine	8.4
		agmatine	4.0
SMc01823	ABC	uracil	8.2
		uridine	4.0
SMc02325	ABC	rhamnose	10.1
SMc02344	ABC	choline	5.1
		glycine betaine	12.0
		sulphur starvation	4.2
SMc02452	ABC	glycine	3.7
SMc02516	ABC	glycerol	13.5
SMc02616	APC	trigonelline	18.3
SMc02773*	ABC	fucose	7.9
		pyruvate	4.6
SMc03061	ABC	turanose	3.6
		maltotriose	5.7
		maltose	2.8
		sucrose	3.8
SMc03807 (*amtB*)	Amt	0.10% glutamine	28.5
		0.5 mM KNO_3	26.6
SMc04147	APC	trigonelline	8.5
SMc04259	ABC	gentiobiose	57.8

		cellobiose	16.7
		dextrin	13.0
		salicin	18.1
		gluconate	9.1
SMc04393	ABC	dextrin	10.9
SMc04407	MFS	taurine	5.5

* indicates more that one fusion to this transport cluster, in these cases the results of one representative fusion were shown
** indicates gene fusions that were induced by the presence of glycerol, in these cases glucose was used as a comparison to calculate the fold increase

ABC: ATP binding cassette family with classification noted in parenthesis
MFS: Major Facilitator Superfamily
TRAP-T: Tripartite ATP-independent Periplasmic Transporter Family
RND: Resistance-Nodulation-Cell Division Superfamily
Amt: Ammonium Transporter Family
APC: Amino Acid-Polyamine-Organocation Family

Chapter 3.2: Phenotypes of Selected Transport Mutants

For several of the transporters for which inducers were found in this study, tests were conducted to investigate whether or not the transporter in question was the sole system translocating that inducer. This was done by using relevant knockout fusions found in the library and testing the strain for its ability to utilize that test compound as the sole carbon and/or nitrogen source. When testing as a carbon source strains were examined for growth on M9 medium containing the relevant carbon source both in solid and liquid media. However if we were testing a nitrogen source, growth curves had to be used as the initial test of whether or not a compound could be utilized by a knockout strain. Table 3-3 lists those transporters that were found to have knockout phenotypes.

Table 3-3. Summary of transport and metabolism gene mutants, which generate a no growth or weak growth phenotype

Gene Fusion	Inducing Compounds	Gene Knock-out	Test Media	Result
SMb21103	Fucose	SMb21112	Fucose	NG
			Glycerol	G
SMb21376	Galactose	SMb21377	Galactose	G
	Galactitol		Galactitol	NG (r.c)
	Tagatose		Tagatose	NG (r.c)
	Sorbose		Sorbose	G
	Lyxose		Lyxose	G
			Glycerol	G
SMb20263	Hydroxyproline	SMb20263	Hydroxyproline	NG
	Allohydroxyproline		Allohydroxyproline	NG
			Glycerol	G
SMc02325	Rhamnose	SMc02325	Rhamnose	NG
			Glycerol	G
SMc01625	Erythritol	SMc01625	Erythritol	NG
	Adonitol		Adonitol	NG (r.c.)
	Sorbitol		Sorbitol	G
	Xylitol		Xylitol	G
			Glycerol	G
SMb21216	Galactosamine	SMb21218	Galactosamine	NG
	Glucosamine		Glucosamine	G
			Glycerol	G
SMb21138	Galactosamine	SMb21137	Galactosamine	NG
	Glucosamine		Glucosamine	G
			Glycerol	G
SMb21151	Galactosamine	SMb21151	Galactosamine	NG
			Glucosamine	G
			Glycerol	G
SMb20328	Maltose	SMb20328	Maltose	G
(*thuK*)	Trehalose		Trehalose	NG
	Maltitol		Maltitol	NG
	Turanose		Glycerol	G
SMc03061	Turanose	SMc03063	Turanose	SG
(*aglE*)	Maltotriose		Maltotriose	SG
	Maltose		Maltose	NG
	Sucrose		Sucrose	SG
			Glycerol	G
SMc04259	Salicin	SMc04255	Salicin*	SG
	Gentiobiose	(*manB*)	Gentiobiose	NG
	Cellobiose		Cellobiose	NG
	Dextrin		Glycerol	G
	Gluconate			

G = growth, referring to normal growth comparable to that of wild type
NG = no growth, referring to the absence of colonies or growth in growth curve.
SG = slower growth, referring to slower formation of colonies or reduced growth in growth curve.
r.c. = the presence of revertant colonies
salicin could not support growth of P110 or fusion 5679 in liquid media, but on plates, 5679 formed colonies slower than wild type.
All of the compounds in the above table were used as carbon sources at a 10 mM final concentration, except hydroxyproline and allohydroxyproline which were used at a final concentration of 5 mM.

Chapter 3-3. Discussion of Screening Results

In this study 405 integrated fusion strains were tested for induction in over 120 test conditions. Of these tested strains, 35 unique transport systems were found to be specifically induced by at least one of the test conditions.

The initial screening of the transportome was done using the automated liquid handling system, MultiProbeII (PerkinElmer). Tests from this screen were considered positive if they had an expression above three-fold compared to the expression of that strain when grown in glycerol. This was an arbitrary value that was considered sufficient and not too lenient or too strict to consider retesting. These putative positives were identified in three separate methods. First and foremost, when carrying out the reactions when a strain showed reporter enzyme activity (yellow colour) in a test media but not in another, it was recorded. Finally, every strain was analyzed visually by making a graph including all of the test media and if any strain showed three-fold induction in any media it was retested.

CHAPTER 4: RESULTS OF THE ABC AND Trap-T TRANSPORTERS

As many of the results concerning the ABC and Trap-T transporters have been published (Mauchline et al., 2006), the remaining results have been divided into two separate chapters: Chapter 4 will focus on those ABC and Trap-T transporters that we have identified inducers for and have gone on to do a more intrinsic analysis. Chapter 5 focuses on the remainder of the transport systems, where I describe the overall findings of those transport systems that were not included in the publication. These transporters include all the non-ABC and Trap-T systems, as well as those ABC and Trap-T systems that were not included in the Mauchline et al. publication (2006).

Chapter 4.1. Caffeine and Theobromine

4.1-1. Introduction

Caffeine and theobromine are alkaloid molecules that belong to a family of molecules known as methylxanthines. These two stimulants are found in coffee and chocolate, respectively. Theobromine is the primary methylxanthine found in products of the cocoa tree, *Theobroma cacao*. Caffeine is naturally produced by several plants, including coffee beans, guarana, yerba maté, cacao beans, and tea. There is a lot of interest in the topic of caffeine and related structures governing human health and how to optimize the yield of the compounds. However, little is known about the metabolism of these compounds by microorganisms.

Caffeine and theobromine were included in the screening of the transportome in hopes of finding a transporter that was induced by either of these compounds. When used as sole nitrogen sources (5mM) these compounds did not support the growth of *S. meliloti*. In fact the average optical density readings at 600 nm wavelength taken with the Tecan Safire spectrophotometer when any *S. meliloti* strain was grown in these compounds was no higher than 0.1. In a subsequent study, 0.5% glycerol was added to the media and growth comparable to that with only 0.5% glycerol was obtained (data not shown).

4.1-2. Results

A single fusion, SmFL4493 (Sma2125::*gusA*) was found to be largely induced by the presence of both caffeine and theobromine when either were used as a sole source of nitrogen (Figure 4-1). Upon further analysis it was discovered that a similar fusion, SmFL2123 (Sma2123::*lacZ*) also shows induction when the alkaloids are available as nitrogen sources (Figure 4-2). The gfp data shows similar results for this fusion (Figure 4-3). A possible explanation for the discrepancy found between the two fusions is that fusion SmFL1501 is a fusion to the very last gene of the operon and may therefore have a lower level of complete transcripts. Also, GusA is a more sensitive reporter enzyme compared to LacZ in *S. meliloti* because it has basal LacZ activity.

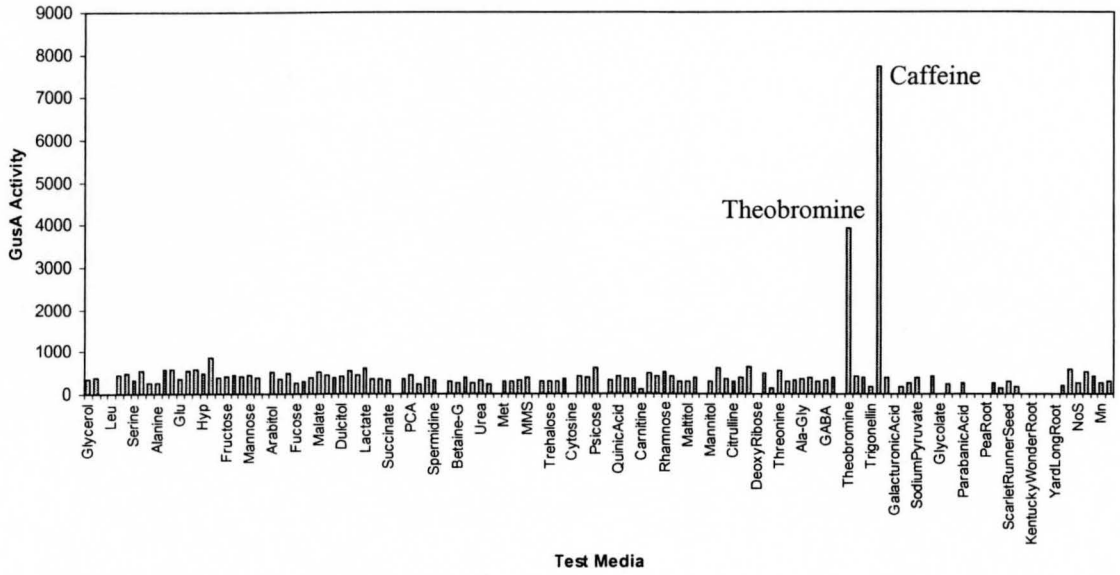

Figure 4-1. β-glucuronidase activity of SmFL4493 (SMa2125::*gusA*) when tested in all test media showing specific induction by theobromine and caffeine.

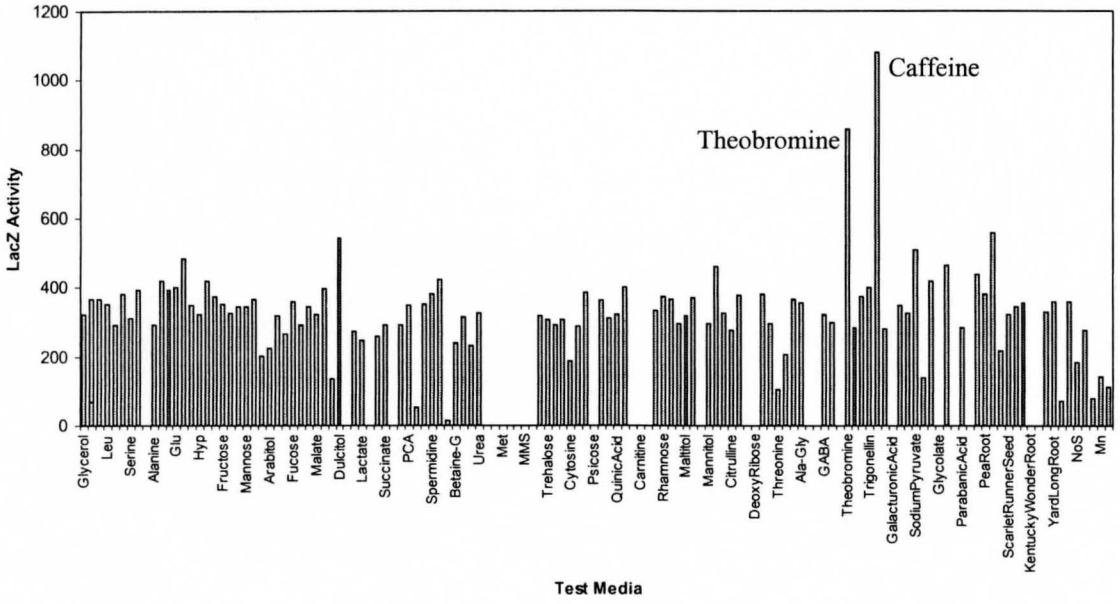

Figure 4-2. β-galactosidase activity of SmFL1501 (SMa2123::*lacZ*) when tested in all the test media showing, to a lesser extent, specific induction by theobromine and caffeine.

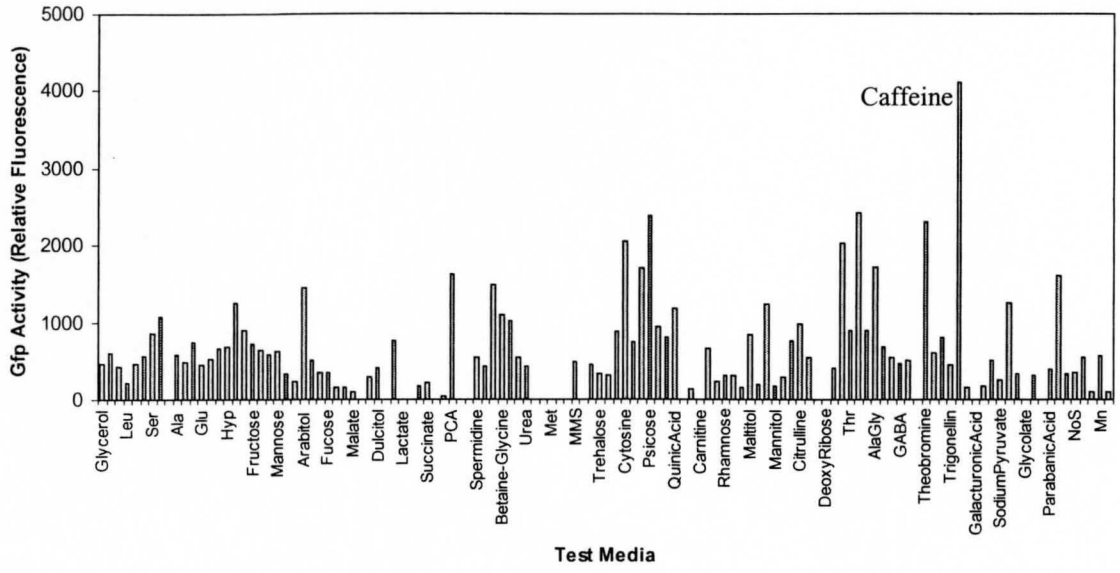

Figure 4-3. Gfp specific activity of SmFL1501 (SMa2123::*lacZ*) tested in all test media showing slight induction in caffeine compared to the other compounds.

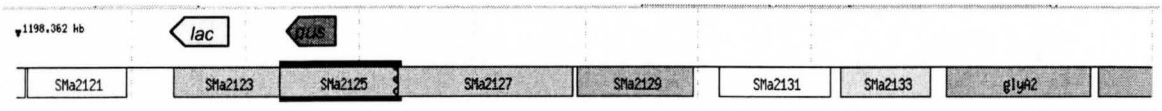

Figure 4-4. Genetic map of the operon induced specifically by caffeine and theobromine. The blue arrow indicates the location of the *gusA* fusion to SMa2125 in SmFL4493. The yellow arrow indicates the location of the *lacZ* and *gfp* fusion to SMa2123 in SmFL1501 (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

As shown in Figure 4-4, the GusA fusion to SMa2125 clearly creates a knockout of the transporter, whereas the LacZ fusion to SMa2123 disrupts the 5' end of the last gene in the operon. This is another possible reason for the discrepency in the fusion data.

In Table 4-1 below shows the retest data for both of the strains, SmFL4493 and SmFL1501, both showing induction in caffeine and theobromine.

Table 4-1. β-glucuronidase and β-galactosidase activities of SmFL4493 (SMa2125::*gusA*) and SmFL1501 (SMa2123::*lacZ*), respectively when tested in caffeine and theobromine as the sole source of nitrogen.

Gene	Fusion	Caffeine	Theobromine	Glycerol
SMa2125	SmFL4493	10407 +/- 120	5264 +/- 87	516 +/- 10
SMa2123	SmFL1501	448 +/- 19	1510 +/- 73	292 +/- 10

Chapter 4.1-3. Discussion

One transport system was induced by the stimulants caffeine and theobromine. SMa2125 and SMa2123 are the ABC permease subunits and SMc2127 is the ATPase, however there is no periplasmic binding protein. However, there is a hypothetical protein, SMa2129, which could be the periplasmic bindig protein. When analysed for homology using the amino acid sequence and the BLAST program, similarity was found with periplasmic binding proteins of other species of bacteria such as *Bordetella avium* 197N, *Roseovarius nubinhibens* ISM, *Roseobacter sp.* MED193, and *Roseovarius sp.* 217, though none of these have identified substrates.

The available data support the operation of a xanthosine → 7-methylxanthosine → 7-methylxanthine → theobromine → caffeine pathway as the major route to caffeine; the first, third and fourth steps being catalyzed by *N*-methyltransferases (NMTs) that use *S*-adenosyl-L-methionine (SAM) as the methyl donor (Kato, 2004).

Caffeine is metabolised along the same pathway in microorganisms as in humans (Madyastha et. al., 1999). Though most of the interest lies in the pathway creating caffeine and in the metabolism of caffeine by animals and humans, the use of bacteria is very useful for several reasons. First, bacteria can effectively be used as model systems for further research and understanding of the metabolic pathway. In 1998, Madyastha and Sridhar identified microbial metabolism of caffeine by a consortium of

microorganisms containing strains belonging to the genera *Klebsiella* and *Rhodococcus* (Madyastha and Sridhar, 1998). Second, microorganisms are now being used as tools to create decaffeinated beverages (Ramarethinam and Rajalakshmi, 2004). For example, Ramarethinam and Rajalakshmi (2004) have identified a *Bacillus* strain, *Bacillus licheniformis*, capable of proliferating on nutrient medium supplemented with 2% leaf extract. They have proposed the use of such bacteria in the decaffeination of tea.

Our collaborating group, led by Dr. P. Poole at Reading University, has taken interest in this finding and has further demonstrated that this cluster is also induced by theophylline, another caffeine-like substance found in black and green tea (personal communication). They have been taking the initiative in using this cluster to identify the presence of caffeine in substances. For example, the fusion strain can be used as an indicator strain to screen for any presence of a caffeine-like substance in a particular liquid. This has obvious benefits for the health organization.

Until present no transport system has been reported for such compounds, making this an exciting, novel, and applicable finding.

Chapter 4.2 β-glucoside Transport

4.2-1. Introduction

Cellulose is a β-1, 4-linked glucose polymer and represents nearly half of the dry weight of plant cell walls. When cellulose is hydrolyzed by a combination of endoglucanase and cellobiohydrolase activities, cellobiose is the primary product (Lai et al., 1997). Cellobiose is a β-glucoside with a β-1, 4 glucosidic linkage. Another

compound found naturally in plants is gentiobiose, a β-glucoside that can be hydrolysed to two β-D-glucose molecules. Arbutin and salicin are other β-glucosides that *S. meliloti* would encounter in the soil (see Figure 4-5 below for structures of all β-glucosides used in this study).

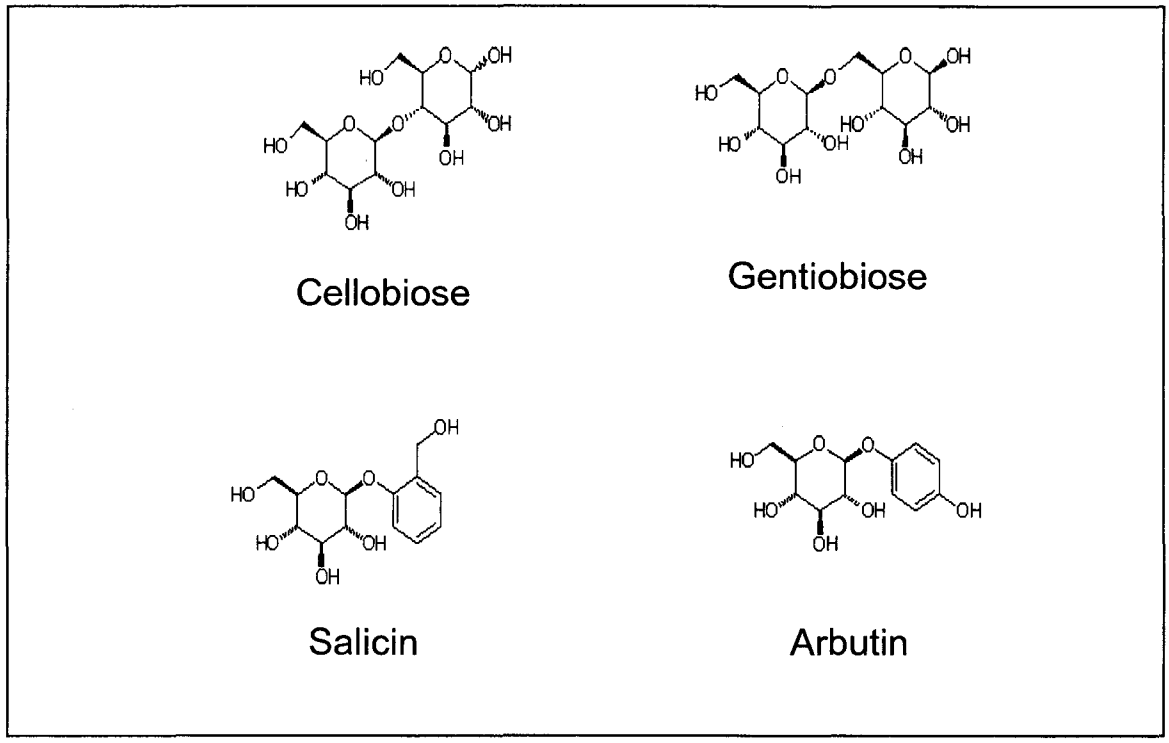

Figure 4-5. Structures of the β-glucosides used in this study. Note that salicin and arbutin are both have aromatic rings and thus are aryl- β-glucosides. (Structures taken from Sigma website).

E. coli has three cryptic operons that when activated allow the utilization of various β-glucosides. The *celABCDFG* operon allows the transport and catabolism of cellobiose, arbutin, and salicin via a PTS transporter system. It has been shown that CelF is responsible for the hydrolysis of β-glucosides, including cellobiose-6P, salicin-6P, and arbutin-6P (Kricker and Hall, 1984). The other cryptic operon, *bglBC* and *bglA* (a separately transcribed arbutin-specific phospho-β-glucosidase A) is a PTS system that transports and metabolizes such aryl β-glucosides as arbutin and salicin (Schaefler et al.,

1967). Furthermore, *arbT* is another cryptic locus that allows for the transport of arbutin (Kricker and Hall, 1987). Similar operons are found in *Streptococcus mutans* and they have been found to be subject to glucose repression (Old et al., 2006).

There have also been reports of ABC-type transport systems importing various β-glucosides into the cell. For example a high-affinity ABC transport system has been identified in *Pyrococcus furiosus*, a hyperthermophilic Archaeon, which transports cellobiose. The periplasmic binding protein was purified and shown to bind not only to cellobiose, but also to cellotriose, cellotetraose, cellopentaose, laminaribinose, laminaritriose, and sophorose, all of which are β-glucoside polymers (Koning et al., 2001). There have also been reports of such similar systems in bacteria. In *Streptomyces reticuli* an ABC-type transport system has been reported of transporting cellobiose and cellotriose (Schlösser et al., 1999).

Cellobiose and other β-glucosides are metabolised by a β-glucosidase into glucose monomers. In *E. coli* the expression of the β-glucosidase gene (and the permease) is better induced by the aryl β-glucosides, arbutin and saslicin (Schlösser et al., 1999), however the opposite is found in such organisms as *Rhodotorula minute* (Duerksen and Halvorson, 1958).

SMc04257, the permease of an ABC transport system, was found to be induced by cellobiose, gentiobiose, gluconate, dextrin, salicin and arbutin, as shown in Table 4-2. A knock-out of the permease, strain SmFL6588 (Smc04257::*gusA*), was unable to grow on gentiobiose, cellobiose and arbutin. Also, SmFL5679 (*manB*::*lacZ*) a knock-out strain of a metabolism gene (*manB*) in this cluster, was found to be unable to grow on gentiobiose and cellobiose. *manB* is annotated as a β-mannosidase, however it is

probably a β-glucosidase since this is the enzyme responsible for catalysing the hydrolysis of terminal non-reducing residues in β-D-glucosides releasing β-glucose as a product.

Chapter 4.2-2. Results

A single transport gene, SMc04251 lies downstream of the putative β-glucoside transport system and is slightly induced by arbutin and salicin. This gene is annotated as a mannitol-binding protein and is not included in the transport DB classification. Other surrounding genes were tested for induction by these compounds. SMc04248, is a hypothetical conserved transmembrane protein located upstream of the transport system, but was not found to be induced by the tested compounds. SMc04247 is annotated as a metabolism gene located upstream of the transport system but was also not found to be induced by the tested compounds. Finally, SMc04254 is a hypothetical conserved protein located in the operon downstream of *manB* and this gene was found to be highly induced by all of the β-glucosides included in this study. Figures 4-6 and 4-7 show the results from the initial screening of SmFL1580 (SMc04259::*gusA*) and the genetic map of the operon in which it lies.

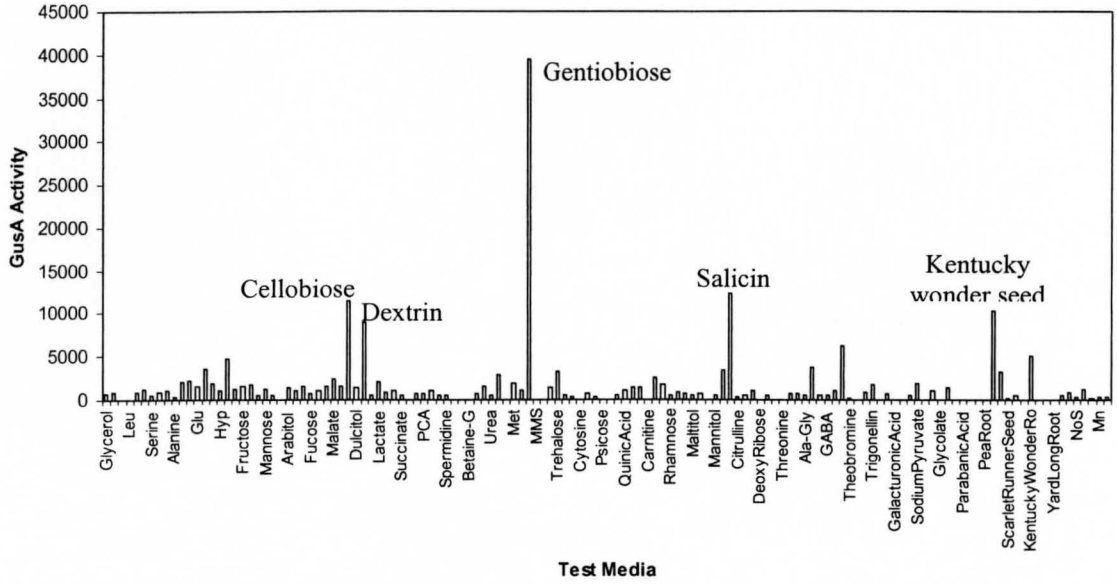

Figure 4-6. β-glucuronidase assay of SmFL1580 (SMc04259::*gusA*) tested for induction in all the test media, showing specific induction in cellobiose, dextrin, gentiobiose, salicin, gluconate, and kentucky wonder seed. The retest data for kentucky wonder seed proved to be negative (data not shown).

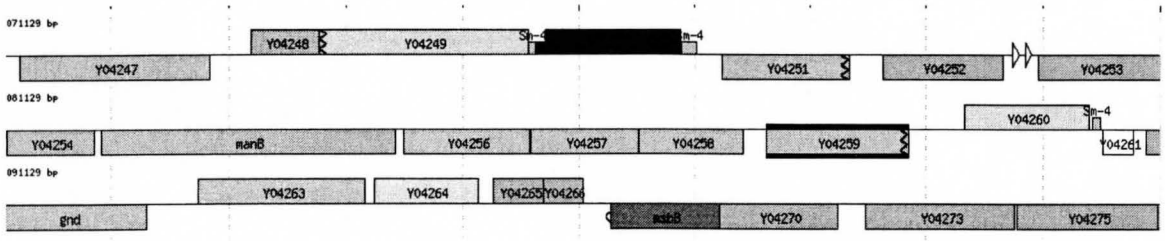

Figure 4-7. Gene map of the operon induced by β-glucosides and the surrounding genes (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

To verify that this operon was induced by various β-glucosides, this fusion, along with several fusions to other genes in potentially the same operon were taken from the pTH1522 library, and retested in the test media. Arbutin was not included as an initial test compound but since it is a β-glucoside and it is known to induce other β-glucoside transporters (Kricker and Hall, 1984), it was included in all the proceding tests. The results from this screening are shown below in Table 4-2.

Table 4-2. β-glucuronidase and β-galactosidase acitivities of different library fusions to the operon induced by β-glucosides and surrounding genes that may also be involved in the transport and metabolism of β-glucosides.

	Gentiobiose	Cellobiose	Salicin	Arbutin	Dextrin	Gluconate	Glycerol
sm04247	770+/-210	789+/-24	783+/-14	1135+/-102	1082+/-12	1324+/-102	522+/-8
smc04248	776+/-27	660+/-8	952+/-13	1535+/-132	1141+/-761	1441+/-8	499+/-8
smc04251	1706+/-57	1825+/-31	2787+/-32	3148+/-182	908+/-94	3696+/-72	762+/-24
smc04254	1760+/-15	12824+/-148	4976+/-159	7837+/-84	5713+/-201	14308+/-463	525+/-22
manB	723+/-216	5742+/-2446	1093+/-97	726+/-58	1694+/-249	3280+/-67	324+/-13
smc04257	2571+/-111	7987+/-109	4203+/-106	2505+/-92	1739+/-119	4431+/-194	198+/-2
smc04259	6679+/-38	6282+/-67	29183+/-117	32355+/-257	12650+/-292	38988+/-358	1106+/-10
smc04260	3147+/-13	2816+/-25	3017+/-19	3311+/-26	1199+/-82	3310+/-128	767+/-4

We were concerned that poor growth and the resulting low optical densities following growth on salicin, dextrin, or gluconate as carbon sources could result in artificial enzyme activity. Also, we were curious on whether induction would still be observed when the strains were grown in cellobiose, gentiobiose, and arbutin even if glycerol was present in the media. Therefore, 0.5% glycerol was added to the test media to i) give the strain the ability to grow to a reasonable optical density (i.e. above 0.1 Abs_{600} using the Tecan plate reader) to ensure that a low O.D. reading does not skew the results and ii) observe the effects of its presence on the inducing ability of the β-glucosides. Another interest lies in the possibility that the presence of 15 mM succinate in the media will cause repression of the system. Thus, to address this possibility, SmFL1580 (smc04259::*gusA*) was tested for induction in the inducing compound alone,

the inducing compound supplemented with 0.5% glycerol, and the inducing compound supplemented with 0.5% glycerol as well as 15 mM succinate.

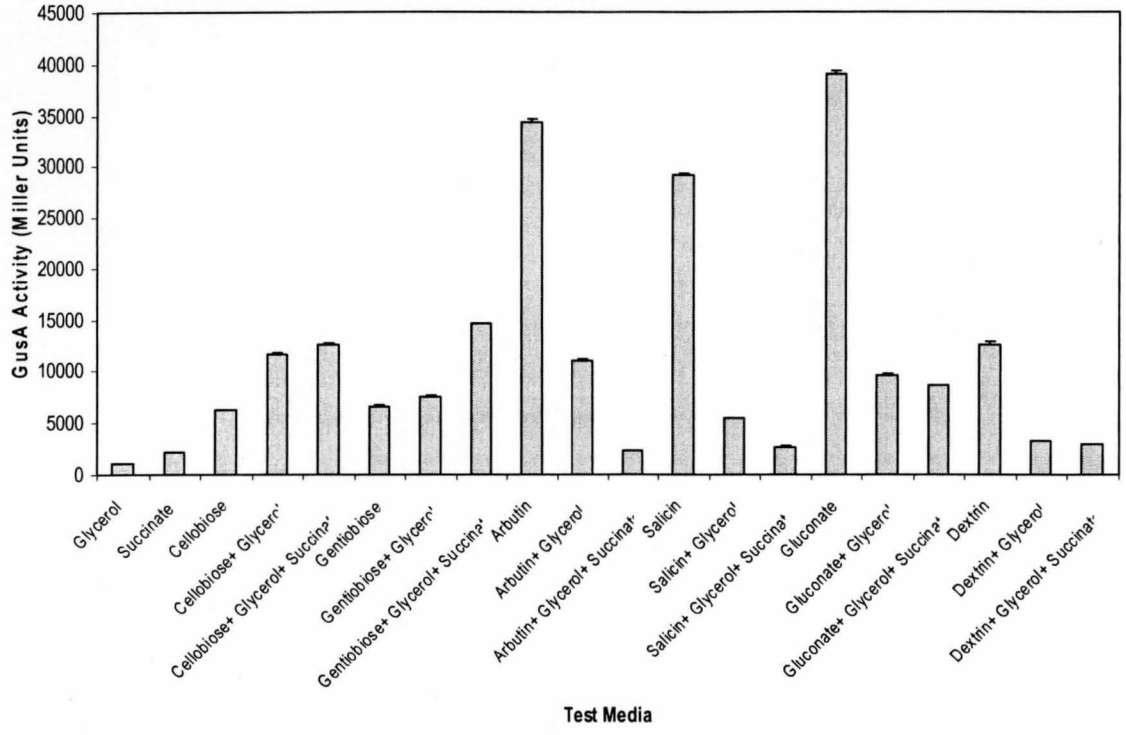

Figure 4-8. β-glucuronidase activity of SmFL1580 (SMc04259::*gusA*) when grown in the indicated test media to test for the effect of low optical density as well as the presence of succinate.

These results show that when glycerol is present in the test media at 0.5%, similar induction is still present from the β-glucosides as when glycerol is not present. Induction in dextrin and gluconate is still present, although at a much lower level, approximately 4-fold decrease in both. Furthermore, succinate repression is seen when present in the media only with the aryl-β-glucosides, arbutin (nearly 5-fold decrease) and salicin (2-fold decrease).

The location of the promoters had not been determined for these genes and whether or not the metabolism genes belong in the same operon as the transporter was

unknown. It was also unknown whether the periplasmic binding protein, SMc04259, was under the control of the same promoter as the three other transport genes, SMc04258, SMc04257, and SMc04256. In an attempt to locate the number and location of promoters involved in this system, several strains were built involving the replicating plasmid pTH1582. In this system, the upstream regions of genes with potential promoters were cloned upstream of a promoter-less *gusA* gene. The promoter regions that were cloned for the analysis are given below in Table 4-3. The results from these experiments are shown in Figure 4-9 below.

Table 4-3. The 5' and 3' ends of the upstream regions of the indicated genes cloned to investigate the presence and regulation of potential promoters. SMc04260 was not included in this experiment but was in the following experiment depicted in Figure 4-10.

	SMc04247	SMc04248	SMc04251	SMc04252
5'	2073271	2073050	2078651	2080066
3'	2072704	2073525	2078052	2079501

	SMc04258	SMc04259	SMc04260	*gnd*
5'	2087894	2089183	2089213	2092727
3'	2087446	2088668	2089786	2092057

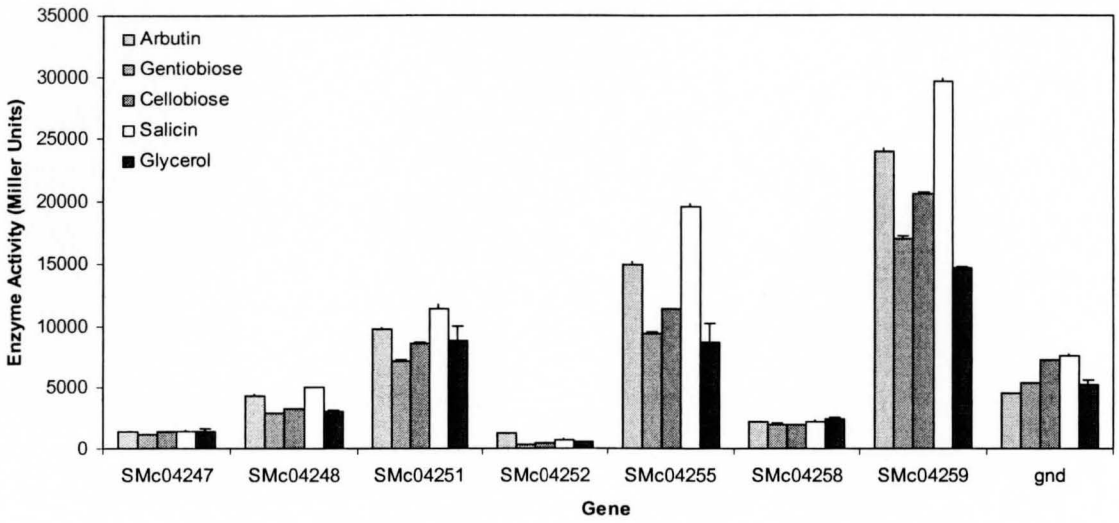

Figure 4-9. β-glucuronidase and β-galactosidase activities of strains of *S. meliloti* with the replicating plasmid pTH1582 containing upstream regions of the specified genes.

Unexpectedly, fusion SMc04259 no longer shows specific induction by the β-glucosides and it should since this particular promoter region is what is cloned into library fusion SmFL1580, which shows specific induction in those test media. For this reason another approach was taken using a plasmid cointegrant system, in which the low copy number plasmid, pTH1508, is integrated with pTH1703 plasmids via the *Streptomyces* ϕC31 *attP/attB* sequences. The same upstream regions were cloned as shown is Table 4-3 above. The results from this experiment are shown below in Figure 4-10.

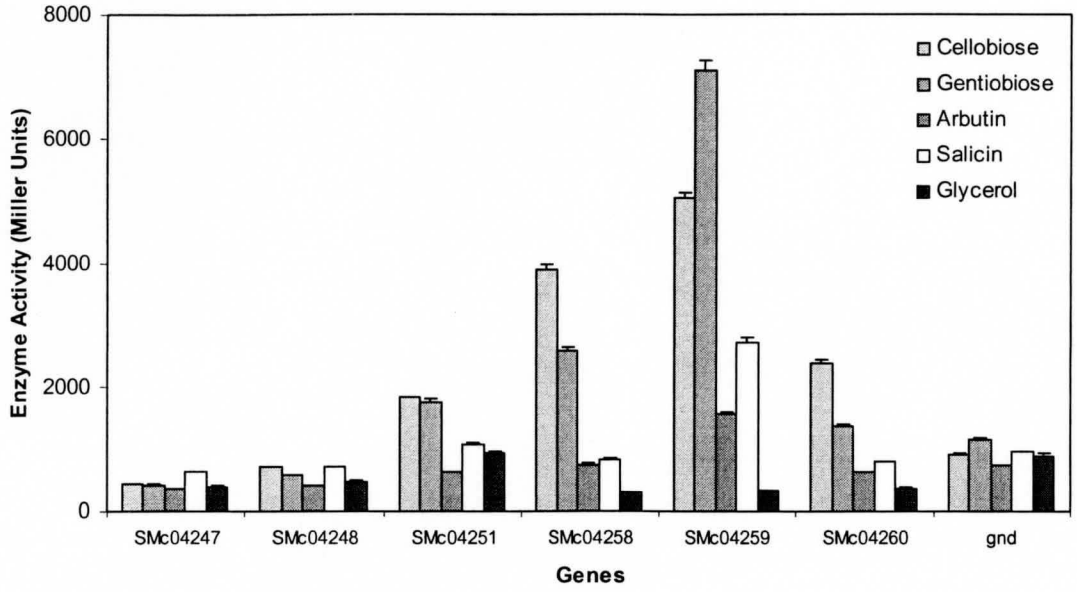

Figure 4-10. β-glucuronidase and β-galactosidase activity of the cointigrant replicating plasmids in *S. meliloti* containing fusions to the indicated genes.

From this experiment it looks as though there is a promoter driving the expression of the periplasmic binding protein (SMc04259) and an additional promoter just downstream driving expression of the remainder of the operon. The promoter upstream of the putative regulator, SMc04260, also shows induction by the β-glucosides. Furthermore, there appears to be a separate promoter for SMc04251, the annotated mannitol binding protein, but it does not appear to be induced by the tested compounds. The same situation of no induction is found with SMc04248 and *gnd*.

The metabolism gene, *manB*, is annotated as a β-mannosidase but it seems likely that the gene product would actually have β-glucosidase activity, as mentioned earlier. To explore this possibility a crude cell extract was prepared from 2 litres each of SmP110 grown in minimal media supplemented with either cellobiose, salicin, or glycerol as the sole source of carbon. Following a protocol adapted from Wulff-Strobel and Wilson

(1995) (also see Materials and Methods), β-glucosidase activity was measured in crude cell extracts of SmP110 in an assay that employed p-nitrophenyl-β-D-glucopyranoside as the substrate. It was interesting to find that activity was present not only in cells grown in minimal media supplemented with cellobiose or salicin as the sole source of carbon, but also when glycerol was used as a sole carbon source (Table 4-4). The assays were carried out at three different temperatures; room temperature, 30^0C, and 37^0C (see Appendix for Bradford data). The highest activity was measured when the reactions were allowed to take place at 37^0C. A deletion mutant of *manB* may be constructed to demonstrate loss of β-glucosidase activity.

Table 4-4. β-glucosidase specific activity of crude cell lysate of *S. meliloti* grown in M9-minimal media with either glycerol, cellobiose, or salicin as the sole source of carbon tested at room temperature, 30^0C, and 37^0C.

	Room Temperature	30 ^{0}C	37 ^{0}C
Glycerol	297 +/- 20	493 +/- 27	527 +/- 45
Cellobiose	286 +/- 53	524 +/- 26	574 +/- 18
Salicin	272 +/- 25	418 +/- 10	453 +/- 26

It was also investigated whether the putative *S. meliloti* transporter and associated metabolism genes would be able to allow for *E. coli* to utilize β-glucosides if the genes were introduced to a wildtype *E. coli* strain. By manipulating the *flp* recombinase system refined by Branka Paduska, the entire *S. meliloti* transport system and associated metabolism genes were transferred to *E. coli* DH5α strain M928. The resultant strain, M1223, was tested for the ability to utilize cellobiose, gentiobiose, salicin, and arbutin as sole sources of carbon but was found still unable to do so. Growth curves in liquid broth were not carried out since the strain was not even able to grow on plates.

Chapter 4.2-3. Discussion

Cellulose is present in nature almost exclusively in plant cell walls with some animals (tunicates) and a few bacteria (*Acetobacter xylinum*) also containing the compound. Rarely is cellulose found in a pure form in nature (except cotton bolls), rather it is found in a crystaline structure embedded in a matrix of lignin and hemicellulose, which are both structural biopolymers (Lynd et al., 2002).

The metabolism of cellulose occurs outside of the cell by secreted metabolic gene products. Three major types of enzymatic activities found are: (i) endoglucanases or 1,4-β-D-glucan-4-glucanohydrolases, (ii) exoglucanases, including 1,4-β-D-glucan glucanohydrolases (also known as cellodextrinases) and 1,4-β-D-glucan cellobiohydrolases (cellobiohydrolases), and (iii) β-glucosidases or β-glucoside glucohydrolases. Endoglucanases cleave the cellulose polysaccharide chain at random at internal amorphous sites, generating oligosaccharides of various lengths and consequently new chain endings. Exoglucanases act in a processive manner on the reducing or nonreducing ends of cellulose polysaccharide chains, liberating either glucose (glucanohydrolases) or cellobiose (cellobiohydrolase) as major products. Exoglucanases can also act on microcrystalline cellulose, presumably peeling cellulose chains from the microcrystalline structure (Lynd et al., 2002). The resulting β-glucoside cellobiose is then transported into the cell where further metabolism occurs by β-glucosidase.

Other β-glucosides such as gentiobiose, salicin, and arbutin are expected to be found in the soil as they are commonly found in plants. For example, salicin is found in high abundance in willow bark and arbutin is a component found in bearberry plant

leaves. It is therefore no surprise that all these compounds are often transported into the cell by the same transport system (Schaefler, 1967, Kricker and Hall, 1984). This seems sensible as they all share the common structure of having a β-linkage to a glucose monomer (see Figure 4-5 above). As mentioned in the results section, this is essentially the case with the two cryptic *E. coli* PTS transporters (Schaefler, 1967, Kricker and Hall, 1984). Also, *Pyrococcus furiosus* has an ABC-type transport system that transports a wide range of β-glucosides (Koning et al., 2001). In some species of bacteria, only cellobiose and cellotriose are transported by a single transporter. For example *Streptomyces reticuli* possesses an ABC-type transport system that imports only cellobiose and cellotriose into the cell (Schlosser et al., 1999). Some bacteria, such as *Agrobacterium tumefacians*, have a known β-glucosidase but a transport system has not yet been identified (Watt et al., 1998).

The operon in *S. meliloti* found to be induced by cellobiose, gentiobiose, salicin, and arbutin is an ABC-type transport system and is located adjacent to the putative metabolism genes (Figure 4-7). SMc04259 is the periplasmic binding protein, suggested by transport DB as having a sugar as a substrate. SMc04258 and SMc04257 are the permease subunits and SMc04256 is the ATPase, all of which are components necessary and sufficient to make up an ABC-type transport system. Just downstream, only 72 nucleotides, lies *manB* an annotated β-mannosidase and located 56 nucleotides downstream from that is a hypothetical conserved gene, SMc04254, which is highly induced by all the test β-glucosides. SMc04253 and SMc04252 are both annotated as oxidoreductases and SMc04253 seems as though it may be located in the operon with the other metabolism genes and the ABC-transporter. SMc04253 does not have any data, as

there is no fusion to this gene in the library, and SMc04252 was found to not be induced by any of the test compounds.

Salicin, dextrin, and gluconate do not support good growth of *S. meliloti* (giving an $O.D._{600}$ of less than 0.1 as read by the Tecan Safire Microtiter Plate Reader) so a further investigation was carried out by adding 0.5% glycerol to the test media. From this experiment it was found that these media do indeed cause induction of the transport system, although not to the extent that it was noticed initially (Table 4-2 and Figure 4-8). This test also showed that the inducing compounds did still induce the transport system even when glycerol was present suggesting that these β-glucosides are a preferred carbon source.

Furthermore, the transport system showed some succinate repression in the presence of 15 mM succinate as well as 0.5% glycerol. Interestingly expression of the system was repressed only when succinate was added with the aryl-β-glucosides, arbutin and salicin. One possibility for the difference between cellobiose, gentiobiose and the aryl-β-glucosides is that *S. meliloti* prefers cellobiose and gentiobiose as a carbon source over the aryl-β-glucosides and succinate. The opical densities of the liquid cultures after 30 hours of incubation is much higher in cellobiose (0.16 in the Tecan Safire) and gentiobiose (0.19) than in arbutin (0.06) and salicin (0.08), yet it is comparable to that of succinate (0.17). However, as shown in Figure 4-8, when the optical densities are comparable (0.5% glycerol present in the media), the induction pattern is quite similar for all the β-glucosides. It seems that if cellobiose and gentiobiose were preferred substrates they would cause a higher induction in the transport system.

A single transport gene, SMc04251 lies downstream of the transport system and associated metabolism genes (Figure 4-7 genetic map) and is slightly induced in arbutin and salicin. This gene is annotated as a mannitol-binding protein and is not included in the transport DB classification. SMc04248, is a hypothetical conserved transmembrane protein located upstream of the transport system, but was not found to be induced by the tested compounds. SMc04247 a metabolism gene located upstream of the transport system was also not found to be induced by the tested compounds.

When homology of these genes is searched using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) the overall appearance of the operon appears to be conserved among several closely related α-proteobacteria including *Agrobacterium tumefaciens*, *Rhizobium etli*, *Mesorhizobium loti*, *Rhizobium etli*, and *Oceanicola granulosus*.

The location of putative promoters was investigated as it seemed likely that the transporter and the downstream *manB*, the hypothetical protein SMc04254, and first oxidoreductase SMc04253, were part of one operon (Figure 4-7 genetic map). It also was possible that the periplasmic binding protein was under the control of its own promoter and a second promoter was located downstream controlling the expression of the remainder of the operon.

By creating a cointegrant replicating plasmid with the promoter region of interest and testing the expression of such plasmids in the test β-glucosides, three promoters were identified. The first obviously being upstream of the periplasmic binding protein, and having the highest expression level. The second promoter was identified downstream of the periplasmic binding protein and evidently controls at least some of the expression of

the rest of the operon. The third promoter is expressed in the opposite orientation of that of the periplasmic binding protein and is just upstream of a single gene annotated as a LacI-type transcriptional regulator SMc04260. Clearly this has potential to be the regulator for the transporter and metabolism genes.

A knock-out mutant of the ABC-transporter and a *manB* mutant both failed to grow on any of the β-glucosides as a sole source of carbon. This strongly suggests that *manB* is a/the metabolism gene necessary for the use of the β-glucosides as a carbon source. It also suggests that this ABC-transporter is transporting the test β-glucosides but it is the only transporter that is capable of transporting these β-glucosides.

Interestingly, there is a metabolism gene associated with the ABC-transport system of a *Pyrococcus furiosus* that is annotated as being a β-mannosidase, but this too may also be an actual β-glucosidase. The physiological role of this gene product is unclear (Bauer et al., 1996).

As mentioned earlier, *E. coli* has cryptic genes but is unable to utilize any β-glucosides. The ability of the entire *S. meliloti* β-glucoside transport system and associated metabolism genes to be sufficient for *E. coli* to utilize β-glucosides was investigated. The *S. meliloti* promoter for the operon was tested for activity in *E. coli* and was found to be functional. Using the *flp* recombinase system modified by Podusk, B. (unpublished data), the system, including the region from SMc04260 to SMc04251, was transferred to *E. coli* but the bacterium was still found to be unable to utilize the test β-glucosides.

Chapter 4.3 Dextrin

Chapter 4.3-1 Introduction

Starch is a common molecule used widely as a storage polysaccharide (Ball and Morell, 2003). It is made of two distinct polysaccharide fractions: amylopectin and amylose. Amylose is a linear molecule of (1→4) linked α-D -glucopyranosyl units. Amylopectin is the highly branched component of starch: it is formed through chains of α-D -glucopyranosyl residues linked together mainly by (1→4) linkages but with 5–6% of (1→6) bonds at the branch points (Buleon et al., 1998). Dextrin is a compound where any one of a number of carbohydrates having the same general formula as starch but a smaller and less complex molecule. α-dextrin is made up of several glucose units joined by an α-1,6 linkage in addition to α-1,4 linkages. Dextrin is hydrolyzed to glucose by α-dextrinase and γ-amylase. Limit dextrinase (LD) releases straight chain dextrins from amylopectin-derived branched dextrins (Stahl et al., 2004).

Chapter 4.3-2. Results

A single fusion strain, SmFL6315 (SMc04393::*gusA*), was found to be specifically induced when 10 mM dextrin was present in the test media as the sole source of carbon (Figure 4-11).

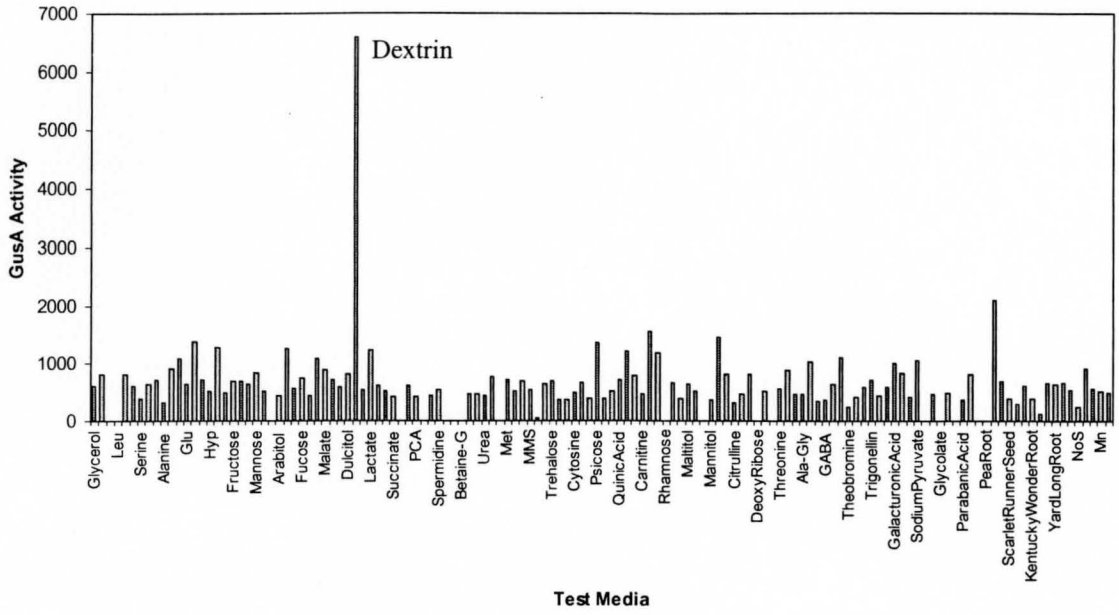

Figure 4-11. β-glucuronidase activity of SmFL39 (SMc04393::*gusA*) grown in all the different test media and showing specific induction when grown in dextrin.

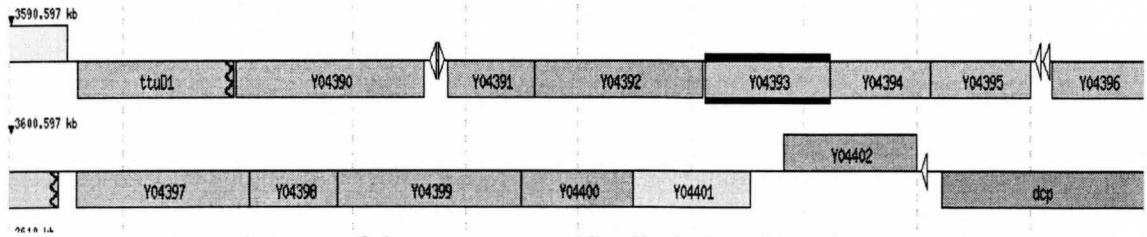

Figure 4-12. Genetic map of the operon specifically induced by dextrin and the surrounding metabolism genes (http://bioinfo.genopoletoulouse.prd.fr/annotation/iANT/bacteria/rhime/).

SMc04393 lies within what appears to be an operon with several transport and metabolism genes (Figure 4-12). In further studies, this fusion and several fusions to other genes in the same operon and surrounding operons were tested for induction in media with dextrin as the sole carbon source. SmFL6315 is a lacZ fusion to SMc04401, which is annotated as a LacI transcriptional regulator. SmFL4635 is a lacZ fusion to SMc04399, annotated as an acyl-CoA transferase. SmFL2208 is a lacZ fusion to

SMc04398, annotated as an enoyl-CoA transferase. SmFL1108 is a gusA fusion to SMc04397, annotated as a NADP-dependent L-sorbosone dehydrogenase. SmFL4583 is a gusA fusion to SMc04396, annotated as a periplasmic binding protein of the ABC-type transporters. SmFL39 is a gusA fusion to SMc04393, annotated as an ATP-binding protein. The other two transport genes in the operon consist of two permeases. SmFL401 is a gusA fusion to SMc04392, annotated as a dehydrogenase. SmFL5009 is a lacZ fusion to SMc04390, annotated as a FAD-dependent L-sorbose dehydrogenase. SmFL2336 is a gusA fusion to *ttuD1*, annotated as a hydroxypyruvate reductase. Figure 4-13 below shows the β-glucuronidase and β-galactosidase activities of these fusions when tested for induction when 0.2% dextrin is used as the sole source of carbon.

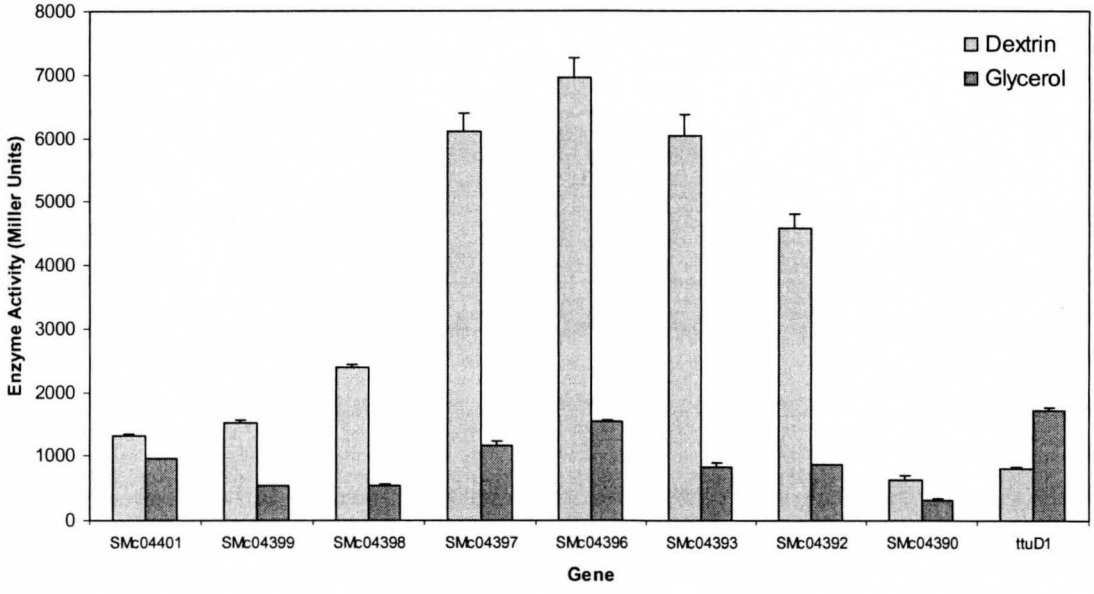

Figure 4-13. β-glucuronidase and β-galactosidase activity of various fusions to the indicated genes when grown in dextrin (0.2%) or glycerol (0.5%) as the sole source of carbon.

Dextrin is not a single compound, rather it is a mixture of similar molecules of varying length. Thus, to further elucidate the nature of this transporter, the dextrin

mixture was separated using an Amicon Ultra-4 Cellulose 10,000 MW cutoff membrane cartridge. Two different fractions were obtained; one with compounds greater than 10 000 dalton MW and the other with compounds less than 10,000 dalton MW (work done by Dr. Summers). A blank (ddH_20) was also run through the column as a negative control and an unfractionated sample of dextrin was used as a positive control. Those fusions that were not specifically induced by dextrin were not included in this particular test. The results from this experiment are shown in Figure 4-14 below.

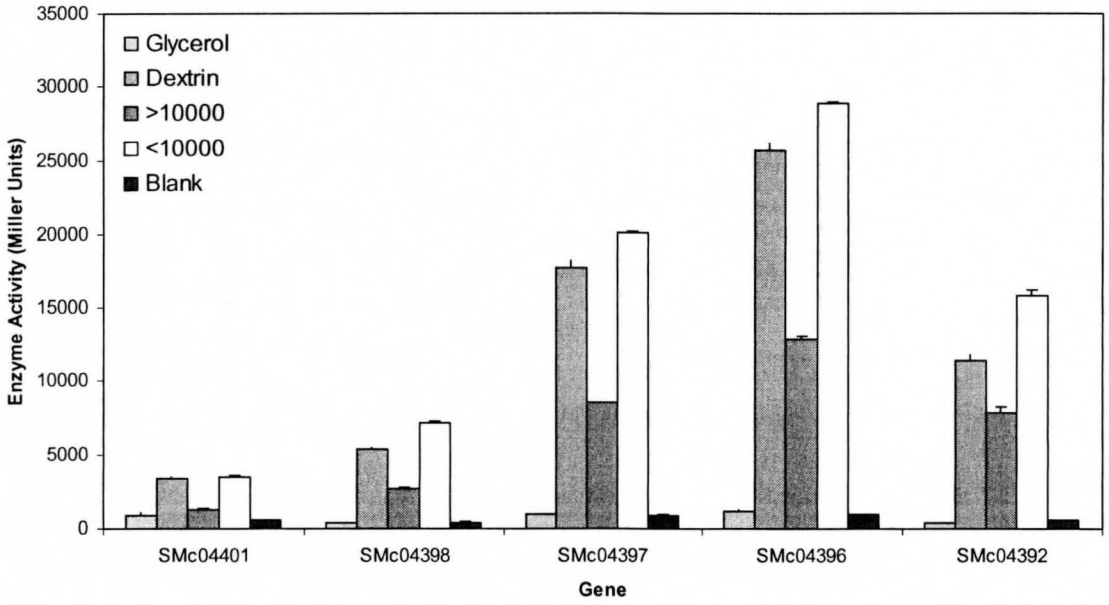

Figure 4-14. β-glucuronidase and β-galactosidase activity of the indicated gene fusions when grown in glycerol, dextrin, or fractions of dextrin as a sole source of carbon. The blank was the ddH_2O that was run through the column as a control. In this case, 0.5% glycerol was added to the media to ensure growth.

Results indicate that dextrin fraction containing molecules smaller than 10 000 dalton was a better inducer. This is probably due to the fact that the dextrin compound used in the study is not pure and a small molecule is the actual inducer.

Chapter 4.3-3. Discussion

One transport system was found to be specifically induced by the starch-like substance, dextrin (Figure 4-11). This cluster is a relatively large group of genes with an ABC-type transporter, a LacI transcriptional regulator and several metabolism genes. All these genes may be part of the same operon but it would not be unlikely if there was at least one more promoter involved in the expression of this system (see Figure 4-12 for genetic map). The first gene, SMc04401, is the regulator and it does not show specific induction by dextrin as a sole source of carbon. The next four genes are metabolism genes; SMc04400 is an oxidoreductase that was not tested because there is no fusion in the library (and it is clearly included in the operon with the surrounding metabolism genes so a fusion was not built), SMc04399 is annotated as an acyl-coA transferase and it was found to be induced three-fold over the glycerol value, SMc04398 is annotated as an enoyl-coA hydratase and was induced four-fold, SMc04397 is annotated as an L-sorbosone dehydrogenase (NADP-dependent) and was induced five-fold. The next genes make up the ABC-type transport system with SMc04396 being the periplasmic binding protein, which was found to be induced five-fold; SMc04395 and SMc04394 are both permease subunits but were not tested for induction due to the absence of relevant fusions in the library (and it is clearly in an operon with SMc04393), SMc04393 is an ATPase and was the fusion that was initially tested and found to be induced seven-fold over glycerol. The next two genes that are most probably included in the operon are SMc04392, a dehydrogenase found to be induced five-fold and SMc04391, an oxidoreductase that was not tested due to lack of fusions in the library. The last two genes are probably not functionally associated with the previous genes and seem to be

dissociated from the operon as well; SMc04390 is an L-sorbose dehydrogenase (FAD-dependent) that was less than two-fold induced, and SMc04389 (*ttuD1*) is a hydroxypyruvate reductase that was 2.5-fold induced in glycerol over dextrin as the carbon source. For this reason, it seems likely that the last two genes are not associated with the transport and metabolism of dextrin, whereas the other tested genes are.

According to the KEGG website, there are several metabolism genes involved in the breakdown of dextrin. First, dextrin is released from strach through the activity of α-amylase, which acts on starch, glycogen and related polysaccharides and oligosaccharides in a random manner; reducing groups are liberated in the alpha-configuration. (The term 'alpha' relates to the initial anomeric configuration of the free sugar group released and not to the configuration of the linkage hydrolysed). The next two steps involve glucose amylase, which performs the hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose, and α-limit dextrinase which hydrolyses the 1,6-alpha-D-glucosidic linkages in some oligosaccharides produced from starch and glycogen alpha-amylase (such as dextrin), and in isomaltose (http://www.genome.jp/dbget-bin/show_pathway?map00500+C00721).

Concern in the purity of the sample of dextrin used led to the use of an Amicon Ultra-4 Cellulose 10,000 MW cutoff membrane cartridge to separate the molecules larger than 10 000 MW from those equal to or less than 10 000 MW. Interestingly the samples with compounds less than 10 000 MW caused a greater induction. This could be a clear example where the sample has a lot of glucose present and the transporter may be transporting glucose as well or glucose di- and tri-peptides. However, in the initial screen

such compounds (e.g. D-glucose and maltose) were used and did not cause any induction at all. Yet there is a great possibility that some small compound in the mixture that was not present in the screen.

The fusion strain SmFL39 (SMc04393::*gusA*) that was used in the initial screen and found to be induced by dextrin is a knockout strain. Clearly a knockout of this transporter does not create a complete knockout phenotype, as it is able to grow in minimal media with dextrin as the sole source of carbon to the same optical density as wildtype strain SmP110. However, the growth of wildtype *S. meliloti*, SmP110, in dextrin as a sole source of carbon is very poor so the investigation into such growth phenotypes is challenging.

Chapter 4.4. Choline and Glycine Betaine Transport

Chapter 4.4-1. Introduction

Choline is a precursor to phosphatidyl choline, which is an important component of cell membranes in *S. meliloti*. It is also oxidized, via choline oxidase, to form the important osmoprotectant glycine betaine (Dupont et al., 2004). Many bacteria respond to high-salt stress by either importing compatible solutes from the external environment or synthesizing osmoprotectants internally. Such compatible solutes confer protection against the deleterious effects of low water activity, aid to maintain the appropriate cell volume, and protect intracellular macromolecules from the effects of high salt (Boncompagni et al., 1999). Though choline itself has been shown to lack

osmoprotectant capabilities, glycine betaine has been shown to be a very useful one in *S. meliloti* (Pocard et al. 1997).

In *S. meliloti opuCBA* make up a putative glycine betaine ABC-type transport system (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/). This system has been identified and shown to transport choline and is expressed in bacteroids (Dupont et. al., 2004). Interestingly this operon is annotated as an *opuCBA* transport system, after the glycine betaine ABC transporter of *E. coli*, but in the literature it is referred to as *choXWV* (Dupont et al., 2004). *choX* (*opuC*) is the periplasmic binding protein, *choW* (*opuB*) is the permease, and *choV* (*opuA*) is the ATP-binding protein. Results from Dupont et al. (2004) showed that this operon was induced specifically by choline and not other betaines or acetylcholine or salt stress (Dupont et al., 2004).

Chapter 4.4-2 Results

From the initial screen, two separate transport systems were found to be induced by choline and glycine betaine when present as the sole source of nitrogen; RmP214 (SMb20571::*gusA*) and SmFL2829 (SMc02344::*lacZ*). The results from the initial screen showing induction of these two fusions are shown in Figures 4-15 and 4-18 below. The Poole group and other studies have used choline and glycine betaine as carbon sources but in this study growth of *S. meliloti* on these betaines as carbon sources was not achieved.

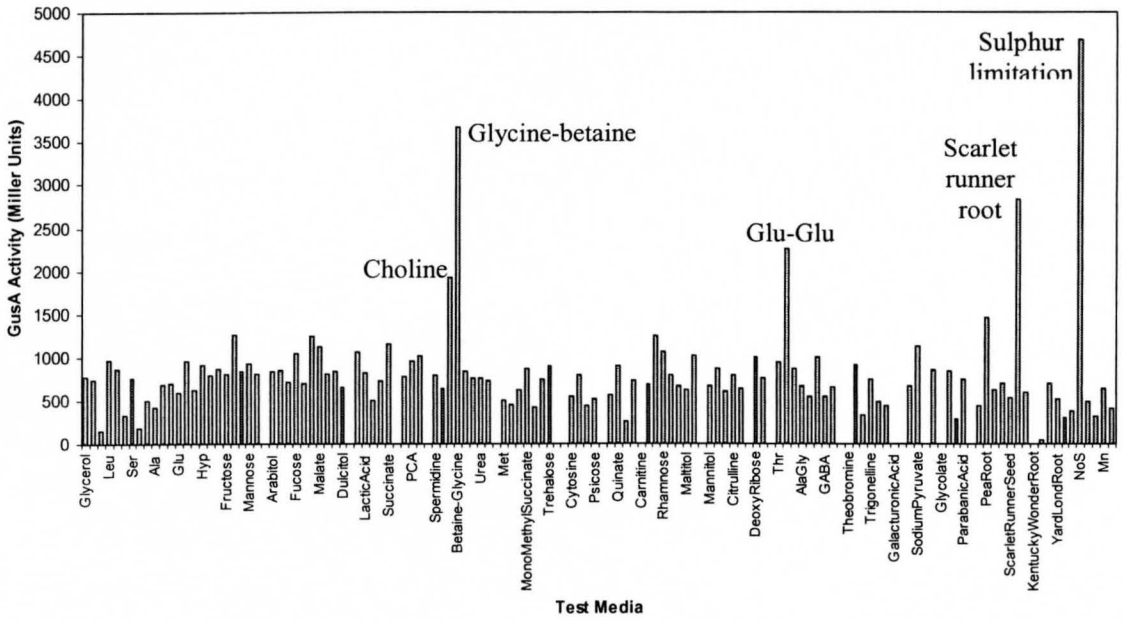

Figure 4-15. β-glucuronidase activity of RmP214 (SMb20571::*gusA*) grown in all the test media, showing induction in several different test conditions including choline, glycine betaine, the dipeptide Glu-Glu, Scarlet runner root (SRR) exudates, and sulphur limitation (no sulphur source added to the media).

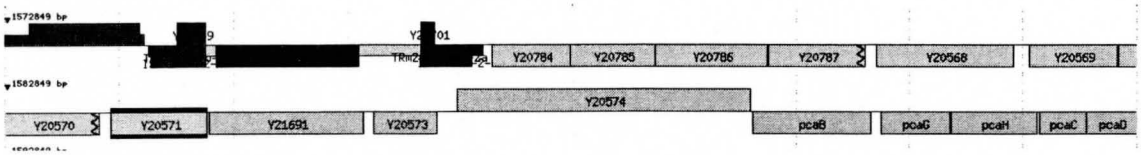

Figure 4-16. Genetic map of SMb20571 and surrounding genes (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

RmP214 (SMb20571::*gusA*) was retested in all the above potential inducers and was found only to be induced by choline, glycine betaine, and sulphur starvation (as achieved by not supplementing the media with a sulphur source) as shown in Figure 4-17. SmFL2829 (SMc02344::*lacZ*) was retested in glycine betaine, choline, and sulphur starvation and found to be induced by glycine betaine as well as choline and sulphur starvation, but to a lesser degree (Figure 4-20).

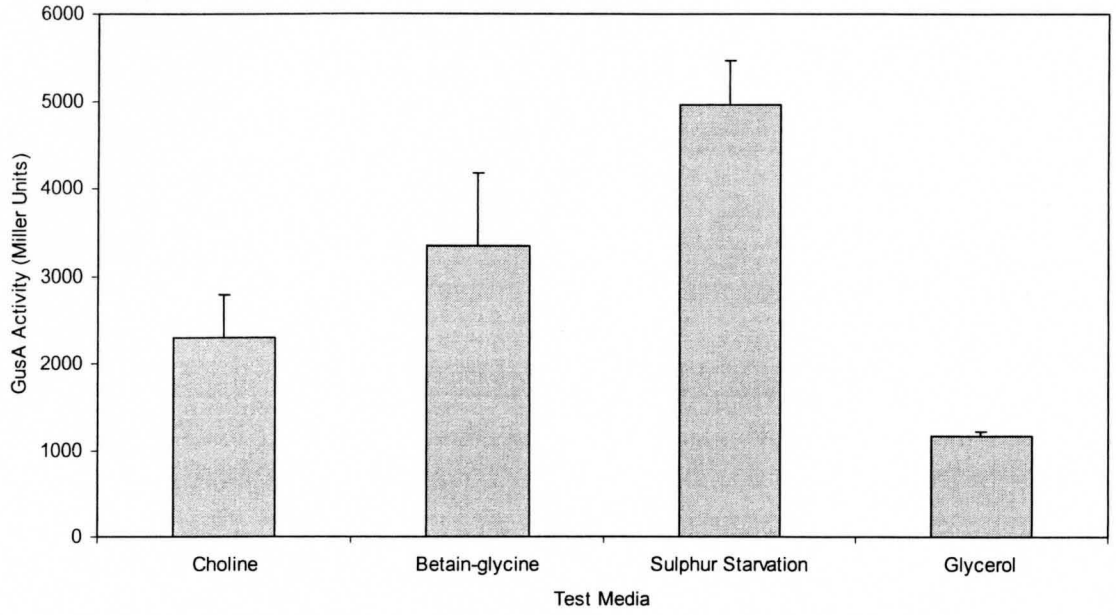

Figure 4-17. β-glucuronidase activity of RmP214 (SMb20571::*gusA*) showing induction by choline, glycine-betaine, and sulphur starvation. As choline and betaine-glycine were tested as nitrogen sources, NH_4Cl served as the reference nitrogen source. In all cases, 0.5% glycerol was added as the carbon source.

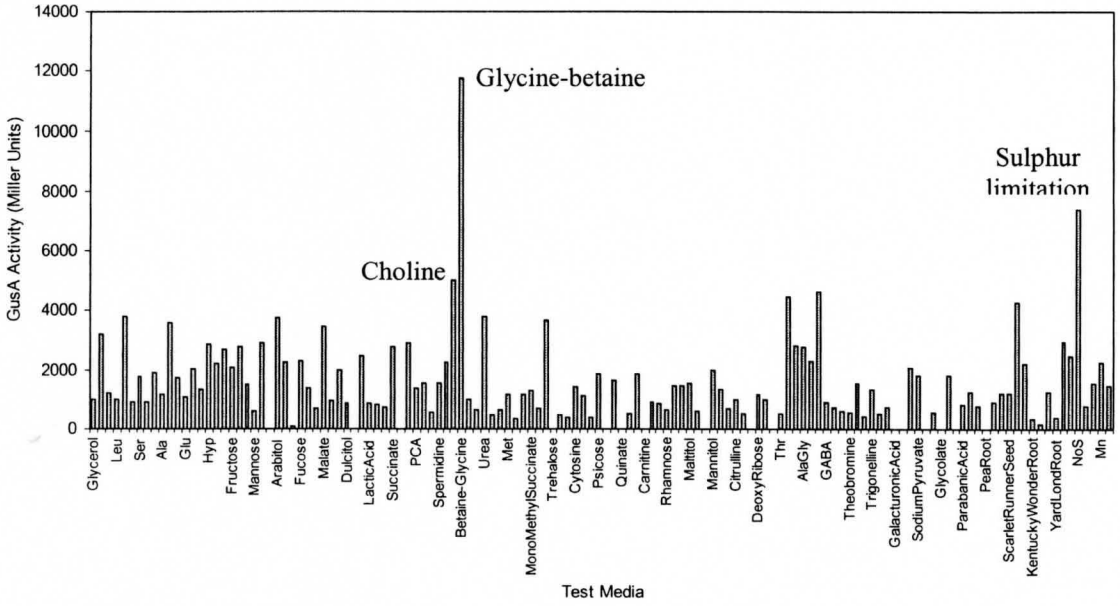

Figure 4-18. β-glucuronidase activity of SmFL2829 (SMc02344::*lacZ*) when tested for induction in all the test media, showed induction by glycine-betaine and to a lesser extent choline when used as nitrogen sources as well as in sulphur limiting conditions.

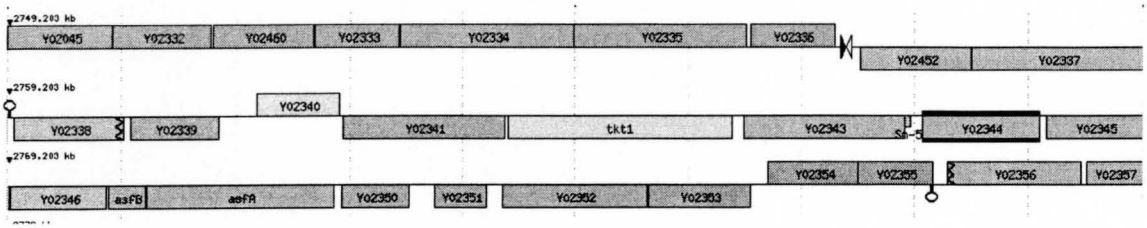

Figure 4-19. Genetic map of SMc02344 and surrounding genes (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

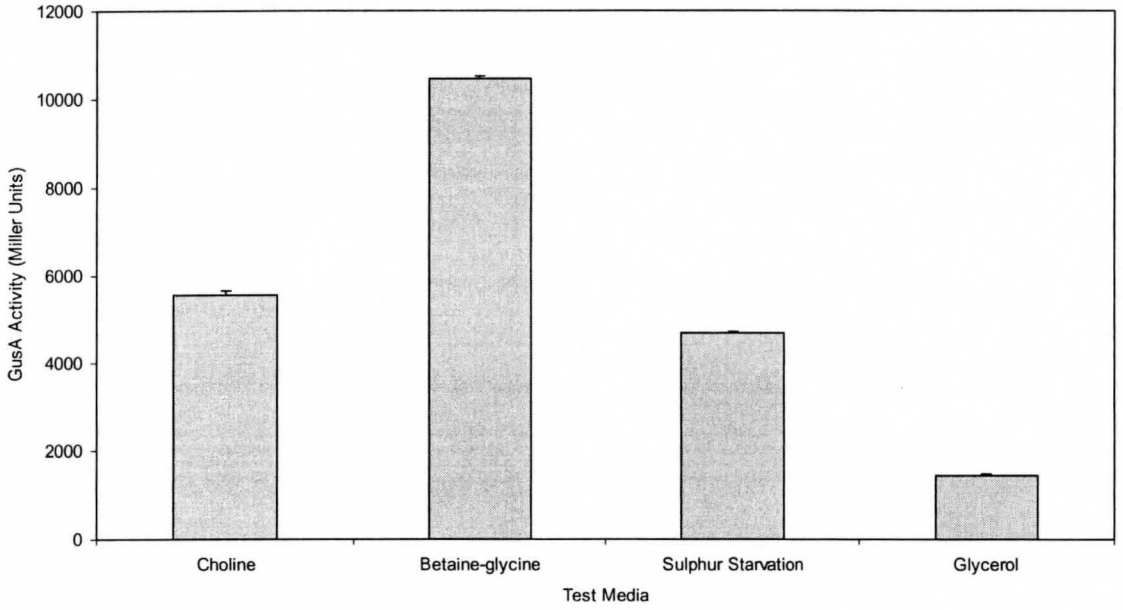

Figure 4-20. β-glucuronidase activity of SmFL2829 (SMc02344::*lacZ*) when retested in the betaines as nitrogen sources (with 0.5% glycerol as the carbon source) and in minimal media without sulphur supplementation.

There is another fusion in the transport library to the exact same gene, SMc02344, except it creates a knock-out and the orientation generates a SMc02344::*lacZ-gfp* reporter fusion. Interestingly this fusion, SmFL2637, did not show any induction with any of the above compounds. Rather it appeared to be induced by pea seed exudates in the initial screen (Figure 4-21). However, in the retest this induction was not observed (retest data not shown).

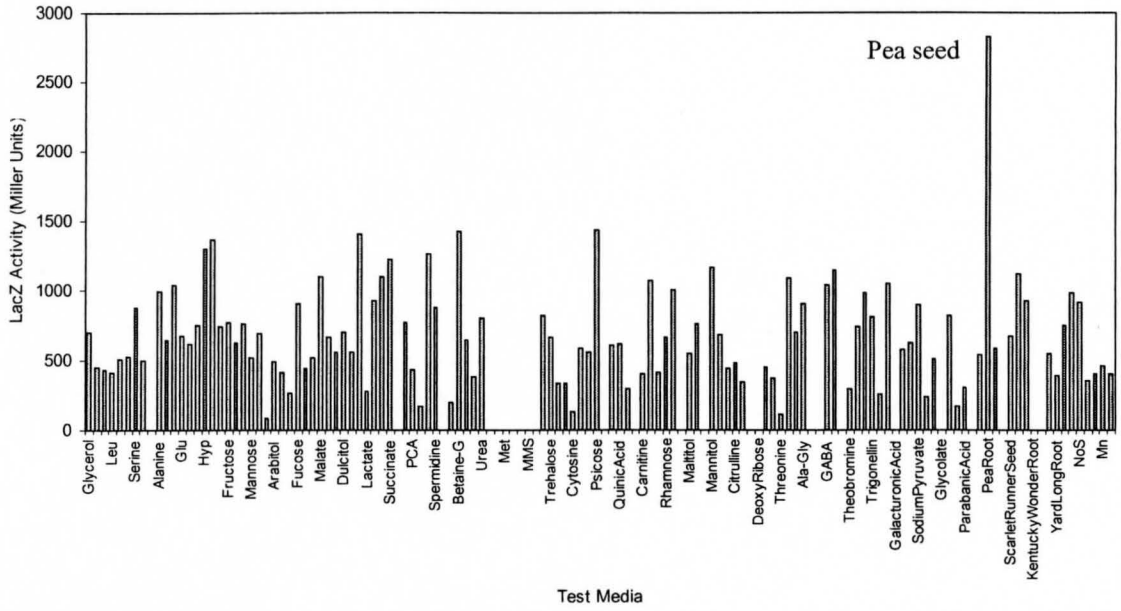

Figure 4-21. β-galactosidase activity of SmFL2637 (SMc02344::*lacZ*) when tested in all the test media showing only induction in pea seed exudate but this was retested and found to be negative (data not shown).

The collaboration with the Poole group allowed for other researchers to analyze our data. According to the Poole group, the SMc02737 (*choXWV*) operon was also induced by choline and glycine betaine (2.2 fold) but our criteria does not agree with that decision (Figure 4-22) (Mauchline et al., 2006). Thus this finding was included in the publication even though we did not consider this an induction. The inclusion of this finding was based on previous findings of the operon being induced by choline (Dupont et al., 2004).

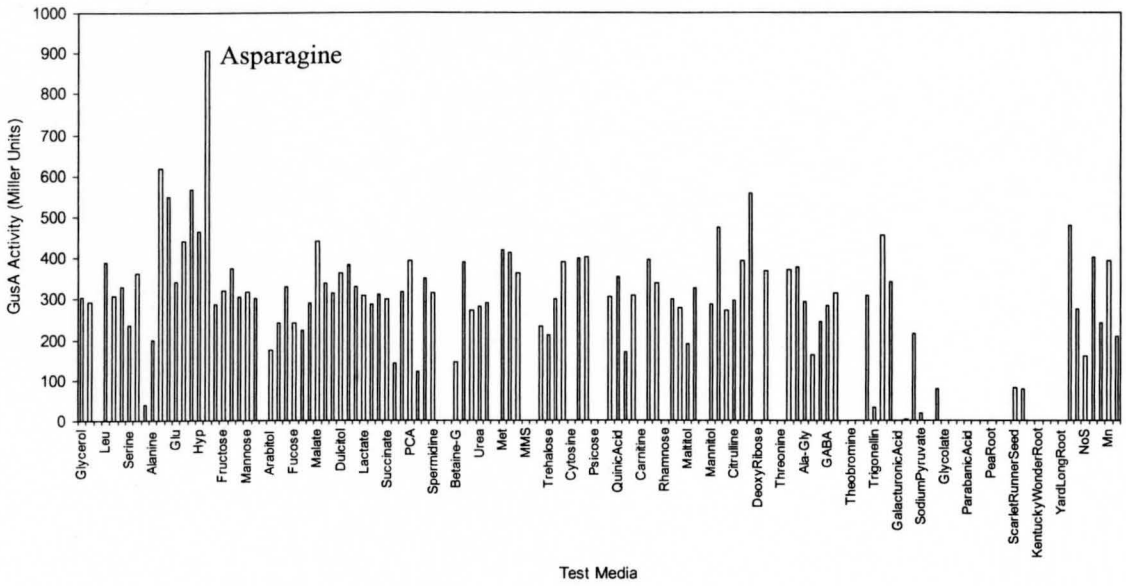

Figure 4-22. β-glucuronidase activity of SmFL4177 (SMc02737::*gusA*) showing no induction in glycine-getaine, choline or sulphur starvation. There appears to be induction in asperagine but when this was retested it was found to be negative (data not shown).

We know that the transport of betaines may be used in response to a high salt concentration since betaines are often used as osmoprotectants (Dupont et al., 2004). Therefore transport systems induced by the betaines were tested for induction when the fusion strains were grown in 0.5M NaCl. As shown in Figure 4-23, neither of the systems appeared to be induced by such a condition. To test SMc02344 for induction, fusion strain SmFL2829 (SMc02344::*lacZ*) was used instead of SmFL2637 (SMc02344::*lacZ*) because it is thought to not create a knock-out and therefore should have more reliable results.

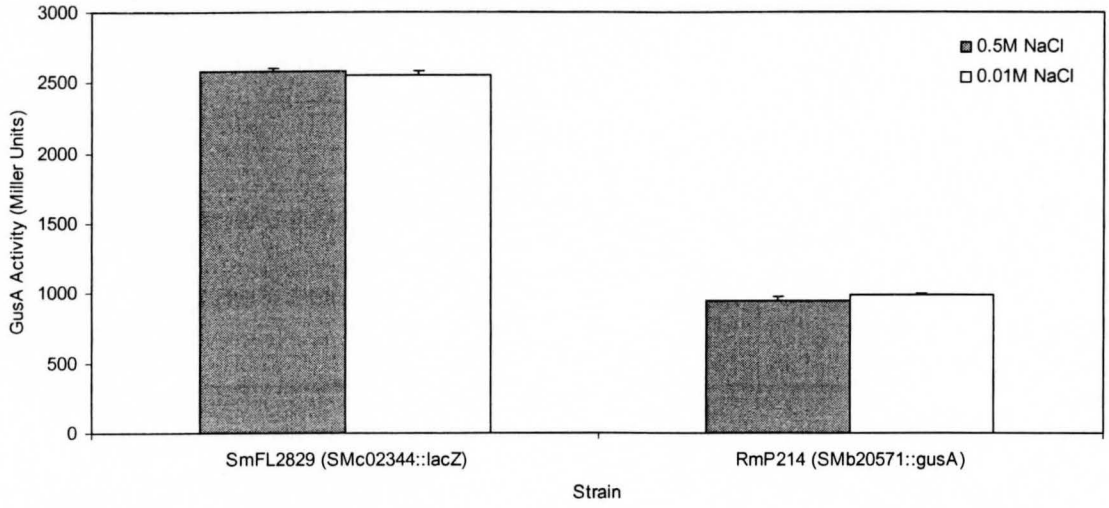

Figure 4-23. β-glucuronidase activities of SmFL2829 (SMc02344::*lacZ*) and RmP214 (SMb20571::*gusA*) when tested for induction by high salt concentration (0.5M NaCl).

These results indicate that SMc02344 and SMc20571 are not induced by salt stress.

Chapter 4.4-3. Discussion

Two fusions strains were found to be induced by choline and glycine betaine (Figures 4-15 and 4-18). One being RmP214 (Smb20571::gusA), putative aliphatic sulfonate uptake ABC transporter permease protein which is a component of a ABC transport system. There are two metabolism genes located directly upstream from the transport genes annotated as a putative nitrilotriacetate monooxygenase component-A and a putative NADH-dependent FMN reductase (Figure 4-16).

The other strain SmFL2829 (SMc02344::gusA/tdimer) which is a fusion to a putative periplasmic binding protein and is also induced by these two compounds. Located down stream from the transport genes are a putative sugar kinase and a probable

transketolase protein (*tkt1*). Located upstream from the transport genes lies a putative ferredoxin ASFB iron-sulfur protein and a putative oxidoreductase protein (Figure 4-19).

In both the above fusion strains glycine betaine induces gene expression about two fold above that of choline. SMb20333 (*betS*) is responsible for glycine betaine and proline betaine uptake and is important in overall betaine uptake under salt stress conditions (Boscari et al, 2002). OpuCBA (*choXWV*) have been shown to be responsible for choline but not betaine transport in *S. meliloti*, where OpuC (choX) binds choline with a high affinity but a mutant of this system is not impaired in growth under standard conditions or on Nod or Fix phenotypes (Dupont et al., 2004). OpuC is the periplasmic binding protein annotated to be a glycine betaine ABC transporter. When blasted for homology using the amino acid sequence, the closest hit was in fact ProX from *S. meliloti*. In *E. coli proX* encodes a periplasmic binding protein that is part of a transport system, ProVWX, which has been found to be dedicated to the transport of glycine betaine and L-proline (Dattananda and Gowrishankar, 1989, Barron et al., 1987). A similar system has also been identified and characterized in *Salmonella typhimurium* (Striling et al., 1989). These systems have both been shown to have osmoprotecting capabilities (Dattananda and Gowrishankar, 1989).

It is possible that both of the gene clusters found in this study could be responsible for the compensation of choline and glycine betaine transport in *S. meliloti*. These two reporter fusions strains were tested for increased gene expression under salt stress conditions but neither were found to be induced and both gene clusters were induced in nodule extracts (unpublished lab data).

Furthermore, both clusters that were found to be induced by choline and glycine betaine were also found to be induced by sulphur starvation (Figures 4-17 and 4-20). This coincides with what is known about choline metabolism since choline-O-sulphate can be converted to choline, which in turn can be oxidized to glycine betaine. It is likely that this transporter is also transporting choline-O-sulphate, but this has yet to be investigated.

CHAPTER 5. TRANSPORT SYSTEMS NOT INCLUDED IN Mauchline et al. (2006) PUBLICATION

ABC-type and Trap-T Transporters Not Included in Collaboration

As mentioned in the Material and Methods chapter, not all the transport systems were represented in the integrated fusion library and approximately 70 fusion strains to the remaining transport systems were made (work done by myself, Jane Fowler, and Alison Cowie). The majority of these fusions were to non-ABC or Trap-T transporters. However, three were of the Trap-T family and 18 were of the ABC-type. As a result, the data obtained from screening these transporters were not included in the Mauchline et al. (2006) paper. These strains and are listed in Table 5-1 below. It should be noted that inducers were detected only for SmFL7032 (SMb20902::*lacZ*), yet this data was found for a fusion to SMb20904, the ABC binding protein. Both of these fusions were found to be induced by mannose, sorbose, glucose, lyxose. However, P206 (SMb20904::*gusA*) did not show induction when grown in xylose, whereas SMFL7032 (SMb20902) showed 7.3-fold induction (see Table 3-2 in Chapter 3).

Table 5-1. Gene fusions to ABC-type and Trap-T transport systems that were not included in the Mauchline et al. (2006) collaboration. All fusions are in the LacZ/Gfp orientation.

Gene	**Fusion**	**Superfamily**	**Inducer** (fold increase over glycerol)	**Median** (Miller Units)
Sma04259	SmFL7023	ABC		344
Sma0527	SmFL7045	ABC		333
Sma1365	SmFL3048	ABC		315
Smb20155	SmFL7064	ABC		323
Smb20263	SmFL7003	ABC		460
Smb20416	SmFL7028	ABC		572
Smb20713	SmFL7047	ABC		366
Smb20813	SmFL7030	ABC		600
Smb20895	SmFL7031	ABC		466
Smb20902	SmFL7032	ABC	Manose (9.2) Glucose (8.1) Lyxose (6.6) D-xylose (7.3)	480
Smb20981	SmFL7048	Trap-T		307
Smb21316	SmFL7036	ABC		328
Smc00265	SmFL7069	Trap-T		535
Smc00550	SmFL7050	ABC		444
Smc00773	SmFL7038	ABC		316
Smc01376	SmFL7053	ABC		701
Smc02169	SmFL481	ABC		643
Smc02418	SmFL3256	ABC		194
Smc04287	SmFL1077	Trap-T		267
Smc04317	SmFL7020	ABC		281
Smc04454	SmFL7056	ABC		646

Non-ABC or Trap-T TRANSPORTERS

As mentioned earlier, this section focuses on those transport systems that were not included in the Mauchline et al. publication (2006). As that paper only included those transport systems belonging to the ABC-type and Trap-T families, this section encompasses all the other systems found in *S. meliloti* (Table A-2 of the Appendix lists all the genes in such transport systems that were analysed in this study). However, only ten such transport systems were found to have inducers and all but two belong to the secondary transporters. The other two transport systems belong to the voltage gated ion-channel (VIC) superfamily (belonging to the Ion Channels) and the P-type ATPase (P-ATPase) superfamily (belonging to the ATP-dependent transporters) (see Table 5-2 below).

Table 5-2. Summary of all positive inducers for gene fusions to non-ABC or Trap-T transporters. The fold increase of LacZ or GusA enzyme activity (Miller Units) in the presence of an inducting compound over the enzyme activity when that fusion was grown in M9 minimal media with 0.5% glycerol and 5 mM NH_4Cl as the carbon and nitrogen sources, respectively, unless otherwise noted.

Gene	Fusion	Family	Inducer	Fold Increase
sma1153	SmFL6159	P-ATPase	$CaCl_2$ Limitation	6.7
sma1447	SmFL4563	MFS	Isoleucine Leucin	25.7 11.9
smb20272	SmFL154	MFS	Glycerol*	6.3
smb20345	SmFL631	RND	Glycerol*	3.9
smb20361	SmFL2301	VIC	Glycine	2.1
smb21486	SmFL3579	MFS	$CaCl_2$ Limitation	4.0
smc02616	SmFL5242	APC	Trigonelline	18.3
smc03807	SmFL3396	Amt	0.10% glutamine 0.5 mM KNO_3	28.5 26.6
smc04147	SmFL4572	APC	Trigonelline	8.5
smc04407	SmFL1286	MFS	Taurine	5.5

* indicates gene fusions that were induced by the presence of glycerol, in these cases glucose was used as a basal level to calculate the fold increase.

As mentioned above, the inducing conditions for only ten fusions to non ABC or Trap-T transport systems were identified in this study. It is possible that the nature of the fusions to the genes being analyzed were causing the induction to be undetectable. That is, perhaps the fusions led to a constitutive activity so that even if the inducing compound were present or absent there would be no induction. To investigate this possibility, histograms were made for both sets of transport fusions, the ABC and Trap-T fusions and the remaining fusions. By analyzing the location of the peaks of the histograms and the spread of the values we can determine whether there was a lot of constitutive activity in one set of fusions. Thus, the histograms of the medians of each fusion strain was made and analyzed. It was found that the average median of these genes was comparable to that of the ABC and Trap-T transport fusions (Figures 5-1 and 5-2 below). In fact, the histogram for the ABC and Trap-T fusions peaks at 500 Miller Units indicating that the majority of fusions have a median expression at around 500 Miller Units. In comparison, the histogram for the remaining transport fusions peaks only at 350 Miller Units indicating that the overall background expression for these fusions is lower than the ABC and Trap-T fusions. Also, the ABC and Trap-T fusions have a slightly wider spread than that of the non-ABC and Trap-T transporter fusions.

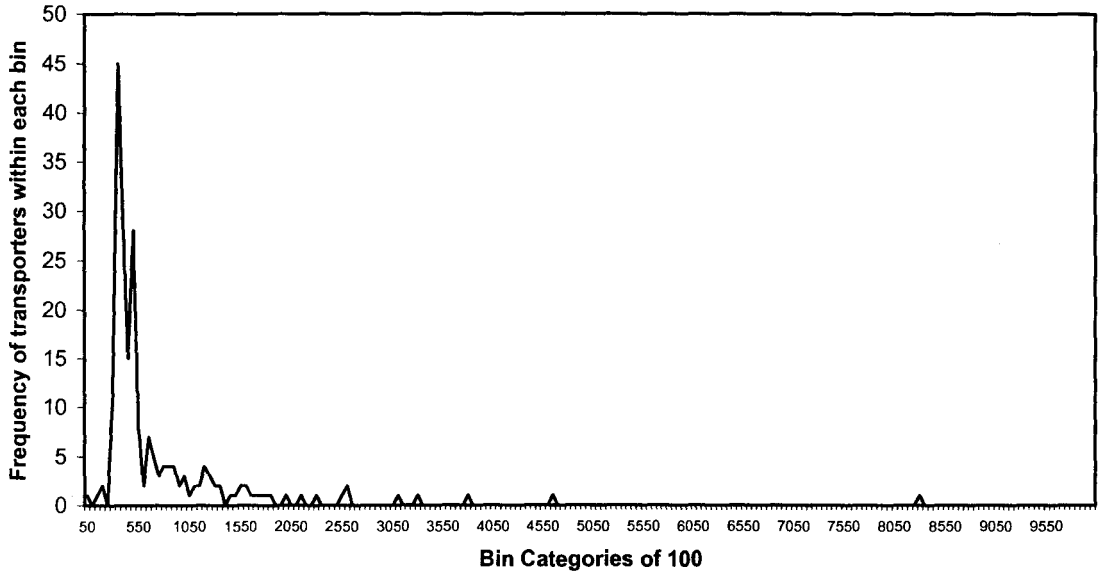

Figure 5-1. Histogram diagramming the distribution of the median for those fusions to transport systems outside of the ABC or Trap-T superfamilies. Each bin on the x-axis refers to the enzyme activity in miller units.

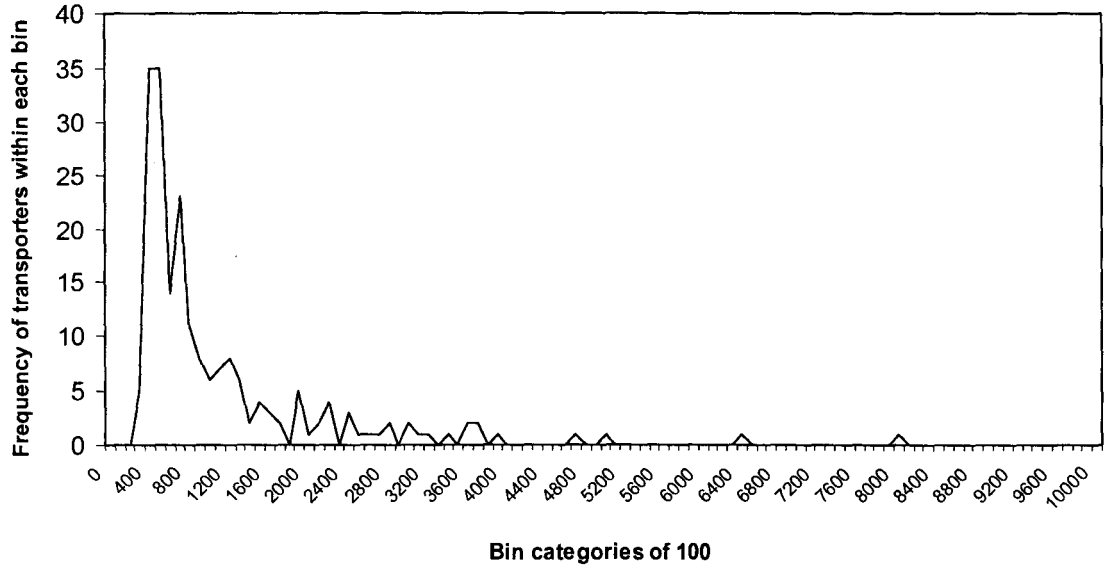

Figure 5-2. Histogram of the distribution of the median for those transport systems belonging to the ABC and Trap-T superfamilies. Each bin on the x-axis refers to the enzyme activity in miller units (Jane Fowler).

Chapter 5.1 Nitrate Transport

Chapter 5.1-1. Introduction

Nitrogen metabolism and transport has been well studied in bacteria over the years. The uptake of nitrogen is tightly regulated by a two component system, the Ntr system. The majority of what is known today about nitrogen control is from studies on *E. coli*, *Klebsiella aerogenes, K. pneumoniae,* and *Salmonella typhimurium.*

Ammonium Transport

There is some evidence of ammonium active transport of ammonium across the bacterial cytoplasmic membranes. In *E. coli* the Amt transporter is a single protein making up either a secondary carrier or a channel that increases the rate of equilibration of NH_3 across the cell membrane (Luzhkov et al., 2006). Amt activity has been shown to be repressed in the presence of high extracellular ammonium concentrations and studies have shown that the expression is Ntr regulated (Jayakumar et al., 1986).

Interestingly, most cyanobacteria take up nitrate/nitrite via an ABC-type transport system, NrtABCD, located in the cytoplasmic membrane. As expected, the expression of these transport genes instantaneously turn off once ammonium is present in the media (Nagore et al., 2006).

Chapter 5.1-2. Results

It was observed that the reporter enzyme activities for strains SmFL1790 (SMb20604::*gusA*), SmFL4232 (Sma0583::*gusA*), and SmFL3396 (SMc03807::*lacZ*) were induced in media containing nitrogen sources other than NH_4Cl (i.e. when NH_4Cl

was absent from the media). Though SMb20604 belongs to an ABC-type transport system these results were combined for simplicity and results obtained from starvation conditions were not a focus of the paper (Mauchlin et al., 2006). The results from the initial screen for each of the fusions are shown below along with the genetic map of the transport systems.

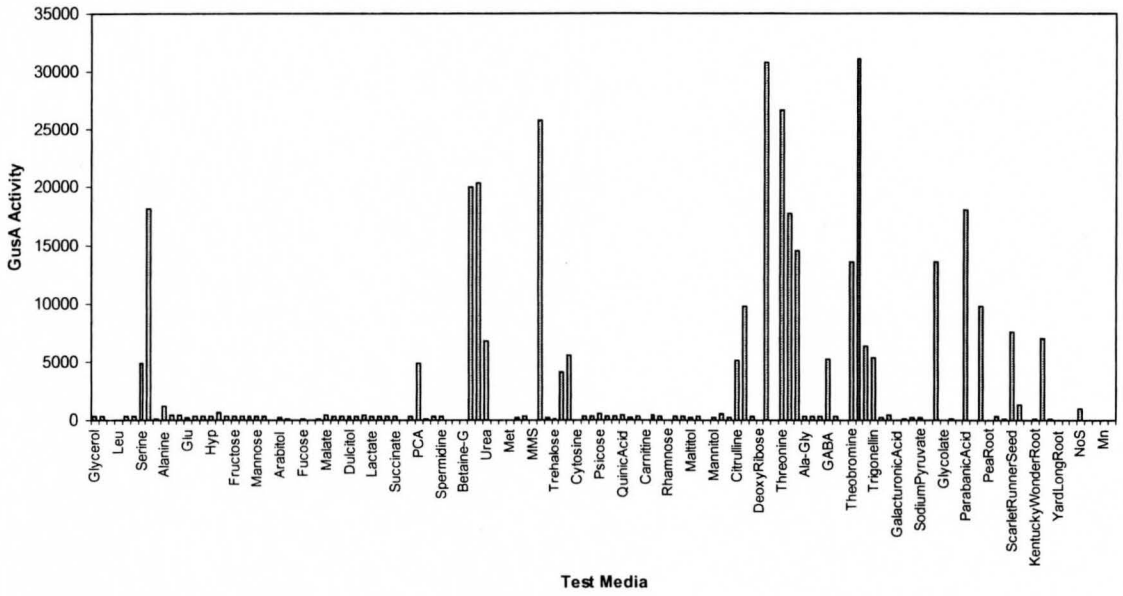

Figure 5-3. SmFL1790 (SMb20604::*gusA*) showing specific induction in test media that does not contain NH_4Cl as a nitrogen source.

Figure 5-4. Gene map of the operon induced by nitrogen limiting conditions (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

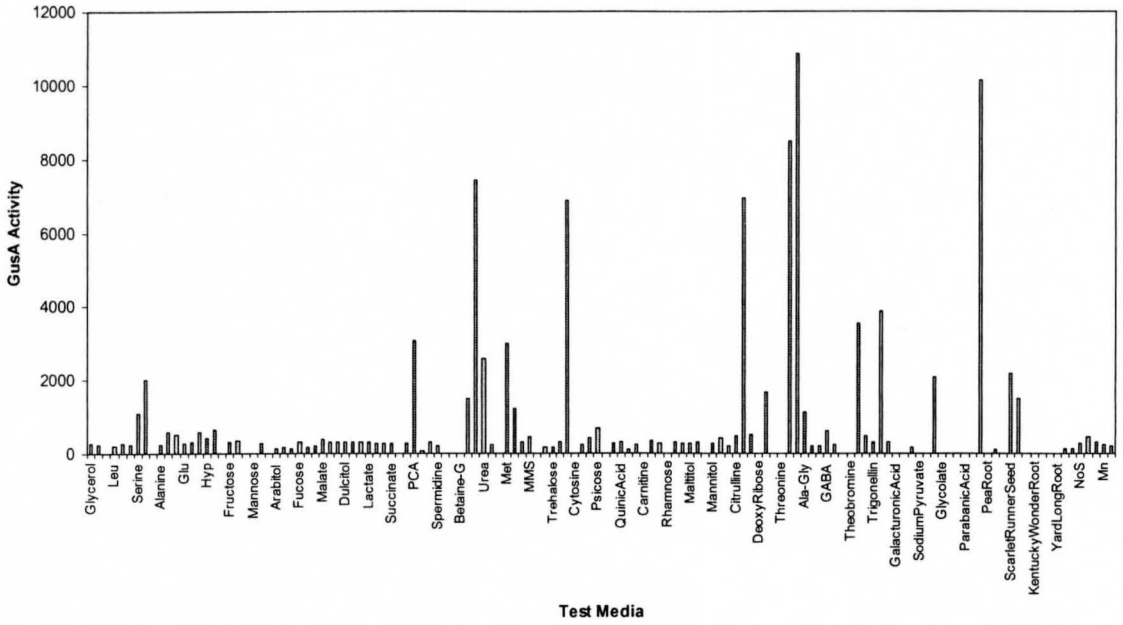

Figure 5-5. SmFL4232 (SMa0583::*gusA*) when grown in all the different test media. Induction is found when this strain is grown without NH_4Cl in the media.

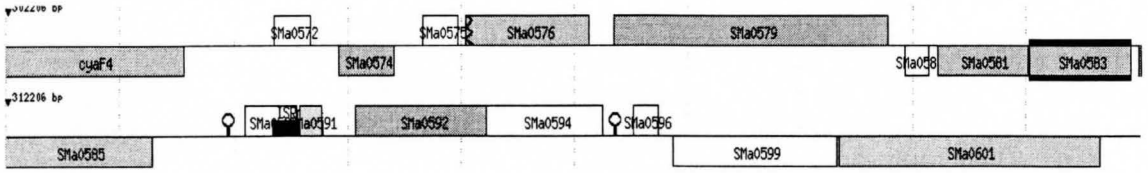

Figure 5-6. Gene map of the operon induced by nitrogen starvation (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

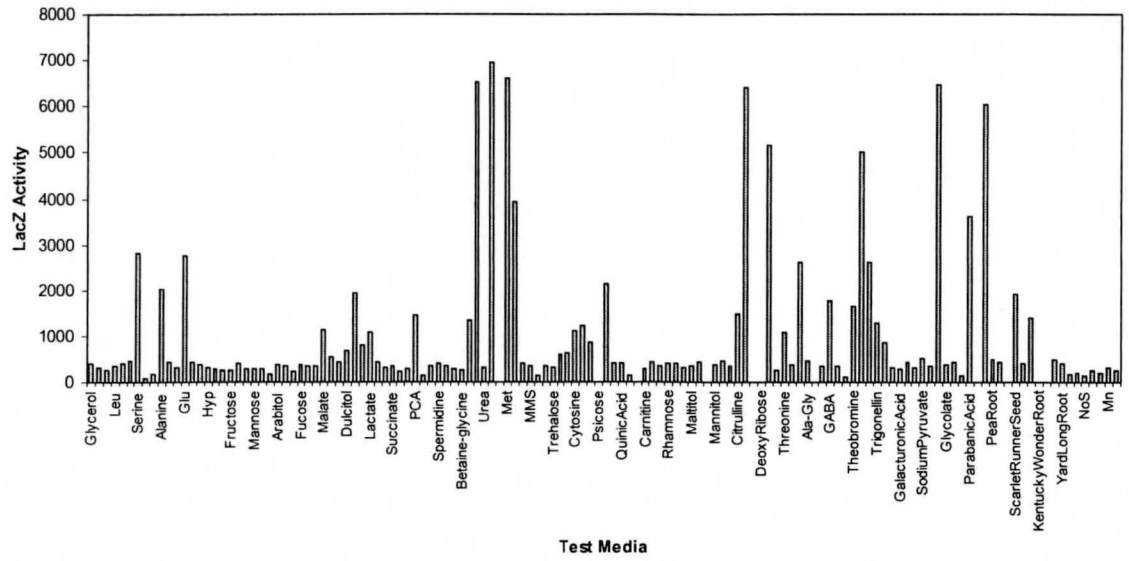

Figure 5-7. SmFL3396 (SMc03807::*lacZ*) showing induction when NH_4Cl is missing from the test media.

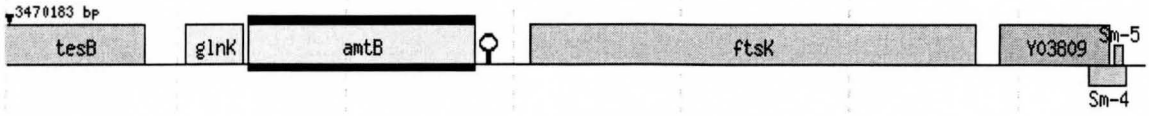

Figure 5-8. Gene map of the operon (*glnK* and *amtB*) induced by nitrogen starvation (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

To further elucidate whether or not these fusions were induced by the lack of NH_4Cl in the media they were retested in 0.01% glutamine, 0.5 mM KNO_3, and 5 mM NH_4Cl as nitrogen sources (with 0.5% glycerol as the carbon source). Though KNO_3 did not allow for good growth of *S. meliloti* (less than 0.1 OD_{600} as read in the Tecan Safire) the results were still useful in the analysis. Library fusion SmFL4410 (*glnII*::*gusA*) was used as a positive control, for expression of *glnII* is known to be induced by nitogen limiting conditions (de Bruijn et al., 1989).

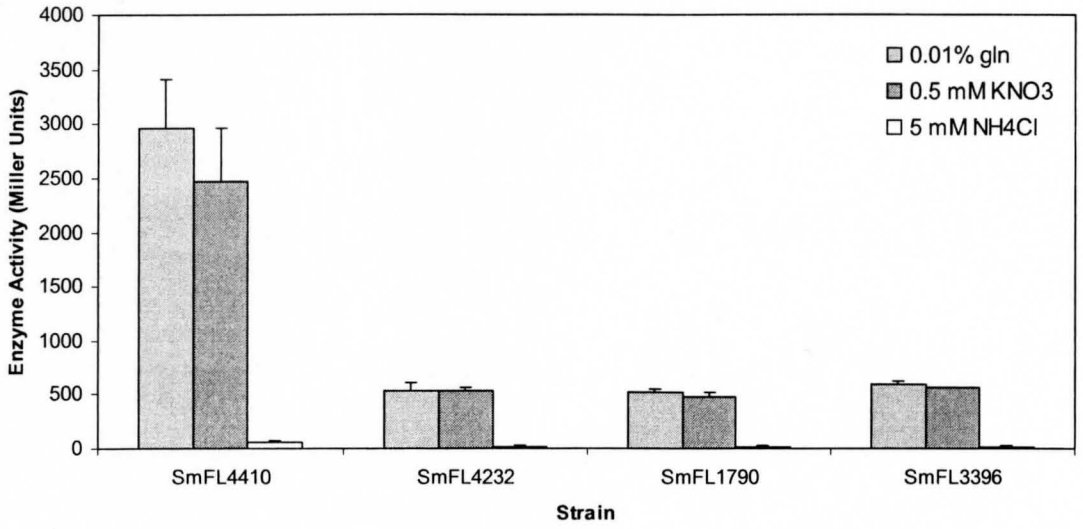

Figure 5-9. β-glucuronidase and β-glucosidase assay of SmFL4232 (SMa0583::*gusA*), SmFL1790 (SMb20604::*gusA*), and SmFL3396 (SMc03807::*lacZ*) when tested for expression when grown in different nitrogen sources with SmFL4410 (*glnII*::*gusA*) as a positive control.

To explore the possibility of these operons being controlled by the ntr system, transductions were carried out to create *ntrC*⁻ and *ntrA*⁻ (rpoN) strains. Each fusion of interest was tested in a wild type background, *ntrA*⁻ background, and a *ntrC*⁻ background.

Table 5-3. Enzyme activities (miller units) of SmFL4232 (SMa0583::*gusA*), SmFL1790 (SMb20604::*gusA*), and SmFL3396 (SMc03807::*lacZ*) when tested for induction with either wildtype, NtrA- or NtrC- backgrounds.

Background	**Nitrogen Source**	**SmFL4232** (SMa0583::*gusA*)	**SmFL3396** (SMc03807::*lacZ*)	**SmFL1790** (SMb20604::*gusA*)
Wildtype	NH_4Cl	211 +/- 21	321 +/- 5	336 +/- 17
	Glutamine	5570 +/- 191	1570 +/- 12	8081 +/- 188
	KNO_3	10730 +/- 93	2469 +/- 43	20762 +/- 405
NtrA-	NH_4Cl	303 +/- 18	336 +/- 11	364 +/- 11
	Glutamine	298 +/- 13	190 +/- 4	3242 +/- 100
	KNO_3	725 +/- 224	205 +/- 79	14030 +/- 413
NtrC-	NH_4Cl	310 +/- 5	331 +/- 14	354 +/- 4
	Glutamine	259 +/- 9	242 +/- 7	1741 +/- 40
	KNO_3	342 +/- 60	159 +/- 52	8605 +/- 2354

These results indicate that operons containing SMa0583 and SMc03807 are regulated by the Ntr system but the SMb20604 transport system is not.

Chapter 5.1-3. Discussion

Three separate transport systems were induced upon nitrogen starvation, SMa0583, SMb20604, and SMc03807. SMa0583 is annotated as NtrB (a nitrate transport permease protein) and lies in an operon with a probable nitrate transport ATP binding protein (SMa0581) and a probable NrtA-type periplasmic nitrate transport binding protein (SMa0585). SMb20604 is annotated as a putative urea/short-chain amide or branched-chain amino acid uptake ABC transporter permease protein and lies in an operon with four other genes that make up all the necessary components of an ABC transport system. In a study which isolated carbon and nitrogen deprivation induced loci in *S. meliloti*, one of the ATP binding proteins of this transporter (SMb21707) was found to be induced during nitrogen deprivation (Milcamps et al., 1999). AmtB is a probable ammonium transporter protein which lies directly downstream from *glnK*. In *R. etli* the expression of these two genes was found to be induced under nitrogen deprivation conditions and down regulated in bacteroids (Tate et al., 1998).

To determine whether these transport systems are regulated by the Ntr system, *ntrC*::Tn5 and *ntrA*::Tn5 mutant alleles were transduced into the fusion strains. It was expected that if the systems were regulated by NtrC, then when either of the Ntr genes was disrupted, induction of the transport systems would no longer be observed in nitrogen starved conditions. As shown in Table 5-3 the induction of SMa0583 and SMc03807 (*amtB*) required the *ntrC* and *ntrA* genes. However, we see that SMb20604

does not show an altered expression when the *ntrA* or *ntrC* genes are knocked out and therefore it is concluded that this particular transport system is not regulated by the Ntr system.

Chapter 5.2 Taurine Transport

Chapter 5.2-1. Introduction

Taurine is a β amino acid and is a very important compound that is found in relatively high abundance in the tissues of many animals, especially those in the sea. Taurine is also an important factor in bile acid formation and osmoregulation (Kendler, 1989).

Taurine transport and metabolism has been studied in *E. coli*, where taurine is taken up and metabolized under sulfate or cysteine starvation conditions. Eichhorn et al. (2000) reported the presence of two gene clusters, *tauABCD* and *ssuEADCB* involved in the transport and utilization of taurine and alkanesulfonates as the sole source of sulfur in *E. coli*. *tauD* and *ssuD* encode an α-ketoglutarate-dependent taurine dioxygenase and a reduced flavin mononucleotide-dependent alkanesulfonate monooxygenase, respectively, which are the enzymes responsible for the desulfonation of taurine and alkanesulfonates. The remaining genes of both clusters make up the components of two separate ABC-transport systems. Through creating chromosomally in-frame deletions of each of the clusters, these two systems were found to be required for the utilization of taurine and alkanesulfonates (Eichhorn et al, 2000).

Chapter 5.2-2. Results

In this study two fusion strains, SmFL627 (*tauC*::*gusA*) and SmFL1286 (SMc04407::*gusA*) were found to be induced by the presence of taurine in the test media as the sole source of carbon and nitrogen. Though *tauABC* is an ABC-type transport system, the data for this system was included in this section for simplicity as SMc04407 is a MFS system and the systems are presented together.

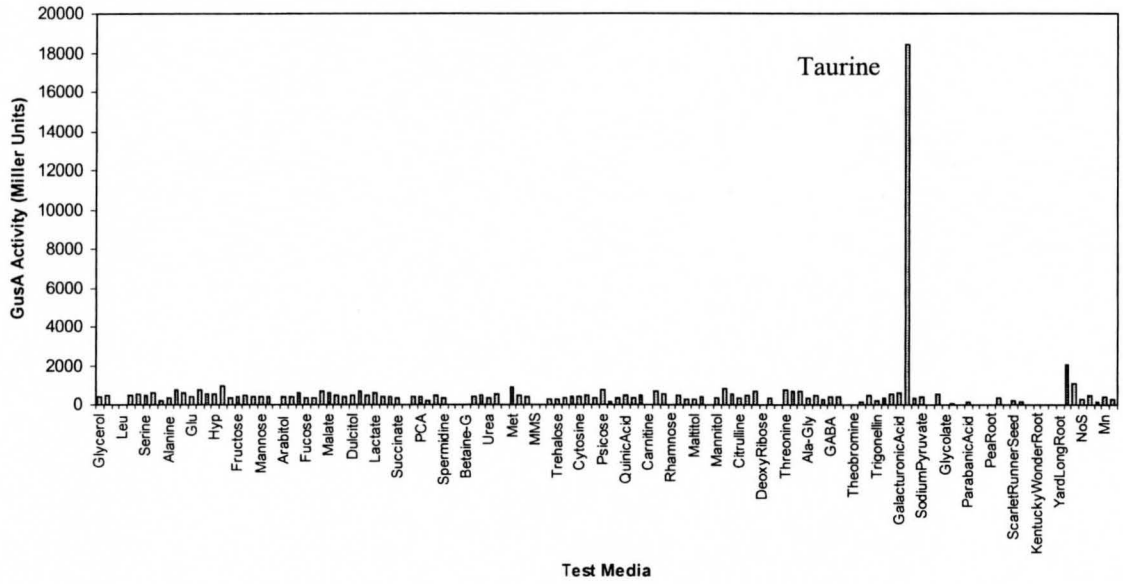

Figure 5-10. β-glucuronidase activity of SmFL627 (*tauC*::*gusA*) grown in all the different test media and showing specific induction by taurine.

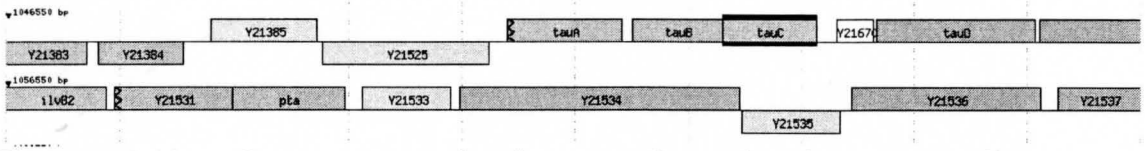

Figure 5-11. Gene map showing *tauC* and the surrounding genes (http://bioinfo.genopoletoulouse.prd.fr/annotation/iANT/bacteria/rhime/).

Table 5-4. β-glucuronidase activity of SmFL627 (*tauC::gusA*) showing induction by over 30-fold when grown in taurine as a sole nitrogen and carbon source versus NH_4Cl and glycerol as the nitrogen and carbon sources, respectively.

Carbon and Nitrogen Source	**Taurine**	**Glycerol and NH_4Cl**
β-glucuronidase activity (Miller Units) +/- standard deviation	26550 +/- 1011	579 +/- 12

From the above data it is evident that *tauC* is induced by taurine. The below figures are of fusion strain SmFL1286 (SMc04407::*gusA*). This fusion was found to be induced by taurine and in the initial screen it also showed induction by isoleucine and methyl-pyruvate.

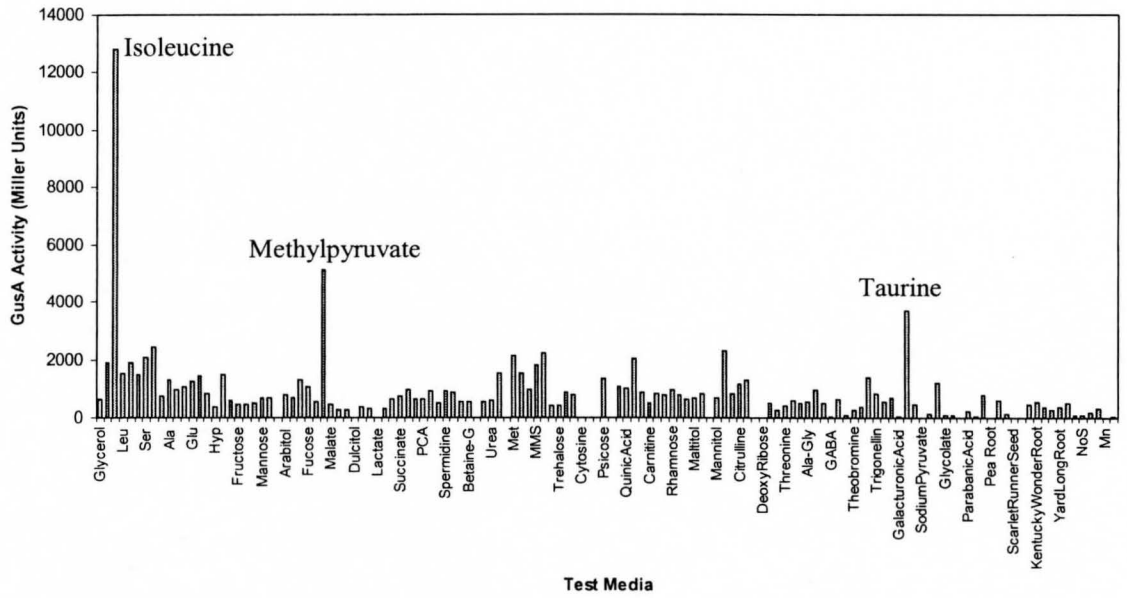

Figure 5-12. β-glucuronidase activity of SmFL1286 (SMc04407::*gusA*) showing a large induction by isoleucine, methylpyruvate, and taurine.

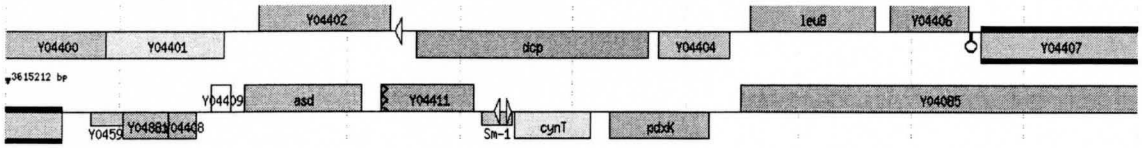

Figure 5-13. Genetic map of SMc04407 and the surrounding genes (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

To verify the initial screening results, fusion strain SmFL1286 (SMc04407::*gusA*) was retested in all three inducing test media; taurine, methylpyruvate, and isoleucine. Analysis of the genetic map surrounding SMc04407 shows two genes of interest that may be associated with the functioning of this transport system. SMc04387 is annotated as an ω-amino acid transporter and SMc04389 is annotated as a hydroxypyruvate reductase. Thus fusion strains to SMc04387 and SMc04388 were also included in this analysis because they were thought to perhaps be involved in the transport and metabolism of pyruvate and taurine. However, in the retest only the fusion to SMc04407 showed induction in methyl-pyruvate and taurine and the other fusions did not show any specific induction by the test media.

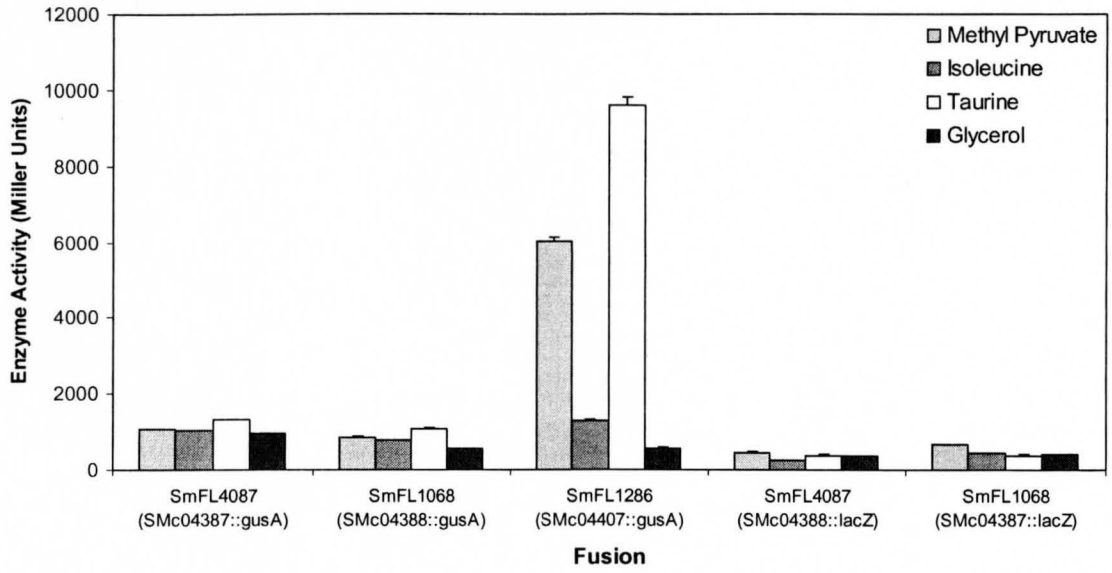

Figure 5-14. β-glucuronidase and β-galactosidase activity of various fusions to the three genes that were suspected to be involved in the transport and metabolism of taurine and pyruvate.

Chapter 5.2-3. Discussion

Two transport systems, SMb21528 and SMc04407, were found to be induced when taurine was added to the medium as the sole source of carbon and nitrogen. SMb21528 is part of a three gene operon that is annotated as being a putative taurine ABC uptake transport system. As mentioned earlier, this system has been studied in *E. coli*, where it was found to be involved in the uptake of sulphur. Furthermore, this system, along with *ssuEADCB*, is only expressed under cysteine and sulfate starvation conditions (Eichhorn et al., 2000). Unlike the systems found in *E. coli* the *S. meliloti tauC* system was not found to be induced under sulphur starvation conditions (Figure 5-10).

The other transport system induced by the presence of taurine, SMc04407, is a single gene annotated by Transport DB as belonging to the major facilitator superfamily

(MSF) of transporters. The fusion to this system, SmFL1286 (SMc04407::*gusA*) also showed a 6-fold induction when grown in minimal media with methylpyruvate as the sole source of carbon. Taurine and pyruvate can react to form L-alanine via taurine-pyruvate aminotransferase. SMc04388, which is located upstream from the MFS transporter, is annotated as being an omega amino acid--pyruvate aminotransferase. This enzyme acts on β-amino acids and taurine is a β-amino acid. Located adjacent to SMc04388 is a metabolism gene, SMc04389, which is annotated as a hydroxypyruvate reductase. From this annotation it seems likely that the transporter and the two metabolism genes would be involved in the transport and metabolism of taurine and methylpyruvate. This hypothesis was investigated by testing several fusions to the above mentioned metabolism genes along SMc04387, annotated as a hydroxypyruvate reductase, as it is adjacent to SMc04388. Fusion strains SmFL4087 (SMc04387::*gusA*, SMc04388::*lacZ*), SmFL1068 (SMc04388::*gusA*, SMc04387::*lacZ*) did not show any induction in either taurine or methylpyruvate indicating that they are indeed not involved in the metabolism of tauring and (methyl) pyruvate.

Like SMb21528 this transporter was not induced in the absence of a sulphur source. Unfortunately knockout fusions were not available in the library nor were such strains built. It would be interesting to explore the growth phenotypes of such knockout strains.

Chapter 5.3. Trigonelline Transport

Chapter 5.3-1. Introduction

The rhizosphere has copious amounts of plant-secreted compounds, namely rhizopines, which are plant metabolites that are found exclusively in the nodules that are utilized by free living rhizobia (Boivin et al., 1990). In this region, there are also bacteria, fungi, and protists that strive on the organic compounds such as amino acids, sugar alcohols, sugars, and polysaccharides (Bringhurst et al., 2001). Alfalfa roots and other legumes that can host nitrogen-fixing bacteria, secrete signal molecules that affect the transcription of *nod* (nodulation) genes. Such compounds are also thought to be used by the bacteria as carbon and nitrogen sources. Trigonelline is a betaine that has been found in rhizobium leguminous hosts and studies with alfalfa seed rinse has identified the presence of trigonelline (Boivin et al., 1990, Phillips et al., 1992).

Trigonelline genes involved in the catabolism of trigonelline have been identified in *S. meliloti* strain RCR2011. The genes were located near *nod-nif* genes on the pSymA megaplasmid (Boivin et al, 1990). Boivin and colleagues made *lacZ* gene fusions to the *trc* genes of RCR2011 and monitored the expression of the genes throughout the various stages of infection and nodulation. From studying free living *S. meliloti* it was found that the metabolism genes were transcribed as four separate transcriptional units and trigonelline was a specific inducer for three of them. They also found that the *trc* genes were highly induced during all stages of nodulation; free living, infection thread, and in bacteroids. However, this study did not yield the identification of a transport system involved in the import of trigonelline into *S. meliloti* (Boivin et al., 1990).

Chapter 5.3-2. Results

From the screen two different transport systems were found to be induced by trigonelline as a sole source of carbon and nitrogen. Interestingly, both of the transport systems have been identified by the Transport DB database as being part of the APC (amino acid-poylamine-organocation) superfamily. SMc02616 and SMc04147 are both single genes, which is characteristic of this type of transport system. It is interesting that both of these transport systems are also induced by red clover seed exudates. Red clover root exudate was also tested for induction of all the strains but like many of the exudates used in this study, all of the concentrations of exudates tested inhibited the growth of *S. meliloti*. Figures 5-15 and 5-17 show the results from the initial screening of SmFL5242 (SMc02616::*gusA*) and SmFL4572 (SMc04147::*lacZ*).

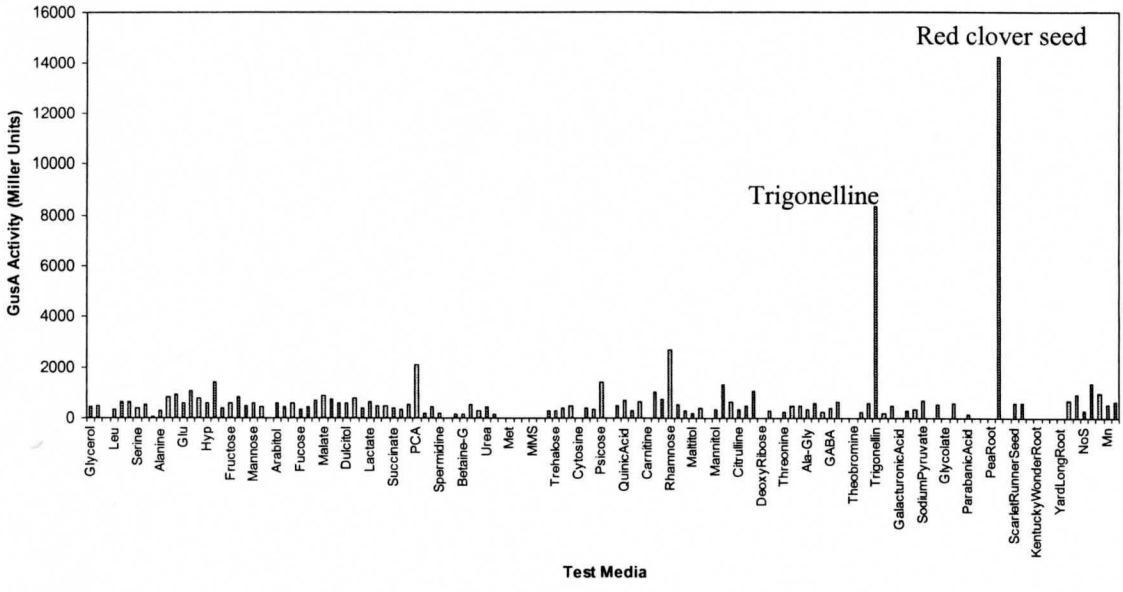

Figure 5-15. β-glucuronidase activity of SmFL5242 (SMc02616::*gusA*) grown in all the test media, showing induction by trigonelline and red clover seed exudates (RCS).

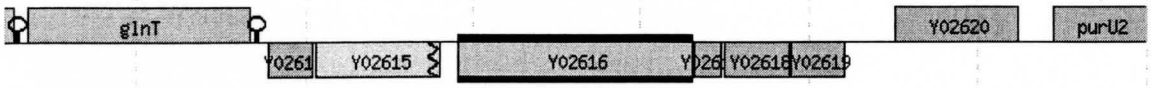

Figure 5-16. Genetic map of SMc02616 and the surrounding hypothetical genes (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

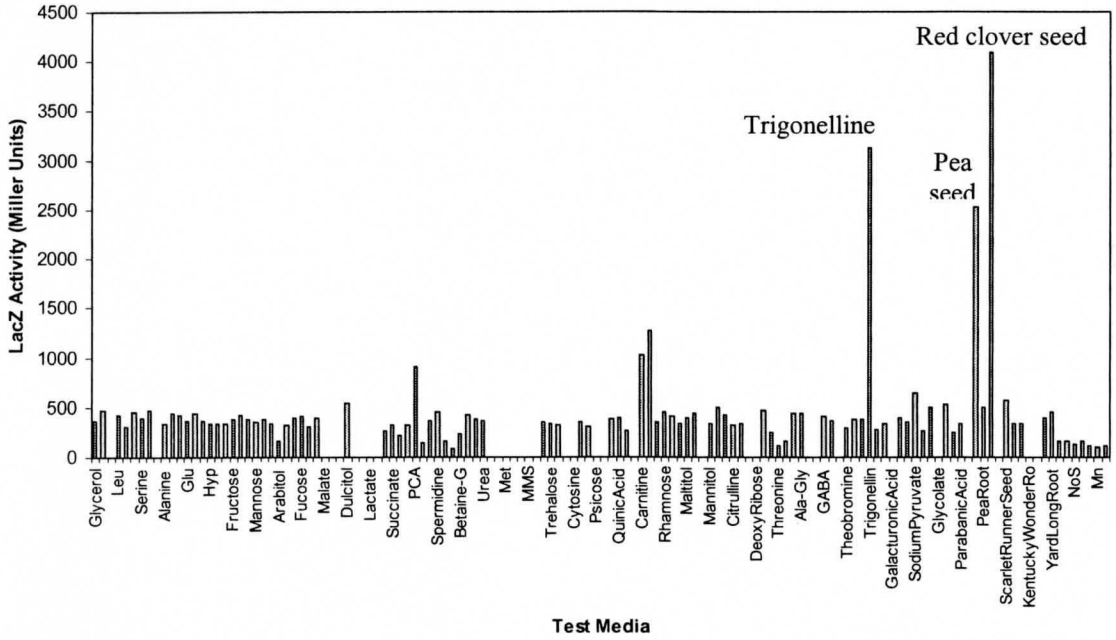

Figure 5-17. β-galactosidase activity of SmFL4572 (SMc04147::*lacZ*) when grown in all the test media showing specific induction in RCS, trigonelline, and also pea seed (PS) exudate.

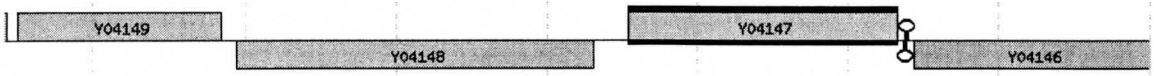

Figure 5-18. Genetic location of SMc04147, the other MFS transporter that was found to be induced by trigonelline and RCS exudates when used as sole sources of both carbon and nitrogen (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

Figure 5-19 shows the retest data of both fusions indicating that both transport systems are indeed induced by trigonelline and red clover seed exudates. SmFL4572 (SMc04147::*lacZ*) was also retested in pea seed exudate, lyxose, carnitine, and PCA

because they appeared to be inducers in the initial screen, but they were found not to induce the system upon retest in triplicate.

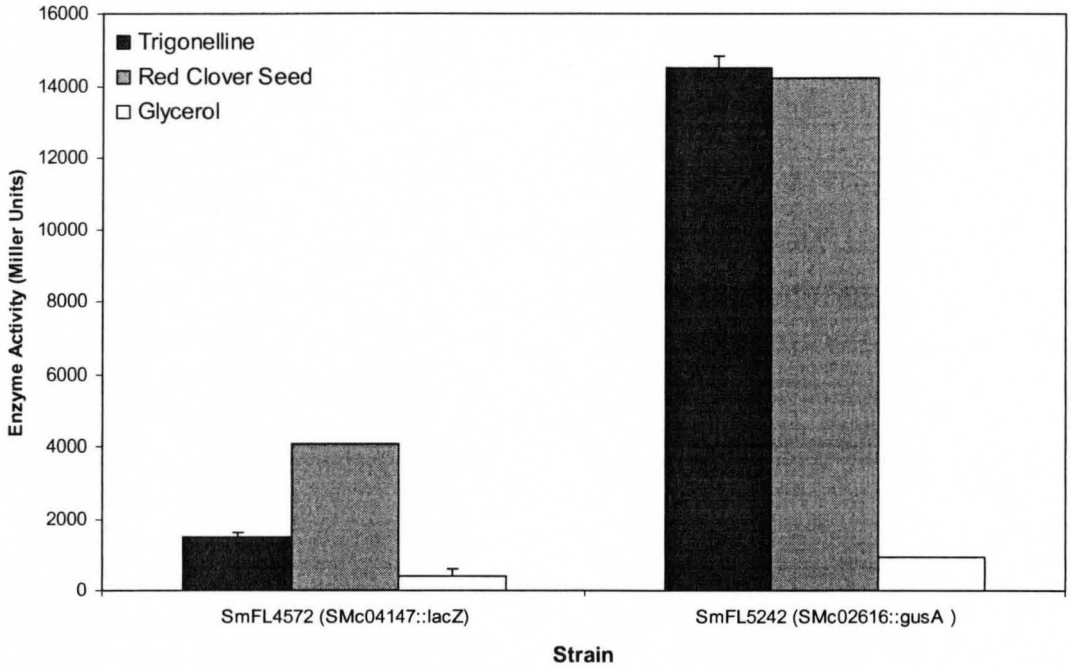

Figure 5-19. β-glucosidase and β-glucuronidase activities of SmFL4572 (SMc04147::*lacZ*) and SmFL5242 (SMc02616::*gusA*) when retested in trigonelline and RCS exudates (as sole sources of carbon and nitrogen) and glycerol and NH_4Cl (carbon and nitrogen sources, respectively) as a negative control.

As shown in Figure 5-18 SMc02616 is surrounded by several hypothetical conserved (grey) genes. To investigate whether these genes are all part of an operon or at least involved in the uptake or metabolism of trigonelline, chromosomal integrated fusions were built for SMc02619, SMc02618, and SMc02615 using pTH1722 as the cloning vector with LacZ and Gfp as the reporters. Three fusions were then tested with trigonelline as the inducing compound.

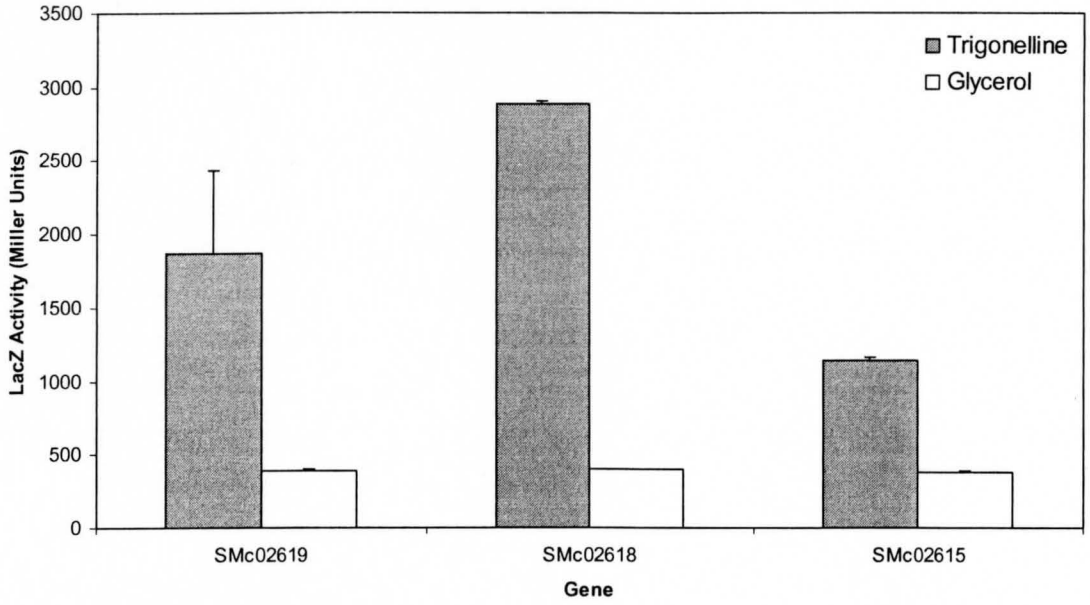

Figure 5-20. β-galactosidase activity from LacZ reporter fusions to the indicated genes, showing specific induction in all three cases by trigonelline.

A previous study has shown that the expression of *S. meliloti* trigonelline catabolism genes are inhibited by the addition of other betaines such as carnitine, choline, and glycine betaine (Boivin, 1990). This finding was applied to the two trigonelline transport systems that were identified in this study. Media containing trigonelline as the sole source of carbon and nitrogen was supplemented with one other betaine (or glycerol as a control) and the expression of the fusion strains were measured (Figure 5-21).

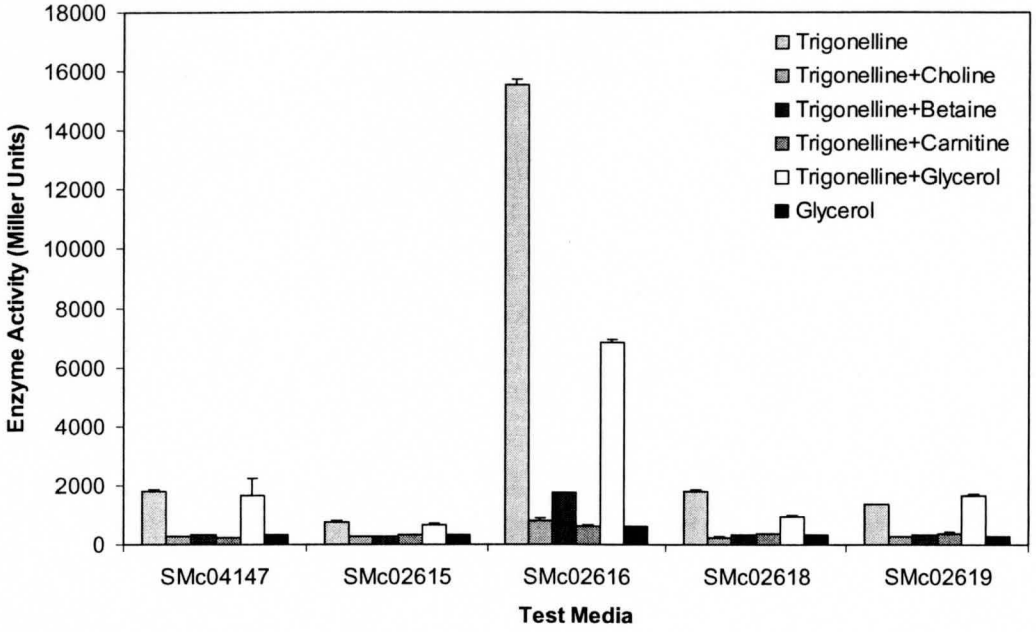

Figure 5-21. β-galactosidase and β-glucuronidase activities of fusions to the indicated genes showing induction only when trigonelline is present in the media without another betaine present.

Results show that the fusion strains have a marked reduction in activity (expression) when another betaine is present in the media, which is consistent with the above mentioned study. Furthermore, to prove that any molecule added to the media does not decrease activity, the addition of glycerol to the media does not have much of an effect on expression.

Another compound of interest is nicotinic acid, due to its structural relatedness to trigonelline (see Figure 5-22). Both transport systems were tested for induction by nicotinic acid, but only SMc04147 was found to show specific induction by both trigonelline and nicotinate.

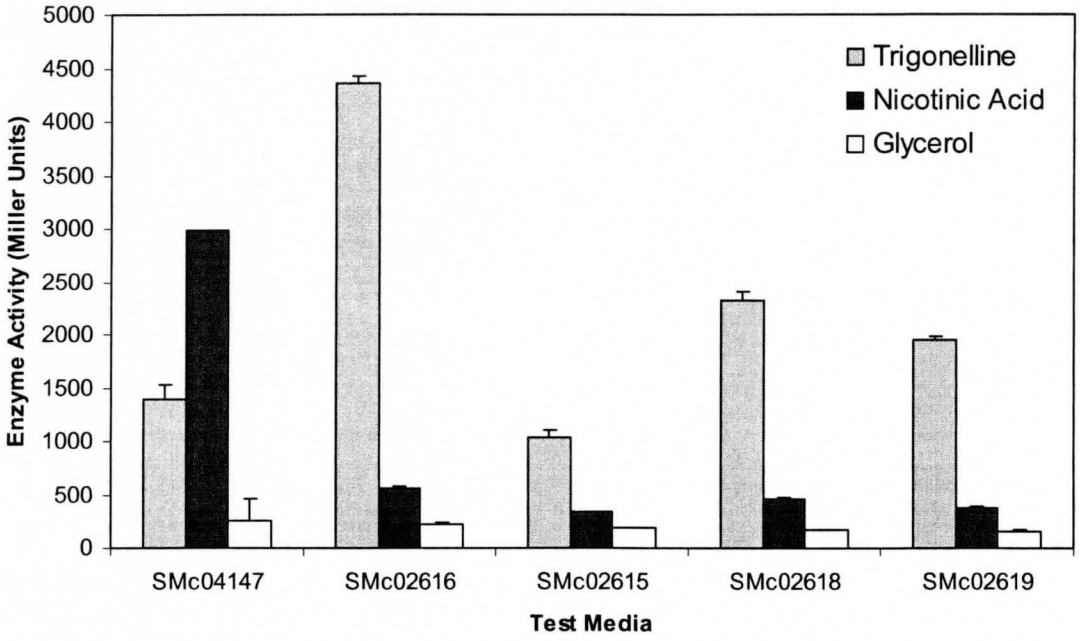

Figure 5-22. β-galactosidase and β-glucuronidase activities of fusions to the indicated genes showing that only the single transport gene SMc04147 is induced by nicotinic acid as well as trigonelline.

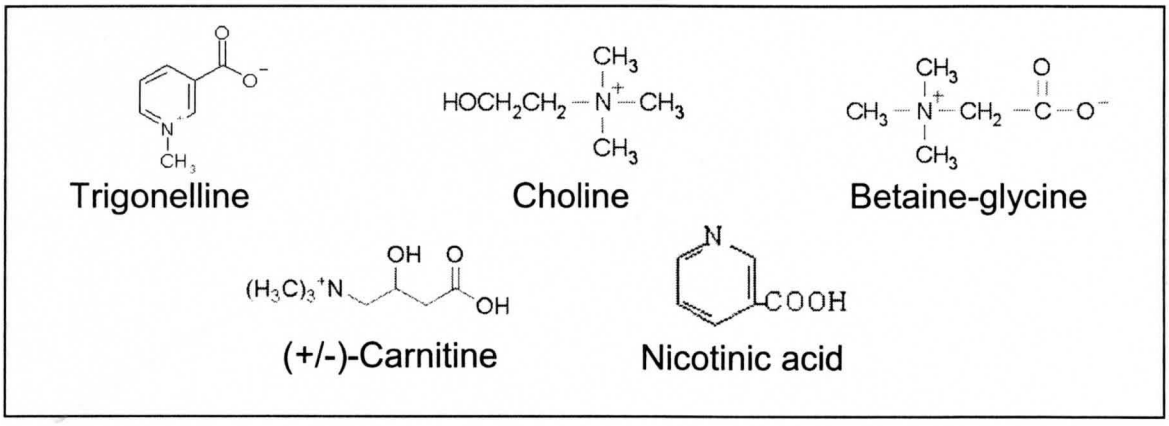

Figure 5-23. Chemical structure of trigonelline and the related compounds that were used in the analysis of the two transport systems induced by trigonelline and RCS exudates (structures taken from Sigma-aldrich website).

Chapter 5.3-3. Discussion

Two separate transport systems were found to be induced by the betaine, trigonelline, and the exudate made from red clover seed. Both SMc02616 (fusion strain SmFL5242) and SMc04147 (fusion strain SmFL4572) are classified by transport DB as amino acid-polyamine-organocation (APC) transporters belonging to the family of major facilitator transport systems and are annotated as having amino acids as their substrates. Though trigonelline does not fit into this category, it is interesting that both transporters induced by trigonelline belong to this family. Another similarity between these systems is that they are completely lacking a regulator. Looking at the genetic maps of both these systems (Figures 5-16 and 5-18) it is noticed that there is no regulator located near either of the transport systems.

Trigonelline has been found to be an inducer of hyphal-branching of mesquite (*Prosopis laevigata*), a semi-arid leguminous plant, during the presymbiotic phase the arbuscular mycorrhizal (AM) fungus, *Gigaspora rosea* (Rojas-Andrade et al., 2003). Furthermore, trigonelline has been identified as a major component of alfalfa seed rinse and an inducer of nodulation gene transcription in *S. meliloti* (Phillips et al., 1992). Therefore it is not surprising that *S. meliloti* would have at least two transport systems dedicated to the transport of such an important molecule. This also corresponds with the identified metabolism genes, which are located adjacent to the *nod-nif* genes of pSymA (Boivin et al., 1990).

It is also not surprising that red clover seed exudate would induce these two transport systems since it is likely that trigonelline is also present in this exudate. Unfortunately alfalfa seed exudate was not only unable to support growth of *S. meliloti*

but seemed to inhibit growth even in the presence of NH_4Cl and 0.5% glycerol as the nitrogen and carbon sources, respectively. Therefore there is no usable data to determine if the exudate would have caused induction of these two transport systems.

SMc02616 is surrounded by hypothetical conserved genes with three being upstream, averaging about 250 nucleotides each in length, and two downstream, averaging about 500 nucleotides in length. There were no clones in the fusion library to any of these small genes, so *lacZ/gfp* fusions were created to three of them and tested for induction in trigonelline. As shown in Figure 5-20 of the results section, all tested fusions showed induction. This may suggest that these are fragmented gene duplications, however ClustalW alignments of the amino acid sequences of these genes do not show there is much similarity between these genes (see Appendix). This may then suggest that these genes are degraded cryptic genes that once participated in the transport and/or metabolism of trigonelline or perhaps still do participate in those functions. When these gene sequences were used in a BLAST search, it was found that the order or proximity of these genes is relatively conserved among *Mesorhizobium loti*, *Magnetospirillum magnetotacticum MS-1* (a water-isolated α-proteobacterium), *Roseovarius nubinhibens ISM* (water-isolated α-proteobacterium, converting dimethylsulfoniopropionate (DMSP) to dimethylsulfide (DMS) (González et al., 2003)), and *Rubrobacter xylanophilus DSM 9941* (gram-positive aquatic thermophile) (Ferreira et al., 1999).

It would be interesting to further investigate these systems by creating a double knockout mutant by deleting both transport systems and testing for the ability of the resulting strain to grow on trigonelline as a sole source of carbon and nitrogen.

Chapter 5.4 Glycerol and Glycerol-3-Phosphate Transport

Chapter 5.4-1. Introduction

The best characterized transporters of glycerol and glycerol-3-phosphate (G3P) are those that have been identified and studied in *E. coli*. The glycerol facilitator of *E. coli*, Glp, is an energy-independent transport system that has been found to be induced also by G3P (Richey and Lin, 1972). This transporter was found to transport glycerol as well as erythritol, pentitols, and hixitols. However, the analogous sugars, erythrose, pentose, and hexose, were not transported by this system (Heller et al., 1980).

In 1964 it was discovered that membranes were not actually impermeable to G3P as was previously thought (Hayashi et al., 1964). An energy-dependent secondary carrier (GlpT) was later identified to be transporting G3P into the cell for the specific utilization of the compound (Larson et al., 1982). This GlpT transporter is an antiporter that exchanges a phosphate ion for a G3P molecule (Lin, 1976). However this is not the only transport system identified as transporting G3P into the cell. The ugp operon, which is highly specific for G3P, is part of the *pho* regulon and is highly induced under phosphate starvation (Argast and Boos, 1980). G3P has also been found to enter the cell via the relatively non-specific hexose phosphate transport system (*uhp*) (Guth et al., 1980).

Chapter 5.4-2. Results

Two separate transport systems, SmFL631 (SMb20345::*gusA*) and SmFL4050 (SMc02516::*gusA*), were found to be induced when glycerol was present in the minimal media as either the sole carbon source or simply present as an alternative carbon source for the culture (Figures 5-24 and 5-26).

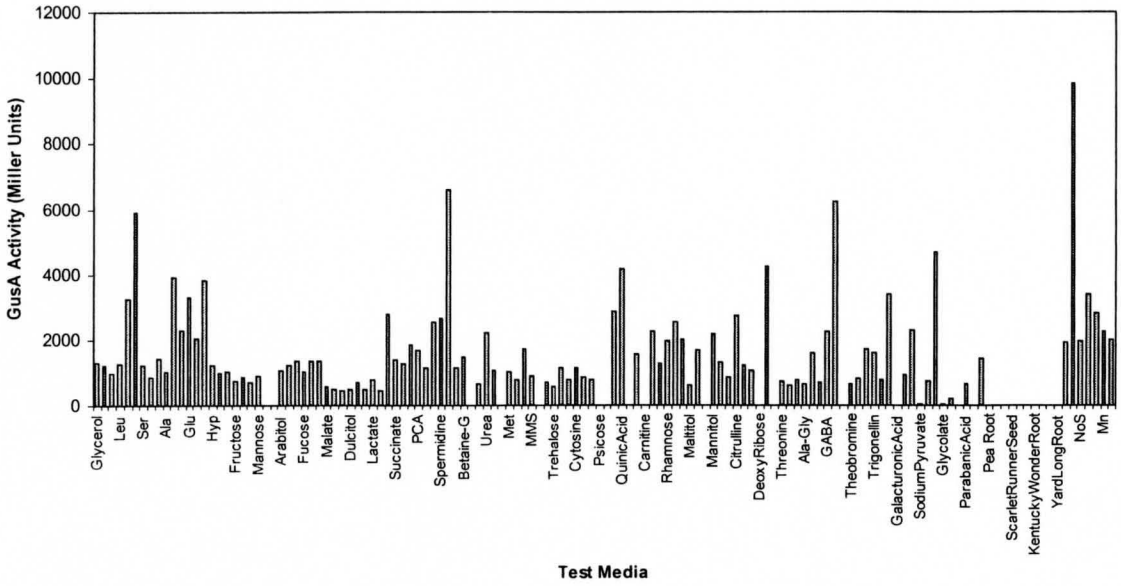

Figure 5-24. β-glucuronidase activity of SmFL631 (SMb20345::*gusA*) when tested for induction in all the test media. This fusion was induced when glycerol was present in the media. Glycerol was added at a concentration of 0.5% to compounds that were used only as a nitrogen source or to those compounds that could only be tested as inducers (see Materials and Methods Table 3-1).

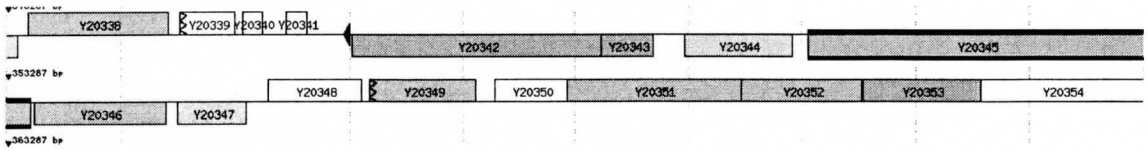

Figure 5-25. Genetic map of the operon iducuced by the presence of glycerol and glycerol-3-phosphate in the test media (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/).

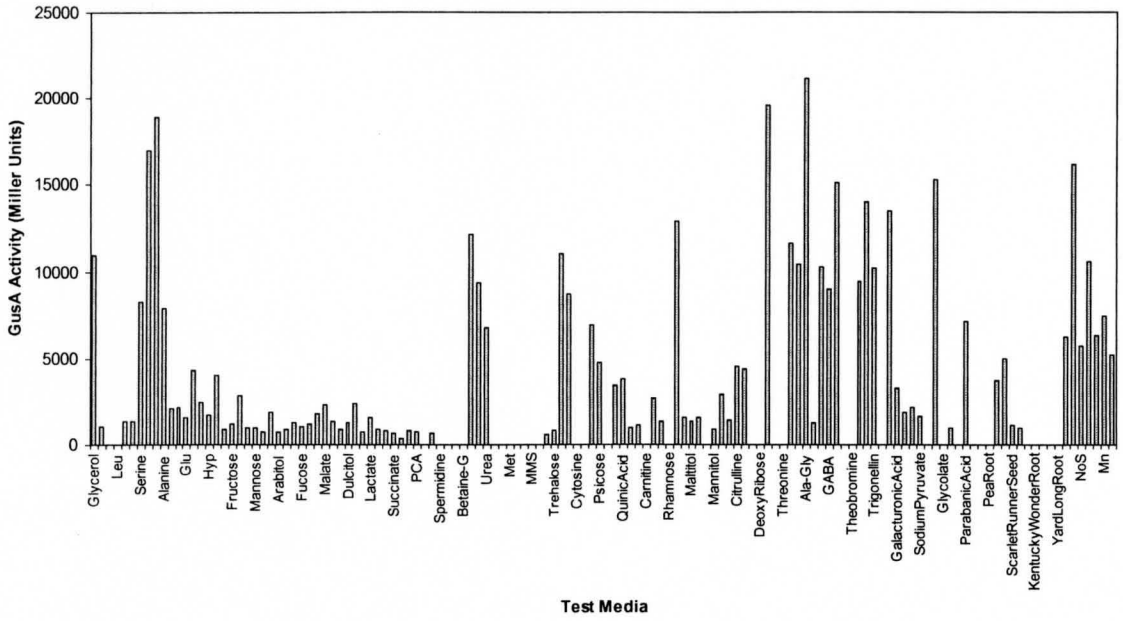

Figure 5-26. β-glucuronidase activity of SmFL4050 (SMc02516::*gusA*) showing induction when glycerol is present in the test media.

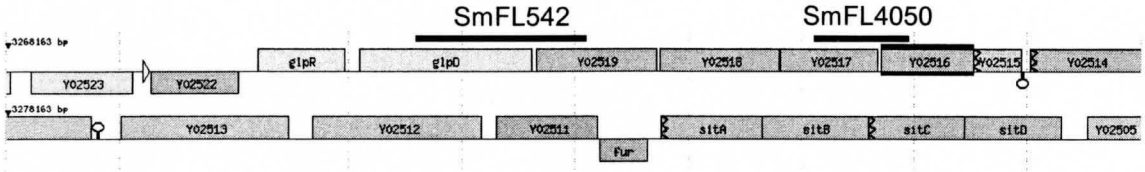

Figure 5-27. Genetic map of the transport system induced by the presence of glycerol and glycerol-3-phosphate in the test media (http://bioinfo.genopole-toulouse.prd.fr/annotation/iANT/bacteria/rhime/). The black bars represent the region that was cloned the indicated fusion.

It seemed likely that these systems, especially SMc02516, because of its proximity to the annotated *glpR* and *glpD* (G3P regulator and dehydrogenase, respectively), would also be induced by glycerol-3 phosphate. Therefore both systems were tested for their induction by this compound. Typically in the testing and retesting, glycerol was used as a negative control but in the case of these two systems, 10 mM glucose was used as the comparison compound. SmFL542 (SMc02519::*lacZ*) was included in the screen along with SmFL4050 (SMc02516::*gusA*) because it is a fusion to

the same operon but it does not seem to create a knock-out as SmFL4050 (SMc02516::*gusA*) does. SmFL542 (SMc02519::*lacZ*) contains the upstream region of SMc02519 and part of *glpD* and thus, as shown in Figure 5-27, this may still create a knockout of the transport system if its expression is not driven by a separate promoter from *glpD and glpR*.

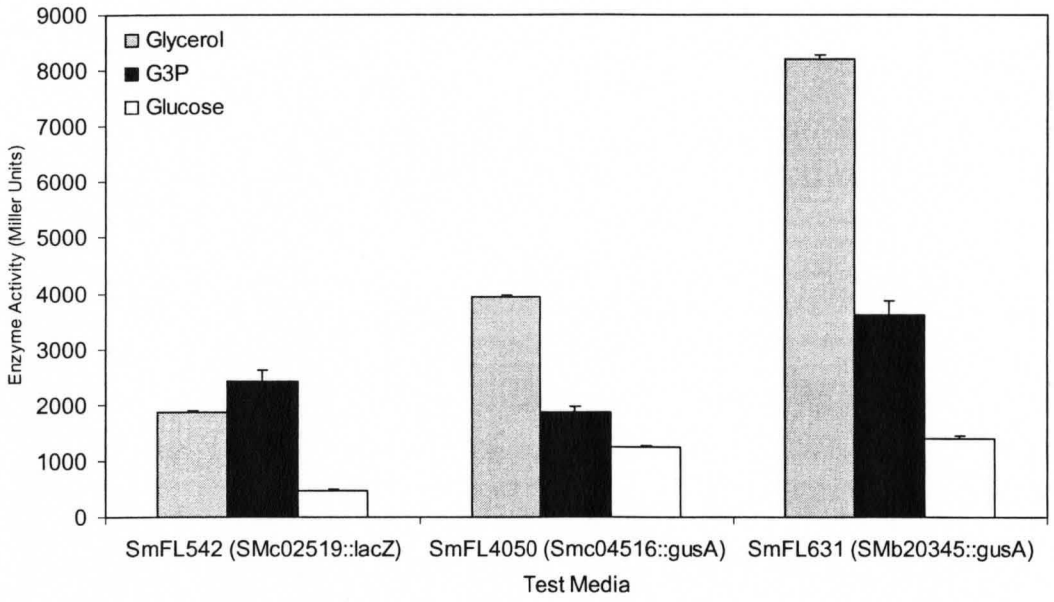

Figure 5-28. β-glucuronidase and β-galactosidase activities of the indicated fusions showing specific induction by glycerol for all three fusions and by glycerol-3-phosphate for SmFL542 (SMc02519::*lacZ*) and SmFL631.

As shown above in Figure 5-28, SmFL542 (SMc02519::*lacZ*) is induced by the presence of glycerol-3-phosphate even though SmFL4050 does not. This finding is most likely a result of the fusion being a knockout of the operon. The discrepancy could also be due to the fusion being to one of the last genes in the operon and therefore will have less complete transcripts. Similarly, SmFL631 (SMb20345::*gusA*) does appear to be moderately induced by this compound as well.

Chapter 5.4-1. Discussion

Two different transporters, SMb20345 and SMc02516, were found to be induced by glycerol. These fusion strains showed induction whenever glycerol was present in the media, whether another carbon or nitrogen source was present as an inducer. SMb20345 (fused to *gusA* in strain SmFL631) is a transmembrane efflux protein that belongs to the Resistance-Nodulation-Cell Division (RND) family of MFS transporters. RND transporters are often associated with periplasmic membrane fusion proteins (MFPs) and outer membrane channels (OMFs). Located directly upstream is SMb20346, which is also a transmembrane efflux protein and has the MFP domain. A fusion strain, SmFL6175 (SMb20346::gusA), to this gene was also found to be induced by glycerol. These two transport proteins are located in between two regulators, SMb20344 and SMb20347. At least one of these two regulators could be involved in the regulation of this transport system.

The second transporter fusion found to be induced by glycerol is SmFL4050 (SMc02516::*gusA*) a fusion to the putative ABC transport system permease subunit. This fusion strain is also induced in nodule extracts (unpublished lab data). Organized in this cluster are four genes encoding two permease subunits and two ATP-binding proteins and upstream of which are the putative glycerol-3-phosphate regulon repressor (*glpR*) and the putative glycerol-3-phosphate dehydrogenase. In *P. aeruginosa* it has been shown that glycerol is transported by a high-affinity binding protein-independent facilitated diffusion system, which is interesting considering the absence of a periplasmic binding protein in this cluster (Williams et al., 1994). Located directly downstream from the transport genes is a hypothetical transmembrane protein.

It was suspected that these two transport systems would also be induced by glycerol-3-phosphate (G3P). Therefore these two systems were tested by adding G3P to minimal media that was also supplemented with pyruvate at a concentration of 10 mM to ensure growth as G3P did not support growth and actually caused some attenuation. Since SmFL4050 was a knock-out of this transport system, SmFL542 (SMc02519::*lacZ*) was included in the retesting of this transporter. Interestingly the latter fusion showed a five-fold induction over glucose background whereas SmFL4050 did not. This is not surprising because due to the knocked-out operon the regulation may not be represented accurately. These results are promising considering the annotation of the upstream metabolism genes, as previously mentioned above, as being involved in the metabolism of G3P. The other transporter, SMb20345 showed a 2-fold induction by the presence of G3P.

There is the *ugpBAEC* operon that is annotated as being an ABC-type transport system with glycerol-3-phosphate being the substrate. Screening of this operon did not show any induction by glycerol as a carbon source (data not shown). However, the Poole group did use G3P in their initial screen and they did not find induction either (Mauchline et al., 2006).

Chapter 6. CONCLUSIONS

This study has demonstrated the value of a random genomic library, employing over 400 strains to nearly as many transport systems. Without the library, this project would not have been quite as successful. As a result of screening all the fusion strains in over 120 different test conditions, inducers were identified for 13% of the 381 transport systems in *S. meliloti*. The largest fraction of transport genes in *S. meliloti* being of the ABC-type (54%), it is not surprising that the majority of transport systems that we found inducers for were also of this family. Perhaps if a wider range of inducing conditions was explored, inducers would have been identified for more of the non-ABC type transport systems. Though some starvation conditions were investigated (Table 3-1 in Materials and Methods) there are an abundance of other conditions that would have been relevant in this study. Perhaps if a wider range of starvation conditions were pursued in this study more inducers would have been identified for secondary transport systems as they are often involved in the transport of ions and small solutes (Leblanc et al., 1989).

The variety of conditions found to induce the transport systems of *S. meliloti* demonstrates the ability to which the bacterium can compete for nutrients in the soil. By having a wide range of sugars, amino acids, organic acids, amino sugars, sugar alcohols that can be transported and metabolized, the bacteria are increasing the competitive edge. Furthermore, by having such transport systems as ABC-type transporters with high affinity periplasmic binding proteins to scavenge and tightly bind a solute, the ability to survive in nutrient-deprived conditions is magnified (Higgins, 1992).

The amount of data generated from this study will prove to be useful for future research endeavors. Not only can one continue to research one of the transport systems

that was explored in the Results section, but one can also choose to characterize a transport system, or set of systems, where only the inducer has been identified. Furthermore, as mentioned above, much more information could be gained by doing additional screening in order to identify the inducers of more of the transport systems.

.

REFERENCES

Alloing, G., Travers, I., Sagot, B., Le Rudulier, D., and L. Dupont. 2006. Proline betaine uptake in *Sinorhizobium meliloti* characterization of Prb, an Opp-like ABC transporter regulated by proline betaine and salinity stress. *J. Bacteriol.* **188(17):**6308-6317.

Argast, M. and W. Boos. 1980. Co-regulation in *Escherichia coli* of a novel transport system for sn-glycerol-3-phosphate and outer membrane protein Ic (e, E) with alkaline phosphatese and phosphate-binding protein. *J. Bacteriol.* **143(1):**142-150.

Ball S.G. and M.K. Morell. 2003. From bacterial cell glycogen to starch: Understanding the biogenesis of the plant starch granule. Ann. Rev. Plant Biology **54**:207-233.

Bardin, S. D., Dan, S., Osteras, M., and T. M. Finan. 1996. A phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti. J. Bacteriol.* **178:**4540-4547.

Barnett, M. J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J., Gurjal, M., Hong, A., Huizar, L., Hyman, R.W., Kahn, D., Kahn, M.L., Kalman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R., Wells, D.H., Yeh, K.C., Davis, R.W., Federspiel, N.A., and S.R. Long. 2001. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc.Natl.Acad.Sci.U.S.A **98**:9883-9888.

Barron, A., Jung, J.U. and M. Villarejo. 1987. Purification and characterization of a glycine betaine binding protein from *Escherichia coli. J. Biol. Chem.* **262(24):**11841-11846.

Bauer, M.W., Bylina, E.J., Swanson, R.V., and R.M. Kelly. 1996. Comparison of a beta-glucosidase and a beta-mannosidase from the hyperthermophilic archaeon *Pyrococcus fusiosus*. Purification, characterization, gene cloning, and sequence analysis. *J. Biol. Chem.* **271(39):**23749-23755.

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

Boncompagni, E., Østerås, M., Poggi, M., and D. le Rudulier. 1999. Occurence of choline and glycine betaine uptake and metabolism in the family *Rhizobiaceae* and their roles in osmoprotection. *Appl Environ Microbiol.* **65(5):**2072-2077.

Boscari, A., Mandon, K., Dupont, Poggi, M., and D. Rudulier. 2002. BetS is a major glycine betaine/Proline betaine transporter required for early osmotic adjustment in *Sinorhizobium meliloti. J. Bacteriol.* **184(10):**2654-2663.

Brandford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72:**248-254.

Bringhurst, R.M., Cardon, Z.G., and D.J. Gage. 2001. Galactosides in the rhizosphere: utilization by *Sinorhizobium meliloti* and development of a biosensor. PNAS **98(8):** 4540-4545.

Buléon, A., Colonna, P., and S. Ball. 1998. Starch granules: structure and biosynthesis. Int. J Biological Macromolecules **23(2):**85-112.

Charles, T. C. and T. M. Finan. 1991. Analysis of a 1600-kilobase Rhizobium meliloti megaplasmid using defined deletions generated in vivo. Genetics **127**:5-20.

Charles, T. C., Newcomb, W., and T. M. Finan. 1991. *ndvF*, a novel locus located on megaplasmid pRmeSU47b (pEXO) of *Rhizobium meliloti* is required for normal nodule development. *J. Bacteriol.* **173:**3981-3992.

Clegg, S., Yu, F., Griffiths, L., and J.A. Cole. 2002. The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters. *Mol Microbiol.* **44:**143-155.

Dattananda, C.S. and J. Gowrishankar. 1989. Osmoregulation in *Escherichia coli*: complementation analysis and gene-protein relationships in the *proU* locus. *J. Bacteriol.* **171(4):**1915-22.

de Bruijn, F.J., Rossbach, S., Schneider, M., Ratet, P., Messmer, S., Szeto, W.W., Ausubel, F.M., and J. Schell. 1989. *Rhizobium meliloti* 1021 has three differentially regulated loci involved in glutamine biosynthesis, none of which is essential for symbiotic nitrogen fixation. *J. Bacteriol.* **171(3):**1673-1682.

Dowling, D.N. and W.J. Broughton. 1986. Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40:** 131-157.

Duerksen, J.D., and H. Halvorson. 1958. Purification and properties of an inducible β-glucosidase of yeast. *J. Biolog. Chem.* **233(5)**1113-1120.

Dupont, L., Garcia, I., Poggi, M., Alloing, G., Mandon, K., and D. le Rudulier. 2004. The *Sinorhizobium meliloti* ABC transporter Cho is highly specific for choline and expressed in bacteroids from *Medicago sativa* nodules. *J. Bacteriol.* **186(18):**5988-5996.

Eichhorn E., van der Ploeg, J.R., and T. Leisinger. 2000. Deletion analysis of the *Escherichia coli* taurine and alkanesulfonate transport systems. *J. Bacteriol.* **182(10):** 2687–2695.

Ferreira, A.C., Nobre, M.F., Moore, E., Rainey, F.A., Battista, J.R., and M.S. de Costa. 1999. Characterization and radiation resistance of new isolates of *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus*. *Extremophiles* **3(4):**235-238.

Finan, T.M., Kunkel, B., De Vos, G.F., and E.R. Singer. 1986. Second symbiotic megaplasmid in Rhizobium meliloti carrying exopolysaccharide and thiamine synthesis genes. *J. Bacteriol.* **167:**66-72.

Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B., and A. Puhler. 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc.Natl.Acad.Sci.U.S.A **98**:9889-9894.

Galibert, F., T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandenbol, F. J. Vorholter, S. Weidner, D. H. Wells, K. Wong, K. C. Yeh, and J. Batut. 2001. The composite genome of the legume symbiont Sinorhizobium meliloti. Science **293**:668-672.

González, J.M, Convert, J.S., Whitman, W.B., Henriksen, J.R., Mayer, F., Scharf, B., Schmitt, R., Buchan, A., Fuhrman, J.A., Kiene, R.P., and A. Moran. 2003. *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments. *Int. J. Syst. Evol. Microbiol.* **53:**1261-1269.

Guth, A., Engel, R., and B.E. Tropp. 1980. Uptake of glycerol-3-phosphate and some of its analogs by the hexose phosphate transport system of *Escherichia coli*. *J. Bacteriol.* **143:**538-539.

Hall, B.G. and P.W. Betts. 1987. Cryptic genes for celobiose utilization in natural isolates of *Escherichia coli*. *Genetics* **115:**431-439.

Hayashi, S., Koch, J.P., and E.C.C. Lin. 1964. Active transport of L-α-glycerolphosphate in *Escherichia coli*. *J. Biological Chemistry* **239(9):**3098-3105.

Higgins, C. F. 1992. ABC transporters: from microorganisms to man. Annu.Rev.Cell Biol **8**:67-113.

Heller, K.B., Lin, E.C.C., and T. Hastings Wilson. 1980. Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. *J. Bacteriol.* **144(1):**274-278.

Jack, D.L., Paulsen, I.T., and M.H. Saier. 2000. The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology* **146(pt 8):**1797-1814.

Jayakumar, A., Schulman, I., MacNeil, D., and E.M. Barnes. 1986. Role of the *Escherichia coli glnALG* operon in regulation of ammonium transport. *J. Bacteriol.* **166(1):**281-284.

Kato, M., and K. Mizuno. 2004. Caffeine synthase and related methyltransferases in plants. *Front Biosci.* **9:**1833–1842.

Kendler, B.S. 1989. Taurine: an overview of its role in preventative medicin. *Prev. Med.* **18(1):**188-198.

Koning, S.M., Elferink, M.G., Konings, W.M., and A.J. Driessen. 2001. Cellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. *J. Bacteriol.* **183(17):** 4979-84.

Kricker, M. and B.G. Hall. 1984. Directed evolution of cellobiose utilization in *Escherichia coli* K12. *Mol. Biol. Evol.* **1(2):**171-182.

Kricker, M. and B.G. Hall. 1987. Biochemical genetics of the cryptic system for cellobiose utilization in *Escherichia coli* K12. *Genetics* **115:**419-429.

Lai X., Davis, F.C., Hespell, R.B., and L.O. Ingram. 1997. Cloning of cellobiose phosphoenolpyruvate-dependent phosphotransferase genes: functional expression in recombinant Escherichia coli and identification of a putative binding region for disaccharides. *Appl. Environ. Microbiol.* **63(2):** 355-63.

Lambert, A., M. Osteras, K. Mandon, M. C. Poggi, and D. Le Rudulier. 2001. Fructose uptake in Sinorhizobium meliloti is mediated by a high-affinity ATP-binding cassette transport system. J.Bacteriol. **183**:4709-4717.

Larson, T.J., Schumacher, G., and W. Boos. 1982. Identification of the glpT-encoded sn-glycerol-3-phosphate permease of Escherichia coli, an oligomeric integral membrane protein. *J. Bacteriol.* **152(3):**1008-1021.

Leblanc, G., Pourcher, T., and M. Bassilana. 1989. Molecular biology and bacterial secondary transporters. *Biochimie.* **71(8):**969-979.

Lin, E.C. 1976. Glycerol dissimilation and its regulation in bacteria. *Annu. Rev. Microbiol.* **30:**535-578.

Linton, K. J. and C. F. Higgins. 1998. The Escherichia coli ATP-binding cassette (ABC) proteins. Mol.Microbiol. **28**:5-13.

Luzhkov, V.B., Almlof, M., Nervall, M., and J. Agvist. 2006. Computational study of the binding affinity and selectivity of the bacterial ammonium transporter AmtB. *Biochemistry* **45(36):**10807-10814.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., and I.S. Pretorius. 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol Mol Biol Rev.* **66(3):**506-577.

Madyastha, K.M. and G.R. Sridhar. 1998. A novel pathway for the metabolism of caffeine by a mixed culture consortium. *Biochemical and biophysical research communications* **249:**178-181.

Madyastha, K.M., Sridhar, G.R., Vadiraja, B.B., and Y.S. Madhavi. 1999. Purification and partial characterization of caffeine oxidase-a novel enzyme from a mixed culture consortium. *Biochemical and biophysical research communications* **263:**460-464.

Mauchline, T.H., Fowler, J.E., East, A.K., Sartor, A.L., Zaheer, R., Hosie, A.H.F., Poole, P.S., and T.M. Finan. 2006. Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *PNAS.* **103(47):**17933-17938.

Meade, H.M., Long, S.R., Ruvkun, G.B., Brown, S.E., and F.M. Ausubel. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of Rhizobium meliloti induced by transposon Tn5 mutagenesis. *J. Bacteriol.* **149:**114-122.

Merrick, M.J., and R.A. Edwards. 1995. Nitrogen control in bacteria. *Microbiol. Rev.* **59(4):**604-622.
Merrick, M., Javelle, A., Durand, A., Severi, E., Thornton, J., Avent, N.D., Conroy, M.J., and P.A. Bullough. 2006. The *Escherichia coli* AmtB protein as a model system for understanding ammonium transport by Amt and Rh proteins. *Transfus. Clin. Biol.* **13(1-2):**97-102.

Milcamps, A., D.M. Ragaz, Lim, P., K.A., Berger, and F.J. de Bruijn. 1998. Isolation of carbon-and nitrogen-deprivation-induced loci of *Sinorhizobium meliloti* 1021 by Tn5-*luxAB* mutagenesis. *Microbiology* **144:**3205-3218.

Nagore, D., Sanz, B., Soria, J., Llarena, M., Llama, M.J., Calvete, J.J., and J.L. Serra. 2006. The nitrate/nitrite ABC transporter of *Phormidium laminosarum*: phosphorylation state of NrtA is not involved in its substrate binding activity. *Biochimica et Biophysica Acta-General Subjects* **1760(2):**172-181.

Old, L.A., Lowes, S., and R.R.B. Russel. 2006. Genomic variation in *Streptococcus mutans*: deletion affecting the multiple pathways of β-glucoside metabolism. *Oral Microbiology Immunology* **21:**21-27.

Phillips, D.A., Joseph, C.M., and C.A. Maxwell. 1992. Trigonelline and stachydrine released from *Alfalfa* seeds activate NodD2 protein in *Rhizobium meliloti. Plant Physiol.* **99(4):**1526-1531.

Pocard, J., Vincent, N., Boncompagni, E., Smith, L.T., Poggi, M., and D. le Rudulier. 1997. Molecular characterization of the *bet* genes encoding glycine betaine synthesis in *Sinorhizobium meliloti* 102F34. *Microbiology* **143:**1369-1379.

Prell, J., Boesten, B., Poole, P., and U.B. Priefer. 2002. The Rhizobium leguminosarum bv. viciae VF39 gamma-aminobutyrate (GABA) aminotransferase gene (gabT) is induced by GABA and highly expressed by bacteroids. *Microbiology* **148:**615-623.

Ramarethinam, S. and N. Rajalakshmi. 2004. Caffeine in tea plants [*Camellia sinensis* (L) O. Kuntze]: in situ lowering by *Bacillus licheniformis* (Weigmann) Chester. *Indian J. Exp. Biol.* **42(6):**575-580.

Ren, Q., K.H. Kang, and I.T. Paulsen. 2004. TransportDB: A relational database of cellular membrane transport systems. Nucleic Acids Research. **32:** D284-D288.

Richey, D.P., and E.C.C. Lin. 1972. Importance of facilitated diffusion for effective utilization of glycerol by *Escherichia coli. J. Bacteriol.* **112(2):**784-790.

Rojas-Andrade, R., Cerda-Garcia-Rojas, C.M., Frias-Hernandez, J.T., Dendooven, L., Olalde-Portugal, V., and A.C. Ramos-Valdivia. 2003. Changes in the concentration of trigonelline in a semi-arid leguminous plant (*Prosopis laevigata*) induced by an arbuscular mycorrhizal fungus during the presymbiotic phase. *Mycorrhiza* **13:**49-52.

Rosenblueth, M., Hynes, M.F., and E. Martinez-Romero. 1998. *Rhizobium tropici teu* genes involved in specific uptake of *Phaseolus vulgaris* bean-exudate compounds. *Mol Gen Genet* **258:**587-598.

Schaefler, S. 1967. Inducible system for the utilization of β-glucosides in *Escherichia coli. J. Bacteriol.* **93(1):**254-263.

Schlösser, A., Jantos, J., Hackmann, K., and H. Schrempf. 1999. Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader Streptomyces reticuli. *Appl. Environ. Microbiol.* **65(6):** 2636-43.

Stahl, Y., Coates, S., Bryce, J.H., and P.C. Morris. 2004. Antisense downregulation of the barley limit dextrinase inhibitor modulates starch granule size, starch composition and amylopectin structure. *The Plant Journal* **39(4):**599-611.

Striling, D.A., Hulton, C.S., Waddell, L., Park, S.F., Stewart, G.S., Booth, I.R. and C.F. Higgins. 1989. Molecular characterization of the proU loci of *Salmonella typhimurium* and *Escherichia coli* encoding osmoregulated glycin betaine transport systems. *Mol. Microbiol.* **3(8):**1025-1038.

Tate, R., Riccio, A., Merrick, M., & Patriarca, E. J. 1998. The Rhizobium etli amtB gene coding for an NH4+ transporter is down-regulated early during bacteroid differentiation. *Mol Plant Microbe Interact.* **11:** 188-198.

Tikhonova, E.B., Wang, Q., and H.I. Zgurskaya. 2002. Chimeric analysis of the multicomponent multidrug efflux transporters from gram-negative bacteria. *J. Bacteriol.* **184(23):** 6499–6507.

Tomii, K. and M. Kanehisa. 1998. A comparative analysis of ABC transporters in complete microbial genomes. Genome Res. **8**:1048-1059.

Unkles, S.E., Hawker, K.L., Grieve, C., Campbell, E.I., Montague, P., and J.R. Kinghorn. 1991. *crnA* encodes a nitrate transporter in *Aspergillus nidulans*. *Proc. Natl. Acad. Sci. USA* **88:**204-208.

Voegeles, R.T., Bardin, S., and T.M. Finan. 1997. Characterization of the *Rhizobium* (*Sinorhizobium*) *meliloti* high- and low-affinity phosphate uptake systems. *J. Bacteriol.* **179(23):**7226-7232.

Watson, R.J., Chan, Y.K., Yang, A.F., and S.H. Han. 1988. *Rhizobium meliloti* genes required for C4-dicarboxylate transport and symbiotic nitrogen fixation are located on a megaplasmid. *J. Bacteriol.* **170(2):**927-34.

Willis, L.B., and G.C. Walker. 1999. A novel *Sinorhizobium meliloti* operon encodes an α-glucosidase and a periplasmic-binding-protein-dependent transport system for α-glucosides. *J. Bacteriol.* **181(14):** 4176–4184.

Winnen, B., Hvorup, R.N., and M.H. Saier. 2003. The tripartite tricarboxylate transporter (TTT) family. *Res. Microbiol.* **154(7):** 457-465.

Wu, Q., and V. Stewart. 1998. NasFED proteins mediate assimilatory nitrate and nitrite transport in *Klebsiella oxytoca* (*pneumoniae*) M5al. *J. Bacteriol.* **180(5):**1311-1322.

Wulff-Strobel, C.R., and D.B. Wilson. 1995. Cloning, sequencing, and characterization of a membrane-associated Prevotella ruminicola B(1)4 beta-glucosidase with cellodextrinase and cyanoglycosidase activities. *Bacteriol.* **177(20):**5884-5890.

Yarosh, O.K., Charles, T.C., and T.M. Finan. 1989. Analysis of C4-dicarboxylate transport genes in *Rhizobium meliloti*. *Mol. Microbiol.* **(6):**813-23.

Yuan, Z., Zaheer, R., and T.M. Finan. Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. *J. Bacteriol.* **188(3):**1089-1102.

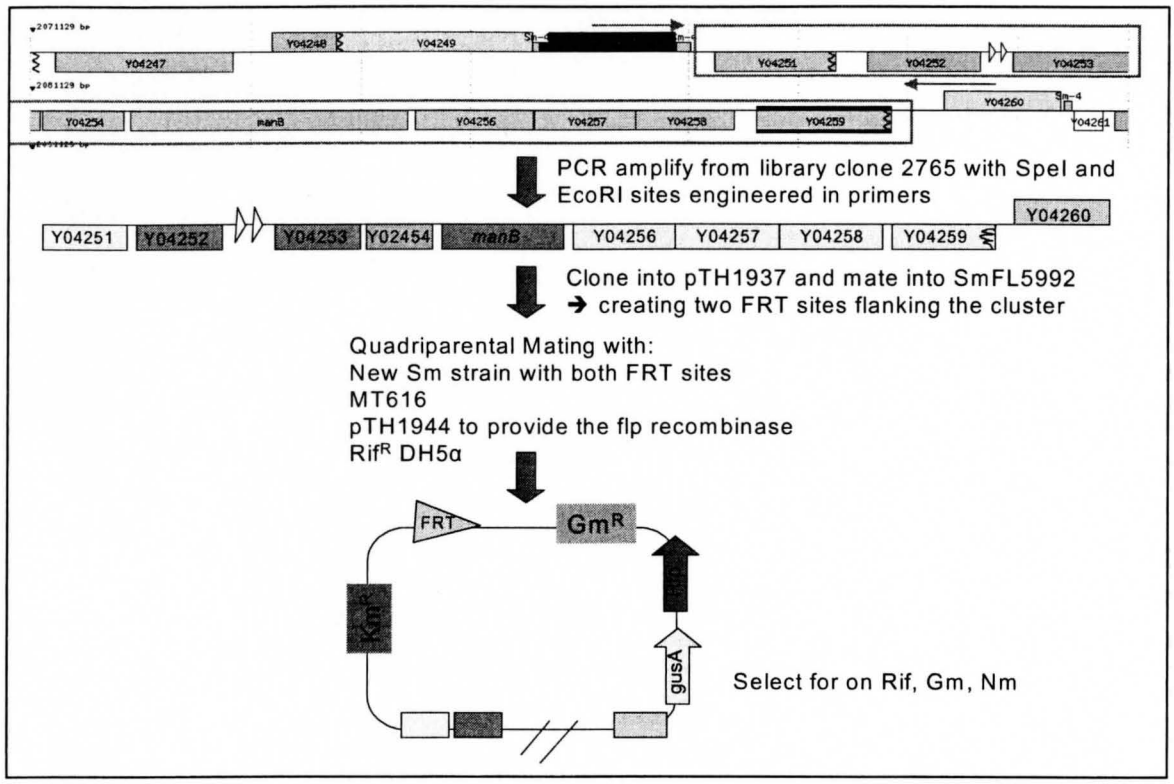

Figure A-1. Schematic diagram illustrating the procedure in creating *E. coli* M1223, carrying the *S. meliloti* genes suspected to be involved in β-glucoside transport and metabolism. See Materials and Methods for description.

Table A-1. Order of compounds used for the X-axis in all the graphs shown throughout the thesis. See Table 3-1 for description of the compounds and the concentrations at which they were used.

Compound	Source	Compound	Source
Glycerol	C	Thymidine	N
Asn	C+N	Thymine	N
Ile	C+N	Cytosine	N
Leu	C+N	Adenine	N
Lys	C+N	Adenosine	N
Orn	C+N	Psicose	C
Ser	N	Alanineamide	N
Val	N	B-Hydroxybutyrate	I
Gly	N	QuinicAcid	I
Ala	N	Uridine	N
Arg	C+N	Lactulose	C
Gln	N	Carnitine	C
Glutamate	N	Lyxose	C
His	C+N	Maltotriose	C
Pro	C+N	Rhamnose	C
Hyp	C+N	Adonitol	C
Asp	C+N	Turanose	C
Palatinose	C	Maltitol	C
Fructose	C	Xylose	C
Arabinose	C	Malonate	C
Sucrose	C	Mannitol	C
Mannose	C	Pyruvate	C
Myo-Inositol	C	Salicin	C
Sorbose	C	Citrulline	N
Arabitol	C	Putrescine	N
Ribose	C	PropionicAcid	N
Galactosamine	C	DeoxyRibose	C
Fucose	C	Xanthosine	N
Glucosamine	N	Canavanine	C
MethylPyruvate	C	Thr	N
Malate	C	GlyGlu	N
Melibiose	C	GlyAsp	N
Cellobiose	C	AlaGly	N
Dulcitol	C	Lactose	C
Dextrin	C	Xanthine	N
Fumerate	C	GABA	N
Lactate	C	Stachyose	I
Raffinose	C	Gluconate	C
Glucose	C	Theobromine	N
Succinate	C	Inosine	N
Xylitol	C	GlycylGlycin	N
Maltose	C	Trigonellin	C+N
PCA	C	Caffeine	N
a-ketoglutarate	I	HydroxyTryptophan	I
Tagatose	C	GalacturonicAcid	C
Spermidine	I	Taurine	C+N

POB	I	Erythritol	C
Choline	N	SodiumPyruvate	C
Glycine Betaine	N	Talose	C
Allantoin	N	Agmatine	N
Uracil	N	Glycolate	C
Urea	N	Galactose	C
Phe	N	Glutarate	C
Tyr	N	ParabanicAcid	C
Met	N	Mg limitation (0 mM)	
Cys	N	Ca Limitation (0 mM)	
Gentiobiose	C	Sulphur limitation (0 mM)	
MMS	C	Excess Fe (50 µM)	
DeoxyAdenosine	N	Escess Zn (2.5 µM)	
Sorbitol	C	Excess Mn (2.5 µM)	
Trehalose	C	Trace Elements	see materials and methods

Bradford Standard Curve Data for β-glucosidase Assay

Protein (μg/μL)	Absorbance (O.D. 595)
0	0.5013
2	0.6813
4	0.8017
6	0.9127
8	1.0156

Equation of the Line: $y = 0.063x + 0.5305$

	Average Absorbance OD595	mg protein/uL
Glycerol	0.6242	0.014873016
Cellobiose	0.7325	0.032063492
Salicin	0.6434	0.017915344

Table A-2. List of all the genes and their associated fusions included in the screen that was not included in the Mauchline et al. publication (2006). The family that each transport gene is a member of is listed as well. Some genes were included in this screen that were either hypothetical conserved genes not included in the Transport DB classification or are metabolism genes associated with predicted transport systems.

Gene	Fusion	Family	Gene	Fusion	Family
sma0185	7022	MFS	smc00317	7008	AEC
sma0224	1084	MFS	smc00350	7058	MFS
sma0383	7062	MFS	smc00381	7049	DMT
sma0627	4612	MIP	smc00422	7009	RhtB
sma0630	4547	MscS	smc00422	7066	RhtB
sma0675	7001	CaCA	smc00423	7037	RhtB
sma0675	7063	CaCA	smc00428	186	DMT
sma0677	1689	APC	smc00476	4182	SulP
sma0677	1689	APC	smc00498	6471	TTT
sma0682	4903	APC	smc00537	7010	MFS
sma0683	54	APC	smc00536	536	MFS
sma0684	535	APC	smc00564	7011	MFS
sma0830	3318		smc00642	4001	DMT
sma0875	97	RND	smc00744	834	MFS
sma0937	2040	MscS	smc00808	3545	CHR
sma1008	7024		smc00813	2	MFS
sma1153	6159	P-ATPase	smc00827	4136	PiT
sma1155	7002	P-ATPase	smc00868	5319	F-ATPase
sma1328	7025	MFS	smc00873	2325	KUP
sma1447	4563	MFS	smc00874	824	MIT
sma1538	781	CPA3	smc00898	7051	CPA2
sma1541	974	CPA3	smc00922	580	NCS1
sma1600	4016	CPA2	smc00937	2813	TrK
sma1641	1190	MFS	smc00954	P228	
sma1662	1751	RND	smc00978	7057	MIT
sma1667	274	APC	smc01141	7012	PTS
sma1668	3335	APC	smc01211	7052	MOP
sma1691	1505	TrK	smc01212	5381	MFS
sma1697	5509		smc01217	2613	MFS
sma1798	2973	KUP	smc01261	3149	MgtE
sma1814	3514	MFS	smc01361	2085	
sma1913	4961	NhaA	smc01368	5379	MFS
sma1916	2853	DASS	smc01457	2957	RND
sma1937	2344	MFS	smc01584	3595	DMT
sma1959	807	MFS	smc01597	1629	APC
sma2337	5272	MFS	smc01600	4016	
sma2377	4088	MFS	smc01729	4154	DMT
smb20025	4077	TTT	smc01829	7013	RND
smb20027	3291	TTT	smc01869	622	MFS
smb20027	P230	TTT	smc01870	622	MIP
smb20030	P190		smc01970	4977	DMT
smb20069	262	AGCS	smc02057	2733	RND
smb20070	7046	SulP	smc02065	1371	TAT
smb20071	1059	MFS	smc02066	2322	TAT

smb20112	265	DMT
smb20112	P212	DMT
smb20128	2321	
smb20134	458	NCS2
smb20153	2589	PNaS
smb20268	6491	
smb20272	154	MFS
smb20289	7027	NCS2
smb20299	661	
smb20333	2487	BCCT
smb20345	631	RND
smb20354	P219	
smb20361	2301	VIC
smb20402	3133	
smb20433	P225	
smb20436	602	MFS
smb20625	2493	MFS
smb20697	P210	MFS
smb20701	7029	MFS
smb20705	7004	DMT
SMb20716	1274	DMT
smb20724	3347	TTT
smb20771	3481	TTT
smb20863	7005	MscS
smb20999	7033	PUP
smb21050	7006	MOP
smb21162	7034	MFS
smb21169	318	MFS
smb21251	4275	MscS
smb21281	7035	NCS2
smb21424	7065	
smb21424	P207	
smb21486	3579	MFS
smb21498	3146	RND
smb21507	7061	RhtB
smb21512	105	MFS
smb21536	7007	PNaS
smb21555	5693	CPA2
smb21575	7060	MFS
smb21578	3376	P-ATPase
smc00028	4438	MscS
smc00044	3814	RhtB
smc00172	2136	MFS
smc00196	2030	
smc00233	135	PNaS
smc02067	4354	TAT
smc02141	2307	
smc02161	1665	MFS
smc02224	1578	CaCA
smc02250	7014	MscL
smc02265	1516	RND
smc02343	7054	MFS
smc02437	7059	PTS
smc02484	7039	RhtB
smc02511	4876	DMT
smc02603	1239	MFS
smc02616	5242	APC
smc02648	2355	MIP
smc02724	3225	CDF
smc02753	7040	PTS
smc02793	P216	
smc02814	7055	MFS
smc02855	556	DMT
smc02861	4512	PiT
smc02867	7015	RND
smc02888	7071	MFS
smc02889	4616	MFS
smc02892	7041	MFS
smc02895	3779	AEC
smc02907	2512	RhtB
smc02910	1043	MFS
smc02981	7070	RhtB
smc03168	7018	MFS
smc03179	1178	CPA3
smc03237	7042	MFS
smc03277	7043	MFS
smc03807	3396	Amt
smc03824	7044	
Smc03827	1090	RhtB
smc03971	746	RND
smc04128	5864	P-ATPase
smc04147	4572	APC
smc04167	7021	CDF
smc04167	214	CDF
smc04179	7019	DASS
smc04179	5514	DASS
smc04362	7067	MFS
smc04404	2556	RhtB
smc04407	1286	MFS

ClustalW alignment of the MFS transporter, SMc02616, and the surrounding conserved hypothetical (grey) genes (see Figure 5-16 for genetic map).

```
SMc02616        MDDTTSAAPEPDRLRLLRVLGPAHVWALGVGIVLVGEYMGWNFSVGKGGMIAGLMACWVA 60
SMc02618        ------------------------------------------------------------
SMc02617        ------------------------------------------------------------
SMc02614        ------------------------------------------------------------
SMc02619        ------------------------------------------------------------
SMc02615        ---------------------------------------------MITTIAFYAFAIA 13

SMc02616        GLLYTCVAMIDSEVTSTVAAAGGQYAQAKHIVGPLMAFNVGLFLVMAYTMLEAANAITVG 120
SMc02618        ------------------------------------------------------------
SMc02617        ------------------------------------------------------------
SMc02614        ------------------------------------------------------------
SMc02619        ------------------------------------------------------------
SMc02615        SGIWMSVRAVRDRRAAIAERRGLLDDAARHFPGARITHGADHFPILAGRLDDGRQVRVEL 73

SMc02616        FLLDTVAGMQGQTGLNQQPFIVLAIMFLAWLNYRGVLATLTFNLVITAIAFLAIVALFVS 180
SMc02618        ---------------------------MAYEPPKQSFAGQIFDVIT----------LLV 22
SMc02617        ------------------------------------------------------------
SMc02614        ----------------------------MAKPINPYLVLMLAAVLP------------- 18
SMc02619        ----------------------MEATLSKLGEDASFSQRIDAMHRRD------------ 25
SMc02615        VPDTLVCRRLPQLWLKLTLLETSPCARLKIGALARPTGAEFYSLVHEMP----------- 122

SMc02616        VQFGASAVPLDFSAITSDPLPYGWVGIVASLHFGLWYYLGIEGTCQAAEEVRSPARSLPY 240
SMc02618        LNIGALYIPL---------------------------YLGLAGAAKVPDPIPDPT----- 50
SMc02617        ------------------------------------------------------------
SMc02614        ---GAG-------------------------------HVALRDAARG------------- 31
SMc02619        ------------------------------------RVCLTAFVVVLWCT--------- 39
SMc02615        ------------------------------------HLLIPPPSGAALLMRGDGN---- 141

SMc02616        GTMAGIMTLLIAATMTWYICSGLMPWEYLGQAGTPLFDAARVTGSTGLMVLLFVGTAFAT 300
SMc02618        -------------------------WEALGQNAT--------------------------- 59
SMc02617        ------------------------MWDIL------------------------------- 5
SMc02614        ----------------------LAFAFF------------------------------- 37
SMc02619        ----------------------LLFALF------------------------------- 45
SMc02615        -------------ASRRQVERAAAMFAKLFADPT-------------------------- 162
                                      :  :

SMc02616        LASANGCINDASRAWFSMSRDRYLPSWFGAVHPVYRTPYRAIVFLVPIALIFALGAPLDQ 360
SMc02618        ----------EQQQWAALG-----------------------------------ITDP 72
SMc02617        ----------EYAAW---------------------------------------- 10
SMc02614        ------------VVFFS------------------------------------- 42
SMc02619        ------------TVWPY------------------------------------- 50
SMc02615        ---------LKEAAITPRG-----------------------------------V 173

SMc02616        VVTFSILSGLLGYTFMTFNMVMFRNKWPLGRIKRGYVHPFHPLPTVVLLILCSTAYFAVF 420
SMc02618        AAANDIITARFDYSFS-------------------------WASLIVMAVLVIGYFVMV 106
SMc02617        -----ALSALFG-----------------------------VLMLADLVRIDTTYDNEL 35
SMc02614        -----VITYMTAP----------------------------PDRSFIGRHAGGIFVWA 67
SMc02619        -IATPAIAVILT-----------------------------VACGLVLLFNTAAIVAM 78
SMc02615        RLVRQAAQGQRGAHLLLR-----------------------QAHFSITAIAPEVIRRT 208
                                                         :

SMc02616        LGYGTQLSAMMCFYIVASLWFHFRRYKFVRRGDQFTMPWPKPHGY 465
SMc02618        VRLSDRE-----YREVIEERFGTERH------------------ 127
SMc02617        L-ISSRE---------GEIEATAERHEI----------------- 53
SMc02614        LSIPDAYR-------RARIRTVMARKS------------------ 87
SMc02619        LRHYEEDK---HFIYSLDLKHLDEMRPQSR--------------- 105
SMc02615        IAEAEALSG-----WLADDEPAYSLPPSAGAFPEVFSGFRTGSA- 247
                :
```