STUDY OF SAPROPHYTIC COMPETENCE IN *SINORHIZOBIUM MELILOTI*

STUDY OF SAPROPHYTIC COMPETENCE IN *SINORHIZOBIUM MELILOTI*

By

ALLYSON M. MACLEAN, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2008) McMaster University

(Biology) Hamilton, Ontario

TITLE: The study of saprophytic competence in *Sinorhizobium meliloti*

AUTHOR: Allyson M. MacLean, B.Sc. (McMaster University)

SUPERVISOR: Professor Turlough M. Finan

NUMBER OF PAGES: xiv, 280

ABSTRACT

This thesis details a study of saprophytic competence in the Gram-negative bacterium *Sinorhizobium meliloti*, and comprises three main areas of research. The β–ketoadipate pathway is required for the catabolism of a wide range of aromatic compounds that are released into soil through the degradation of lignin. We demonstrate that *S. meliloti* encodes enzymes associated with the protocatechuate branch of the β–ketoadipate pathway within two operons (*pcaDCHGB* and *pcaIJF*) whose expression is regulated by the LysR-protein PcaQ and the IclR-type regulator PcaR, respectively. We show that purified PcaQ recognizes a motif with partial dyad symmetry (5' ATAACC-N_4-GGTTAA 3') positioned upstream of the *pcaD* promoter, and that this site is required for the regulated expression of *pcaD in vivo*. We report that PcaQ also regulates the expression of a protocatechuate-inducible ABC-type transport system that we infer is involved in the uptake of this aromatic acid, and we extend this analysis to identify PcaQ-binding motifs in the genomes of α-, β-, and γ-proteobacteria.

In addition to protocatechuate, *S. meliloti* may utilize hydroxyproline as an energy source, as this amino acid is released into soil during the natural decay of plant tissue. We demonstrate that *S. meliloti* encodes a hydroxyproline-inducible ABC-type transport system that mediates the uptake of *trans*-4-hydrox-L-proline, as determined via growth and transport assays.

As a more comprehensive method of examining saprophytic competence, we assayed the growth of *S. meliloti* upon inoculation into sterile bulk soil. We screened 40

S. meliloti strains carrying deletions within the pSymA or pSymB megaplasmids for growth in soil, and report that the majority of strains establish a stable population ($\geq 10^8$ cells g^{-1} soil) that persists for several weeks. In contrast, two *S. meliloti* strains exhibited a decreased ability to colonize soil, indicating that loci within the deleted regions play a role in saprophytic competence.

ACKNOWLEDGEMENTS

I wish to offer a sincere thanks to my supervisor Dr. Turlough Finan for his support, advice, and assistance, during my years as a graduate student. I am particularly grateful for the freedom and encouragement he extended in allowing me to pursue my numerous (and often unrelated) projects in the lab – I would not have gained as much experience without this support, nor would I have enjoyed myself as much as I have.

I gratefully acknowledge the guidance and support of my supervisory committee, Dr. Elizabeth Weretilnyk and Dr. Christian Baron over the last five years. To this I must also add thanks to Dr. Dick Morton for his input as offered during many lab meetings.

I thank the many members of the Finan lab that I have been fortunate to work with, including Branka Poduska, Dr. J. Cheng, Alison Cowie, Chris Sibley, Ye Zhang, Dr. Catharine White, Vladmir Jokic, Michelle Anstey, Ann Kim and Derrick Leach. I also thank Jane Fowler, Andrea Sartor, and Laura Smallbone for their friendship outside of the lab.

Two individuals have played a key role in my education as a scientist, and deserve special recognition. I thank Dr. Shawn MacLellan and Dr. Rahat Zaheer for the many hours of helpful advice and encouragement they have offered. In particular, I owe Shawn enormous thanks for his support in all matters relating to pigeons and kittens.

I wish to thank my girls Abha Ahuja, Maria Abou Chakra, Danya Lottridge, Melanie Huntley (Sr.), Melanie Lou (Jr.), and Andrea Morash for many coffee breaks and much laughter. I thank Dr. Wilfried Haerty for his amazing support and kindness (and patience) while writing this thesis.

Finally, I wish to thank my family, Audrey Webb, Douglas MacLean, and Aaron MacLean for their emotional (and financial) support over these years, and for always wanting to know 'exactly what is Ally working on'. It is to my grandparents, Stella MacLean and David N. Webb, that I dedicate this work.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGEMENTS v

LIST OF FIGURES ix

LIST OF TABLES xii

LIST OF ABBREVIATIONS xiv

CHAPTER 1

Literature Review 1

Microbial Survival and Growth in a Soil Habitat 2

Nutrient Sources Available in Soil 2

Aromatic Acids and Other Lignin-derived Compounds 2

Neutral and Amino Sugars 5

Nucleic Acids 6

Amino Acids 6

The Rhizosphere is a Unique and Dynamic Soil Environment 8

Soil as a Stressful Environment 9

Factors Affecting Bacterial Soil Diversity 11

The β–ketoadipate Pathway 14

Distribution of Aromatic Acids in Soil 14

The β-ketoadipate Pathway: a Historical Perspective 15

The Protocatechuate Branch of the β–ketoadipate Pathway 16

Protocatechuate Catabolism in *Rhizobiaceae* 19

Protocatechuate Catabolism in *Roseobacter* 21

Protocatechuate Catabolism in *Acinetobacter* 21

Protocatechuate Catabolism in *Pseudomonas* 22

The Catechol Branch of the β-ketoadipate Pathway 23

Characteristics of Aromatic Acid Catabolism Genes:

Gene Organization and Regulation 25

Aromatic Acid Transport Mechanisms 27

Aromatic Acids Induce a Chemotactic

Response in Motile Bacteria 28

Biodegradative Pathways Associated with the

β–ketoadipate Pathway 29

The Metabolism of Hydroxyproline in Bacteria 31

Hydroxyproline-rich Proteins in Plants and Legumes 31

The Hydroxyproline Metabolic Pathway in Bacteria 33
Hydroxyproline Transport in *Pseudomonas* 36
The Genetics and Regulation of Hydroxyproline Uptake and Catabolism 37
The Relationship Between Hydroxyproline and Proline 38
References 41
Figures 58

CHAPTER 2
Characterization of the β-ketoadipate pathway in *Sinorhizobium meliloti*. 63
Preface 64
Abstract 65
Introduction 66
Materials and Methods 67
Results and Discussion 74
References 86
Tables and Figures 93

CHAPTER 3
Binding site determinants for the LysR-type transcriptional regulator PcaQ in the legume endosymbiont *Sinorhizobium meliloti*. 103
Preface 104
Abstract 105
Introduction 106
Materials and Methods 107
Results 113
Discussion 117
References 122
Tables and Figures 128

CHAPTER 4
Regulation of an ABC-type transport system by the protocatechuate LysR-protein PcaQ in *S. meliloti*. 138
Preface 139
Abstract 140
Introduction 141
Materials and Methods 142
Results 148
Discussion 154
References 162
Tables and Figures 167

CHAPTER 5
The legume endosymbiont *Sinorhizobium meliloti* encodes an ATP-binding cassette (ABC) transport system involved in the uptake of hydroxyproline. 182
Preface 183
Abstract 184
Introduction 185
Materials and Methods 187
Results 194
Discussion 199
References 205
Tables and Figures 212

CHAPTER 6
The growth of *Sinorhizobium meliloti* in sterile, bulk soil: A study of saprophytic competence 223
Preface 224
Abstract 225
Introduction 226
Materials and Methods 228
Results 233
Discussion 237
References 247
Tables and Figures 253
Appendix 270

Conclusions 278

LIST OF FIGURES

Figure 1.1 Compounds which may be metabolized through the β-ketoadipate pathway 58

Figure 1.2 Schematic depiction of genes involved in the catabolism of the aromatic acid protocatechuate via the β–ketoadipate pathway 59

Figure 1.3 The β–ketoadipate pathway is involved in the conversion of protocatechuate and catechol to tricarboxylic acid cycle intermediates....60

Figure 1.4 Stereoisomers of hydroxyproline and L-proline 61

Figure 1.5 Metabolism of *trans*-4-hydroxy-L-proline, as described in *Pseudomonas*..62

Figure 2.1 Protocatechuate catabolism in *S. meliloti*....99

Figure 2.2 Analysis of the *pcaD* promoter in *S. meliloti*....100

Figure 2.3 Purification of β–ketoadipate succinyl-CoA transferase in *S. meliloti*.....101

Figure 2.4 Analysis of the *pcaI* promoter in *S. meliloti*....102

Figure 3.1 Schematic diagram of the 94-bp intergenic region located between *pcaQ* and *pcaDCHGB* on the pSymB megaplasmid of *S. meliloti*....133

Figure 3.2 Electrophoretic mobility shift assay for PcaQ binding to the *S. meliloti* intergenic region....134

Figure 3.3 Identification of the sequence involved in PcaQ binding within the *S. meliloti pcaDQ* intergenic region....135

Figure 3.4 Alignment of *pcaDQ* intergenic regions of S. *meliloti, R. leguminosarum, A. tumefaciens, M. loti,* and *R. etli*....136

Figure 3.5 Effect of site-directed mutations within a putative PcaQ binding site on the ability of PcaQ to bind upstream of *pcaD* in *S. meliloti*....137

Figure 4.1 Analysis of expression (A) and organization (B) of genes encoding a putative ABC-type transport system involved in the uptake of the aromatic acid protocatechuate in *S. meliloti* 175

Figure 4.2 Identification and analysis of a promoter associated with an ABC-type transport system encoded by *smb20568-smb20784*. 176

Figure 4.3 Electrophoretic mobility shift assay for PcaQ binding to the intergenic region upstream of *smb20568* 177

Figure 4.4 Identification of a putative PcaQ-binding site located upstream of *smb20568* in *S. meliloti* 178

Figure 4.5 Analysis of *smb20568*::*gfpuv* expression in *S. meliloti* wild-type and *pcaQ*::Ω strains 179

Figure 4.6 Analysis of the PcaQ-binding motif 180

Figure 4.7 A comparison of a directed mutagenesis of PcaQ-binding sites upon the regulation of *pcaD* and *smb20568* expression in *S. meliloti* 181

Figure 5.1 Hydroxyproline transport in *S. meliloti* 217

Figure 5.2 Growth of *S. meliloti* strains subcultured into M9-minimal medium with (A) 10 mM succinate; (B) 5 mM *trans*-4-hydroxy-L-proline; or (C) 5 mM *cis*-4-hydroxy-D-proline as sole carbon sources 218

Figure 5.3 Analysis of the *hypM* promoter 219

Figure 5.4 Uptake of labeled *trans*-4-hydroxy-L-proline (nmols/mg protein) in *S. meliloti* 220

Figure 5.5 Histochemical staining for β–glucuronidase activity in alfalfa nodules 221

Figure 5.6 Phylogenetic tree of GntR family member proteins 222

Figure 6.1 Growth of *S. meliloti* wild-type strain RmP110 in a soil microcosm 261

Figure 6.2 Influence of soil moisture content upon growth of *S. meliloti* wild-type strain RmP110 in a soil microcosm 262

Figure 6.3 Growth of *S. meliloti* wild-type strain RmP110 in a sterile soil microcosm.... 263

Figure 6.4 Growth of *S. meliloti* pSymA deletion strains in a sterile soil microcosm.... 264

Figure 6.5 Growth of *S. meliloti* pSymB deletion strains in a sterile soil microcosm.... 265

Figure 6.6 Growth of *S. meliloti* pSymB deletion strains in a sterile soil microcosm.... 266

Figure 6.7 Growth of *S. melilot* strains in sterile soil 267

Figure 6.8 Schematic depiction of deletions generated in *S. meliloti* strains 268

Figure 6.9 A putative purine catabolic pathway in *S. meliloti*.....269

LIST OF TABLES

Table 2.1 Bacterial strains and plasmids used in this study 93

Table 2.2 Expression of *pcaD-lacZ* fusion in *S. meliloti* strains 95

Table 2.3 β-ketoadipate succinyl-CoA transferase activity in the presence of adipate 96

Table 2.4 Expression of *pcaF-gusA* fusion in *S. meliloti* wild-type and PcaR⁻ backgrounds 97

Table 2.5 Expression of *pcaR-gusA* fusion in *S. meliloti* wild-type and PcaR⁻ backgrounds 98

Table 3.1 Bacterial strains and plasmids used in this study 128

Table 3.2 Analysis of expression of *pcaD::gfpuv* in *S. meliloti* 130

Table 3.3 Quantification of PcaQ binding to the *pcaD* regulatory region *in vitro* 131

Table 3.4 Expression of *pcaQ::lacZ* in *S. meliloti* 132

Table 4.1 Strains and plasmids used in this study 167

Table 4.2 Expression of *pca* genes in wild-type and regulator-minus strains of *S. meliloti* 170

Table 4.3 Expression of *smb20568* as regulated by PcaQ·His in *S. meliloti* 171

Table 4.4 Analysis of expression of *smb20568::gfpuv* in *S. meliloti* 172

Table 4.5 Description of putative PcaQ-binding motifs in Proteobacteria 173

Table 5.1 Strains and plasmids used in this study 212

Table 5.2 Expression of *hypM::gfp+/lacZ* in *S. meliloti* wild-type and *hypR*::Ω backgrounds 214

Table 5.3 Effect of competitors on *trans*-4-hydroxy-L-proline uptake 215

Table 5.4 Expression of *hypM::gusA* in young and senescent alfalfa root nodules...216

Table 6.1 Bacterial strains and plasmids used in this study 253

Table 6.2 Physiochemical properties of soil used in this study.....256

Table 6.3 Growth of *S. meliloti* strains in sterile soil microcosms.....257

Table 6.4 Description of genes deleted in *S. meliloti* strain RmP801 (ΔB116) 258

Table 6.5 Growth of *S. meliloti* strains inoculated into a sterile soil microcosm.....259

Table 6.6 Description of genes deleted in *S. meliloti* strain RmP798 (ΔB122) 260

LIST OF ABBREVIATIONS

Abs	absorbance
bp	base pair
dpi	days post inoculation
DTT	dithiothreitol
EDTA	ethylenedinitrilotetraacetic acid
Gfp	green fluorescent protein
His	hexahistidine tag
Hyp	hydroxyproline
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase
kGy	kilogrey
LB	luria bertani
nt	nucleotide
NT	not tested
OD	optical density
Pca	protocatechuate
PCR	polymerase chain reaction
Pob	*p*-hydroxybenzoate
SD	standard deviation
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
Thi	thiamine

CHAPTER ONE

Literature Review

MICROBIAL SURVIVAL AND GROWTH IN A SOIL HABITAT

1.1.1 Nutrient sources available in soil

Soils may offer a rich source of diverse organic molecules such as polysaccharides, amino acids, vitamins, aromatic and humic acids, for saprophytic and soil-dwelling microorganisms. However, a scavenging bacterium must compete for these nutrient sources against other species of bacteria and prokaryotes, fungi, and even plants, which may also utilize the same energy source. As well, abiotic factors further limit the availability of many organic molecules. In the consideration of soil as a growth substrate, it is essential to appreciate that the vast majority of organic molecules present may not be in a form that is readily accessible (or bio-available) to microbes.

Aromatic Acids and Other Lignin-derived Compounds

Lignin represents one of the largest biotic reserves of carbon on this planet, and it is generally accepted that this high molecular mass compound contributes a significant proportion (>70%) of organic carbon in soils (Grandy & Neff, 2008). Lignin is one of the most abundant components of woody or vascularized plants where it accumulates in the plant cell walls of roots and shoots as a structural support mechanism (Zhong & Ye, 2007). This polymer is synthesized via random cross-linking reactions between phenylpropanol units, which are covalently joined via carbon-carbon and ether linkages (Ferrer, et al., 2008). Accordingly, the heterogeneous and complex structure of lignin makes this macromolecule exceptionally difficult to degrade via enzymatic conversion, and only a few species possess this ability.

Chemical extraction of lignin yields a range of aromatic acids, including *p*-hydroxybenzoate, vanillic acid, syringic acid, *p*-coumaric acid, ferulic acid, and their related aldehydes (Kogel & Bochter, 1985). The composition of these phenolics in lignin varies between plant species; gymnosperms yield predominately vanillin and vanillic acid whereas angiosperms contain a greater proportion of syringic acid and syringaldehyde (Kogel & Bochter, 1985). In contrast, non-woody plants tend to accumulate cinnamyl

phenols (ferulic and *p*-coumaric acids), and the relative concentrations of these aromatic compounds in soil reflects the plants species that previously occupied (or currently occupy) a particular site (Hautala, et al., 1997).

Lignin degradation is mediated almost exclusively by basidiomycetous fungi, with white rot fungi exhibiting an ability to completely metabolize lignin to CO_2 via oxidation, demethylation/demethoxylation, and aromatic ring-cleavage reactions (Filley, et al., 2000; Hofrichter, et al., 1999). Brown rot fungi are unable to completely mineralize lignin however this group may initiate a partial dissimilation of the aromatic polymer while preferentially metabolizing cellulose and hemicelluloses (Yelle, et al., 2008).

Although the degradation of lignin may release a range of phenolic monomers into soil, not all of these are readily available to soil-inhabiting microbes, as the monomers may be subjected to abiotic and biotic transformations. Humic acids are composed, in part, of a range of phenolic constituents including vanillic, *p*-hydroxybenzoic, and syringic acids that have polymerized to form a high molecular mass compound (Burges, et al., 1964). Monomeric phenols may form abiotic associations with humic acids in soil, with the strength of the interaction depending upon the chemical nature of the aromatic compound and the pH of the surrounding medium (Vinken, et al., 2005). For example, aromatic monomers with two hydroxyl groups (diphenols such as catechol and protocatechuate) tend to form stronger covalent interactions with humic acids than aromatics with a single hydroxyl group (monophenols such as ferulic acid). A study of chlorinated phenols indicates that similar associations may be catalyzed in the presence of various peroxidases; the enzymes oxidize the phenol, and the unstable and reactive product polymerizes or binds covalently to humic acids (Dec, et al., 2003). Clays may catalyze the abiotic ring-cleavage and polymerization of phenolics with one another or with amino acids (Wang & Huang, 2003). Under certain conditions, phenols such as catechol, *p*-coumaric and caffeic acids, will similarly undergo oxidative coupling reactions to form dimers or trimers (Smejkalova, et al., 2006). These associations into

high molecular mass polymers afford lignin-derived phenolics a measure of protection from microbial degradation.

A survey of surface soil solutions obtained from a mature spruce forest in Switzerland revealed a range of phenolic monomers, such as protocatechuate (57 nmols l^{-1}), *p*-hydroxybenzoate (146 nmols l^{-1}), vanillic acid (243 nmols l^{-1}), *p*-hydroxyacetophenone (44 nmols l^{-1}) and others at lower concentrations (caffeic and ferulic acids, catechol) (Gallet & Keller, 1999). Such studies demonstrate that lignin-associated phenolics may be available for microbial degradation at low levels, though it is unclear to what degree these are metabolized (Gallet & Keller, 1999). Many soil-dwelling bacteria convert a range of these aromatic acids to one of two common intermediates (protocatechuate or catechol), which are subsequently metabolized to succinate and acetyl-CoA via the β–ketoadipate pathway (Harwood & Parales, 1996). The β–ketoadipate pathway is encoded by a wide range of soil-dwelling microorganisms, including basidiomycetous fungi, Actinobacteria (*Streptomyces*, *Rhodococcus*, *Corynebacterium*), and Proteobacteria (*Pseudomonas*, *Acinetobacter*, *Agrobacterium*) such as the marine organism *Roseobacter* (Parke & Ornston, 1986; Harwood & Parales, 1996; Eulberg, et al., 1998; Grund & Kutzner, 1998; Buchan, et al., 2004; Brinkrolf, et al., 2006).

A certain proportion of unavailable or refractory aromatics consist of charcoal or 'black carbon', a term which refers to residues resulting from the incomplete combustion of carbon compounds released during (for example) forest fires. Black carbon is prevalent throughout much of the Earth's surface and may represent a significant percentage (up to 80%) of total organic carbon in soils (Poirier, et al., 2000; Lehmann, et al., 2008), however this form of carbon is quite recalcitrant to microbial degradation and is generally considered to be unavailable to organisms as an energy or carbon source (Seiler & Crutzen, 1980).

Neutral and Amino Sugars

An alternative source of carbon and energy is available in the form of a variety of neutral and amino sugars. Neutral sugars comprise a significant (albeit labile) fraction of organic matter in soils, representing roughly 10% total organic matter in certain soil types. For example, one analysis of soils yielded total neutral sugar concentrations in the range of 1622 to 5270 mg kg^{-1} soil (Zhang, et al., 2007); other studies have reported comparable results (Murata, et al., 1999). Arabinose, xylose, glucose, mannose, and galactose include the most commonly identified soil sugars. There is evidence that the age of a forest may influence the quantity of sugars (and soluble phenolic acids) in surrounding soils; for example, Johnson and Pregitzer report low sugar concentrations (217 to 281 mg kg-1) associated with 9 year old maple, aspen, and birch groves (Johnson & Pregitzer, 2007).

Amino sugars derive from various sources, including the chitin (a polymer of N-acetylglucosamine) that comprises insect exoskeletons, and polysaccharides present in the cell walls of fungi and bacteria. In particular, N-acetylglucosamine and N-acetylmuramic acid are associated with the peptidoglycan layer in bacteria. The accumulation of these sugars varies with soil type, however one study of eight soils reported total amino sugar levels between 281 to 2639 mg kg^{-1} (North American prairie soils) and 1310 to 6509 mg kg^{-1} (German forest soils) (Zhang & Amelung, 1996). The ratios of hydrolyzed amino sugars may be used to estimate the relative contribution of fungi, bacteria, and actinomycetes, to the microbial biomass in different soils (Glaser, et al., 2004). For example, the prevalence of bacterial species may be estimated by determining the ratio of glucosamine to muramic acid released into soil upon chloroform fumigation (Glaser, et al., 2004). Glucosamine, galactosamine, and mannosamine also accumulate in humic acids where they may represent as much as 5% of total N (Coelho, et al., 1997). However, amino sugars have a rapid turn-over rate in soils due to transformation by the microbial community ($t_{1/2}$ = 1-3 hrs; (Roberts, et al., 2007)).

Nucleic Acids

Ribonucleosides and deoxyribonuleosides have been isolated from rhizosphere and bulk soils using a methanol-based extraction technique, coupled with high performance liquid chromatography (Phillips, et al., 1997). This study identified cytidine, uridine, guanosine, and adenosine, with reported concentrations typically > 100 μmols kg^{-1} soil. Although it is likely that a certain proportion of living cells were lysed during the extraction process (thereby contributing to the pool of nucleosides), the authors demonstrated that cell lysis could not account for more than 1% of the recovered nucleosides. Interestingly, two studies identifying genes with up-regulated expression in soil independently isolated genes predicted to encode nucleotide permeases in *Pseudomonas fluorescens* strains SBW25 and Pf0-1 (Silby & Levy, 2004; Gal, et al., 2003). Nonetheless, it is unclear what proportion of the total pool of nucleosides and nucleotides is biologically available in soils; these compounds adhere tightly to clays and soils under certain conditions (Cortez & Schnitzer, 1981; Phillips, et al., 1997), and thus may be sequestered from scavenging microbes.

Amino Acids

Soils contain free or biologically available amino acids in quantities sufficient for detection, as originating from animals, plants, and microorganisms. For example, an assay of the bio-available amino acid content in a soil (C_{tot}: 3.17%, N_{tot}: 0.31%) yielded mean concentrations in the range of 40 to 1900 ng g^{-1} soil, depending upon the individual amino acid assayed (Formanek, et al., 2005). Other studies have yielded comparable values with vastly different soil types (Kielland, 1995; Warren, 2008). An assay of rich soils obtained from a eucalyptus forest in Australia reported a free amino acid concentration of >100 μM, corresponding to 64% of total soluble nitrogen (Warren, 2008). Another assay determined that the soluble nitrogen concentration of free amino acids averaged 24 ± 8 μM over seven soil sampling sites; in comparison, concentrations of soluble nitrogen averaged 39 and 67 μM, for ammonium and nitrate, respectively (Jones, et al., 2002). These and other studies indicate that free amino acids comprise a

significant proportion of total soluble nitrogen, and in certain environments (such as the arctic tundra; (Kielland, 1995)), amino acids represent the greatest source of soluble nitrogen in soil (Kielland, 1995; Warren, 2008; Jones & Kielland, 2002; Jones, et al., 2002).

Nonetheless, the amount of amino acids that are readily available for uptake by bacteria is likely only a small percentage of the total amino acid content present in soil. The majority of amino acids is likely contained within high molecular mass products such as proteins and root mucilage. Free amino acids may react with sugars or phenolics and thus be either sequestered or otherwise altered, with certain amino acids exhibiting a greater reactivity (basic amino acids) than others. The term 'melanoidin' refers to insoluble high molecular mass compounds originating from randomly occurring condensation reactions between amino acids and sugars; these compounds may comprise a significant fraction of unavailable organic matter (Poirier, et al., 2000). As well, free amino acids have a high turn-over rate, with a half life that varies per individual amino acid but likely falls within the range of several hours (Kielland, 1995; Jones & Kielland, 2002; Jones, et al., 2005). Much of this rapid turn-over is likely due to microbial activity (Jones, et al., 2005) however there is also evidence that abiotic factors alone (Fe, Mn, Al oxides) are sufficient to catalyze the deamination and subsequent polycondensation of amino acids with phenolic compounds in soil (Wang & Huang, 2003).

As a general rule, the amino acid composition (percentage each individual amino acid contributes to the total amount) is relatively constant between soils of varying type although the overall quantity of amino acids present may vary (Friedel & Scheller, 2002). For example, the relative molar distribution (mol%) of 16 amino acids extracted from eight soil types yielded similar values for each amino acid between soil samples, despite variations in soil pH, moisture content, texture, and presence of vegetation at the site of sampling (Friedel & Scheller, 2002). However, there is also evidence that the absolute levels of a particular amino acid may fluctuate over a period of several months (Johnson & Pregitzer, 2007). A comparison of the overall amino acid composition in algae, fungi, bacteria, and yeasts indicates that bacteria may act as major contributors to the pool of

amino acids in soil, as their mean amino acid composition most closely matches that observed in soil samples (Sowden, 1977). Consistent with this hypothesis, certain amino acids detected in soil are most closely associated with bacteria (example, diaminopimelic acid). However it has also been proposed that the greatest amino acid input originates from plants, in the form of decaying leaf litter and root exudates (Jones & Kielland, 2002). It is likely that the influence of any particular source upon soil amino acid composition (such as microbial biomass versus plants) will vary depending upon the conditions associated with a given soil site.

1.1.2 The rhizosphere is a unique and dynamic soil environment

The rhizosphere is defined as the zone of soil that immediately surrounds and encompasses a plant root (Hinsinger, et al., 2005). Due to its intimate association with the root, the rhizosphere represents a unique and dynamic soil environment that differs from bulk soil with respect to its biological, chemical, and physical parameters. Plant roots continually exude a diverse range of organic molecules, including vitamins, amino acids, proteins and enzymes, ions, and sugars, which are enriched in rhizosphere soils (Kumar, et al., 2007; Bringhurst, et al., 2001; Jaeger, et al., 1999; Nguyen, 2003; Hinsinger, et al., 2005; Weisskopf, et al., 2008). Additionally, plant roots secrete mucilage and root-tip slime, and constantly slough off root epidermal cells (Nguyen, 2003). The emergence of lateral roots from a mature root may also result in some leakage of amino acids (Jaeger, et al., 1999) and sugars (Bringhurst, et al., 2001). Accordingly, the nutrient-rich rhizosphere may support a greater bacterial cell density than the surrounding bulk soil (Jaeger, et al., 1999; Smalla, et al., 2001). In fact, plants lose a considerable proportion of reduced carbon via rhizodeposition (>15% of net C fixed during photosynthesis; Nguyen, 2003) and while this may appear to be a complete (and energetically expensive) loss for the plant, the input of nitrogen- and carbon-containing organics into the soil stimulates the growth of bacteria and fungi, which may in turn be of direct benefit the plant. Examples include colonization of the rhizosphere

with a symbiotically relevant fungal or bacterial species or with plant growth promoting bacteria (Avis, et al., 2008).

The designation of rhizosphere versus bulk soil is necessarily subjective however plant roots exert an influence extending several millimeters into the surrounding soil (Hinsinger, et al., 2005). Rhizosphere soil is a spatially and temporally heterogeneous environment, as a gradient is established in which the concentration of a given compound decreases with a decreased proximity to the root cells; as well, the distribution of a particular compound may be irregular within the rhizosphere (Jaeger, et al., 1999; Bringhurst, et al., 2001). For example, the relative concentrations of tryptophan and sucrose vary greatly along the length of a growing root; as well, the distribution patterns of the amino acid and sugar differ from one another, with tryptophan being detected primarily in the older root tissue and sucrose being secreted from newly developing tissue (Jaeger, et al., 1999). In addition, factors such as pH, water potential, partial pressures of gases (pCO_2), and concentration of various ions, may differ along the length of a single root system, as influenced by the species, age, and relative health of the plant, type and age of root, season of year, and even time of day (Hinsinger, et al., 2005). The ability of motile microorganisms to sense and respond to these chemical and physical gradients may contribute greatly to their ability to competitively colonize the rhizosphere (Matilla, et al., 2007; de Weert, et al., 2002; Ramos-Gonzalez, et al., 2005).

1.1.3 Soil as a stressful environment

As a habitat, soil is not invariably benign, and soil-dwelling microorganisms may encounter significant physical and chemical stressors that require the adoption of active mechanisms to survive. Soil inhabitants may experience frequent fluctuations in soil moisture content, temperature, nutrient availability, and even soil pH may vary over time. A thorough examination of bacterial species diversity in Antarctic soils concurrently documented various stresses inevitably encountered by microbes living in this particularly harsh environment (Aislabie, et al., 2006). Analyses of soil samples revealed high concentrations of salts and very low (> 5%) moisture contents; accordingly, soil

inhabitants need to be desiccation resistant and salt tolerant. As well, soil temperatures often registered below freezing for most of the year, while increasing temperatures during the summer months resulted in many freeze-thaw cycles. Finally, sources of available nitrogen and carbon are relatively scarce and of patchy distribution in Antarctica, with soils yielding low concentrations of organic material.

Antarctica represents an extreme environment for any species however a soil saprophyte may expect to encounter these and other stresses in temperate or tropical soils as well. For example, the rhizosphere is particularly enriched in plant-derived organics that support a high bacterial cell density, yet these soils also contain predatory protozoa that graze upon the local bacterial population (Bringhurst, et al., 2001). A scavenging bacterium may be exposed to soils contaminated by heavy metals, polycyclic aromatic hydrocarbons, or other pollutants. Residence in acidic or basic soils may impose a significant stress; a survey performed across North and South American soils identified pH as the most important environmental factor influencing bacterial species diversity, as many species are unable to maintain viable populations in acidic soils (Fierer & Jackson, 2006). Soil inhabitants may experience sporadic or long-term periods of drought (or flooding), requiring the expression of desiccation-resistance genes to mediate an appropriate physiological response (Cytryn, et al., 2007). Similarly, soils may contain high concentrations of salts; increased soil salinity levels may require (for example) the synthesis or accumulation of compatible solutes within the cell as a means of counteracting the effects of the ensuing osmotic stress (Paul & Nair, 2008). Microorganisms located within or near the soil surface may be exposed to high levels of solar ultraviolet radiation (UVR), which may result in DNA damage, and possibly cell death. Soil fungi have adapted to this stress through the production of pigmented melanins; a survey of *Aspergillus niger* strains isolated from two sites established a correlation between UVR exposure and melanin content (Singaravelan, et al., 2008). In other words, strains inhabiting soils associated with high levels of UVR exhibited a greater conidial melanin concentration, and a greater resistance to UVR, than strains

isolated from the shady slope of the canyon. That this fungal species has evolved such a trait emphasizes the deleterious effects of UV exposure upon soil-dwelling inhabitants.

Microarray analyses performed with *P. putida* reveal that multiple genes associated with adaptive responses to abiotic stresses are up-regulated in a maize rhizosphere (Matilla, et al., 2007). Particularly, genes encoding functions involved in oxidative stress responses (such as a glutathione peroxidase) were identified as exhibiting induced expression in this environment. In vivo expression technology (IVET) has been used to identify genes whose expression in *Pseudomonas* is induced in bulk soil and in maize, rice, and sugar beet rhizospheres (Ramos-Gonzalez, et al., 2005; Silby & Levy, 2004; Gal, et al., 2003; Rainey, 1999; Rediers, et al., 2003). In addition to various genes relevant to nutrient acquisition, each of these studies independently identified genes predicted to encode products associated with stress responses, including proteins implicated in heavy metal resistance and oxidative stress response.

1.1.4 Factors Affecting Bacterial Soil Diversity

The utilization of PCR-mediated DNA amplification offers an unparalleled opportunity to probe microbial species diversity in various habitats in a manner that is culture-independent. The adoption of a culture-independent method is particularly relevant to the description of soil microorganisms, as the vast majority of these species are not readily cultivated (Torsvik, et al., 1990; Amann, et al., 1995). One of the earlier studies to address species diversity in soils employed the (then) novel technique of PCR-amplification of total genomic DNA extracted from soil samples using primers specific to 16S rRNA genes (Liesack & Stackebrandt, 1992). Results obtained from an analysis of Australian soil samples found a predominance of proteobacteria, most closely related to a class of nitrogen-fixing α–proteobacteria.

Alternatively, total ribosomal RNA may be extracted from soils in lieu of genomic DNA (Miethling, et al., 2000). As ribosomes are only produced by viable cells, such analyses offer insights into the fraction of the community that is metabolically active; an important distinction as DNA and RNA-based profiles may differ significantly,

particularly with respect to bulk soil samples (Weisskopf, et al., 2008). Fatty acid methyl ester analysis (FAME) offers an alternative, culture-independent, method to examine microbial diversity through an analysis of the fatty acid profile of soils (Miethling, et al., 2000).

Soil bacterial diversity varies considerably between different soils and environmental conditions. At one extreme, the cold, desert soils of the Antarctica contain a low species diversity with a few main groups of bacteria comprising the majority of the prokaryotic population (Aislabie, et al., 2006); in this instance, species diversity is likely limited due to the (primarily) abiotic stresses encountered in such an environment and population levels are similarly limited ($<10^6$ to 10^8 cells g^{-1} soil). A contrasting scenario is offered by forest soils, which are typically rich in plant-derived organics; in these soils, a gram may contain as many as 10^5 to 10^6 different species and carry a total of 10^9 cells (Torsvik, et al., 1990; Gans, et al., 2005).

However, the view that nutrient availability is the greatest variable to influence species diversity is overly simplistic. For example, it is known that plants exert an influence upon the microbial community that occupies the rhizosphere, and there is evidence that plant species may occasionally exert a greater impact upon the composition of microbial communities than the type, origin, or characteristics of soil *per se* (Miethling, et al., 2000; Smalla, et al., 2001). As well, seasonal effects may be manifest, in terms of overall species diversity and composition within a given sampling site (Smalla, et al., 2001). The release of heavy metals, toxins, or pollutants into a soil environment may significantly alter the microbial community by selecting for species with enhanced tolerance of such compounds (Li, et al., 2006). For example, bacterial species are generally more sensitive to heavy metal pollution than fungal populations and the introduction of such metals may alter the relative ratio of bacterial/fungal species in contaminated soils (Rajapaksha, et al., 2004). One study estimated bacterial diversity to be reduced by 99.9% in the presence of heavy metal (Cd, Cu, Ni, and Zn) contamination (Gans, et al., 2005). Also, the spatial distribution of competing species within soils may have a considerable effect upon diversity, as influenced by soil moisture content (Treves,

et al., 2003). According to this model, increased water content facilitates migration and increases the degree of inter-species interactions. Certain species are unable to compete successfully for common resources and are removed from the population, thereby decreasing species diversity. In contrast, drier soils afford protection from intense competition through the establishment of microenvironments where isolated populations of bacteria can survive (Treves, et al., 2003).

The rhizosphere is inhabited by a dynamic population of soil bacteria and fungi, including mycorrhizal fungi. As a result of the rich assortment of organic compounds originating from plant roots, the rhizosphere may contain a greater diversity of species than adjacent bulk soils (Smalla, et al., 2001); however not all studies support this observation (Weisskopf, et al., 2008). Nonetheless, the complexity of microbial communities in the rhizosphere is such that similar methodologies may yield apparently contrasting results. Community profiles as determined by DNA-based methods (i.e., 16S rRNA amplification of genomic DNA) indicate species diversity may actually increase with distance from the root system (Weisskopf, et al., 2008). Analyses performed using RNA-based profiling of the same samples (i.e., amplification of reverse transcribed 16S rRNA) suggests that species diversity increases with proximity to the root system (Weisskopf, et al., 2008). An obvious interpretation of these studies is that while the overall species diversity may increase with distance from roots, the metabolically active cells are located in close proximity to the plant. Recent studies employing stable isotope probing offer a method to distinguish between plant-associated species that directly (and rapidly) assimilate carbon from root exudates and co-existing soil saprophytes that are less reliant upon reduced photosynthate (Vandenkoornhuyse, et al., 2007). In this study, $^{13}CO_2$ was provided in a short pulse to plants growing within turfs of grassland and peatland. The extraction and analysis of labeled [^{13}C]RNAs from the co-habiting microbial community permitted the identification of plant-associated phylotypes such as arbuscular mycorrhizal fungi and β–proteobacteria (*Burkholderiales*) that catabolize (labeled) photosynthetic carbon via root exudates. In contrast, saprophytic bacteria that

do not directly assimilate plant photosynthate were identified through the analysis of unlabeled 16S rRNA.

In summary, soils constitute a unique and challenging environment inhabited by a diverse assemblage of bacteria, fungi, and other microorganisms. Soils comprise a rich variety of nutrients, including amino acids and proteins, sugars, organic and phenolic acids, as originating from plant, animal, and anthropogenic sources. However, the availability of these compounds is limited by biotic (i.e., inter-species competition) and abiotic factors. The complexity of interactions occurring between soil microorganisms and their environment is most evident in the rhizosphere, a highly dynamic microenvironment where the input of plant-derived organic (and inorganic) compounds influences the microbial community. Nonetheless, soil may also be regarded as a harsh habitat and soil-borne organisms must be adapted to survive fluctuations in water and nutrient availability, increases in salinity and soil acidification, and the presence of pollutants such as heavy metals. These environmental factors shape the soil microbial community by influencing the composition and diversity of soil-dwelling species.

THE β–KETOADIPATE PATHWAY

1.2.1 Distribution of aromatic acids in soil

A diverse range of aromatic compounds are present in soil and represent a potential source of carbon and energy to microorganisms that inhabit this environment (Gallet & Keller, 1999). The majority of these phenolics are of plant origin (primarily in the form of lignin), however the utilization of herbicides, pesticides, and other pollutants, has resulted in the introduction of anthropogenic aromatic hydrocarbons into soils as well.

The β-ketoadipate pathway is a metabolic pathway of broad taxonomic distribution that mediates the catabolism of a wide range of aromatic compounds via conversion to one of two common intermediates: protocatechuate (3,4-

dihydroxybenzoate) or catechol (1,2-dihydroxybenzene) (Figure 1.1). These diphenolic metabolites are substrates for the β-ketoadipate pathway and are consequently oxidized in parallel and convergent branches of the pathway to yield succinate and acetyl-CoA (Figure 1.2), which are funneled into the tricarboxylic acid cycle thereby affording a source of energy.

1.2.2 The β-ketoadipate pathway: a historical perspective

β-ketoadipate was first identified as an intermediate in the metabolism of aromatic compounds by a 'Gram-negative vibrio' (later classified as *Moraxella*, and subsequently re-classified as *Acinetobacter*) (Kilby, 1948; Kilby, 1951; Stanier, et al., 1950), however the steps through which aromatic acids are metabolized were obscure and the earliest studies erroneously identified several plausible compounds as intermediates in what is now known as the β-ketoadipate pathway (for example, see Evans, et al., 1949).

One of the most puzzling aspects addressed during this time was the method utilized by bacteria to attack or initiate metabolism of a chemically stable aromatic ring and an initial hypothesis involved the sequential hydroxylation of the ring, resulting in the production of mono-hydroxy-derivatives (Stanier, 1948). The description of an enzyme (referred to as 'pyrocatechase') partially purified from *Pseudomonas* by Hayaishi and Hashimoto revealed that the oxidation of catechol occurred via a ring cleavage reaction, resulting in the production of *cis*, *cis*-muconate (Hayaishi & Hashimoto, 1950). This work was extended by a study demonstrating the conversion of *cis*, *cis*-muconate to β–ketoadipate (Hayaishi & Stanier, 1951)[1]; the catabolism of β–ketoadipate to acetyl-CoA and succinate was proposed and demonstrated soon after the initial description of β–ketoadipate as a pathway intermediate (Kilby, 1951).

[1]The phenomenon of 'simultaneous adaptation' was proposed by Stanier as a means of identifying common intermediates in a metabolic pathway; it was noted that bacteria previously 'adapted' to a growth substrate would not exhibit a lag in growth or metabolism upon exposure to the same compound or pathway intermediates. That this adaptation was actually due to the inducible synthesis of enzymes was first demonstrated by Hayaishi and Stanier; prior to this study, the linkage between adaptation and induction of enzymes had not been formally demonstrated for the metabolism of any compound.

It was quickly realized that although the oxidation of both catechol and protocatechuate resulted in the production of β–ketoadipate (Stanier, et al., 1950), these two phenolics were metabolized in separate and distinct pathways (Sleeper & Stanier, 1949). The oxidation of protocatechuate to β–carboxy-*cis*, *cis*-muconate was demonstrated using a partially purified dioxygenase from *Pseudomonas* (Stanier & Ingraham, 1954), however these studies were significantly hindered by the instability of the tricarboxylic acid and its rapid isomerization to a biologically inactive *cis-trans* form (MacDonald, et al., 1954).

A series of landmark papers by Ornston (Ornston & Stanier, 1966; Ornston, 1966b; Ornston, 1966c; Ornston, 1966d) describes the isolation and identification of the remaining intermediates in the metabolism of catechol and protocatechuate to β–ketoadipate (Ornston & Stanier, 1966) and the purification of the enzymes involved in *Pseudomonas* (Ornston, 1966b; Ornston, 1966c). Prior to this work, it had been speculated that the biochemical reactions involved in the conversion of protocatechuate to β-ketoadipate varied between *Pseudomonas* and *Acinetobacter*, however a more thorough and complete description of pathway intermediates revealed that all steps of the pathway were conserved within prokaryotes (Ornston & Stanier, 1966). These studies also provided the first evidence that the parallel branches of catechol and protocatechuate catabolism in (most) prokaryotes converge in the production of the common intermediate β-ketoadipate enol-lactone as opposed to β–ketoadipate.

1.2.3 The protocatechuate branch of the β–ketoadipate pathway

The first step in the metabolism of protocatechuate to succinate and acetyl-CoA via the β-ketoadipate pathway involves an oxidative ring-opening reaction, as mediated by protocatechuate-3,4-dioxygenase (MacDonald, et al., 1954; Stanier & Ingraham, 1954) (Figure 1.2). This reaction requires the incorporation of molecular oxygen into the aromatic acid by the dioxygenase, to yield β-carboxy-*cis*, *cis*-muconate. Protocatechuate-3,4-dioxygenase has been purified (Ornston, 1966b) and more than a dozen crystal structures of the dioxygenase complexed with various substrates and their analogues have

been obtained and analyzed (Vetting, et al., 2000; Lange & Que, 1998). This enzyme consists of two non-identical subunits (α and β; encoded by *pcaG* and *pcaH*) of an equivalent number that varies between species; for example, the protocatechuate-3,4-dioxygenase of *Agrobacterium radiobacteri* is comprised of $(\alpha\beta Fe^{3+})_2$ whereas its counterparts in *A. baylyi* and *P. putida* employ an enzyme of $(\alpha\beta Fe^{3+})_{12}$.

Protocatechuate-3,4-dioxygenase contains a nonheme ferric centre which comprises an active site in conjunction with two tyrosines, two histidines, and one arginine that are highly conserved (Lange & Que, 1998; Iwagami, et al., 2000). In contrast to catechol-1,2-dioxygenase, the dihydroxybenzoate-specific enzyme includes positively charged, basic amino acids that surround the entrance to the active site; these amino acids contribute a positive electrostatic potential that may aid in directing the negatively charged substrate into the active site (Vetting, et al., 2000). In a complex series of interactions, the substrate (protocatechuate) is activated to form an intermediate with the enzyme and molecular oxygen; oxygen attacks a carbon (C4) of the aromatic ring and the collapse of this transition state intermediate results in protocatechuate ring cleavage and the production of a muconic acid (Vetting, et al., 2000; Lange & Que, 1998).

In the subsequent step, β-carboxy-*cis*, *cis*-muconate is converted to the highly unstable intermediate γ-carboxymuconolactone through the action of a β-carboxy-*cis*, *cis*-muconate lactonizing enzyme (encoded by *pcaB*), which has been purified and characterized in *P. putida* (Ornston, 1966b). β-carboxy-*cis*, *cis*-muconate is a highly toxic metabolite, and this has been exploited through the use of mutant strains lacking the lactonizing enzyme activity; these mutants cannot metabolize β-carboxy-*cis*, *cis*-muconate, and thus accumulate levels of this compound that are inhibitory (Lorite, et al., 1998; Parke, 2000). Growth of such a mutant upon protocatechuate (or a related compound) provides a strong selection method for cells in which secondary mutations have been spontaneously generated in genes encoding products that act upstream in the pathway (i.e., transport system, dioxygenase, regulatory genes). In this manner, strains in which the activity of the protocatechuate-3,4-dioxygenase is compromised have been

isolated in *A. tumefaciens* and *A. baylyi*, allowing the identification of amino acids critical to the function of this enzyme (Parke, 2000; Gerischer & Ornston, 1995; D'Argenio, et al., 1999). As well, the structural enzyme PobA (encoding a hydroxylase involved in conversion of *p*-hydroxybenzoate to protocatechuate) (Hartnett, et al., 1990), and regulation of *pca* (D'Argenio, et al., 2001) genes have been examined using this positive selection method.

The chemical instability of the succeeding intermediate, γ-carboxymuconolactone (half life of 30 min at 30°C; (Ornston & Stanier, 1966)), is due to the presence of a labile carboxyl group; spontaneous decarboxylation of this functional group yields β–ketoadipate enol-lactone. Despite the probability of a non-enzymatic synthesis of β–ketoadipate enol-lactone from γ-carboxymuconolactone under physiological conditions, *P. putida* nonetheless employs a specific γ-carboxymuconolactone decarboxylase (encoded by *pcaC*) to catalyze this same reaction at an enhanced rate (Ornston, 1966b). This particular step in the pathway is notable because the catechol and protocatechuate branches of the β–ketoadipate pathway converge in the production of β–ketoadipate enol-lactone in most prokaryotes (Figure 1.2).

β–ketoadipate enol-lactone is next hydrolyzed to β–ketoadipate, as mediated by the β–ketoadipate enol-lactone hydrolase (encoded by *pcaD*), which has been purified in *P. putida* (Ornston, 1966b). Rarely, some species employ an enzyme that catalyzes both the decarboxylation of the γ-carboxymuconolactone and subsequent conversion of the enol-lactone to β–ketoadipate. This enzyme (PcaL), which appears to be the result of a fusion event occurring between the genes encoding the two enzymes (*pcaC* and *pcaD*), was first described in *Rhodococcus opacus* (Eulberg, et al., 1998) and has since been reported in *Streptomyces* (Iwagami, et al., 2000). It has been proposed that the organization of the two enzymatic functions within separate genes in proteobacteria represents the ancestral state, and that the fused protein corresponds to a trait recently adopted by Gram-positive species (Eulberg, et al., 1998). The recent report of a comparable protein in the Gram-negative *A. baumannii* indicates that *pcaL* homologues are present in atleast some members of γ-proteobacteria however (Park, et al., 2006).

Similarly, *pcaC* and *pcaD* are encoded as separate functions in the Gram-positive actinomycete *Terrabacter* sp. DBF63 (Habe, et al., 2005), and thus the association of *pcaL* as a Gram-positive trait may be more reflective of its initial description than the manifestation of a group-specific characteristic.

The penultimate step in the catabolism of protocatechuate to succinate and acetyl-CoA is catalyzed by a β–ketoadipate succinyl-CoA transferase, which activates β–ketoadipate via the transfer of coenzyme A (CoA) from succinyl-CoA. This enzyme is comprised of two non-identical subunits ($\alpha_2\beta_2$) encoded by *pcaI* and *pcaJ* (Parales & Harwood, 1992; MacLean, et al., 2006), and has been purified from *Pseudomonas*, *Sinorhizobium* and *Acinetobacter* (Yeh & Ornston, 1981; MacLean, et al., 2006). In the final step of the pathway, β–ketoadipyl-CoA undergoes a thiolytic cleavage via β–ketoadipyl-CoA thiolase (encoded by *pcaF*) generating acetyl-CoA and succinyl-CoA (Harwood, et al., 1994). As coenzyme A was initially transferred to β–ketoadipate from succinyl-CoA, the net yield as arising from the catabolism of protocatechuate is acetyl-CoA and succinate, which are funneled into the tricarboxylic acid cycle.

1.2.4 Protocatechuate catabolism in *Rhizobiaceae*

An auxanographic screening of *Sinorhizobium*, *Rhizobium*, *Agrobacterium* and *Bradyrhizobium* strains for growth at the expense of various aromatic compounds revealed a wide range of metabolic capabilities, with a few species utilizing the majority of substrates tested (*Bradyrhizobium*) and others demonstrating only a limited ability to catabolize these compounds (*Sinorhizobium*) (Parke & Ornston, 1984). Nonetheless, the protocatechuate branch of the β–ketoadipate pathway appears to be universal in all species of rhizobia examined in this and other studies (Parke & Ornston, 1984; Chen, et al., 1984; Parke & Ornston, 1986; Parke, et al., 1991; Parke, 1995; MacLean, et al., 2006), suggesting the presence of selective constraints or pressures acting within this group of bacteria to maintain the ability to utilize this aromatic acid.

Protocatechuate catabolism in α-proteobacteria has been best-described in the plant pathogen *Agrobacterium tumefaciens* and the endosymbiont *S. meliloti* (Parke,

1993; Parke, 1995; Parke, 1996a; Parke, 1996b; MacLean, et al., 2006; MacLean, et al., 2008). In these species, the genes encoding structural enzymes relevant to the pathway are organized into two operons (*pcaDCHGB* and *pcaIJF*) that are regulated by distinct transcriptional regulators. The LysR-type transcriptional regulator encoded by *pcaQ* is located adjacent to, and divergently transcribed from, the *pcaDCHGB* operon (Figure 1.3). PcaQ regulates expression of this operon, and induces gene expression in the presence of β-carboxy-*cis*, *cis*-muconate and γ–carboxymuconolactone (Parke, 1993; Parke, 1996a; MacLean, et al., 2006; MacLean, et al., 2008). It had initially been proposed that the adoption of metabolites as inducing agents by bacteria was necessarily restricted to the chemically stable compounds such as protocatechuate and β–ketoadipate, which perform this function in γ–proteobacteria (Stanier & Ornston, 1973). The characterization of PcaQ, and corresponding identification of the two least stable pathway intermediates as coeffectors in *A. tumefaciens* and *R. leguminosarum*, has since refuted this hypothesis (Parke, et al., 1991; Parke, 1993; Parke, 1996a).

The gene encoding PcaQ is conserved in several species of rhizobia, including *S. fredii, R. etli, R. tropici* and *R. leguminosarum* (Parke, 1996b), and homologues have been identified in the genome sequences of many members of the class α-proteobacteria, including the marine bacteria *Roseobacter* (Buchan, et al., 2004). *S. meliloti* PcaQ has been purified and *in vitro* and *in vivo* assays indicate that this regulator binds a sequence of partial dyad symmetry located (at positions -72 to -57) upstream of the *pcaD* promoter (MacLean, et al., 2008).

The operon *pcaDCHGB* in *A. tumefaciens* and *S. meliloti* encodes the structural enzymes required for the conversion of protocatechuate to β–ketoadipate; further metabolism to succinate and acetyl-CoA is encoded by the *pcaIJF* operon, which is subject to regulation by the IclR-type transcriptional regulator PcaR (Parke, 1995; MacLean, et al., 2006). Expression of these genes in *A. tumefaciens*, *S. meliloti* and *R. leguminosarum* is induced in the presence of β–ketoadipate (Parke, et al., 1991; Parke, 1995; MacLean et al., 2006).

Protocatechuate catabolism has been studied in other *Rhizobiaceae* as well, and expression of *pca* genes in fast-growing rhizobia is typically inducible whereas slow-growing bradyrhizobia primarily express *pca* genes constitutively (Parke & Ornston, 1986, Parke, et al., 1991; Parke, 1993; Parke, 1995; Parke, 1996a; MacLean et al., 2006). In fact, *Bradyrhizobium* represents the only known exception to the rule that the activity of β–ketoadipate pathway enzymes is strictly inducible. Parke and Ornston propose that the energetic demands inherent in the synthesis of a regulatory system exceed those required for the maintenance of a low level of constitutive enzyme activity, and thus slow-growing oligotrophic species such as *B. japonicum* have not evolved an elaborate regulatory mechanism (Parke and Ornston, 1986).

1.2.5 Protocatechuate catabolism in *Roseobacter*

In addition to *Sinorhizobium* and *Agrobacterium*, protocatechuate catabolism has been the subject of study in the α–proteobacteria *Roseobacter*. This group of bacteria inhabits coastal regions such as salt marshes, and studies of this clade offer a unique opportunity to examine aromatic acid catabolism in a marine environment. Several isolates of the *Roseobacter* group can utilize a range of aromatic compounds including benzoate, vanillate, ferulate, and protocatechuate (Buchan, et al., 2000). These studies have been extended through the use of *pcaH* (encoding the well conserved β-subunit of protocatechuate-3,4-dioxygenase) as a marker to estimate the sequence diversity present in species associated with decaying marsh grass and in enrichment cultures (Buchan, et al., 2001). The results obtained reveal a *pcaH* gene pool of high sequence diversity (of 149 *pcaH* clones, 85 unique sequences were retrieved), likely reflecting a correspondingly diverse taxonomic distribution of the β–ketoadipate pathway in marine bacteria (Buchan, et al., 2001).

1.2.6 Protocatechuate catabolism in *Acinetobacter*

Acinetobacter have the ability to degrade an unusually wide range of phenolic compounds, which are generally funneled into the β–ketoadipate pathway via enzymatic

conversion to protocatechuate or catechol. *Acinetobacter* also readily undergoes natural transformation thus facilitating genetic studies, and this feature, coupled with its impressive metabolic portfolio, has made this genus the best characterized with respect to aromatic acid catabolism via the β–ketoadipate pathway.

Protocatechuate catabolism genes in *A. baylyi* are organized into a single large operon with genes involved in the dissimilation of quinate and shikimate (precursors of protocatechuate; *pcaIJFBDKCHGquiBCXA*) (Dal, et al., 2005). The *pcaIJFBDCHG* genes specify structural enzymes as described in a previous section; *pcaK* encodes a transport permease implicated in the uptake of *p*-hydroxybenzoate and protocatechuate (Kowalchuk,et al., 1994; D'Argenio, et al., 1999b).

Expression of this 14 kb operon is modulated by PcaU, a member of the IclR family of transcriptional regulators (Gerischer, et al., 1998); in the absence of protocatechuate, PcaU represses expression of the metabolic genes, whereas interaction of the regulator with protocatechuate induces gene expression (Trautwein & Gerischer, 2001). Purified PcaU has been demonstrated to bind a 45 bp site that is composed of three conserved 10 bp repeats located upstream of *pcaI*; interaction of the protein with this site is necessary for both activating and repressing functions (Popp, et al., 2002). As with most regulators, PcaU auto-regulates expression of its cognate promoter, in a manner that is influenced by the carbon sources available (Gerischer, et al., 1998; Trautwein and Gerischer, 2001; Siehler, et al., 2007).

1.2.7 Protocatechuate catabolism in *Pseudomonas*

In *P. putida*, genes involved in the dissimilation of protocatechuate are organized into four gene clusters (*pcaHG*, *pcaRKF*, *pcaIJ*, *pcaTBDCP*) that specify structural enzymes, a dedicated transport system and chemoreceptor (*pcaK*), and a transcriptional regulator (*pcaR*) (Jimenez, et al., 2002). As with most transport systems implicated in the uptake of aromatic compounds, PcaK is a proton-symporter belonging to the major facilitator superfamily, and participates in the uptake of *p*-hydroxybenzoate and protocatechuate (Harwood, et al., 1994; Nichols & Harwood, 1997; Ditty & Harwood,

1999). PcaR participates in the activation of *pcaIJ* and *pcaTBDC* expression in concert with the coeffector β–ketoadipate (Romero-Steiner, et al., 1994); in contrast, expression of *pcaHG* is not dependent upon β-ketoadipate as a coinducer. PcaT is unique to *Pseudomonas* thus far, and may act as a scavenging transport system involved in the uptake of β–ketoadipate from the environment (Ornston & Parke, 1976; Parke, et al., 2000). The presence of this system is remarkable given a demonstrable permeability barrier to β-ketoadipate in *P. putida* however it has been proposed that this system may be more important in mediating chemotaxis to β–ketoadipate or as a means of regulating the intracellular concentration of this coinducing metabolite than as a means of obtaining a source of energy (Ondrako & Ornston, 1980).

1.2.8 The catechol branch of the β-ketoadipate pathway

Protocatechuate and catechol differ by the presence of a single carboxyl group that is present in protocatechuate and lacking in catechol (Figure 1.1). However, these two substrates are acted upon by enzymes entirely specific to each branch of the pathway, and the catabolic proteins involved are not interchangeable though they may catalyze analogous reactions (Ornston, 1966b; Ornston, 1966c). This observation emphasizes a specificity that is characteristic of enzymes in general, but also underscores a real difference in the chemistry of each compound that belies the apparent structural similarities.

Catechol is metabolized to succinate and acetyl-CoA through a series of reactions that are chemically analogous but distinct to those specific to the protocatechuate branch of the β–ketoadipate pathway (Ornston & Stanier, 1966; Ornston 1966c). Catechol is cleaved via a ring-opening step involving the incorporation of oxygen by a catechol-1,2-dioxygenase (encoded by *catA*) to yield *cis, cis*-muconate. This metabolite is acted upon by CatB (*cis, cis*-muconate cycloisomerase enzyme, also referred to as muconate lactonizing enzyme), resulting in the production of a muconolactone. In the final step specific to the catechol branch, muconolactone isomerase (specified by *catC*) catalyzes

the migration of the double bond in the muconolactone to form β–ketoadipate enol-lactone.

In *A. baylyi*, the LysR-type regulator CatM induces expression of *catBCIJFD* and *catA* in concert with the pathway metabolite *cis, cis*-muconate (Romero-Arroyo, et al., 1995). Unusually, CatM shares partially overlapping regulatory roles with a paralogue (BenM) required for the expression of genes involved in benzoate catabolism. BenM recognizes *cis,cis*-muconate and benzoate as coeffectors (Collier et al., 1998), and can complement for loss of CatM with respect to *catA* expression (Romero-Arroyo, et al., 1995; Collier, et al., 1998). Similarly, CatM may compensate for the absence of BenM regarding the activation of *benABCDE* expression, under conditions in which levels of *cis,cis*-muconate are allowed to accumulate (Cosper, et al., 2000). These proteins also jointly regulate expression of two genes (*benPK*) that encode a benzoate transport system (Clark, et al., 2002).

A LysR-type transcriptional regulator is also required for the activation of *catBCA* expression in *P. putida* (Rothmel, et al., 1990; Rothmel, et al., 1991; Aldrich & Chakrabarty, 1988). This regulatory protein (designated as CatR) similarly recognizes *cis, cis*-muconate as a coeffector molecule (Parsek, et al., 1992). This regulatory system represents one of the rare examples of a LysR-type protein repressing expression of a gene through interaction with a site located within the target gene itself. In addition to an activating function, CatR binds a site located within *catB*, thereby decreasing expression of the gene (Chugani, et al., 1998). Binding of CatR to this internal site is facilitated through co-operative interactions with CatR molecules occupying sites located in the upstream promoter region. While a repressive interaction under inducing conditions seems counterintuitive, it has been proposed that this method allows a fine-tuning of *catBCA* expression, thus preventing gratuitous synthesis of the enzymes (Chugani, et al., 1998).

In contrast, the regulator designated as CatR in *Rhodococcus* belongs to the IclR family of transcriptional regulators (Vesely, et al., 2007). Unlike the LysR proteins, CatR in this genus functions as a repressing agent; expression of *catA* is increased 3- to 4-fold

upon deletion of *catR* in *R. erythropolis*, however the authors have suggested the possibility of a second regulator required for the activation of *catA* expression (Vesely, et al., 2007).

The catechol branch of the β–ketoadipate pathway does not have as wide a taxonomic distribution as the protocatechuate dissimilatory branch, and the ability to catabolize catechol may vary even between strains of a given species (Parke & Ornston, 1984; Chen, et al., 1984). For example, certain strains of *R. leguminosarum* biovar *trifolii* can utilize catechol as a carbon source and others fail to do so (Chen, et al., 1984); the catechol branch is similarly present in only certain strains of *A. tumefaciens*. When both branches are present, strains may also differ with respect to whether the conversion of β–ketoadipate enol-lactone to succinate and acetyl-CoA is performed by either *cat* or *pca* (or both) gene products (Figure 1.2). For example, *A. baylyi* encodes *pcaDIJF* and *catDIJF* as independently regulated transcriptional units, even though these isofunctional enzymes catalyze identical reactions associated with the last three steps of the β–ketoadipate pathway. In contrast, *P. putida* strain KT2440 does not encode *catIJF*; rather, the catabolism of β–ketoadipate as originating from both branches of the pathway is catalyzed by PcaIJ and PcaF (Jimenez, et al., 2002).

1.2.9 Characteristics of aromatic acid catabolism genes: gene organization and regulation

Supraoperonic clustering refers to the grouping of genes or operons that encode physiologically relevant functions in comparatively tight clusters; the organization of aromatic acid catabolic genes represents one of the best examples of supraoperonic clustering in prokaryotes. Genes involved in the β–ketoadipate pathway form supraoperons in *A. tumefaciens* (Parke, 1995), *A. baylyi* (Elsemore & Ornston, 1994), and *P. putida* (Jimenez, et al., 2002); remarkably, *A. baylyi* encodes 14 structural enzymes associated with the dissimilation of protocatechuate, quinate, shikimate, and *p*-hydroxybenzoate in a single cluster. The polycyclic aromatic hydrocarbon (PAH)-degrading bacterium *Mycobacterium vanbaalenii* encodes all enzymes involved in PAH

dissimilation (including β–ketoadipate genes) within a relatively small region of its genome (Kim, et al., 2008). Protocatechuate catabolism genes are linked to genes involved in the conversion of phthalate and fluorine in the Gram-positive *Terrabacter* (Habe, et al., 2005). One hypothesis to account for such a genetic arrangement is that the organization of related metabolic genes into discrete units greatly facilitates the horizontal transfer of entire pathways (Stanier & Ornston, 1973; Buchan, et al., 2004).

The large number of related but disparate compounds that may be enzymatically converted to catechol or protocatechuate necessitates a finely tuned regulatory mechanism to enable the most energy efficient metabolism. The importance of this is underscored by the consideration that not all compounds afford an equivalent net energy yield, and that many aromatics may be simultaneously available for catabolism. Several species have demonstrated an ability to selectively and sequentially utilize various aromatic substrates (Nichols & Harwood, 1995; Brzostowicz et al., 2003). Atleast three layers of regulation govern the expression of the *pcaIJFBDKCHGquiBCXA* operon in *A. baylyi.* The most direct method of regulation involves the repression and activation of *pca* expression, as modulated by PcaU (Gerischer, et al., 1998; Trautwein & Gerischer, 2001) in response to the availability of protocatechuate. However, when substrates of both protocatechuate and catechol branches of the β–ketoadipate pathway are presented, *A. baylyi* preferentially metabolizes benzoate, shikimate, quinate, protocatechuate and anthranilate over *p*-hydroxybenzoate (Brzostowicz, et al., 2003; see also Siehler, et al., 2007). This cross-regulation involves the participation of CatM and BenM, which act to induce expression of catechol and benzoate catabolism genes while simultaneously repressing expression of *pca* genes (Brzostowicz, et al., 2003). Finally, *pca*, *pob*, and *van* genes are subject to carbon catabolite repression and thus expression is negatively influenced by the presence of certain organic acids (particularly succinate and acetate) regardless of the occurrence of aromatic substrates (Dal, et al., 2002). A similar scenario has been described in *P. putida*, where Crc (global catabolite repression protein) has been demonstrated to bind to the 5'end of *benR* mRNA; this interaction presumably inhibits translation and the decreased levels of BenR exert a pleiotropic effect upon the

expression of *ben*, *cat*, and *pcaIJ* genes (Moreno & Rojo, 2008). In these ways, species prioritize the metabolism of certain compounds over others; presumably this differential response has evolved as an active mechanism to maximize the amount of enuergy (or carbon) that may be extracted from a particular environment.

1.2.10 Aromatic acid transport mechanisms

As a general rule, transport systems related to aromatic acid catabolic pathways belong to the major facilitator superfamily (MFS) of transporter proteins, a symporter family of permeases that utilizes proton motive force as a source of energy for solute transport (Pao, et al., 1998; Saier, et al., 1999). Proteins involved in the uptake of benzoate (BenK; Collier, et al., 1997), 2,4-dichlorophenoxyacetate (TfdK; Leveau, et al., 1998; Hawkins & Harwood, 2002), vanillate (VanK; D'Argenio, et al., 1999; Chaudhry, et al., 2007) and the β–ketoadipate metabolite *cis,cis*-muconate (MucK; Williams & Shaw, 1997) in part comprise the aromatic acid/H+ symporter subfamily within the MFS. The protocatechuate transport system (PcaK) is the founding member of this subfamily, and uptake of the aromatic acid has been the focus of study in *A. baylyi* (D'Argenio, et al., 1999), *Corynebacterium glutamicum* (Chaudhry, et al., 2007) and most extensively in *P. putida* (Harwood, et al., 1994; Nichols & Harwood, 1997; Ditty & Harwood, 1999; Ditty & Harwood, 2002). PcaK consists of twelve membrane-spanning domains that catalyze the uptake of protocatechuate, and also *p*-hydroxybenzoate in *P. putida* (Harwood, et al., 1994; Nichols & Harwood, 1997) and benzoate in *A. baylyi*. Expression of *pcaK* is modulated by the transcriptional regulators PcaU (Gerischer, et al., 1998) and PcaR (Nichols & Harwood, 1995), which also participate in the regulation of *pca* genes encoding catabolic enzymes for the β-ketoadipate pathway in *A. baylyi* and *P. putida*, respectively. Unusually, PcaK is directly involved in mediating a chemotactic response towards *p*-hydroxybenzoate and benzoate in the motile *P. putida* (Harwood, et al., 1994), a role that is unique to this protein and TfdK (Hawkins & Harwood, 2002) amongst members of MFS.

Aromatic acids such as protocatechuate (pK_a 4.48), benzoate (pK_a 4.19), and *p*-hydroxybenzoate (pK_a 4.48) may diffuse across a cell membrane as an undissociated acid, and thus an active transport system may seem superfluous. Indeed, disruption of *pcaK* results in a phenotype that is difficult to discern when cells are grown with these aromatic compounds at a neutral pH, as the concentration of undissociated acid is sufficient to permit growth in the absence of an active transport mechanism (Harwood, et al., 1994; Nichols & Harwood, 1997). However, the presence of an aromatic acid transport system may be important in a natural environment such as soil where these compounds are often present at low concentrations and a foraging bacterium must compete against other organisms for access to an energy source (Harwood, et al., 1994; Nichols & Harwood, 1997). A premise that PcaK makes a significant contribution towards the ability of a cell to effectively scavenge is underscored by the observation that *pcaK* expression in *P. putida* is repressed in the presence of benzoate, thereby allowing this species to sequentially and preferentially metabolize benzoate instead of *p*-hydroxybenzoate (Nichols & Harwood, 1995).

1.2.11 Aromatic acids induce a chemotactic response in motile bacteria

Aromatic compounds have been documented to stimulate chemotaxis in several soil-dwelling species of bacteria, including *Pseudomonas*, *Bradyrhizobium*, *Agrobacterium* and *Rhizobium* (Parke et al., 1985; Parke, et al., 1987; Harwood, et al., 1994). The threshold concentration required to elicit a chemotactic response (10^{-7} M; response of *B. japonicum* to many aromatic acids (Parke, et al., 1985)) is within a range that might be expected to accumulate in the rhizosphere surrounding the root of a plant, and thus secretion of these phenolics may aid in recruiting free-living and motile bacteria. In particular, β–ketoadipate is a strong chemoattractant for several species, and as this metabolite is produced during the metabolism of most aromatic acids, it is possible that it may serve as a signal to motile cells indicating the location of a potential source of energy (Parke, et al., 1985). Indeed, it has been demonstrated that cells will respond to aromatic compounds they do not have the ability to dissimilate (Parke, et al., 1985; Parke,

et al., 1987); this disconnect between chemotactic and metabolic capabilities may reflect the importance of these compounds as indicators signaling the availability of related compounds that may be more readily metabolized. Chemotaxis to aromatic acids appears to be a constitutive trait in *A. tumefaciens* and (in some cases) in *B. japonicum*, and it has been suggested that this unregulated expression is consistent with these compounds playing an important signaling role in the survival of these cells as saprophytes (Parke, et al., 1987). In contrast, chemotaxis by *P. putida* is inducible by growth with (for example) benzoate and *p*-hydroxybenzoate, as mediated by the transporter PcaK (Harwood, et al., 1994).

1.2.12 Biodegradative pathways associated with the β–ketoadipate pathway

The ability of soil-dwelling microorganisms to degrade or neutralize aromatic-containing pollutants is of obvious interest given the extensive usage of phenolic compounds in toxic chemicals such as solvents, pesticides, herbicides, and the persistence of such chemicals due to the thermodynamic stability of the aromatic ring. As such anthropogenic chemicals have only recently been introduced into the environment, the ability to catabolize these likely evolved relatively recently from pre-existing metabolic pathways as the result of a broadened or lax substrate specificity exhibited by the enzymes towards structurally familiar chemicals. The ability to dissimilate 3-chlorobenzoate in *P. putida* is plasmid-borne (pAC27), and it has been hypothesized that the genes involved (*clcABD*) may have originated as encoding products involved in catechol catabolism (*catBCA*) (Ngai & Ornston, 1988).

The degradation of many phenolic pollutants proceeds via *meta*-cleavage of the aromatic ring, and thus by-pass the β–ketoadipate pathway (which is initiated via *ortho*-cleavage), nonetheless, there are a few examples of bioremediation processes involving this catabolic pathway.

Polycyclic aromatic hydrocarbons (PAHs) are generated from industrial processes and may be highly toxic and carcinogenic to a variety of organisms (Figure 1.1). High molecular weight PAHs (such as pyrene and derivatives) are particularly difficult to dissimilate, however species such as *Mycobacterium vanbaalenii* encode the ability to

degrade these, thereby generating intermediates that are further processed via the protocatechuate branch of the β–ketoadipate pathway (Kim, et al., 2008). In fact, the degradation of PAHs phthalate, phenanthrene, pyrene, and fluorene, has been linked to the protocatechuate catabolic pathway in *M. vanbaalenii* and related species (Habe, et al., 2005; Patrauchan, et al., 2005; Kim, et al., 2008). *Rhodococcus* sp. strain RHA1 dissimilates biphenyls via the *bph* pathway, yielding 2-hydroxy-penta-2,4-dienoate and benzoate (Masai, et al., 1995), which is subsequently metabolized to tricarboxylic acids via the catechol branch of the β–ketoadipate pathway (Patrauchan, et al., 2005).

In summary, aromatic acids and related compounds accumulate in soil as originating primarily from plant-associated lignin. The β–ketoadipate pathway mediates the catabolism of a diverse range of aromatics via conversion to protocatechuate or catechol and is of broad taxonomic distribution in soil-dwelling bacteria. The pathway entails parallel and convergent branches whereby protocatechuate and catechol are acted upon by specific and non-interchangeable enzymes, resulting in the production of tricarboxylic acids acetyl-CoA and succinate. While the biochemical reactions catalyzed within the β–ketoadipate pathway are identical in all prokaryotes, the organization and regulation of associated genes may differ considerably between species. Nonetheless, the protocatechuate branch of the pathway is particularly well-represented in members of the α–proteobacteria, suggesting the presence of selective pressures acting to conserve protocatechuate catabolism in this class of bacteria. Aromatic acids represent an important carbon and energy source for soil microorganisms, as reflected by the ability of motile bacteria to mediate a chemotactic response to these compounds.

THE METABOLISM OF HYDROXYPROLINE IN BACTERIA

1.3.1 Hydroxyproline-rich proteins in plants and legumes

Hydroxyproline-rich glycoproteins (HRGPs) are extracellular or membrane-associated proteins that accumulate in plant cell walls and undergo significant post-translational modifications (glycosylation, proline hydroxylation, cross-linking) that allow these proteins to participate in a wide variety of functions such as providing structural support and facilitating intercellular communication (Wu, et al., 2001; Khashimova, 2003).

Hydroxyproline-rich extensins comprise the major structural protein of root cell walls and root nodules in *Medicago truncatula* (Frueauf, et al, 2000). Brewin and colleagues recognized a specific subclass of extensins (termed root nodule extensins) that are unique to legumes, and that are expressed in both root cells and nodules (Rathbun et al, 2002). These extensins are particularly associated with the matrix of the infection threads and infection droplets, as well as the intercellular spaces between uninfected nodule parenchyma cells and uninfected plant tissue located at the apex of the nodule (Rae, et al., 1991; 1992, Rathbun et al., 2002). Nodule extensins have been detected in empty infection threads in peas (Rae, et al., 1991), in the peribacteroid membrane and peribacteroid space of isolated pea symbiosomes and intact bean root nodules (Benhamou, 1991; Olsson, 2002), and are secreted in root cortex cells that are located in close proximity to infection threads in *Vicia* (Rae, et al., 1992). The distribution pattern of these hydroxyproline-rich proteins has led to the proposal that the accumulation of extensins occurs very early during infection with *Rhizobium* and that their enhanced secretion in surrounding host tissue is a direct response to colonization by the symbiont (Rae, et al., 1991; Rae, et al., 1992). Similarly, nodule extensins (as recognized by the monoclonal antibody MAC 265) are largely absent in pea nodules induced by a Fix- strain of *Rhizobium*, leading to the hypothesis that intercellular signaling events between plant and microbe may be required for expression of these proteins in root nodules (Olsson, et al., 2002); at the least, regulation of these extensins is modulated by the

nitrogen-fixing ability of the nodule. These proteins are also secreted from pea root tip cells of uninfected plants, and likely play a more general role in cell wall growth of legumes (Wisiniewski, et al., 2000; Rathbun, et al., 2002).

In addition to its association with plant cell walls in nodules, roots, and other plant tissues, high levels of hydroxyproline have been reported in soybean seed coats (5.2 μg Hyp/mg dry plant weight) with lower levels detected throughout the remaining soybean plant (Cassab, et al., 1985). HPGPs (primarily in the form of arabinogalactan proteins) are also enriched in the reproductive organs of plants including the extracellular matrix of the pistil and pollen tubes, the exudates released from the stigma, the ovaries, and pollen grains (Wu, et al., 2001).

Hydroxyproline-rich proteins are thus common to many plants species, and have been reported in all major plant organs, including roots, leaves, and flowers. Although HRGPs may comprise the principle structural proteins of root nodules in legumes (Frueauf, et al., 2000) and appear to be intimately associated with the symbiosome (Benhamou, et al., 1991; Olsson, et al., 2002), it would be highly speculative to conclude that rhizobia are exposed to free hydroxyproline in a form that is available as an energy source *in planta*, and we are not aware of any literature that offers direct and unambiguous evidence that supports this assumption. However, the prevalence of hydroxyproline in plants (and animals, in the form of collagen) implies the possibility that soil-dwelling microorganisms such as *Rhizobium* may encounter free hydroxyproline in a soil environment as arising from the decay of plant and animal matter. Several studies assaying the amino acid composition of soils have documented the presence of hydroxyproline in various soil types and depths, and hydroxylated proline may be particularly enriched in humic acids (Griffith, et al., 1976; Sowden, et al., 1976; Morita & Sowden, 1981). Low levels of hydroxyproline have likewise been reported in aspen and spruce leaf litter (Lahdesmaki & Phspanen, 1989). More particularly, the abundance of hydroxyproline-rich glycoproteins in the nodules, seeds, and roots of legumes may even result in an enrichment of hydroxyproline in soils where these plants predominate, in the form of senescing nodules for example. These are the environments where rhizobia

proliferate and the ability to catabolize hydroxyproline (or any other legume or plant-associated compound) may confer a selective advantage to cells that encode this metabolic pathway.

1.3.2 The hydroxyproline metabolic pathway in bacteria

Hydroxyproline catabolism has been documented in several species of soil-dwelling bacteria, including *Pseudomonas* (Adams, 1959; Jayaraman & Radhakrishnan, 1965a; Jayaraman & Radhakrishnan, 1965b; Thacker, 1968; Gryder & Adams, 1969; Manoharan, 1980). The metabolic pathway in prokaryotes was elucidated from one strain of *P. putida* that was isolated from soil in a garden at the NIH via selective culturing for growth with hydroxyproline as a sole source of nitrogen and carbon (Adams, 1959; Adams & Frank, 1980). This strain readily oxidized all four isomers of hydroxyproline (Figure 1.4), however initial studies revealed that only growth with *trans*-4-hydroxy-L-proline or *cis*-4-hydroxy-D-proline led to the induction of hydroxyproline catabolic enzymes (Adams, 1959). Few studies have further examined the metabolism of either *cis*-4-hydroxy-L-proline or *trans*-4-hydroxy-D-proline in prokaryotes (Jayaraman & Radhakrishnan, 1965a).

The metabolism of *trans*-4-hydroxy-L-proline (*trans*-hyp) in bacteria involves four steps and ultimately yields ammonia, and α-ketoglutarate, which is subsequently funneled into the tricarboxylic acid cycle (Figure 1.5) (Adams, 1973; Adams & Frank, 1980). The initial step involves the epimerization of *trans*-hyp to *cis*-4-hydroxy-D-proline (*cis*-hyp) and is catalyzed by a hydroxyproline-2-epimerase (Adams, 1959; Adams & Norton, 1964). The epimerase is the most extensively studied of the hydroxyproline-specific enzymes, and has been purified in *P. putida* (Adams, 1959; Adams & Norton, 1964; Zervos & Adams, 1975; Ramaswamy, 1984) and more recently the epimerases of *P. aeruginosa*, *Burkholderia pseudomallei*, and *Brucella* species have been purified as recombinant proteins through the use of affinity chromatography (Goytia, et al., 2007).

At the time of its initial purification, the hydroxyproline-inducible epimerase of *P. putida* was the first reported example of an amino acid isomerase obtained as a pure protein (Adams & Norton, 1964). Unexpectedly, the purified epimerase was not associated with pyridoxal phosphate (Adams & Norton, 1964); previously, this coenzyme had been shown to be essential for the activity of several partially purified amino acid racemases. It has since been reported that proline racemase also lacks pyridoxal phosphate (Cardinale & Abeles, 1968); accordingly, it has been proposed that racemases which act upon primary amino acids require pyridoxal phosphate as a cofactor whereas the racemases/epimerases which recognize secondary amino acids do not utilize any such coenzyme (Adams, 1973).

The epimerase active site includes two conserved and catalytic cysteine residues that are essential for enzyme activity; substitution of either of these cysteines with serine abolishes hydroxyproline epimerase activity in *P. aeruginosa* (Goytia, et al., 2007). The epimerization reaction likely involves the cysteine sulfhydryls acting as acceptor/donors of the α-hydrogen of hydroxyproline, catalyzing a proton exchange reaction that results in the isomerization of the substrate (Ramaswamy, 1984).

The D-isomer of *cis*-4-hydroxyproline is recognized by an oxidase that converts the modified amino acid to a cyclic ketimine via a dehydrogenation reaction (Adams, 1959). It was immediately apparent that the hydroxyproline oxidase activity was sequestered within the insoluble pellet upon centrifugation of cell-free lysate and attempts to purify this enzyme were frustrated by the inability to isolate the protein in a soluble form (Adams, 1959; Yoneya & Adams, 1961; Adams, 1973). Despite this, the end product resulting from the dehydrogenation of *cis*-hyp was correctly identified as Δ^1-pyrroline-4-hydroxy-2-carboxylic acid (Adams, 1959; Yoneya & Adams, 1961). Unusually, the *cis*-hyp oxidase described by Adams did not exhibit the broad substrate specificity characteristic of other D-amino acid oxidases; for example, it is possible to substitute the hydroxyproline-inducible oxidase of *P. putida* with hog kidney D-amino acid oxidase as this protein recognizes a range of substrates, however the hydroxyproline

catabolic enzyme was unable to efficiently catalyze the oxidation of any other D-amino acids (Adams, 1959).

The penultimate step in the hydroxyproline metabolic pathway involves an inducible deaminase which catalyzes the conversion of Δ^1-pyrroline-4-hydroxy-2-carboxylic acid to yield an α–ketoglutaric semialdehyde and ammonia (Singh & Adams, 1964; Singh & Adams, 1965a; Singh & Adams, 1965b). The identification of α-ketoglutaric semialdehyde (referred to as 2,5-dioxovalerate) as a biologically relevant metabolite was unprecedented (Adams, 1959; Singh & Adams, 1964), however this intermediate was soon isolated from other metabolic pathways (Dagley & Trudgill, 1965). The deaminase was purified from extracts of *Pseudomonas* cells grown upon hydroxyproline; enzyme assays performed with purified protein and a range of plausible substrates indicated that enzyme was specific in its recognition of Δ^1-pyrroline-4-hydroxy-2-carboxylic acid as a substrate (Singh & Adams, 1965a).

The oxidation of α-ketoglutaric semialdehyde to α-ketoglutarate in *Pseudomonas* may be catalyzed by any one of several isoenzymes, thus hindering the purification of the hydroxyproline-inducible dehydrogenase. In particular, the similarity between a glucarate-inducible α-ketoglutaric semialdehyde dehydrogenase and the hydroxyproline enzyme (in terms of physical and kinetic parameters) made it particularly difficult to discern whether these disparate pathways involved one common enzyme or distinct proteins (Adams & Rosso, 1967). However, purification of enzyme activity obtained from cells grown with either hydroxyproline or glucarate revealed that genetically, immunochemically, and physically distinct enzymes were associated with each pathway (Koo & Adams, 1974).

Intriguingly, the purified dehydrogenase exhibited an unexpected ability to catalyze the deamination of Δ^1-pyrroline-4-hydroxy-2-carboxylate, yielding α–ketoglutaric semialdehyde which was sequentially acted upon by the enzyme and converted to α–ketoglutarate (Koo & Adams, 1974). The possibility that the one enzyme performed dual functions (i.e., acted as a deaminase and semialdehyde dehydrogenase) cannot be excluded based upon published reports describing either hydroxyproline-

inducible enzyme. An enzyme demonstrating Δ^1-pyrroline-4-hydroxy-2-carboxylate deaminase activity was purified and characterized in *P. putida*, however the molecular size of the protein (~60 kDa (Singh & Adams, 1965a) is comparable to that estimated per subunit of the hydroxyproline dehydrogenase (which may consist of two 60 kDa subunits (Koo & Adams, 1974)). As well, fractions containing the 'purified' deaminase likewise exhibited some dehydrogenase activity (Singh & Adams, 1965a; Koo & Adams, 1974). Finally, the report of a mutant strain of *P. putida* (strain M-14) lacking α-ketoglutaric semialdehyde dehydrogenase activity supports a linkage between the two enzyme activities as this strain was also demonstrated to lack the relevant deaminase activity, although the activity of the remaining hydroxyproline catabolic enzymes was unaffected (Koo & Adams, 1974). The nature of the mutation was not conclusively demonstrated however it likely resulted in the synthesis of a non-functional form of the dehydrogenase (i.e., structural mutation) as opposed to a loss of hydroxyproline-inducible expression (i.e., regulatory mutation) that might be expected to simultaneously affect the activity levels of more than one enzyme. Conversely, the description of a series of hydroxyproline mutants of *P. aeruginosa* failed to identify any which lacked both deaminase and dehydrogenase activities, although multiple strains lacking only one of the enzyme activities were obtained (Manoharan, 1980). Thus there appear to be two distinct genes encoding separate deaminase and dehydrogenase enzyme activities in *P. aeruginosa*.

1.3.3 Hydroxyproline transport in *Pseudomonas*

The systems that participate in the uptake or transport of hydroxyproline in bacteria have not been as extensively studied as the hydroxyproline metabolic enzymes, however a system dedicated to the uptake of this amino acid has been described in *P. putida* (Gryder & Adams, 1969; Gryder & Adams, 1970). Early studies had indicated that growth of an epimerase-minus strain of *P. putida* upon *cis*-hyp as a sole source of nitrogen and carbon was inhibited by the addition of relatively low levels of *trans*-hyp (which the strain could not metabolize) (Gryder & Adams, 1969). This phenotype could

not be attributed to the production of a toxic metabolite or intermediate, as growth of the mutant was restored upon addition of alternative sources of nitrogen and carbon. Subsequent analyses revealed that the affinity of the transport system for *trans*-hyp was considerably greater than that for *cis*-hyp (K_m values of 30 μM and 1 mM, respectively) (Gryder & Adams, 1970). Consequently, the growth inhibition previously observed in the epimerase-mutant was a reflection of this difference in affinities between the two epimers; the uptake system would bind *trans*-hyp much more effectively than *cis*-hyp however the mutant cells were unable to further metabolize *trans*-hyp and starved.

Competition experiments examining uptake of labeled *trans*- and *cis*-hyp in *P. putida* confirmed that the presence of (unlabeled) *trans*-hyp strongly reduces uptake of (labeled) *cis*-hyp. While the reciprocal is also true, *cis*-hyp is considerably less effective in reducing uptake of *trans*-hyp (Gryder & Adams, 1969; Gryder & Adams, 1970). *P. aeruginosa* strain PAO is completely impermeable to *cis*-hyp and cells cannot utilize this compound as a sole carbon source unless made artificially permeable by the addition of EDTA; in contrast, these cells readily transport *trans*-hyp (Manoharan, 1980).

1.3.4 The genetics and regulation of hydroxyproline uptake and catabolism

Although the hydroxyproline catabolic enzymes have been (mostly) purified in *Pseudomonas*, our understanding of the genetics underlying the hydroxyproline metabolic pathway is not well developed. Genes encoding enzymes with hydroxyproline-2-eprimerase and hydroxyproline-related α-ketoglutaric semialdehyde dehydrogenase activities have been identified in several species (Goytia, et al., 2007; Watanabe, et al., 2007), however the genetic identity of the remaining enzymes (*cis*-4-hydroxy-D-proline oxidase and Δ^1-pyrroline-4-hydroxy-2-carboxylate deaminase) is unknown. Also, genes encoding a hydroxyproline-uptake system have yet to be described in any species.

The hydroxyproline catabolism and transport enzyme activities are induced in cells grown in the presence of hydroxyproline (Adams, 1959; Yoneya & Adams, 1961; Jayaraman & Radhakrishnan, 1965a; Jayaraman & Radhakrishnan, 1965b; Gryder &

Adams, 1969; Singh & Adams, 1965a; Koo & Adams, 1974; Manoharan, 1980). The identification of the inducing metabolite is complicated by the rapid interconversion of *trans*- and *cis*-forms of hydroxyproline via the hydroxyproline-2-epimerase. In *P. putida*, the study of an epimerase-negative mutant revealed that *trans*-hyp was an intrinsic activator of hydroxyproline gene expression, as growth of the mutant in the presence of this compound resulted in fully induced levels of enzyme activities (Gryder & Adams, 1969). Similarly, growth with *cis*-hyp indicates that this epimer may also act as an inducer, however metabolism of *cis*-hyp to α-ketoglutarate is not impaired in an epimerase-minus strain and thus it is formally possible that another downstream metabolite is responsible for inducing gene expression (Figure 1.5). Comparable studies of hydroxyproline mutant strains of *P. aeruginosa* lead to the conclusion that *trans*-hyp induced gene expression (Manoharan, 1980).

The expression of hydroxyproline catabolic genes does not appear to be influenced by catabolite repression in *P. putida* (Gryder & Adams, 1969). Hydroxyproline metabolism was measured indirectly (via O_2 consumption) and directly (assays of enzyme activity) in cells grown with hydroxyproline in the presence or absence of glucose, succinate, glutamate, and α-ketoglutarate, with no repressive effect evident.

Genes specifying hydroxyproline metabolism enzymes were determined not to be plasmid-borne in *P. aeruginosa* (Manoharan, 1980) and have been mapped to the chromosome of this species as a gene cluster (Manoharan & Jayaraman, 1979). It is likely that the genes encoding the metabolic enzymes and transport system are organized in an operon or share a common regulator in *P. putida*, as evidenced by a regulatory mutant strain in which the up-regulation of both groups is simultaneously abolished (Gryder & Adams, 1970).

1.3.5 The Relationship Between Hydroxyproline and Proline

Many species of bacteria including members of the genus *Pseudomonas* have the ability to utilize proline as a primary source of nitrogen, carbon, and energy (Vilchez, et

al., 2000). The physical and chemical similarities inherent between proline and hydroxyproline (Figure 1.4) raise the possibility of a shared transport system and/or cross-induction of catabolic genes specific to either pathway. Reports on this matter in *Pseudomonas* have been contradictory however, and it is unclear to what degree the presence of proline influences hydroxyproline gene expression.

There is some evidence to support a hypothesis that hydroxyproline and proline share a common transport system in *Pseudomonas*. The addition of L-proline as a competitor strongly reduces uptake of labeled *trans*-hyp in *P. putida* (Gryder & Adams, 1970), and the reciprocal example of *trans*-hyp blocking L-proline transport in *P. aeruginosa* has also been described (Manoharan, 1980). Likewise, *P. aeruginosa* cells grown upon *trans*-hyp induce an uptake system that effectively transports L-proline (Manoharan, 1980). A hydroxyproline-2-epimerase mutant of *P. putida* grows readily upon proline as an organic source yet the addition of *trans*-hyp to growth medium (which the strain is unable to metabolize) inhibits growth, presumably due to competition between the two substrates for binding of a saturable uptake system (Gryder & Adams, 1969). These reports imply that proline and hydroxyproline may utilize a common uptake system in *Pseudomonas*, however this data is not conclusive and it is uncertain whether similar overlap exists in other species.

Data regarding possible cross-induction of hydroxyproline and proline catabolic genes are even less clear. *P. aeruginosa* proline-grown cells exhibit a low rate of hydroxyproline oxidation (that is nonetheless above background levels) and *trans*-hyp grown cells likewise metabolize L-proline to a degree that is consistent with a low level of induction; these results were also confirmed by direct enzyme assays (Manoharan, 1980). Yoneya and Adams reported that the activity of the *cis*-4-hydroxy-D-proline oxidase was induced at low levels in *P. putida* when cultured with L-proline but not L-glutamate (Yoneya & Adams, 1961). This is in disagreement with a finding of no induction of Δ^1-pyrroline-4-hydroxy-2-carboxylate deaminase activity in the same strain grown under comparable conditions (Singh & Adams, 1965). Initial reports detailing hydroxyproline catabolism in *Pseudomonas* and *Achromobacter* are also inconsistent

with induction of hydroxyproline gene expression via L-proline (Adams, 1959). Possibly, the presence of small amounts of contaminating hydroxyproline in prepared samples of L-proline (and vice versa) may account for the disparity evident in these early reports. At the least, the degree to which hydroxyproline and proline transport and catabolic functions overlap is ambiguous.

In summary, hydroxyproline is prevalent in plant tissue in the form of hydroxyproline-rich glycoproteins, which perform a variety of functions that include the provision of structural support in cell walls. Hydroxyproline is present in soil and decaying plant material and the soil microbe *Pseudomonas* has acquired the ability to utilize this amino acid as a carbon and nitrogen source. The pathway by which *trans*-4-hydroxy-L-proline is metabolized has been elucidated in *P. putida* and results in the formation of ammonia and α–ketoglutarate. Although the biochemical reactions associated with the catabolic pathway have been described, the genetics underlying hydroxyproline metabolism in prokaryotes remain poorly characterized.

1.4 References:

Adams, E. (1959). Hydroxyproline metabolism. I. Conversion to alpha-ketoglutarate by extracts of *Pseudomonas*. The Journal of Biological Chemistry, **234**:2073-2084.

Adams, E. (1973). The metabolism of hydroxyproline. Molecular and Cellular Biochemistry, **2**:109-119.

Adams, E., & Frank, L. (1980). Metabolism of proline and the hydroxyprolines. Annual Review of Biochemistry, **49**:1005-1061.

Adams, E., & Norton, I. (1964). Purification and properties of inducible hydroxyproline 2-epimerase from *Pseudomonas*. The Journal of Biological Chemistry, **239**:1525-1535.

Adams, E., & Rosso, G. (1967). Alpha-ketoglutaric semialdehyde dehydrogenase of *Pseudomonas*. Properties of the purified enzyme induced by hydroxyproline and of the glucarate-induced and constitutive enzymes. The Journal of Biological Chemistry, **242**:1802-1814.

Aislabie, J.M., Chhour, K.L., Saul, D.J., Miyauchi, S., Ayton, J., Paetzold, R.F., Balks, M.R. (2006). Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. Soil Biology and Biochemistry, **38**:3041-3056.

Aldrich, T.L., & Chakrabarty, A.M. (1988). Transcriptional regulation, nucleotide sequence, and localization of the promoter of the *catBC* operon in *Pseudomonas putida*. Journal of Bacteriology, **170**:1297-1304.

Amann, R.I., Ludwig, W., & Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological Reviews, **59**:143-169.

Avis, T.J., Gravel, V., Antoun, H., & Tweddell, R.J. (2008). Multifaceted beneficial effects of rhizosphere microorganisms on plant health and productivity. Soil Biology and Biochemistry, **40**:1733-1740.

Benhamou, N., Lafontaine, P.J., Mazau, D., & Esquerre-Tugaye, M.T. (1991). Differential accumulation of hydroxyproline-rich glycoproteins in bean root nodule cells infected with a wild-type strain or a C4-dicarboxylic acid mutant of *Rhizobium leguminosarum* bv. *phaseoli*. Planta, **184**:457-467.

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

Brzostowicz, P.C., Reams, A.B., Clark, T.J., & Neidle, E.L. (2003). Transcriptional cross-regulation of the catechol and protocatechuate branches of the β-ketoadipate pathway contributes to carbon source-dependent expression of the *Acinetobacter* sp. strain ADP1 *pobA* gene. Applied and Environmental Microbiology, **69**:1598-1606.

Buchan, A., Collier, L.S., Neidle, E.L., & Moran, M.A. (2000). Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. Applied and Environmental Microbiology, **66**:4662-4672.

Buchan, A., Neidle, E.L., & Moran, M.A. (2004). Diverse organization of genes of the β-ketoadipate pathway in members of the marine *Roseobacter* lineage. Applied and Environmental Microbiology, **67**:1658-1668.

Buchan, A., Neidle, E.L., & Moran, M.A. (2001). Diversity of the ring-cleaving dioxygenase gene *pcaH* in a salt marsh bacterial community. Applied and Environmental Microbiology, **70**:5801-5809.

Burges, N.A., Hurst, H.M., & Walkden, B. (1964). The phenolic constituents of humic acid and their relation to the lignin of the plant cover. Geochimica et Cosmochimica Acta, **28**:1547-1554.

Cardinale, G.J., & Abeles, R.H. (1968). Purification and mechanism of action of proline racemase. Biochemistry, **7**:7-11.

Cassab, G.I., Nieto-Sotelo, J., Cooper, J.B., van Holst, G.J., & Varner, J.E. (1985). A Developmentally Regulated Hydroxyproline-Rich Glycoprotein from the Cell Walls of Soybean Seed Coats. Plant Physiology, **77**:532-535.

Chaudhry, M., Huang, Y., Shen, X., Poetsch, A., Jiang, C., & Liu, S. (2007). Genome-wide investigation of aromatic acid transporters in *Corynebacterium glutamicum*. Microbiology, **153**:857-865.

Chen, Y.P., Glenn, A.R., & Dilworth, M.J. (1984). Uptake and oxidation of aromatic substrates by *Rhizobium leguminosarum* MNF 3841 and *Rhizobium trifolii* TA1. FEMS Microbiology Letters, **21**:201-205.

Chugani, S.A., Parsek, M.R., & Chakrabarty, A.M. (1998). Transcriptional repression mediated by LysR-type regulator CatR bound at multiple binding sites. Journal of Bacteriology, **180**:2367-2372.

Clark, T.J., Momany, C., & Neidle, E.L. (2002). The *benPK* operon, proposed to play a role in transport, is part of a regulon for benzoate catabolism in *Acinetobacter* sp. strain ADP1. Microbiology, **148**:1213-1223.

Coelho, R.R.R., Sacramento, D.R., & Linhares, L.F. (1997). Amino sugars in fungal melanins and soil humic acids. European Journal of Soil Science, **48**:425-429.

Collier, L.S., Gaines, G.L., & Neidle, E.L. (1998). Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. Journal of Bacteriology, **180**:2493-2501.

Collier, L.S., Nichols, N.N., & Neidle, E.L. (1997). *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. Journal of Bacteriology, **179**:5943-5946.

Cortez, J., & Schnitzer, M. (1981). Reactions of nucleic acid bases with inorganic soil constituents. Soil Biology and Biochemistry, **13**:173-178.

Cosper, N.J., Collier, L.S., Clark, T.J., Scott, R.A., & Neidle, E.L. (2000). Mutations in *catB*, the gene encoding muconate cycloisomerase, activate transcription of the distal *ben* genes and contribute to a complex regulatory circuit in *Acinetobacter* sp. strain ADP1. Journal of Bacteriology, **182**:7044-7052.

Cytryn, E.J., Sangurdekar, D.P., Streeter, J.G., Franck, W.L, Chang, W.S, Stacey, G., Emerich, D.W., Joshi, T., Xu, D., & Sadowski, M.J. (2007). Transcriptional and physiological responses of *Bradyrhizobium japonicum* to desiccation-induced stress. Journal of Bacteriology, **189**:6751-6762.

Dagley, S., & Trudgill, P.W. (1965). The metabolism of galactarate, D-glucarate and various pentoses by species of *Pseudomonas*. Biochemical Journal, **95**:48-58.

Dal, S., Steiner, I., & Gerischer, U. (2002). Multiple Operons Connected with Catabolism of Aromatic Compounds in *Acinetobacter* sp. Strain ADP1 are under Carbon Catabolite Repression. Journal of Molecular Microbiology and Biotechnology, **4**:389-404.

Dal, S., Trautwein, G., & Gerischer, U. (2005). Transcriptional Organization of Genes for Protocatechuate and Quinate Degradation from *Acinetobacter* sp. Strain ADP1. Applied and Environmental Microbiology, **71**:1025-1034.

D'Argenio, D.A, Segura, A., Bunz, P.V., & Ornston, L.N. (2001). Spontaneous mutations affecting transcriptional regulation by protocatechuate in *Acinetobacter*. FEMS Microbiology Letters, **201**:15-19.

D'Argenio, D.A., Segura, A., Coco, W.M., Bunz, P.V., & Ornston, L.N. (1999b). The physiological contribution of *Acinetobacter* PcaK, a transport system that acts upon protocatechuate, can be masked by the overlapping specificity of VanK. Journal of Bacteriology, **181**:3505-3515.

D'Argenio, D.A., Vetting, M.W., Ohlendorf, D.H., & Ornston, L.N. (1999a). Substitution, insertion, deletion, suppression, and altered substrate specificity in functional protocatechuate 3,4-dioxygenases. Journal of Bacteriology, **181**:6478-6487.

de Weert, S., Vermeiren, H., Mulders, I.H.M., Kuiper, I., Hendrickx, N., Bloemberg, G.V., Vanderleyden, J., De Mot, R., Lugtenberg, B.J.J. (2002). Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens*. Molecular Plant and Microbe Interactions, **15**:1173-1180.

Dec, J., Haider, K., & Bollag, J.M. (2003). Release of substituents from phenolic compounds during oxidative coupling reactions. Chemosphere, **52**:549-556.

Ditty, J.L., & Harwood, C.S. (2002). Charged amino acids conserved in the aromatic acid/H+ symporter family of permeases are required for 4-hydroxybenzoate transport by PcaK from *Pseudomonas putida*. Journal of Bacteriology, **184**:1444-1448.

Ditty, J.L., & Harwood, C.S. (1999). Conserved cytoplasmic loops are important for both the transport and chemotaxis functions of PcaK, a protein from *Pseudomonas putida* with 12 membrane-spanning regions. Journal of Bacteriology, **181**:5068-5074.

Elsemore, D.A., & Ornston, L.N. (1994). The *pca-pob* supraoperonic cluster of *Acinetobacter calcoaceticus* contains *quiA*, the structural gene for quinate-shikimate dehydrogenase. Journal of Bacteriology, **176**:7659-7666.

Eulberg, D., Lakner, S., Golovleva, L.A., & Schlomann, M. (1998). Characterization of a protocatechuate catabolic gene cluster from *Rhodococcus opacus* 1CP: evidence for a merged enzyme with 4-carboxymuconolactone-decarboxylating and 3-oxoadipate enol-lactone-hydrolyzing activity. Journal of Bacteriology, **180**:1072-1081.

Evans, R., Parr, W., & Evans, W. (1949). The bacterial oxidation of aromatic compounds. Biochemical Journal, 8.

Ferrer, J.L., Austin, M.B., Stewart, C., & Noel, J.P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiology and Biochemistry, **46**:356-370.

Fierer, N., & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Science USA, **103**:626-631.

Filley, T.R., Hatcher, P.G., Shortle, W.C., & Praseuth, R.T. (2000). The application of C-labeled tetramethylammonium hydroxide (C-TMAH) thermochemolysis to the study of fungal degradation of wood. Organic Geochemistry, **31**:181-198.

Formanek, P., Klejdus, B., & Vranova, V. (2005). Bio-available amino acids extraction from soil by demineralized water and 0.5 M ammonium acetate. Amino Acids, **28**:427-429.

Friedel, J.K., & Scheller, E. (2002). Composition of hydrolysable amino acids in soil organic matter and soil microbial biomass. Soil Biology and Biochemistry, **34**:315-325.

Frueauf, J.B., Dolata, M., Leykam, J.F., Lloyd, E.A., Gonzales, M., VandenBosch, K., Kieliszewski, M.J. (2000). Peptides isolated from cell walls of *Medicago truncatula* nodules and uninfected root. Phytochemistry, **55**:429-438.

Gal, M., Preston, G.M., Massey, R.C., Spiers, A.J., & Rainey, P.B. (2003). Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. Molecular Ecology, **12**:3109-3121.

Gallet, C., & Keller, C. (1999). Phenolic composition of soil solutions: comparative study of lysimeter and centrifuge waters. Soil Biology and Biochemistry, **31**:1151-1160.

Gans, J., Wolinsky, M., & Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science, **309**:1387-1390.

Gerischer, U., & Ornston, L.N. (1995). Spontaneous mutations in *pcaH* and *-G*, structural genes for protocatechuate 3,4-dioxygenase in *Acinetobacter calcoaceticus*. Journal of Bacteriology, **177**:1336-1347.

Gerischer, U., Segura, A., & Ornston, L.N. (1998). PcaU, a transcriptional activator of genes for protocatechuate utilization in *Acinetobacter*. Journal of Bacteriology, **180**:1512-1524.

Glaser, B., Turrion, M.B., & Alef, K. (2004). Amino sugars and muramic acid: biomarkers for soil microbial community structure analysis. Soil Biology and Biochemistry, **36**:399-407.

Goytia, M., Chamond, N., Cosson, A., Coatnoan, N., Hermant, D., Berneman, A., Minoprio, P. (2007). Molecular and structural discrimination of proline racemase and hydroxyproline-2-epimerase from nosocomial and bacterial pathogens. *PLoS ONE* , **2**:e885.

Grandy, A.S., & Neff, J.C. (2008). Molecular C dynamics downstream: The biochemical decomposition sequence and its impact on soil organic matter structure and function. Science of the Total Environment, **404**:297-307.

Griffith, S.M., Sowden, F.J., & Schnitzer, M. (1976). The alkaline hydrolysis of acid-resistant soil and humic acid residues. Soil Biology and Biochemistry, **8**:529-531.

Gryder, R.M., & Adams, E. (1969). Inducible degradation of hydroxyproline in *Pseudomonas putida*: pathway regulation and hydroxyproline uptake. Journal of Bacteriology, **97**:292-306.

Gryder, R.M., & Adams, E. (1970). Properties of the inducible hydroxyproline transport system of *Pseudomonas putida*. Journal of Bacteriology, **101**:948-958.

Habe, H., Chung, J.S., Ishida, A., Kasuga, K., Ide, K., Takemura, T., Nojiri, H., Yamane, H., Omori, T. (2005). The fluorene catabolic linear plasmid in *Terrabacter* sp. strain DBF63 carries the beta-ketoadipate pathway genes, *pcaRHGBDCFIJ*, also found in proteobacteria. Microbiology, **151**:3713-3722.

Hartnett, G.B., Averhoff, B., & Ornston, L.N. (1990). Selection of *Acinetobacter calcoaceticus* mutants deficient in the *p*-hydroxybenzoate hydroxylase gene (*pobA*), a member of a supraoperonic cluster. Journal of Bacteriology, **172**:6160-6161.

Harwood, C.S., & Parales, R.E. (1996). The beta-ketoadipate pathway and the biology of self-identity. Annual Review of Microbiology, **50**:553-590.

Harwood, C.S., Nichols, N.N., Kim, M.K., Ditty, J.L., & Parales, R.E. (1994). Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. Journal of Bacteriology, **176**:6479-6488.

Hautala, K., Peuravuori, J., & Pihlaja, K. (1997). Estimation of origin of lignin in humic DOM by CuO-oxidation. Chemosphere, **35**:809-817.

Hawkins, A.C., & Harwood, C.S. (2002). Chemotaxis of *Ralstonia eutropha* JMP134 (pJP4) to the herbicide 2,4-dichlorophenoxyacetate. Applied and Environmental Microbiology, **68**:968-972.

Hayaishi, O., & Hashimoto, K. (1950). Pyrocatecase a new enzyme catalyzing oxidative breakdown of pyrocatechin. The Journal of Biochemistry, **37**:371-374.

Hayaishi, O., & Stanier, R. (1951). The bacterial oxidation of tryptophan. III. Enzymatic activities of cell-free extracts from bacteria employing the aromatic pathway. Journal of Bacteriology, **62**:691-709.

Hinsinger, P., Gobran, G.R., Gregory, P.J., & Wenzel, W.W. (2005). Rhizosphere geometry and heterogeneity arising from root-mediated physical and chemical processes. New Phytologist, **168**:293-303.

Hofrichter, M., Vares, T., Kalsi, M., Galkin, S., Scheibner, K., Fritsche, W., et al. (1999). Production of manganese peroxidase and organic acids and mineralization of C-labelled lignin (C-DHP) during solid state fermentation of wheat straw with the white rot fungus *Nematoloma frowardii*. Applied and Environmental Microbiology, **65**:1864-1870.

Iwagami, S.G., Yang, K., & Davies, J. (2000). Characterization of the protocatechuic acid catabolic gene cluster from *Streptomyces* sp. strain 2065. Applied and Environmental Microbiology, **66**:1499-1508.

Jaeger, C.H., Lindow, S.E., Miller, W., Clark, E., & Firestone, M.K. (1999). Mapping of sugar and amino acid availability in soil around roots with bacterial sensors of sucrose and tryptophan. Applied and Environmental Microbiology, **65**:2685-2690.

Jayaraman, K., & Radhakrishnan, A.N. (1965a). Bacterial metabolism of hydroxyproline. I. Metabolism of L-allohydroxyproline by a *Pseudomonas*. Indian Journal of Biochemistry, **2**:145-153.

Jayaraman, K., & Radhakrishnan, A.N. (1965b). Bacterial metabolism of hydroxyproline. II. Metabolism of L-hydroxyproline in *Achromobacter*. Indian Journal of Biochemistry, **2**:153-158.

Jimenez, J.I., Minambres, B., Garcia, J.L., & Diaz, E. (2002). Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environmental Microbiology, **4**:824-841.

Johnson, R., & Pregitzer, K. (2007). Concentration of sugars, phenolic acids, and amino acids in forest soils exposed to elevated atmospheric CO_2 and O_3. Soil Biology and Biochemistry, **39**:3150-3166.

Jones, D.L., Kemmitt, S.J., Wright, D., Cuttle, S.P., Bol, R., & Edwards, A.C. (2005). Rapid intrinsic rates of amino acid biodegradation in soils are unaffected by agricultural management strategy. Soil Biology and Biochemistry, **37**:1267-1275.

Jones, D.L., & Kielland, K. (2002). Soil amino acid turnover dominates the nitrogen flux in permafrost-dominated taiga forest soils. Soil Biology and Biochemistry, **34**:209-219.

Jones, D.L., Owen, A.G., & Farrar, J.F. (2002). Simple method to enable the high resolution determination of total free amino acids in soil solutions and soil extracts. Soil Biology and Biochemistry, **34**:1893-1902.

Khashimova, Z.S. (2003). Hydroxyproline-containing plant proteins. Chemistry of Natural Compounds, **39**:229-236.

Kielland, K. (1995). Landscape patterns of free amino acids in arctic tundra soils. Biogeochemistry, **31**:85-98.

Kilby, B.A. (1951). The formation of beta-ketoadipic acid by bacterial fission of aromatic rings. Biochemical Journal, **49**:671-674.

Kilby, B.A. (1948). The bacterial oxidation of phenol to beta-ketoadipic acid. Biochemical Journal, **43**:5-6.

Kim, S.J., Kweon, O., Jones, R.C., Edmondson, R.D., & Cerniglia, C.E. (2008). Genomic analysis of polycyclic aromatic hydrocarbon degradation in *Mycobacterium vanbaalenii* PYR-1. Biodegradation, **19**:859-881.

Kogel, I., & Bochter, R. (1985). Characterisation of lignin in forest humus layers by high-performance liquid chromatography of cupric oxide oxiation products. Soil Biology and Biochemistry, **17**:637-640.

Koo, P.H., & Adams, E. (1974). Alpha-ketoglutaric semialdehyde dehydrogenase of *Pseudomonas*. Properties of the separately induced isoenzymes. The Journal of Biological Chemistry, **249**:1704-1716.

Kowalchuk, G.A., Hartnett, G.B., Benson, A., Houghton, J.E., Ngai, K.L., & Ornston, L.N. (1994). Contrasting patterns of evolutionary divergence within the *Acinetobacter calcoaceticus pca* operon. Gene, **146**:23-30.

Kumar, R., Bhatia, R., Kukreja, K., Behl, R.K., Dudeja, S.S., & Narula, N. (2007). Establishment of *Azotobacter* on plant roots:chemotactic response, development and analysis of root exudates of cotton (*Gossypium hirsutum* L.) and wheat (*Triticum aestivum* L.). Journal of Basic Microbiology, **47**:436-439.

Lahdesmaki, P., & Phspanen, R. (1989). Changes in concentrations of free amino acids during humification of spruce and aspen leaf litter. Soil Biology and Biochemistry, **21**:975-978.

Lange, S.J., & Que, L. (1998). Oxygen activating nonheme iron enzymes. Current Opinion in Chemical Biology, **2**:159-172.

Leveau, J.H., Zehnder, A.J., & van der Meer, Jr. (1998). The *tfdK* gene product facilitates uptake of 2,4-dichlorophenoxyacetate by *Ralstonia eutropha* JMP134 (pJP4). Journal of Bacteriology, **180**:2237-2243.

Li, Z., Xu, J., Tang, C., Wu, J., Muhammed, A., & Wang, H. (2006). Application of 16S rDNA-PCR amplification and DGGE fingerprinting for detection of shift in microbial community diversity in Cu-, Zn-, and Cd-contaminated paddy soils. Chemosphere, **62**:1374-1380.

Liesack, W., & Stackebrandt, E. (1992). Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. Journal of Bacteriology, **174**:5072-5078.

Lobe, I., Du Preez, C.C., & Amelung, W. (2002). Influence of prolonged arable cropping on lignin compounds in sandy soils of the South African Highveld. European Journal of Soil Science, **53**:553-562.

Lorite, M.J., Sanjuan, J., Velasco, L., Olivares, J., & Bedmar, E.J. (1998). Characterization of *Bradyrhizobium japonicum pcaBDC* genes involved in 4-hydroxybenzoate degradation. Biochimica et Biophysica Acta, **1397**:257-261.

MacDonald, D.L., Stanier, R.Y., & Ingraham, J.L. (1954). The enzymatic formation of beta-carboxymuconic acid. Journal of Biological Chemistry, **210**:809-820.

MacLean, A.M., Anstey, M.I., & Finan, T.M. (2008). Binding site determinants for the LysR-type transcriptional regulator PcaQ in the legume endosymbiont *Sinorhizobium meliloti.* Journal of Bacteriology, **190**:1237-1246.

MacLean, A.M., MacPherson, G., Aneja, P., & Finan, T.M. (2006). Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti.* Applied and Environmental Microbiology, **72**:5403-5413.

Manoharan, H.T, & Jayaraman, K. (1979). Mapping of the Loci Involved in the Catabolic Oxidation of L-Hydroxyproline in *Pseudomonas aeruginosa* PAO. Molecular and General Genetics, **172**:99-105.

Manoharan, H.T. (1980). On the regulation of L-hydroxyproline dissimilatory pathway in *Pseudomonas aeruginosa* PAO . Journal of Biosciences, **2**:107-120.

Masai, E., Yamada, A., Healy, J.M., Hatta, T., Kimbara, K., Fukuda, M., Yano, K. (1995). Characterization of biphenyl catabolic genes of gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. Applied and Environmental Microbiology, **61**:2079-2085.

Matilla, M.A., Espinosa-Urgel, M., Rodriguez-Herva, J.J., Ramos, J.L., & Ramos-Gonzalez, M.I. (2007). Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. Genome Biology, **8**:R179.

Miethling, R., Wieland, G., Backhaus, H., & Tebbe, C.C. (2000). Variation of microbial rhizosphere communities in response to crop species, soil origin, and inoculation with *Sinorhizobium meliloti* L33. Microbial Ecology, **40**:43-56.

Moreno, R., & Rojo, F. (2008). The target for the *Pseudomonas putida* Crc global regulator in the benzoate degradation pathway is the BenR transcriptional regulator. Journal of Bacteriology, **190**:1539-1545.

Morita, H., & Sowden, F. (1981). Effect of decomposition on the distribution of amino compounds in the acid hydrolysates of organic soils. Soil Biology and Biochemistry, **13**:327-329.

Murata, T., Tanaka, H., Yasue, S., Hamada, R., Sakagami, K., & Kurokawa, Y. (1999). Seasonal variations in soil microbial biomass content and soil neutral sugar composition in grassland in the Japanese Temperate Zone. Applied Soil Ecology, **11**:253-259.

Ngai, K.L., & Ornston, L.N. (1988). Abundant expression of *Pseudomonas* genes for chlorocatechol metabolism. Journal of Bacteriology, **170**:2412-2413.

Nguyen, C. (2003). Rhizodeposition of organic C by plants: mechanisms and controls. Agronomie, **23**:375-396.

Nichols, N.N., & Harwood, C.S. (1997). PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*. Journal of Bacteriology, **179**:5056-5061.

Olsson, P.A., Kjellbom, P., & Rosendahl, L. (2002). *Rhizobium* colonization induced changes in membrane-bound and soluble hydroxyproline-rich glycoprotein composition in pea. Physiologia Plantarum, **114**:652-660.

Ondrako, J.M., & Ornston, L.N. (1980). Biological distribution and physiological role of the beta-ketoadipate transport system. Journal of General Microbiology, **120**:199-209.

Ornston, L. N. (1966b). The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida* II Enzymes of the protocatechuate pathway. The Journal of Biological Chemistry, **241**:3787-3794.

Ornston, L.N. (1966c). The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. III. Enzymes of the catechol. The Journal of Biological Chemistry, **241**:3795-3799.

Ornston, L.N. (1966d). The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. IV. Regulation. The Journal of Biological Chemistry, **241**:3800-3810.

Ornston, L., & Parke, D. (1976). Properties of an inducible uptake system for beta-ketoadipate in *Pseudomonas putida*. Journal of Bacteriology, **125**:475-488.

Ornston, L.N., & Stanier, R.Y. (1966). The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. The Journal of Biological Chemistry, **241**:3776-3786.

Pao, S.S., Paulsen, I.T., & Saier, M.H. (1998). Major facilitator superfamily. Microbiology and Molecular Biology Reviews, **62**:1-34.

Parales, R.E., & Harwood, C.S. (1992). Characterization of the genes encoding beta-ketoadipate: succinyl-coenzyme A transferase in *Pseudomonas putida*. Journal of Bacteriology, **176**:4657-4666.

Park, S.H., Kim, J.W., Yun, S.H., Leem, S.H., Kahng, H.Y., & Kim, S.I. (2006). Characterization of beta-ketoadipate pathway from multi-drug resistance bacterium, *Acinetobacter baumannii* DU202 by proteomic approach. The Journal of Microbiology, **44**:632-640.

Parke, D. (2000). Positive selection for mutations affecting bioconversion of aromatic compounds in *Agrobacterium tumefaciens*: analysis of spontaneous mutations in the protocatechuate 3,4-dioxygenase gene. Journal of Bacteriology, **182**:6145-6153.

Parke, D. (1996a). Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds, from *Agrobacterium tumefaciens*. Journal of Bacteriology, **178**:266-272.

Parke, D. (1996b). Conservation of PcaQ, a transcriptional activator of pca genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. Journal of Bacteriology, **178**:3671-3675.

Parke, D. (1995). Supraoperonic clustering of *pca* genes for catabolism of the phenolic compound protocatechuate in *Agrobacterium tumefaciens*. Journal of Bacteriology, **177**:3808-3817.

Parke, D. (1993). Positive regulation of phenolic catabolism in *Agrobacterium tumefaciens* by the pcaQ gene in response to beta-carboxy-cis,cis-muconate. Journal of Bacteriology, **175**:3529-3535.

Parke, D., & Ornston, L.N. (1986). Enzymes of the beta-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium spp.* and constitutive in *Bradyrhizobium spp.* Journal of Bacteriology, **165**:288-292.

Parke, D., & Ornston, L.N. (1984). Nutritional diversity of *Rhizobiaceae* revealed by auxanography. Journal of General Microbiology, **130**:1743-1750.

Parke, D., D'Argenio, D.A., & Ornston, L.N. (2000). Bacteria are not what they eat: that is why they are so diverse. Journal of Bacteriology, **182**:257-263.

Parke, D., Ornston, L.N., & Nester, E.W. (1987). Chemotaxis to plant phenolic inducers of virulence genes is constitutively expressed in the absence of the Ti plasmid in *Agrobacterium tumefaciens*. Journal of Bacteriology, **170**:5336-5338.

Parke, D., Rivelli, M., & Ornston, L.N. (1985). Chemotaxis to aromatic and hydroaromatic acids: comparison of *Bradyrhizobium japonicum* and *Rhizobium trifolii*. Journal of Bacteriology, **163**:417-422.

Parke, D., Rynne, F., & Glenn, A. (1991). Regulation of phenolic catabolism in *Rhizobium leguminosarum* biovar *trifolii*. Journal of Bacteriology, **173**:5546-5550.

Parsek, M.R., Shinabarger, D.L., Rothmel, R.K., & Chakrabarty, A.L. (1992). Roles of CatR and *cis,cis*-muconate in activation of the *catBC* operon, which is involved in benzoate degradation in *Pseudomonas putida*. Journal of Bacteriology, **174**:7798-7806.

Patrauchan, M.A., Florizone, C., Dosanjh, M., Mohn, W.W., Davies, J., & Eltis, L.D. (2005). Catabolism of benzoate and phthalate in *Rhodococcus sp.* strain RHA1: redundancies and convergence. Journal of Bacteriology, **187**:4050-4063.

Paul, D., & Nair, S. (2008). Stress adaptations in a Plant Growth Promoting *Rhizobacterium* (PGPR) with increasing salinity in the coastal agricultural soils. Journal of Basic Microbiology, **48**:1-7.

Paul, E., & Clark, F. (1989). *Soil Biology and Biochemistry.* San Diego, California: Academic Press.

Phillips, D.A., Joseph, C.M., & Hirsch, P.R. (1997). Occurrence of flavonoids and nucleosides in agricultural soils. Applied and Environmental Microbiology, **63**:4573-4577.

Poirier, N., Derenne, S., Rouzaud, J., Largeau, C., Mariotti, A., Balesdent, J., & Maquet, J. (2000). Chemical structure and sources of the macromolecular, resistant, organic fraction isolated from a forest soil (Lacadee, south-west France). Organic Geochemistry, **31**:813-827.

Popp, R., Kohl, T., Patz, P., Trautwein, G., & Gerischer, U. (2002). Differential DNA binding of transcriptional regulator PcaU from *Acinetobacter sp.* strain ADP1. Journal of Bacteriology, **184**:1988-1997.

Rae, A., Bonfante-Fasolo, P., & Brewin, N. (1992). Structure and growth of infection threads in the legume symbiosis with *Rhizobium leguminosarum*. The Plant Journal, **2**:385-395.

Rae, A., Perotto, S., Knox, J., Kannenberg, E., & Brewin, N. (1991). Expression of extracellular glycoproteins in the uninfected cells of developing pea nodule tissue. Molecular plant-microbe interactions, **4**:563-570.

Rainey, P.B. (1999). Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. Environmental Microbiology, **1**:243-257.

Rajapaksha, R.M, Tobor-Kaplon, M.A., & Baath, E. (2004). Metal toxicity affects fungal and bacterial activities in soil differently. Applied and Environmental Microbiology, **70**:2966-2973.

Ramaswamy, S.G. (1984). Hydroxyproline 2-epimerase of *Pseudomonas*. Subunit structure and active site studies. The Journal of Biological Chemistry, **259**:249-254.

Ramos-Gonzalez, M.I., Campos, M.J., & Ramos, J.L. (2005). Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: in vivo expression technology capture and identification of root-activated promoters. Journal of Bacteriology, **187**:4033-4041.

Rasse, D.P., Dignac, M-F, Bahri, H., Rumpel, C., Mariotti, A., & Chenu, C. (2006). Lignin turnover in an agricultural field: from plant residues to soil-protected fractions. European Journal of Soil Science, **57**:530-538.

Rathbun, E.A., Naldrett, M.J., & Brewin, N.J. (2002). Identification of a family of extensin-like glycoproteins in the lumen of rhizobium-induced infection threads in pea root nodules. Molecular Plant-Microbe Interactions, **15**:350-359.

Rediers, H., Bonnecarrere, V., Rainey, P., Hamonts, K., Vanderleyden, J., & De Mot, R. (2003). Development and application of a *dapB*-based in vivo expression technology system to study colonization of rice by the endophytic nitrogen-fixing bacterium *Pseudomonas stutzeri* A15. Applied and Environmental Microbiology, **69**:6864-6874.

Roberts, P., Bol, R., & Jones, D.L. (2007). Free amino sugar reactions in soil in relation to soil carbon and nitrogen cycling. Soil Biology and Biochemistry, **39**:3081-3092.

Romero-Arroyo, C.E., Schell, M.A, Gaines, G.L, & Neidle, E.L. (1995). *catM* encodes a LysR-type transcriptional activator regulating catechol degradation in *Acinetobacter calcoaceticus*. Journal of Bacteriology, **177**:5891-5898.

Romero-Steiner, S., Parales, R.E., Harwood, C.S., & Houghton, J.E. (1994). Characterization of the *pcaR* regulatory gene from *Pseudomonas putida*, which is required for the complete degradation of *p*-hydroxybenzoate. Journal of Bacteriology, **176**:5771-5779.

Rothmel, R.K., Aldrich, T.L., Houghton, J.E., Coco, W.M., Ornston, L.N., & Chakrabarty, A.M. (1990). Nucleotide sequencing and characterization of *Pseudomonas putida catR*: a positive regulator of the *catBC* operon is a member of the LysR family. Journal of Bacteriology, **172**:922-931.

Rothmel, R.K., Shinabarger, D.L., Parsek, M.R., Aldrich, T.L., & Chakrabarty, A.M. (1991). Functional analysis of the *Pseudomonas putida* regulatory protein CatR: transcriptional studies and determination of the CatR DNA-binding site by hydroxyl-radical footprinting. Journal of Bacteriology, **173**:4717-4724.

Saier, M.H., Beatty, J.T., Goffeau, A., Harley, K.T., Heijne, W.H., Huang, S.C., Jack, D.L., Jahn, P.S., Lew, K., Liu, J., Pao, S.S., Paulsen, I.T., Tseng, T.T., & Virk, P.S. (1999). The major facilitator superfamily. Journal of Molecular Microbiology and Biotechnology, **1**:257-279.

Seiler, W., & Crutzen, P.J. (1980). Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. Climatic Change, **2**:207-247.

Siehler, S.Y., Dal, S., Fischer, R., Patz, P., & Gerischer, U. (2007). Multiple-level regulation of genes for protocatechuate degradation in *Acinetobacter baylyi* includes cross-regulation. Applied and Environmental Microbiology, **73**:232-242.

Silby, M.W., & Levy, S.B. (2004). Use of in vivo expression technology to identify genes important in growth and survival of *Pseudomonas fluorescens* Pf0-1 in soil: discovery of expressed sequences with novel genetic organization. Journal of Bacteriology, **186**:7411-7419.

Singaravelan, N., Grishkan, I., Beharav, A., Wakamatsu, K., Ito, S., & Nevo, E. (2008). Adaptive melanin response of the soil fungus *Aspergillus niger* to UV radiation stress at "Evolution Canyon", Mount Carmel, Israel. PLoS ONE, **3**:e2993.

Singh, R.M, & Adams, E. (1964). Alpha-ketoglutaric semialdehyde: a metabolic intermediate. Science, **144**:67-68.

Singh, R.M., & Adams, E. (1965a). Enzymatic deamination of delta-1-pyrroline-4-hydroxy-2-carboxylate to 2,5-dioxovalerate (alpha-ketoglutaric semialdehyde). The Journal of Biological Chemistry, **240**:4344-4351.

Singh, R.M., & Adams, E. (1965b). Isolation and identification of 2,5-dioxovalerate, an intermediate in the bacterial oxidation of hydroxyproline. The Journal of Biological Chemistry, **240**:4352-4356.

Sleeper, B.P., & Stanier, R.Y. (1950). The bacterial oxidation of aromatic compounds I Adaptive patterns with respect to polyphenolic compounds. **59**:117-127.

Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S., Roskot, N., Heuer, H., & Berg, G. (2001). Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. Applied and Environmental Microbiology, **67**:4742-4751.

Smejkalova, D., Piccolo, A., & Spiteller, M. (2006). Oligomerization of humic phenolic monomers by oxidative coupling under biomimetic catalysis. Environmental Science and Technology, **40**:6955-6962.

Sowden, F.J. (1977). Distribution of nitrogen in representative soils. Canadian Journal of Soil Science, **57**:445-456.

Sowden, F.J., Griffith, S.M, & Schnitzer, M. (1976). The distribution of nitrogen in some highly organic tropical volcanic soils. Soil Biology and Biochemistry, **8**:55-60.

Stanier, R.Y. (1948). The Oxidation of Aromatic Compounds by Fluorescent Pseudomonads. Journal of Bacteriology, **55**:477-494.

Stanier, R.Y., & Ingraham, J.L. (1954). Protocatechuic acid oxidase. Journal of Biological Chemistry, **210**:799-808.

Stanier, R.Y., & Ornston, L.N. (1973). The beta-ketoadipate pathway. **9**:89-149.

Stanier, R.Y., Sleeper, B.P., Tsuchida, M., & MacDonald, D.L. (1950). The bacterial oxidation of aromatic compounds; III. The enzymatic oxidation of catechol and protocatechuic acid to beta-ketoadipic acid. Journal of Bacteriology, **59**:137-151.

Thacker, R.P. (1969). Conversion of L-hydroxyproline to glutamate by extracts of strains of *Pseudomonas convexa* and *Pseudomonas fluorescens*. Archiv fur Mikrobiology, **64**:235-238.

Torsvik, V., Goksoyr, J., & Daae, F.L. (1990). High diversity in DNA of soil bacteria. Applied and Environmental Microbiology, **53**:782-787.

Trautwein, G., & Gerischer, U. (2001). Effects exerted by transcriptional regulator PcaU from *Acinetobacter sp.* strain ADP1. Journal of Bacteriology, **183**:873-881.

Treves, D.S., Xia, B., Zhou, J., & Tiedje, J.M. (2003). A two-species test of the hypothesis that spatial isolation influences microbial diversity in soil. Microbial Ecology, **45**:20-28.

Vandenkoornhuyse, P., Mahe, S., Ineson, P., Staddon, P., Ostle, N., Cliquet, J.B., Francez, A.J., Fitter, A.H., & Young, J.P. (2007). Active root-inhabiting microbes identified by rapid incorporation of plant-derived carbon into RNA. Proceedings of the National Academy of Science USA, **104**:16970-16975.

Vesely, M., Knoppova, M., Nesvera, J., & Patek, M. (2007). Analysis of *catRABC* operon for catechol degradation from phenol-degrading *Rhodococcus erythropolis*. Applied Microbiology and Biotechnology, **76**:159-168.

Vetting, M.W., D'Argenio, D.A., Ornston, L.N, & Ohlendorf, D.H. (2000). Structure of *Acinetobacter* strain ADP1 protocatechuate 3, 4-dioxygenase at 2.2 A resolution: implications for the mechanism of an intradiol dioxygenase. Biochemistry, **39**:7943-7955.

Vilchez, S., Molina, L., Ramos, C., & Ramos, J.L. (2000). Proline catabolism by *Pseudomonas putida*: cloning, characterization, and expression of the *put* genes in the presence of root exudates. Journal of Bacteriology, **182**:91-99.

Vinken, R., Schaffer, A., & Ji, R. (2005). Abiotic association of soil-borne monomeric phenols with humic acids. Organic Geochemistry, **36**:583-593.

Wang, M.C, & Huang, P.M. (2003). Cleavage and polycondensation of pyrogallol and glycine catalyzed by natural soil clays. Geoderma, **112**:31-50.

Warren, C.R. (2008). Rapid and sensitive quantification of amino acids in soil extracts by capillary electrophoresis with laser-induced fluorescence. Soil Biology and Biochemistry, **40**:916-923.

Watanabe, S., Yamada, M., Ohtsu, I., & Makino, K. (2007). alpha-ketoglutaric semialdehyde dehydrogenase isozymes involved in metabolic pathways of D-glucarate, D-galactarate, and hydroxy-L-proline. Molecular and metabolic convergent evolution. The Journal of Biological Chemistry, **282**:6685-6695.

Weisskopf, L., Le Bayon, R.C., Kohler, F., Page, V., Jossi, M., Gobat, J.M., Matinola, E., & Aragno, M. (2008). Spatio-temporal dynamics of bacterial communities associated with two plant species differing in organic acid secretion: A one-year microcosm study on lupin and wheat. Soil Biology and Biochemistry, **40**:1772-1780.

Williams, P.A., & Shaw, L.E. (1997). *mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow on exogenous *cis,cis*-muconate as the sole carbon source. Journal of Bacteriology, **179**:5935-5942.

Wisniewski, J.P., Rathbun, E.A., Knox, J.P., & Brewin, N.J. (2000). Involvement of Diamine Oxidase and Peroxidase in Insolubilization of the Extracellular Matrix: Implications for Pea Nodule Initiation by *Rhizobium leguminosarum*. Molecular Plant-Microbe Interactions, **13**:413-420.

Wu, H., de Graaf, B., Mariani, C., & Cheung, A.Y. (2001). Hydroxyproline-rich glycoproteins in plant reproductive tissues: structure, functions and regulation. Cellular and Molecular Life Sciences, **58**:1418-1429.

Yeh, W.K., & Ornston, L.N. (1981). Evolutionarily homologous $\alpha_2\beta_2$ oligomeric structures in beta-ketoadipate succinyl-CoA transferases from *Acinetobacter calcoaceticus* and *Pseudomonas putida*. The Journal of Biological Chemistry, **256**:1565-1569.

Yelle, D.J., Ralph, J., Lu, F., & Hammel, K.E. (2008). Evidence for cleavage of lignin by a brown rot basidiomycete. Environmental Microbiology, **10**:1844-1849.

Yoneya, T., & Adams, E. (1961). Hydroxyproline metabolism. V. Inducible allohydroxy-D-proline oxidase of *Pseudomonas*. The Journal of Biological Chemistry, **236**:3272-3279.

Zervos, C., & Adams, E. (1975). Hydroxyproline-2-epimerase of *Pseudomonas*: active-site peptides. Molecular and Cellular Biochemistry, **8**:113-121.

Zhang, W., He, H., & Zhang, X. (2007). Determination of neutral sugars in soil by capillary gas chromatography after derivatization to aldononitrile acetates. Soil Biology and Biochemistry, **39**:2665-2669.

Zhang, X., & Amelung, W. (1996). Gas chromatographic determination of muramic acid, glucosamine, mannosamine, and galactosamine in soils. Soil Biology and Biochemistry, **28**:1201-1206.

Zhong, R., & Ye, Z.H. (2007). Regulation of cell wall biosynthesis. Current Opinion in Plant Biology, **10**:564-572.

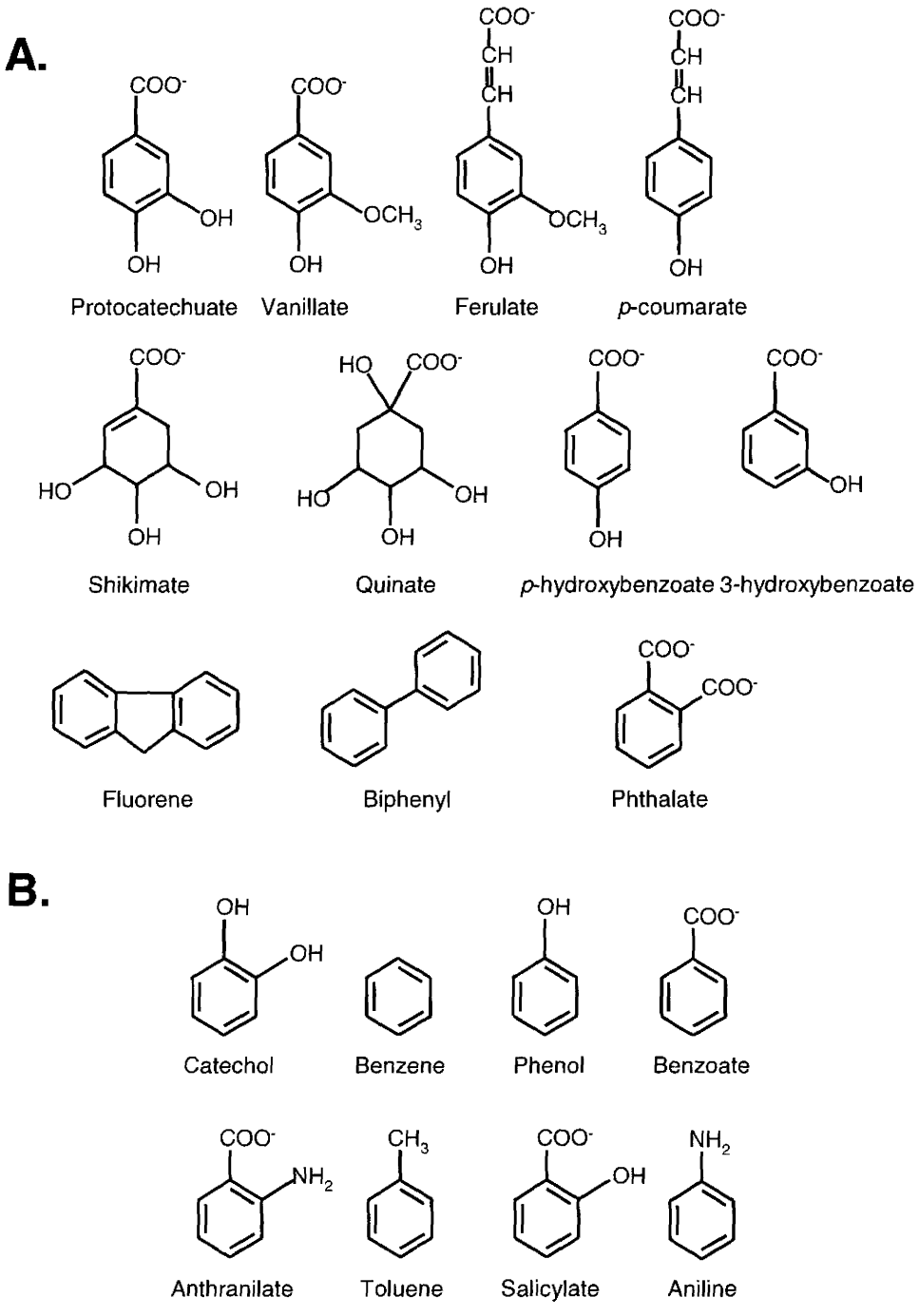


Figure 1.1. Compounds which may be metabolized through the β-ketoadipate pathway via conversion to (A) protocatechuate or (B) catechol.

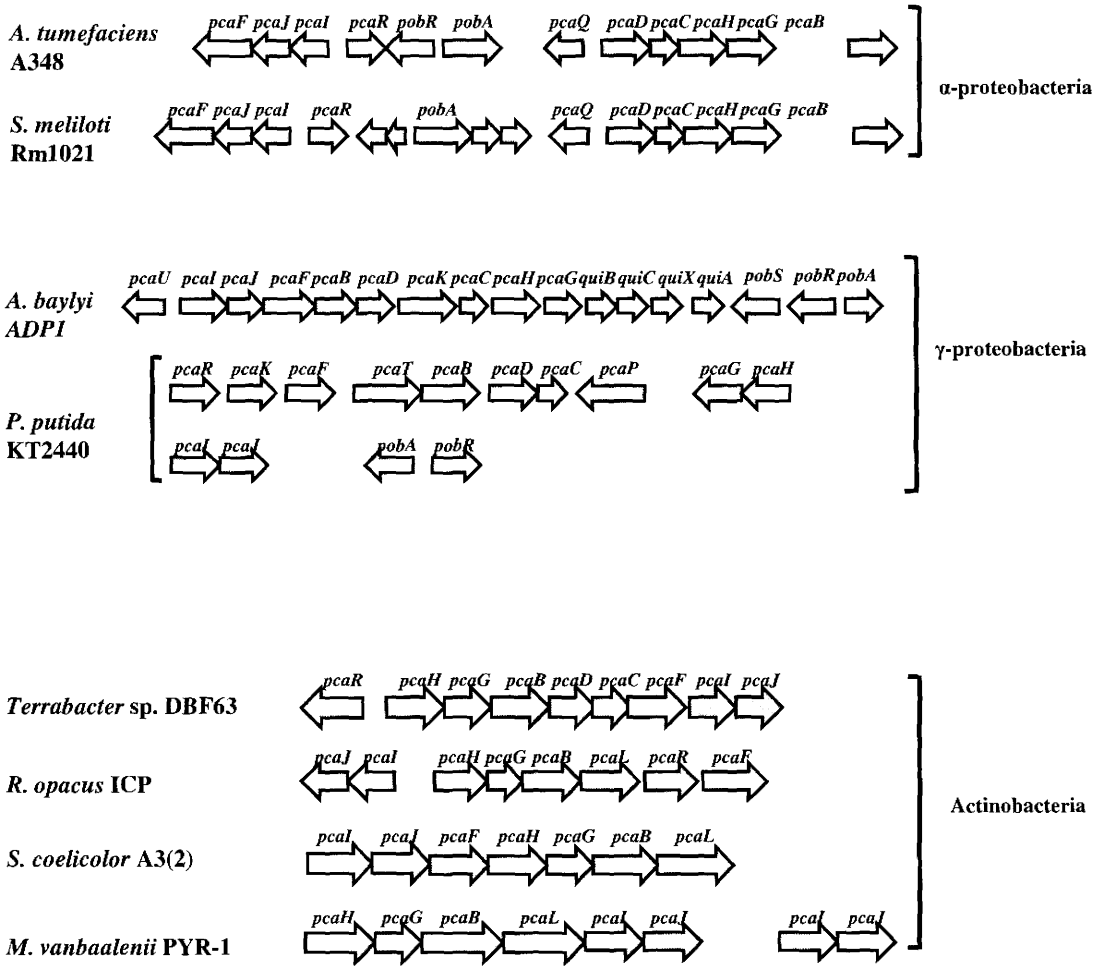


Figure 1.2. Schematic depiction of genes involved in the catabolism of the aromatic acid protocatechuate via the β–ketoadipate pathway. Protocatechuate metabolism to acetyl-CoA and succinate requires five structural enzymes encoded by *pcaHG* (protocatechuate-3,4-dioxygenase), *pcaB* (β-carboxy-*cis*, *cis*-muconate lactonizing enzyme), *pcaC* (γ-carboxymuconolactone decarboxylase), *pcaD* (β–ketoadipate enol-lactone hydrolase), *pcaIJ* (β–ketoadipate succinyl-CoA transferase) and *pcaF* (β–ketoadipyl-CoA thiolase). *pcaL* encodes a fusion protein that exhibits decarboxylase and hydrolase activities. *pcaU*, *pcaR*, *pobR*, and *pcaQ* specify transcriptional regulators associated with the pathway. *pcaK*, *pcaT*, and *pcaP* encode aromatic acid transporter permeases. *qui* genes are associated with the conversion of quinate and shikimate to protocatechuate; *pobA* encodes a hydroxylase involved in the conversion of *p*-hydroxybenzoate to protocatechuate.

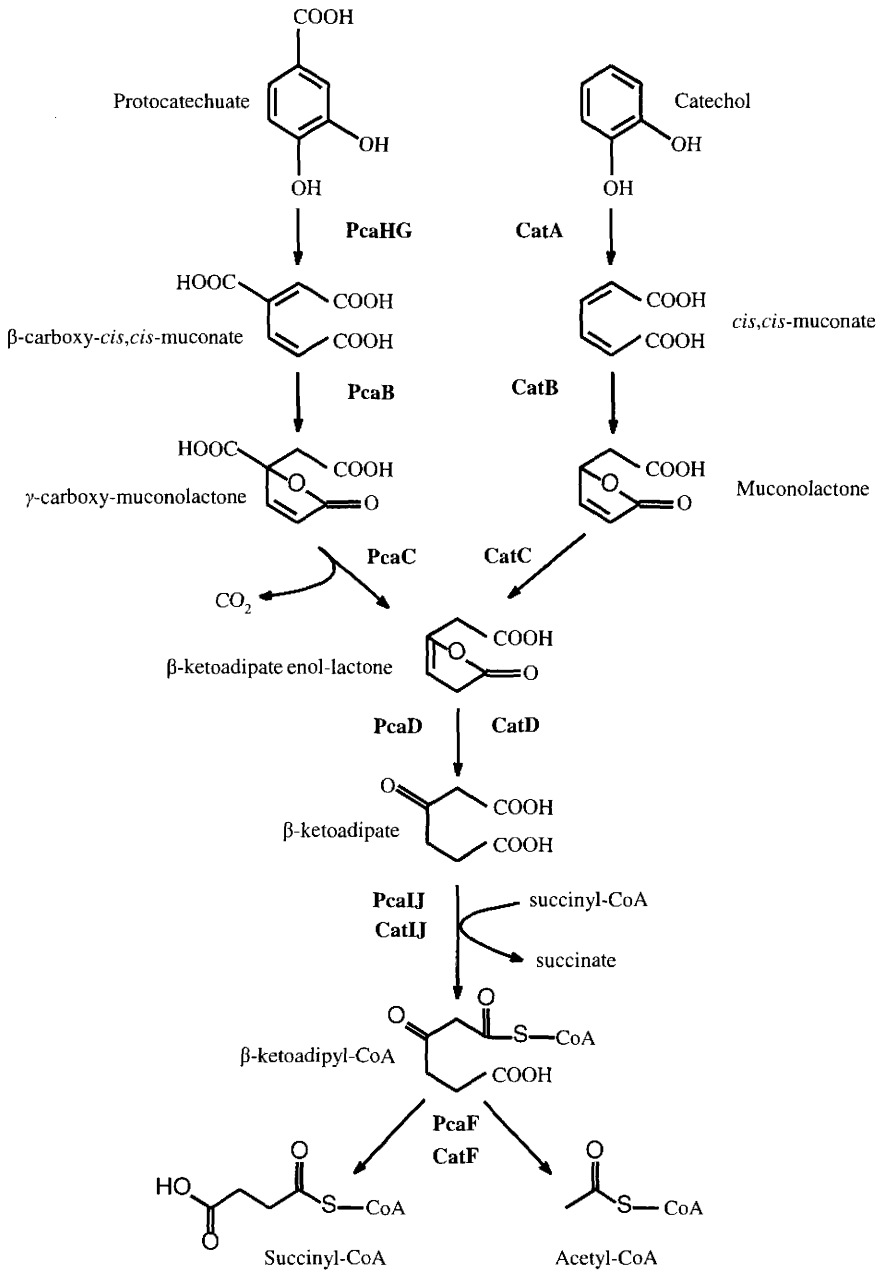


Figure 1.3. The β–ketoadipate pathway is involved in the conversion of protocatechuate and catechol to tricarboxylic acid cycle intermediates.

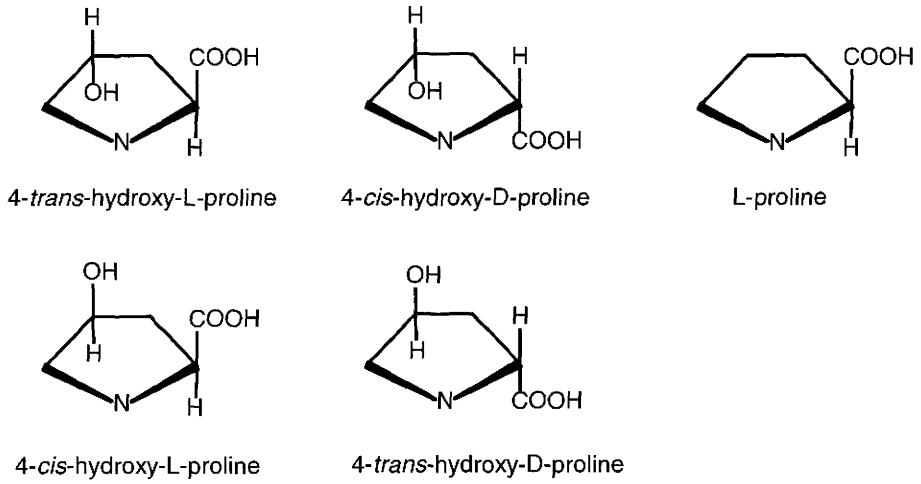


Figure 1.4. Stereoisomers of hydroxyproline and L-proline.

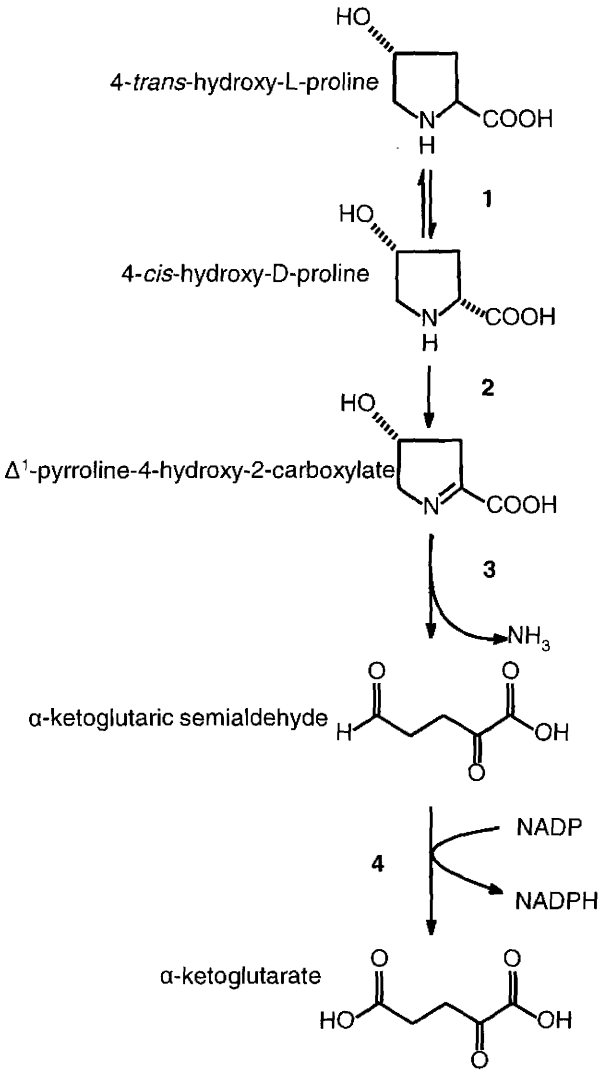


Figure 1.5. Metabolism of *trans*-4-hydroxy-L-proline, as described in *Pseudomonas* (Adams, 1959). The catabolism of *trans*-4-hydroxy-L-proline yields ammonia and α–ketoglutarate, and thus this amino acid may serve as a sole source of carbon and nitrogen. Enzyme 1: hydroxyproline-2-epimerase; enzyme 2: *cis*-4-hydroxy-D-proline oxidase; enzyme 3: Δ1-pyrroline-4-hydroxy-2-carboxylic acid deaminase; enzyme 4: α-ketoglutaric semialdehyde dehydrogenase.

CHAPTER TWO

Characterization of the β–ketoadipate pathway in *Sinorhizobium meliloti*

Preface

This chapter describes a genetic and enzymatic characterization of the protocatechuate branch of the β–ketoadipate pathway in *S. meliloti*. Gordon MacPherson performed the Tn5 mutagenesis experiments, and a characterization of the Pca-negative mutant strains RmG867 and RmG879, including the Rothera test and protocatechuate 3,4-dioxygenase enzyme assays. As well, Gordon constructed a plasmid (pTH468) that was used to monitor the expression of *pcaD*::*lacZ*. Punita Aneja performed the primer extension experiment that determined the transcriptional start site of *pcaD*. I performed all experiments relating to the regulation of *pca* gene expression by PcaQ and PcaR. I identified (via primer extension) the transcriptional start site associated with *pcaI*. As well, I purified β–ketoadipate succinyl-CoA transferase activity from *S. meliloti*, obtaining sufficiently purified protein as to permit identification by mass spectrometry (performed by McMaster Regional Centre for Mass Spectrometry). As primary author, I wrote the manuscript in its entirety, with editing by Shawn MacLellan and Turlough Finan. This chapter has been published in the journal Applied and Environmental Microbiology, and is reprinted with permission from the American Society for Microbiology (license number 2082190868906).

2.1 Abstract

Aromatic compounds represent an important source of energy for soil-dwelling organisms. The β-ketoadipate pathway is a key metabolic pathway involved in the catabolism of the aromatic compounds protocatechuate and catechol and here we show through enzymatic analysis and mutant analysis that genes required for growth and catabolism of protocatechuate in the soil-dwelling bacterium *Sinorhizobium meliloti* are organized on the pSymB megaplasmid in two transcriptional units designated *pcaDCHGB*, and *pcaIJF*. The *pcaD* promoter was mapped by primer extension and expression from this promoter is demonstrated to be regulated by the LysR-type protein PcaQ. β-ketoadipate succinyl-CoA transferase activity in *S. meliloti* was shown to be encoded by *smb20587* and *smb20588* and these genes have been renamed *pcaI* and *pcaJ*. These genes are organized in an operon with a putative β-ketoadipyl-CoA thiolase gene (*pcaF*) and expression of the *pcaIJF* operon is shown to be regulated by an IclR-type transcriptional regulator, *smb20586*, which we have named PcaR. We show that *pcaR* transcription is negatively autoregulated and that PcaR is a positive regulator of *pcaIJF* expression and is required for growth of *S. meliloti* on protocatechuate as carbon source. The characterization of the protocatechuate catabolic pathway in *S. meliloti* offers an opportunity for comparison with related species, including *Agrobacterium tumefaciens*. Differences observed between *S. meliloti* and *A. tumefaciens pcaIJ* offer the first evidence of *pca* genes that may have been acquired after speciation in these closely related species.

2.2 Introduction

Aromatic acids constitute an important source of carbon and energy for soil-dwelling microorganisms and accumulate primarily as the result of the degradation of plant-derived molecules, including lignin. Many aromatic compounds may be converted to one of two common intermediates, protocatechuate or catechol, which are metabolized to tricarboxylic acid intermediates via the β-ketoadipate pathway (Figure 2.1A) (18). In *Agrobacterium tumefaciens*, genes encoding enzymes involved in protocatechuate catabolism are organized into two distinct operons (36). Expression of the *pcaDCHGB* operon is induced by pathway metabolites β-carboxy-*cis*,*cis*-muconate and γ-carboxymuconolactone, via the LysR-type transcriptional regulator protein PcaQ (35, 37). Genes encoded in this operon are involved in the conversion of protocatechuate to the pathway intermediate β-ketoadipate. Expression of the *pcaIJF* operon is induced in the presence of β-ketoadipate (36), and these genes mediate the conversion of β-ketoadipate to the end-products succinate and acetyl-CoA. The transcriptional regulator involved in modulating expression of the *pcaIJF* operon in *A. tumefaciens* is an adjacent IclR-type regulator encoded by *pcaR* (39).

The gram-negative bacterium *Sinorhizobium meliloti* forms a symbiotic relationship with alfalfa through the establishment of root nodules. The β-ketoadipate pathway is present in many members of *Rhizobiaceae* examined to date, emphasizing the importance of aromatic acid catabolism in this family (41, 42). The publication of the *S. meliloti* genome has facilitated the identification and characterization of many metabolic pathways (10) and here we report the characterization of the protocatechuate branch of the β-ketoadipate pathway in *S. meliloti*. Except for *pcaIJ* orthologues, we demonstrate that the *pca* genes are organized, function and are regulated in a similar manner in *S. meliloti* to that previously established for *A. tumefaciens*. Unexpectedly the *S. meliloti* genes *smb20587* and *smb20588* were found to encode proteins with low sequence similarity to the two protein subunits of β-ketoadipate succinyl-CoA transferase (PcaI and PcaJ) in *A. tumefaciens, Acinetobacter baylyi* strain ADP1, and *Pseudomonas putida*. Through overexpression of *smb20587* and *smb20588* in *S. meliloti* followed by

purification of β-ketoadipate succinyl-CoA transferase activity, these two genes are demonstrated to encode β-ketoadipate succinyl-CoA transferase activity.

2.3 Materials and Methods

Bacterial strains and culture conditions

The bacterial strains and plasmids used throughout this study are described in Table 2.1. *Escherichia coli* was grown at 37°C in LB broth. *Sinorhizobium meliloti* was grown at 30°C in M9-minimal medium (Difco) or LB broth supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ (LBmc). M9-minimal medium was supplemented with 1.0 mM $MgSO_4$, 0.25 mM $CaCl_2$, 1 μg/mL D-biotin, and 10 ng/mL $CoCl_2$. Unless otherwise specified, carbon sources were added to M9-minimal medium as follows: 0.5% (v/v) glycerol, 15 mM arabinose, 30 mM adipate (Sigma-Aldrich), or 5 mM protocatechuate (Sigma-Aldrich). For *S. meliloti*, antibiotics were used at the following concentrations (μg/mL): streptomycin: 200; neomycin: 200; gentamicin: 60; spectinomycin: 200; tetracycline: 5.

Transposon mutagenesis

Tn5 mutagenesis of *S. meliloti* was performed by mating the suicide plasmid pRK602 into wild-type derivative strain Rm1021. Neomycin resistant colonies were patched onto M9-minimal media with protocatechuate or glucose as sole carbon sources. Mutants unable to grow on protocatechuate (Pca^-) were examined for ability to grow with succinate as a sole carbon and energy source to eliminate mutants deficient in succinate metabolism.

The pLAFR1 clone bank of *S. meliloti* Rm1021 DNA (12) was screened to isolate clones capable of complementing Pca^- strains. Spot matings were performed with Pca^- mutants, using the clone bank, and strains carrying the complementing cosmids were selected on M9-minimal media with protocatechuate as a sole carbon source. DNA sequencing was provided by Mobix (McMaster University, Hamilton, Ontario).

Rothera test

An overnight LBmc culture was centrifuged and washed with M9-minimal medium. Cells were subcultured into M9-minimal medium supplemented with 0.1% arabinose and 5 mM protocatechuate for overnight incubation at 30°C. Cells were centrifuged and resuspended in 0.02M Tris-HCl, pH 8.0 to an O.D. of 1.0. 0.5 mL toluene was added to 2 mL resuspended cells, which was incubated at 30°C with shaking for 1 hour. 1 gram $(NH_4)_2SO_4$ was added, and the mixture was vortexed. 1 drop of a fresh aqueous sodium nitroprusside (1%) solution was added, followed by the addition of 1 drop of concentrated NH_3 (29%), and the mixture was vortexed. Development of a purple colour within 5 minutes following the addition of NH_3 was considered a positive test for the presence of β-ketoadipate (20).

Protocatechuate 3,4-dioxygenase activity assays

Overnight cultures grown in LBmc were washed and subcultured into M9-minimal medium supplemented with 0.1% arabinose and 5 mM protocatechuate. Tetracycline was included in the growth medium for strains carrying pTH178. Upon harvesting, cells were centrifuged, washed, and resuspended into 4 mL buffer (20 mM Tris-HCl, 1 mM $MgCl_2$; pH 7.8) per gram cells. Aliquots of cells were frozen at -80°C until used in the assay.

Prior to use, 10 μL of 0.1M dithiothreitol was added (per mL) as the aliquots thawed on ice. Cells were disrupted via sonication and extracts were centrifuged to remove intact cells and cellular debris. The dioxygenase assay was performed as previously described (8), with the following modifications. The temperature of the assay was maintained at 30°C and the reaction was monitored by following the reduction in absorbance at 293 nm using a Contron Uvikon 930 double-beam spectrophotometer. Protein concentrations were determined using Biorad protein assay reagent, with bovine serum albumin as the standard.

Construction of an S. meliloti pcaQ::Ω strain

An Ω cassette encoding gentamicin resistance from pHP45Ω*aac* (3) was introduced into a *Pst*I site located 35 bp downstream of the predicted *pcaQ* translational start site as follows. A 893 bp fragment centered upon the *Pst*I site was PCR amplified using *S. meliloti* Rm1021 genomic DNA as a template. This fragment was cloned into the suicide vector pJQ200 uc-1 (44) via *Not*I to create plasmid pTH1577. The Ω cassette was PCR amplified and cloned into pTH1577 via *Pst*I, yielding pTH1592. The *Not*I fragment, encompassing *pcaQ*::Ω, was subcloned from pTH1592 into a derivative of pVO155 (30) to create pTH1882. This pVO155 derivative (pTH1883) lacks the *gusA* reporter gene present in the parental vector, and was selected for use because it is unable to replicate in *S. meliloti* and carries a gene specifying neomycin resistance. pTH1882 was mated into *S. meliloti lac* strain RmG212, and recombinants were selected for by plating onto LB agar supplemented with streptomycin + gentamicin. Colonies were patched onto LB agar + neomycin to identify recombinants in which the suicide vector had recombined out of the genome, leaving the Ω cassette behind. Southern hybridization was performed on DNA extracted from Gm^rNm^s recombinants to confirm the location of the antibiotic cassette and absence of the suicide vector. In this case, an *Eco*RV digest of genomic DNA isolated from putative RmG212 *pcaQ*::Ω mutants and RmG212 was hybridized with a labeled (Random primed DNA labeling kit; Roche) probe encompassing the *pcaQ Pst*I site. A shift corresponding to a ~ 2 kb increase (compared to wild-type) was noted in the putative *pcaQ* mutants, consistent with the incorporation of the 1.8 kb Gm^r cassette into the *Pst*I site and subsequent the excision of the integrating vector. As well, hybridization of the *Eco*RV genomic DNA digest with labeled probe corresponding to the (Gm^r) Ω cassette indicated the presence of the antibiotic cassette within the shifted bands observed in the putative mutants. The Ω probe failed to hybridize with the *Eco*RV digested RmG212 genomic DNA.

Construction of an S. meliloti pcaR::Ω strain

An Ω cassette encoding streptomycin/spectinomycin resistance from pHP45Ω (44) was introduced into a *Sac*II site located 76 bp downstream of the predicted *pcaR* translational start site. A 963 bp fragment encompassing the *Sac*II site within *pcaR* was PCR amplified using *S. meliloti* Rm1021 DNA as a template and Vent DNA polymerase (New England Biolabs). This blunt-ended fragment was cloned into pUC119 at a *Sma*I site to create plasmid pTH1338. The Ω cassette was PCR amplified and cloned into the *Sac*II site in pTH1338 to produce pTH1340. A *Not*I fragment, encompassing the *pcaR*::Ω, was subcloned from pTH1340 into the suicide vector pJQ200 to yield pTH1351. This plasmid was mated into *S. meliloti* Rm1021 and recombinants were selected for by plating onto LB agar supplemented with streptomycin + gentamicin. A single Gm^r colony was inoculated into LBmc and grown in the absence of antibiotic selection. The overnight culture was plated onto LB agar supplemented with 5% sucrose and spectinomycin. Sp^r colonies were patched onto LB agar + gentamicin to confirm excision of the suicide plasmid. Southern hybridization was performed on DNA extracted from Sp^rGm^s $sucrose^r$ colonies to confirm the location of the Ω cassette and verify the loss of the integrating plasmid. Briefly, separate *Xho*I and *Sal*I digests of genomic DNA isolated from putative Rm1021 *pcaR*::Ω mutants and Rm1021 were hybridized with labeled (Random primed DNA labeling kit; Roche) probe encompassing the *pcaR Sac*II site. In each case, a shift corresponding to a ~ 2 kb increase (compared to wild-type) was noted in the putative *pcaR* mutants, consistent with the incorporation of the 2.1 kb antibiotic cassette into the *Sac*II site and the excision of the integrating vector. As well, hybridization of separate *Xho*I and *Sal*I genomic DNA digests with labeled probe corresponding to the (Sm^r/Sp^r) Ω cassette indicated the presence of the antibiotic cassette within the shifted bands observed in the putative mutants. The Ω probe failed to hybridize with either *Sal*I or *Xho*I digested Rm1021 genomic DNA.

Construction of an S. meliloti pcaF::gusA strain

A transcriptional fusion between the annotated *pcaF* gene on the pSymB megaplasmid and a promoterless *gusA* was created in *S. meliloti* wild-type derivative Rm1021 and PcaR-minus strain RmK1014. The fusion was designed such that *pcaF* was not disrupted, and this was verified by growth with protocatechuate as a sole carbon source. A 431 bp fragment spanning the 3' end of *pcaF* was PCR amplified and cloned into the pVO155 (30) derivative pTH1360 to create pTH1559. In pTH1360, the original *gusA* reporter gene present in pVO155 has been replaced by the *gusA* gene present in pFus1, which has a superior ribosome binding site and is expressed more efficiently than its pVO155 counterpart (R. Zaheer and T. M. Finan, unpublished data). pTH1559 was mated into Rm1021 and RmK1014 and recombination of the vector into the *S. meliloti* genome was selected by plating cells onto LB agar supplemented with streptomycin + neomycin.

Purification of β-ketoadipate succinyl-CoA transferase activity in S. meliloti

β-ketoadipate succinyl-CoA transferase was purified according to Kaschabek et al (24), with some modifications. An overnight culture of RmK927 was subcultured into 4 L LBmc supplemented with tetracycline and cells were grown with shaking at 30°C. Expression of genes *smb20587* and *smb20588* was induced at O.D. 0.3 to 0.4 with the addition of 1 mM IPTG and 5 mM protocatechuate. After four hours induction, cells were harvested in the late-exponential growth phase (O.D. 1.2 to 1.4). The pellet was resuspended in 30 mL buffer (100 mM Tris-HCl, 0.5 mM dithiothreitol; pH 7.0), then lysed via five passages through a French pressure cell at 110 MPa. The cell extract was cleared by centrifugation at 100,000 X *g* for 60 minutes. Solid $(NH_4)_2SO_4$ was added to give 75% saturation and the precipitate was collected by centrifugation at 8,000 X *g* for 20 min. The pellet was redissolved in a buffer B1 (50 mM Tris-HCl, 1 M $(NH_4)_2SO_4$, 1 mM EDTA; pH 7.0) and the extract was cleared by centrifugation. The supernatant was loaded onto a Phenyl Sepharose CL-4B (Amersham Biosciences) column preequilibrated with buffer B1 and protein was eluted in a linear gradient of $(NH_4)_2SO_4$ from 1 to 0 M at

a flow rate of 0.3 mL/min. Nine 2 mL fractions with β-ketoadipate succinyl-CoA transferase activity were pooled and dialyzed into buffer A1 (50 mM Tris-HCl, 0.5 mM EDTA; pH 7.0).

The dialyzed fractions were loaded onto a Source 30Q (Amersham Biosciences) column preequilibrated with buffer A1 and protein was eluted in a linear gradient of NaCl from 0 to 1 M at a flow rate of 0.3 mL/min. Twelve 0.5 mL fractions with β-ketoadipate succinyl-CoA transferase activity were pooled and dialyzed into 5 mM potassium phosphate; pH 7.0.

The fractions were loaded onto a CHT Ceramic Hydroxyapatite (BioRad) column preequilibrated with 5 mM potassium phosphate; pH 7.0. Elution occurred in a linear gradient of potassium phosphate from 10 to 400 mM at a flow rate of 0.3 mL/min. Twelve 1 mL fractions with β-ketoadipate succinyl-CoA transferase activity were collected and pooled. An attempt was made to concentrate the enzyme within the pooled extract using Nanosep microconcentrators, however the majority of the purified enzyme was lost at this step, as revealed by subsequent β-ketoadipate succinyl-CoA transferase assays and an SDS-PAGE stained with Coomassie Brilliant Blue.

β-ketoadipate succinyl-CoA transferase assays

β-ketoadipate succinyl-CoA transferase assays were performed as previously described with the following modifications (54). Briefly, the enzyme assays were performed in UV-Star (flat bottom) 96 well microtiter plates (greiner bio-one) using a Safire microplate reader (Tecan). To start the reaction, protein samples were added to a buffered reaction mixture (200 mM Tris-HCl, 40 mM $MgCl_2$, 10 mM β-ketoadipate (Sigma-Aldrich), 0.4 mM succinyl-CoA; pH 8.0) to a final volume of 0.2 mL (path length: 0.52 cm). The formation of β-ketoadipyl-CoA:Mg^{2+} was monitored at 305 nm over a temperature range of 22 to 23°C. An extinction coefficient of 16,300 M^{-1} cm^{-1} was used to calculate the formation of the β-ketoadipyl-CoA:Mg^{2+} complex (24). 1 Unit of activity is defined as the amount of enzyme required to convert 1 umol of substrate to product in 1 minute under the conditions of the assay.

Protein identification via mass spectrometry

Mass spectrometry analyses were provided by the McMaster Regional Centre for Mass Spectrometry (McMaster University, Hamilton, Ontario). The following peptides were used in the identification of the proteins. Molecular mass: 37.6 kDa, peptides: R.NGNVLIEGIVGVQK.E; R.MTPDILYDQLIGVGAAR.G; R.IMSLAEAVEENVR.D. Molecular mass: 35.4 kDa, peptides: R.NGNVLIEGIVGVQK.E; R.MTPDILYDQLIGVGAAR.G. Molecular mass: 28.5 kDa, peptides: R.FANLNTTVVGPYDHPK.V; K.FAETVIETPAPTETELVVLR.D ; R.ITTGFLGGAQIDR.F.

Isolation of total RNA from S. meliloti

Cultures were grown with shaking at 30°C in LBmc ± 5 mM protocatechuate to an O.D. of 0.6 to 0.7. Total RNA was isolated from *S. meliloti* Rm1021 using a hot phenol method as previously described (27).

Primer extension

Primer extension reactions were performed using 50 μg of total *S. meliloti* RNA as previously described (27). The following primers were used for extension reactions. For the identification of the *pcaD* start site: (5' GAAATCCGTGCCGAGCGAGTTGATGAAGAC 3'). For the identification of the *pcaI* start site: (5' GAGAGACATTATCCGCGCCATCG 3') and (5' CGTCTCTGACATTCTCCTCTACCG 3'). Sequencing reactions were performed using the Sequenase Version 2.0 DNA Sequencing kit (USB). The same primer was used in sequencing and primer extension reactions.

β-galactosidase enzyme assays

S. meliloti cultures were grown overnight at 30°C in LBmc and washed with 0.85% saline prior to subculture. Cells were subcultured into M9-minimal media supplemented with a carbon source as indicated and grown with shaking at 30°C for four

hours. 50 to 200 μL aliquots of cells were added directly to Z-buffer with 5% chloroform and 0.0025% SDS. The reaction was started with the addition of 200 μL 2-nitrophenyl β-D-galactopyranoside (ONPG) (4 mg/mL) and stopped upon addition of 500 μL 1M Na_2CO_3. β-galactosidase activities were calculated according to Miller (29).

Enzyme assays were performed using derivatives of RmG212 (Rm1021 *lac*) to reduce background LacZ enzyme activity. RmK948 was created through the transduction of *pcaG*::Tn5 into RmG212, using RmG879 as a donor strain.

β-glucuronidase enzyme assays

Overnight LBmc cultures of *S. meliloti* cultures were grown and washed with 0.85% saline. Cells were subcultured into M9-minimal media supplemented with a carbon source as indicated and grown with shaking at 30°C for four hours. Cultures were centrifuged and resuspended into a buffer consisting of: 50 mM sodium phosphate, 50 mM DTT, and 1 mM EDTA; pH 7.0. Enzyme assays were performed according to Reeve et al. (45), with the exception that assays were performed at room temperature (20 to 22°C).

2.4 Results and Discussion

Isolation and characterization of pca mutants in Rm1021

To directly identify genes involved in protocatechuate catabolism in *S. meliloti*, we screened a transposon Tn5 insertion library for mutants able to grow with succinate but unable to grow with protocatechuate (Pca^-) as a sole carbon source. The precise insertion sites of the transposon in the *S. meliloti* genome were determined by DNA sequencing (see Materials and Methods). Strains RmG867 and RmG879 were found to carry Tn5 within genes annotated to encode β-ketoadipate enol-lactone hydrolase (PcaD) and protocatechuate 3,4-dioxygenase α subunit (PcaG), respectively. Analysis of the *S. meliloti* genome sequence suggests that *pcaD* and *pcaG* are organized in an operon with a γ-carboxymuconolactone decarboxylase (*pcaC*), protocatechuate 3,4-dioxygenase β

subunit (*pcaH*), and β-carboxy-*cis,cis*-muconate cycloisomerase (*pcaB*) (Figure 2.1B) (10).

To determine whether β-ketoadipate accumulated from protocatechuate metabolism in either the *pcaD* or *pcaG* mutant, we performed the Rothera test which detects the presence of β-ketoadipate and thus indicates whether protocatechuate has been metabolized to this pathway intermediate. The wild-type strain Rm1021 exhibited a Rothera-positive phenotype, whereas the mutants RmG867(*pcaD*) and RmG879(*pcaG*) were Rothera-negative indicating that these strains failed to metabolize protocatechuate to β-ketoadipate (Figure 2.1A). Protocatechuate 3,4-dioxygenase catalyzes the first step in the degradation of protocatechuate, the conversion of protocatechuate to β-carboxy-*cis,cis*-muconate (Figure 2.1A) (31). Examination of *S. meliloti* extracts from cultures grown in minimal medium containing arabinose and protocatechuate, revealed protocatechuate 3,4-dioxygenase activity was readily detected in the parent strain Rm1021 while no activity was detected in RmG867(*pcaD*) and RmG879(*pcaG*) (data not shown). Since the dioxygenase consists of protein subunits encoded by two genes (*pcaHG*), disruption of either of these genes (as in RmG879) would therefore result in a corresponding loss of protocatechuate 3,4-dioxygenase activity. The enzyme β-ketoadipate enol-lactone hydrolase (PcaD) mediates the conversion of β-ketoadipate enol-lactone to β-ketoadipate (31), and a mutation in *pcaD* (as in RmG867) would block this step of the pathway, preventing the production of β-ketoadipate (Figure 2.1A). Organization of the *pcaDCHGB* operon ensures that insertion of a transposon within *pcaD* would also disrupt expression of downstream *pca* genes, including *pcaHG* (Figure 2.1B). The loss of protocatechuate 3,4-dioxygenase activity in RmG867(*pcaD*) presumably results from the polar nature of the mutation in this strain and supports the assumption that the genes in question are organized as a single transcriptional unit.

An *S. meliloti* Rm1021 pLAFR1 clone (12), pTH178, was isolated on the basis of its ability to complement the Pca$^-$ phenotype of the RmG867 and RmG879 mutants. The presence of the pTH178 plasmid in strains RmG867 and RmG879 restored a Rothera-positive phenotype and protocatechuate 3,4-dioxygenase activity to the RmG867 and

RmG879 mutant strains (data not shown). Consistent with its ability to complement *pcaD* and *pcaG* mutant phenotypes, DNA sequencing of pTH178 revealed that this cosmid carries the *pcaDCHGB* region in its entirety.

Regulation of expression of the pcaDCHGB operon

In *S. meliloti*, the *pcaDCHGB* operon is located adjacent to, and transcribed divergently from, a gene encoding a product with similarity to the *A. tumefaciens* LysR-type transcriptional regulator PcaQ. To examine the possible involvement of PcaQ in regulating *pcaDCHGB* expression, an *S. meliloti* interposon knock-out mutant of *pcaQ* was constructed. The *pcaQ*-minus strain was unable to utilize protocatechuate as a sole carbon source, however growth of the mutant strain was comparable to wild-type in media containing glucose or glycerol as carbon sources (data not shown).

To monitor expression from the *pcaD* promoter, the *pcaD-pcaQ* intergenic region was cloned into pMP220 (48), with the *pcaD* promoter in the same orientation as a promoterless *lacZ* reporter gene. The resulting replicating plasmid, pTH468, was conjugated into RmG212 (*lac*$^-$ Rm1021 derivative), RmP134 (*pcaQ*::Ω derivative of RmG212) and into RmK948 (a *pcaG*::Tn5 derivative of RmG212), a strain that is incapable of metabolizing protocatechuate to β-carboxy-*cis*,*cis*-muconate. Expression of *pcaD* (as measured by β-galactosidase activity) was examined following growth in minimal medium containing glycerol as a carbon source and in medium containing glycerol plus protocatechuate. In the wildtype background, expression of *pcaD* was induced greater than 10-fold in the presence of protocatechuate while *pcaD* expression did not increase under similar growth conditions in RmP134(*pcaQ*) (Table 2.2). This suggests that a product encoded by *pcaQ* is required for induction of *pcaD* expression. The low level of expression of *pcaD* in both the uninduced wildtype and *pcaQ* mutant backgrounds suggests PcaQ does not act as a repressor of *pcaD* transcription. Expression of *pcaD* was also not induced in the *pcaG* mutant background (RmK948) and this suggests that protocatechuate itself does not act as an inducing agent. Presumably an intermediate in the β-ketoadipate pathway is required to induce *pcaDCHGB*. Given the

similarity that exists in genetic organization and regulation of the β-ketoadipate pathways in *S. meliloti* and *A. tumefaciens* (38), it is likely that β-carboxy-*cis*,*cis*-muconate and γ-carboxymuconolactone serve as coinducers in the PcaQ-regulated expression of the *pcaDCHGB* operon in *S. meliloti*, however this matter was not examined further.

Identification of pcaDCHGB transcriptional start sites

Primer extension analysis was performed on RNA isolated from *S. meliloti* strain Rm1021 grown in LBmc in the presence and absence of protocatechuate prior to RNA isolation. A 30-mer oligonucleotide complementary to the 5' *pcaD* coding region was used to prime the extension reaction. Two major extension products were obtained only with template RNA isolated from cells grown in the presence of protocatechuate, indicating transcriptional start sites at G and C residues, located 14 and 15 nucleotides upstream of the predicted *pcaD* translational start site, respectively (see Figure 2.2A).

The sequence upstream of the *pcaD* transcriptional start sites has AT-rich sequences centered at -10 and -35 hexanucleotide regions (Figure 2.2B). Alignment of the sequences upstream of the *pcaD* genes from *S. meliloti* and *A. tumefaciens* with the region upstream of the annotated *M. loti pcaD* gene indicates these regions are conserved amongst the three species (Figure 2.2B). However the *M. loti* sequence that corresponds to the inferred *S. meliloti* -10 hexanucleotide includes the annotated *M. loti pcaD* translational start site. Alignment of the PcaD amino acid sequence from *S. meliloti*, *A. tumefaciens*, and *M. loti*, reveals that the *M. loti* protein consists of an additional 8 amino acids at the N-terminal that are absent in the other two species (Figure 2.2C). These results suggest that the correct start codon for *pcaD* in *M. loti* is located 24 bp downstream of the annotated site. Interestingly, this would mean that the *M. loti* gene employs a GTG translational start codon, as is predicted for *pcaD* in *S. meliloti*.

Examination of the *pcaD* promoter region also reveals potential PcaQ binding sites, several of which have also been conserved amongst *S. meliloti*, *A. tumefaciens*, and *M. loti*. As a member of the LysR family of transcriptional regulators, PcaQ may be expected to recognize and bind elements established upon a (T-N_{11}-A) consensus binding

motif (16, 47). Examination of the *pcaD-pcaQ* intergenic region from *S. meliloti* revealed seven T-N_{11}-A motifs. Of these, two T-N-A-N_9-A motifs span the -35 region in *S. meliloti* and are conserved in *A. tumefaciens* and *M. loti* (Figure 2.2B). As *pcaD* and *pcaQ* are separated by a 94 bp intergenic region, it is possible that binding of PcaQ to a given site(s) may simultaneously exert positive and negative effects with respect to the expression of *pcaD* and *pcaQ* respectively as has been shown for the *crgA-mdaB* genes in *Neisseria meningitidis* (6, 21). PcaQ-mediated auto-regulation has been described in *A. tumefaciens* (37), and we similarly have evidence of auto-regulation in *S. meliloti* (data not shown). This work is being pursued using purified PcaQ to facilitate the identification of PcaQ binding sites.

Identification and purification of PcaIJ

Genes encoding the β-ketoadipate succinyl-CoA transferase proteins (PcaIJ) have not been identified in the *S. meliloti* genome (13). Two genes (*smb20587* and *smb20588*) are annotated as encoding subunits of a coenzymeA transferase, and these are located approximately 10 kb from the *pcaDCHGB* operon. The *smb20587* and *smb20588* genes lie upstream of the *pcaF* gene annotated to encode β-ketoadipyl CoA thiolase. The low amino acid sequence similarity of *smb20587* and *smb20588* with other PcaIJ proteins prevented their annotation as PcaIJ orthologues.

To establish whether *smb20587* and *smb20588* encode β-ketoadipate succinyl-CoA transferase activity, these genes were cloned into pTH1227, an IPTG-inducible expression vector carrying the *tac* promoter, to give pTH1459. *S. meliloti*(pTH1459) cells were induced with both protocatechuate and IPTG. Whole cell lysate obtained from induced cultures exhibited β-ketoadipate succinyl-CoA transferase activity as detected spectrophotometrically by the increase in absorbance at 305 nm that accompanies formation of the β-ketoadipyl-CoA:Mg^{++} complex (see Materials and Methods). This activity was sequentially purified to near homogeneity using a combination of ammonium sulphate precipitation and chromatography with columns containing Phenyl Sepharose CL-4B, Source 30Q, and CHT Ceramic Hydroxyapatite (see materials and

methods for details) (Figure 2.3). SDS-PAGE analysis demonstrated the presence of a 28 kDa band that was consistent with the predicted size (27.7 kDa) of the β-subunit encoded by *smb20588*. A second higher molecular weight protein was present as a doublet of 36 and 38 kDa bands which we thought could be anomalously migrating species of the α-subunit (predicted size: 31.2 kDa) encoded by *smb20587*. To positively identify these polypeptide species, all 3 bands were subjected to a trypsin digest and tandem mass spectrometry. The 28 kDa protein was confirmed as the β-subunit of a CoA-transferase encoded by gene *smb20588*. Both the 36 and 38 kDa proteins were identified as the α-subunits of a CoA-transferase encoded by gene *smb20587*. We have not explored why the α-subunit is expressed as a doublet but it may be because expression was induced both from the pSymB megaplasmid (using protocatechuate as an inducer) and from an IPTG-inducible expression vector. Possibly, an alternative start codon was used in translation from the expression vector. Examination of nucleotide sequence encoding the α-subunit reveals two potential in-frame start codons downstream of the annotated start site. Translation initiation from these alternative start sites would generate proteins differing by either 4 or 22 amino acids relative to the full length protein. The difference in molecular mass between the doublet proteins as estimated from SDS-PAGE is ~ 2 kDa (or approximately 20 amino acids) and this difference might therefore be explained by the use of two distinct translational start sites. In any case, polypeptides encoded by *smb20587* and *smb20588* are enriched to near homogeneity in a protein sample purified solely on the basis of β-ketoadipate succinyl-CoA transferase activity.

The transfer of coenzymeA to β-ketoadipate has previously been documented as resulting from the non-specific activity of an adipate succinyl-CoA transferase (5, 19). To eliminate the possibility that an adipate succinyl-CoA transferase had inadvertently been purified, we performed enzyme assays with β-ketoadipate and with increasing concentrations of adipate to examine substrate specificity (Table 2.3). Addition of equimolar amounts of adipate did not result in a significant decrease in enzyme activity (7% decrease in activity), however a five-fold greater concentration of adipate (relative to β-ketoadipate) in a reaction reduced activity by 56%. These results indicate that although

adipate might compete with β-ketoadipate as a substrate when present in a greater concentration, β-ketoadipate is the preferred substrate of this enzyme. Enzyme activity was also dependent upon the presence of β-ketoadipate, succinyl-CoA, and Mg^{++}, and omission of any one of these reagents from the reaction mixture abolished activity (<0.01 μmols/min/mg) (data not shown). We therefore concluded that *smb20587* and *smb20588* encode subunits of a β-ketoadipate succinyl-CoA transferase and these genes were named *pcaI* and *pcaJ* respectively.

The β-ketoadipate pathway is composed of protocatechuate and catechol branches, and either *pca* or *cat* genes (or both) may be present within a given species. Accordingly, β-ketoadipate succinyl-CoA transferase activity may be encoded by either *pca* and/or *cat* genes (*pcaIJ* and *catIJ*, respectively), and these may share sequence similarity (15, 25). Amino acid sequence identity between the *S. meliloti* PcaIJ and that of *A. baylyi* (25) (PcaI: 21%; PcaJ: 20%), *P. putida* (33) (PcaI: 21%; PcaJ: 23%), *B. japonicum* (23) (PcaI: 17%; PcaJ: 20%) and even *A. tumefaciens* (53) (PcaI: 19%; PcaJ: 21%) is limited. In contrast, the sequence identity that exists between the *S. meliloti* PcaIJ and that of *M. loti* (22) (PcaI: 63%; PcaJ: 71%), *P. aeruginosa* PAO1 (49) (PcaI: 70%; PcaJ: 60%) and *Pseudomonas* sp. strain B13 CatIJ (15) (CatI: 68%; CatJ: 60%) is extensive. Likewise, signature sequences typically present in PcaI and PcaJ of many species are absent or modified in their *S. meliloti* counterparts, a situation comparable to that previously described in Göbel et al (15). In PcaI, an N-terminal glycine cluster ([DN]-[GN]-x[2]-[LIVMFA][3]-G-G-F-x[3]-G-x-P) (52) present in *A. tumefaciens, B. japonicum, A. baylyi*, and *P. putida* proteins has been modified in *S. meliloti* by the deletion of one glycine residue and the substitution of another with glutamic acid. This modification has previously been reported in CatI of *Pseudomonas* sp. strain B13 (15), and may also be observed in *P. aeruginosa* and *M. loti pcaIJ*. In PcaJ, an N-terminal signature sequence ([LF]-[HQ]-S-E-N-G-[LIVF][2]-[GA]) (33) present in *B. japonicum, A. baylyi*, and *P. putida* is absent in *S. meliloti*. The E-S-G motif reported in CatJ of *Pseudomonas* sp. strain B13 (15), and present in *P. aeruginosa*, *M loti*, and a glutaconate-CoA transferase of *Acidaminococcus fermentans*, is also conserved in *S. meliloti*. Thus

the differences observed between β-ketoadipate succinyl-CoA transferases of one group of species (*A. baylyi*, *P. putida*, *A. tumefaciens*, *B. japonicum*) versus another (*Pseudomonas* sp. strain B13, *P. aeruginosa*, *M. loti*, *S. meliloti*) are striking, especially since closely related organisms such as *A. tumefaciens* and *S. meliloti* carry different forms of the enzyme.

Identification of pcaI transcriptional start site

To facilitate our analysis of transcriptional regulation of *pcaIJF*, we determined the transcriptional start site of the operon using primer extension analysis. Two different primers were used to map the start site, and each yielded a single extension product only with RNA derived from cells grown in the presence of protocatechuate (Figure 2.4A). The transcriptional start site of the *pcaIJF* operon mapped to an A residue located 20 nucleotides upstream of the predicted PcaI start codon. Figure 2.4B shows the sequence of the *pcaI* promoter including the inferred -10 and -35 hexanucleotide regions. Examination of the promoter region revealed the presence of a 9 base pair palindromic sequence (5' GCTTATTCG 3') coincident with the predicted -10 region. Comparison of *pcaIp* with other *S. meliloti* promoters reveals some sequence similarity, particularly with respect to the inferred -35 regions (1, 2, 11, 26, 27, 32).

Regulation of the pcaIJF operon requires an IclR-type regulator encoded by smb20586

The regulatory systems associated with the protocatechuate branch of the β-ketoadipate pathway have been described in a few species. In *P. putida*, expression of *pca* genes (with the exception of *pcaHG*) is regulated by the IclR-type transcriptional regulator PcaR, with β-ketoadipate serving as an inducing metabolite (17, 34). In *A. baylyi*, expression of the *pcaIJFBDKCHG* operon is regulated by the IclR-type protein PcaU (14), which acts as both a repressor and activator of the operon, depending upon the presence of the inducer protocatechuate (51). In *A. tumefaciens* and *R. leguminosarum*, β-ketoadipate succinyl-CoA transferase activity is likewise induced by β-ketoadipate, and

in *A. tumefaciens* expression of *pcaIJ* is subject to regulation by an IclR-type protein (36, 39).

In *S. meliloti*, the *pcaIJF* operon is located beside a gene (*smb20586*) encoding a putative IclR-type protein. To determine whether this putative regulator is involved in the regulation of *pcaIJF* gene expression, a streptomycin/spectinomycin antibiotic cassette was used to inactivate *smb20586*, generating strain RmK1014 (as verified by Southern hybridization). RmK1014 was unable to grow with protocatechuate but grew like wild-type with glycerol or glucose as sole carbon source however, showing that *smb20586* is essential for protocatechuate metabolism (data not shown).

A promoterless *gusA* reporter gene was inserted at the 3' end of the *pcaIJF* operon in the genome in order to monitor expression of this operon. The fusion was designed such that *pcaF* was not disrupted, as was verified by growth of the fusion strain with protocatechuate as a sole carbon source. In the wild-type Rm1021 background, expression of *pcaF*::*gusA* increased 10-fold and 5-fold following growth in the presence of protocatechuate and adipate, respectively (Table 2.4). In this instance, adipate was used as an analogue of β-ketoadipate. In contrast, expression of *pcaF*::*gusA* in the *smb20586* mutant background was minimal regardless of whether protocatechuate or adipate was in the growth media. This demonstrated that a product encoded by *smb20586* was required for *pcaIJF* expression.

The regulator encoded by *smb20586* from *S. meliloti* shares 59% amino acid identity with PcaR of *A. tumefaciens* (53) and 41% identity with PcaR from *P. putida* (46). In *P. putida* and *A. tumefaciens*, adipate is utilized as an inducer analogue of β-ketoadipate (36, 40) and based on the inducing activity of adipate as revealed in Table 2.4, we conclude that β-ketoadipate is the *in vivo* metabolite responsible for *pcaIJF* expression in *S. meliloti*. Based upon its amino acid sequence similarity and role in the regulation of *pcaIJF* expression, we have renamed *smb20586* as *pcaR*.

Transcriptional regulator PcaR participates in auto-regulation

To examine whether PcaR could also auto-regulate its own synthesis, the region upstream of the *pcaR* translational start site was amplified and cloned into the *gusA* reporter vector pFus1 (45) to create pTH1335. This plasmid was then conjugated into *S. meliloti* strains Rm1021 and a *pcaR*::Ω derivative (RmK1014) and promoter activity was monitored via β-glucuronidase enzyme assays (Table 2.5). In both strains, reporter enzyme activities are comparable in cells grown with and without protocatechuate, indicating that *pcaR* expression is not influenced by the presence of this compound. Expression of *pcaR*::*gusA* in the low-copy-number plasmid pTH1335 is relatively low (comparable to expression of the *pcaF*::*gusA* fusion in uninduced cells), as expected of a regulatory gene. However, expression from the *pcaR* promoter was increased 5-fold in the *pcaR* mutant compared to Rm1021 (Table 2.5). These results demonstrate that PcaR expression is negatively auto-regulated.

Analysis of the β-ketoadipate pathways in S. meliloti and A. tumefaciens

Many aspects regarding the organization and regulation of genes encoding enzymes involved in the upper portion of the β-ketoadipate pathway (metabolism of protocatechuate to β-ketoadipate) are conserved between *S. meliloti* and *A. tumefaciens.* In addition to amino acid sequence similarity between homologues of the two species (which ranges from 60 to 77% identity), the organization of genes into a single operon (*pcaDCHGB*) is identical. Likewise, regulation of the operon in both species is mediated by a LysR-type transcriptional regulator. Identification of the *S. meliloti pcaD* transcriptional start site, and comparison with *M. loti* and *A. tumefaciens* sequences, reveals that the *pcaD* promoter is likely conserved amongst these species. Similarly, previous work utilizing an *A. tumefaciens pcaD*::*lacZ* fusion indicates that the PcaQ binding sites of this species are recognized by the *S. meliloti* homologue (38). Although the evidence is limited, it is also quite likely that β-carboxy-*cis,cis*-muconate and γ-carboxymuconolactone serve as coinducing metabolites required for the PcaQ-regulated expression of the *pcaDCHGB* operon in *S. meliloti* as has been shown in *A. tumefaciens.*

With respect to genes whose products participate in the lower portion of the pathway (conversion of β-ketoadipate to succinate and acetyl-CoA) certain similarities may also be observed between *S. meliloti* and *A. tumefaciens*. In both species, genes encoding subunits of a β-ketoadipate succinyl-CoA transferase (*pcaI* and *pcaJ*) are organized into an operon whose expression is regulated by an IclR-type transcriptional regulator (PcaR), with β-ketoadipate serving as a coeffector. On the other hand, amino acid sequence identity between PcaIJ of *S. meliloti* and *A. tumefaciens* is quite low and signature sequences present in the *A. tumefaciens* PcaIJ are absent or modified in the *S. meliloti* protein.

Comparison of the β-ketoadipate pathways in *S. meliloti* and *A. tumefaciens* can be extended to include PobA (4-hydroxybenzoate hydroxylase), an enzyme involved in the catalysis of 4-hydroxybenzoate to protocatechuate. In *S. meliloti*, a gene annotated as *pobA* is situated between the two *pca* operons. Despite the presence of this gene, *S. meliloti* strain Rm1021 is unable to utilize 4-hydroxybenzoate as a sole carbon source and we were unable to isolate an Rm1021 mutant that acquired this capability (data not shown). Likewise, an auxanographic study of *Rhizobiaceae* reported *S. meliloti* incapable of growing upon this compound (41). Moreover using two independent transcriptional fusions to the *S. meliloti pobA* gene (utilizing *gfp* and *gusA* as reporters) we detected only basal *pobA* expression which did not increase upon addition of 4-hydroxybenzoate or protocatechuate to the growth medium (data not shown). In contrast, *A. tumefaciens* is able to grow at the expense of 4-hydroxybenzoate (41), and *pobA* expression in this species is induced by 4-hydroxybenzoate via PobR (39). In *A. tumefaciens* (and *R. leguminosarum*), an AraC-family transcriptional regulator (PobR) regulates *pobA* expression (39) but there are no *araC* homologues located nearby in the *S. meliloti* genome. A putative LysR-type regulator (*smb20582*) is encoded directly downstream of *pobA*, and although members of this family of regulators are typically transcribed divergently from a target gene, it is possible that this gene encodes a *pobA* regulator. In *A. baylyi*, an IclR-type regulator positively regulates expression of the *pobA* gene (7) however the only close IclR gene in *S. meliloti* has been identified as *pcaR* (this

work) because it regulates *pcaIJF* expression. Expression of *pobA* (as determined by reporter enzyme assays) in wild-type and *pcaR*-minus backgrounds is comparable and we have concluded that PcaR is not involved in the regulation of *pobA* gene expression. Consequently, the identity of a regulator of *pobA* expression in *S. meliloti*, if one exists, remains obscure.

The supraoperonic organization of genes whose products participate in the catabolism of protocatechuate and related compounds has been well documented (4, 18, 39). In *A. tumefaciens* and *S. meliloti*, the two *pca* operons are clustered in close proximity, flanking the putative *pobA* gene (4). It has been proposed that this supraoperonic organization in *A. tumefaciens* arose as the result of the acquisition of these genes as a unit and that the protocatechuate pathway evolved prior to the divergence of *Agrobacterium* and *Rhizobium* species (39). Although much of the genetic organization and regulation in these systems has been conserved, the differences observed (particularly) with respect to PcaIJ of *S. meliloti* and *A. tumefaciens* are inconsistent with a shared history. It may be that while the protocatechuate catabolic pathway was established prior to *Agrobacterium* and *Rhizobium* speciation, the *pcaIJ* genes present today were acquired at some point afterwards independently in one or both genera. Possibly, this punctuated assembly of the two *pca* operons resulted in the loss of the *pobA* regulator, leaving *S. meliloti* unable to efficiently metabolize 4-hydroxybenzoate.

2.5 References

1. **Albright, L. M., C. W. Ronson, B. T. Nixon, and F. M. Ausubel**. 1989. Identification of a gene linked to *Rhizobium meliloti ntrA* whose product is homologous to a family to ATP-binding proteins. J.Bacteriol. **171**:1932-1941.

2. **Bae, Y. M., E. Holmgren, and I. P. Crawford**. 1989. *Rhizobium meliloti* anthranilate synthase gene: cloning, sequence, and expression in *Escherichia coli*. J.Bacteriol. **171**:3471-3478.

3. **Blondelet-Rouault, M., J. Weiser, A. Lebrihi, P. Branny, and J. Pernodet**. 1997. Antibiotic resistance gene cassettes derived from the interposon for use in *E. coli* and *Streptomyces*. Gene **190**:315-317.

4. **Buchan, A., E. L. Neidle, and M. A. Moran**. 2004. Diverse organization of genes of the beta-ketoadipate pathway in members of the marine *Roseobacter* lineage. Appl.Environ.Microbiol **70**:1658-1668.

5. **Canovas, J. L. and R. Y. Stanier**. 1967. Regulation of the enzymes of the beta-ketoadipate pathway in *Moraxella calcoacetica*. 1. General aspects. Eur.J.Biochem. **1**:289-300.

6. **Deghmane, A. E., D. Giorgini, L. Maigre, and M. K. Taha**. 2004. Analysis in vitro and in vivo of the transcriptional regulator CrgA of *Neisseria meningitidis* upon contact with target cells. Mol Microbiol **53**:917-927.

7. **DiMarco, A. A., B. Averhoff, and L. N. Ornston**. 1993. Identification of the transcriptional activator *pobR* and characterization of its role in the expression of *pobA*, the structural gene for p-hydroxybenzoate hydroxylase in *Acinetobacter calcoaceticus*. J.Bacteriol. **175**:4499-4506.

8. **Durham, D. R., L. A. Stirling, L. N. Ornston, and J. J. Perry**. 1980. Intergeneric evolutionary homology revealed by the study of protocatechuate 3,4-dioxygenase from *Azotobacter vinelandii*. Biochemistry **19**:149-155.

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

10. **Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, and A. Puhler**. 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*. Proc.Natl.Acad.Sci.U.S.A **98**:9889-9894.

11. **Fisher, R. F., H. L. Brierley, J. T. Mulligan, and S. R. Long**. 1987. Transcription of *Rhizobium meliloti* nodulation genes. Identification of a *nodD* transcription initiation site in vitro and in vivo. J.Biol Chem. **262**:6849-6855.

12. **Friedman, A. M., S. R. Long, S. E. Brown, W. J. Buikema, and F. M. Ausubel**. 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene **18**:289-296.

13. **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.

14. **Gerischer, U., A. Segura, and L. N. Ornston**. 1998. PcaU, a transcriptional activator of genes for protocatechuate utilization in *Acinetobacter*. J.Bacteriol. **180**:1512-1524.

15. **Gobel, M., K. Kassel-Cati, E. Schmidt, and W. Reineke**. 2002. Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: cloning, characterization, and analysis of sequences encoding 3-oxoadipate:succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase. J.Bacteriol. **184**:216-223.

16. **Goethals, K., M. Van Montagu, and M. Holsters**. 1992. Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. Proc.Natl.Acad.Sci.U.S.A **89**:1646-1650.

17. **Harwood, C. S., N. N. Nichols, M. K. Kim, J. L. Ditty, and R. E. Parales**. 1994. Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. J.Bacteriol. **176**:6479-6488.

18. **Harwood, C. S. and R. E. Parales**. 1996. The beta-ketoadipate pathway and the biology of self-identity. Annu.Rev.Microbiol **50**:553-590.

19. **Hoet, P. P. and R. Y. STANIER**. 1970. Existence and functions of two enzymes with beta-ketoadipate: succinyl-CoA transferase activity in *Pseudomonas florescens*. Eur.J.Biochem. **13**:71-76.

20. **Holding, A. J. and J. G. Collee**. 1971. Routine biochemical tests, p. 1-32. *In* J. R. Norris and D. W. Ribbons (eds.), Methods in Microbiology. Academic Press, London.

21. **Ieva, R., C. Alaimo, I. Delany, G. Spohn, R. Rappuoli, and V. Scarlato**. 2005. CrgA is an inducible LysR-type regulator of *Neisseria meningitidis*, acting both as a repressor and as an activator of gene transcription. J.Bacteriol. **187**:3421-3430.

22. **Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, and S. Tabata**. 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7**:331-338.

23. **Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada, and S. Tabata**. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. **9**:189-197.

24. **Kaschabek, S. R., B. Kuhn, D. Muller, E. Schmidt, and W. Reineke**. 2002. Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: purification and characterization of 3-oxoadipate:succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase. J.Bacteriol. **184**:207-215.

25. **Kowalchuk, G. A., G. B. Hartnett, A. Benson, J. E. Houghton, K. L. Ngai, and L. N. Ornston**. 1994. Contrasting patterns of evolutionary divergence within the *Acinetobacter calcoaceticus pca* operon. Gene **146**:23-30.

26. **Leong, S. A., P. H. Williams, and G. S. Ditta**. 1985. Analysis of the 5' regulatory region of the gene for delta-aminolevulinic acid synthetase of *Rhizobium meliloti*. Nucleic Acids Res. **13**:5965-5976.

27. **MacLellan, S. R., L. A. Smallbone, C. D. Sibley, and T. M. Finan**. 2005. The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of alpha-proteobacterial plasmids. Mol Microbiol **55**:611-623.

28. **Meade, H. M., S. R. Long, G. B. Ruvkun, S. E. Brown, 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.

29. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y.

30. **Oke, V. and S. R. Long**. 1999. Bacterial genes induced within the nodule during the *Rhizobium*-legume symbiosis. Mol Microbiol **32**:837-849.

31. **Ornston, L. N. and R. Y. Stanier**. 1966. The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. J.Biol Chem. **241**:3776-3786.

32. **Osteras, M., B. T. Driscoll, and T. M. Finan**. 1995. Molecular and expression analysis of the *Rhizobium meliloti* phosphoenolpyruvate carboxykinase (*pckA*) gene. J.Bacteriol. **177**:1452-1460.

33. **Parales, R. E. and C. S. Harwood**. 1992. Characterization of the genes encoding beta-ketoadipate: succinyl-coenzyme A transferase in *Pseudomonas putida*. J.Bacteriol. **174**:4657-4666.

34. **Parales, R. E. and C. S. Harwood**. 1993. Regulation of the *pcaIJ* genes for aromatic acid degradation in *Pseudomonas putida*. J.Bacteriol. **175**:5829-5838.

35. **Parke, D.** 1993. Positive regulation of phenolic catabolism in *Agrobacterium tumefaciens* by the *pcaQ* gene in response to beta-carboxy-cis,cis-muconate. J.Bacteriol. **175**:3529-3535.

36. **Parke, D.** 1995. Supraoperonic clustering of *pca* genes for catabolism of the phenolic compound protocatechuate in *Agrobacterium tumefaciens*. J.Bacteriol. **177**:3808-3817.

37. **Parke, D.** 1996. Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds, from *Agrobacterium tumefaciens*. J.Bacteriol. **178**:266-272.

38. **Parke, D.** 1996. Conservation of PcaQ, a transcriptional activator of *pca* genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. J.Bacteriol. **178**:3671-3675.

39. **Parke, D.** 1997. Acquisition, reorganization, and merger of genes: novel management of the beta-ketoadipate pathway in *Agrobacterium tumefaciens*. FEMS Microbiology Letters **146**:3-12.

40. **Parke, D. and L. N. Ornston**. 1976. Constitutive synthesis of enzymes of the protocatechuate pathway and of the beta-ketoadipate uptake system in mutant strains of *Pseudomonas putida*. J.Bacteriol. **126**:272-281.

41. **Parke, D. and L. N. Ornston**. 1984. Nutritional diversity of *Rhizobiaceae* revealed by auxanography. Journal of General Microbiology **130**:1743-1750.

42. **Parke, D. and L. N. Ornston**. 1986. Enzymes of the beta-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium* spp. and constitutive in *Bradyrhizobium* spp. J.Bacteriol. **165**:288-292.

43. **Prentki, P. and H. M. Krisch**. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene **29**:303-313.

44. **Quandt, J. and M. F. Hynes**. 1993. Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. Gene **127**:15-21.

45. **Reeve, W. G., R. P. Tiwari, P. S. Worsley, M. J. Dilworth, A. R. Glenn, and J. G. Howieson**. 1999. Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. Microbiology **145**:1307-1316.

46. **Romero-Steiner, S., R. E. Parales, C. S. Harwood, and J. E. Houghton**. 1994. Characterization of the *pcaR* regulatory gene from *Pseudomonas putida*, which is required for the complete degradation of p-hydroxybenzoate. J.Bacteriol. **176**:5771-5779.

47. **Schell, M. A.** 1993. Molecular biology of the LysR family of transcriptional regulators. Annu.Rev.Microbiol **47**:597-626.

48. **Spaink, H. P., J. H. Robert, C. A. Okker, E. P. Wijffelman, and B. J. J. Lugtenberg**. 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Plant Molecular Biology **9**:27-39.

49. **Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson**. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature **406**:959-964.

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

51. **Trautwein, G. and U. Gerischer**. 2001. Effects exerted by transcriptional regulator PcaU from *Acinetobacter* sp. strain ADP1. J.Bacteriol. **183**:873-881.

52. **Wierenga, R. K., P. Terpstra, and W. G. Hol**. 1986. Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. J.Mol Biol **187**:101-107.

53. **Wood, D. W., J. C. Setubal, R. Kaul, D. E. Monks, J. P. Kitajima, V. K. Okura, Y. Zhou, L. Chen, G. E. Wood, N. F. Almeida, Jr., L. Woo, Y. Chen, I. T. Paulsen, J. A. Eisen, P. D. Karp, D. Bovee, Sr., P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutyavin, R. Levy, M. J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z. Y. Zhao, M. Dolan, F. Chumley, S. V. Tingey, J. F. Tomb, M. P. Gordon, M. V. Olson, and E. W. Nester**. 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science **294**:2317-2323.

54. **Yeh, W. K. and L. N. Ornston**. 1981. Evolutionarily homologous alpha 2 beta 2 oligomeric structures in beta-ketoadipate succinyl-CoA transferases from *Acinetobacter calcoaceticus* and *Pseudomonas putida*. J.Biol Chem. **256**:1565-1569.

Table 2.1 Bacterial strains and plasmids used in this study.

Strain or plasmid	Relevant characteristics	Source or reference
S. meliloti		
Rm1021	Sm^r derivative of wild-type strain SU47; Sm^r	(28)
Rm5004	Rm1021 *recA*::Tn5; Sm^r, Nm^r	Strain collection
RmG212	Rm1021 *lac*$^-$; Sm^r	Strain collection
RmG867	Rm1021 *pcaD*::Tn5; Sm^r, Nm^r	This study
RmG879	Rm1021 *pcaG*::Tn5; Sm^r, Nm^r	This study
RmK927	Rm5004 (pTH1459); Sm^r, Tc^r	This study
RmK948	RmG212 *pcaG*::Tn5; Sm^r, Nm^r	This study
RmK1014	Rm1021 *pcaR*::Ω; Sm^r, Sp^r	This study
RmK1015	Rm1021 *pcaF*::*gusA*; Sm^r, Nm^r	This study
RmK1016	Rm1021 *pcaF*::*gusA pcaR*::Ω; Sm^r, Nm^r, Sp^r	This study
RmP134	RmG212 *pcaQ*::Ω; Sm^r, Gm^r	This study
RmP135	RmG212 (pTH468); Sm^r, Tc^r	This study
RmP136	RmP134 (pTH468); Sm^r, Gm^r, Tc^r	This study
RmP892	Rm1021 (pTH1335); Sm^r, Tc^r	This study
RmP893	RmK1014 (pTH1335); Sm^r, Tc^r	This study
RmP894	RmK948 (pTH468); Sm^r, Nm^r, Tc^r	This study
E. coli		
DH5α	F$^-$ Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17*(r_K^- m_K^-) *supE44* λ$^-$ *thi-1 gyrA96 relA1*	Bethesda Research Laboratories, Inc.
BL21	*E. coli* B F$^-$ *ompT hsdS*(r_B^- m_B^-) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte [*argU proL* Cam^r]	Stratagene
Plasmids		
pLAFR1	IncP cosmid cloning vector; Tc^r	(12)
pFus1	Broad host range (bhr), *gusA* transcriptional reporter plasmid; Tc^r	(45)
pMP220	Bhr, *lacZ* transcriptional reporter plasmid; Tc^r	(48)
pHP45Ω	pBR322 derivative with Ω element; Ap^r, Sp/Sm^r	(43)
pHP45Ω*aac*	pBR322 derivative with Ω element; Ap^r, Gm^r	(3)
pRK602	pRK600Ω::Tn5, suicide vector used in transposon mutagenesis of Rm1021; Cm^r, Nm^r	(9)
pJQ200 uc1	Suicide vector with *sacB* to select for plasmid excision; Gm^r	(44)
pVO155	Suicide vector with promoterless *gusA*; Ap^r, Nm^r	(30)

pTH178	pLAFR1 derivative complementing Pca$^-$ phenotype of RmG867 and RmG879; Tc^r	This study
pTH468	434 bp *Eco*RI-*Xba*I PCR product encompassing *pcaD-pcaQ* intergenic region in pMP220 (*pcaD*::*lacZ*); Tc^r	This study
pTH1227	Bhr, derivative of pFus-1 with P_{tac} promoter inserted upstream of *gusA*; Tc^r	(J. Cheng and T.M. Finan, unpublished data)
pTH1335	153 bp *Eco*RI-*Pst*I PCR product encompassing *pcaR-pcaI* intergenic region in pFus-1 (*pcaR*::*gusA*); Tc^r	This study
pTH1338	963 bp blunt-ended PCR product encompassing *SacII* site within *pcaR*, cloned into pUC119 via *Sma*I; Ap^r	This study
pTH1340	ΩSm/Sp^r from pHP45Ω into pTH1338 via *Sac*II; Ap^r, Sm^r, Sp^r	This study
pTH1351	*pcaR*::Ω from pTH1340 into pJQ200 uc1 via *Not*I; Sm^r, Sp^r, Gm^r	This study
pTH1360	pVO155 derivative with *gusA* from pFus-1; Ap^r, Nm^r	(R. Zaheer and T.M. Finan, unpublished data)
pTH1459	1701 bp *Eco*RI-*Pst*I PCR product encompassing *pcaIJ* in expression vector pTH1227; Tc^r	This study
pTH1559	431 bp *Spe*I-*Xba*I PCR product encompassing 3' end of *pcaF* into pTH1360 (*pcaF*::*gusA*); Ap^r	This study
pTH1577	893 bp *Not*I PCR product encompassing *Pst*I site within *pcaQ* cloned into pJQ200 uc1; Gm^r	This study
pTH1592	ΩGm^r from pHP45Ω into pTH1577 via *Pst*I site; Gm^r	This study
pTH1882	*pcaQ*::Ω from pTH1592 into pTH1883 via *Not*I; Ap^r, Nm^r, Gm^r	This study
pTH1883	pTH1360 with *gusA* removed via *Not*I digest; Ap^r, Nm^r	This study

Table 2.2 Expression of *pcaD-lacZ* fusion in *S. meliloti* strains.

Strain	Relevant genotype	Growth conditions	β-galactosidase activity[a] (Miller units ± SD[b])
RmP135	Rm1021 *lac*⁻ (pTH468)	Glycerol	590 (18.5)
		Glycerol + PCA[c]	8210 (288.6)
RmP136	Rm1021 *lac*⁻ *pcaQ*::Ω (pTH468)	Glycerol	530 (39.6)
		Glycerol + PCA	311 (8.6)
RmP894	Rm1021 *lac*⁻ *pcaG*::Tn5 (pTH468)	Glycerol	468 (36.3)
		Glycerol + PCA	550 (20.2)

[a] Average of values obtained from three independent cultures grown overnight in LBmc and subcultured into M9-minimal medium with 0.5% glycerol ± 5 mM protocatechuate at 30°C for four hours.
[b] SD, standard deviation
[c] PCA, protocatechuate

Table 2.3 β-ketoadipate succinyl-CoA transferase activity in the presence of adipate.

Assay Condition	Adipate:β-ketoadipate (ratio)	Transferase activity[a] (mU/mg protein ± SD[b])	Transferase activity (Percentage ± SD)
0 mM adipate	---	539 (12.6)	100 (2.3)
2 mM adipate	1:5	562 (24.4)	104 (4.5)
10 mM adipate	1:1	503 (24.2)	93 (4.5)
50 mM adipate	5:1	236 (24.0)	44 (4.5)

[a] Enzyme assays were performed using purified enzyme. 1 Unit of enzyme is the amount required to convert 1 μmol of β-ketoadipate to β-ketoadipyl-CoA in 1 minute under assay conditions.

[b] SD, standard deviation

Table 2.4 Expression of *pcaF-gusA* fusion in *S. meliloti* wild-type and PcaR-minus backgrounds.

Strain	Relevant Genotype	Growth conditions	β-glucuronidase activity[a] (Miller units ± SD[b])
RmK1015	Rm1021 *pcaF*::*gusA*	Glycerol	27 (2.7)
		Glycerol + adipate	131 (11.7)
		Glycerol + PCA[c]	259 (18.6)
RmK1016	Rm1021 *pcaF*::*gusA*, *pcaR*::Ω	Glycerol	52 (3.0)
		Glycerol + adipate	48 (4.1)
		Glycerol + PCA	66 (6.8)

[a] Average of values obtained from three independent cultures grown overnight in LBmc and sub-cultured into M9-minimal medium with 0.5% glycerol ± 5 mM protocatechuate or 30 mM adipate at 30°C for four hours.
[b] SD, standard deviation
[c] PCA, protocatechuate

Table 2.5 Expression of *pcaR-gusA* fusion in *S. meliloti* wild-type and PcaR-minus backgrounds.

Strain	Relevant Genotype	Growth conditions	β-glucuronidase activity (Miller units ± SD)
RmP892	Rm1021 (pTH1335)	Glycerol	62 (2.2)
		Glycerol + PCA[a]	42 (2.6)
RmP893	Rm1021 *pcaR*::Ω (pTH1335)	Glycerol	291 (8.2)
		Glycerol + PCA	261 (3.1)

[a] Average of values obtained from three independent cultures grown overnight in LBmc and subcultured into M9-minimal medium with 0.5% glycerol ± 5 mM protocatechuate at 30°C for four hours.
[b] SD, standard deviation
[c] PCA, protocatechuate

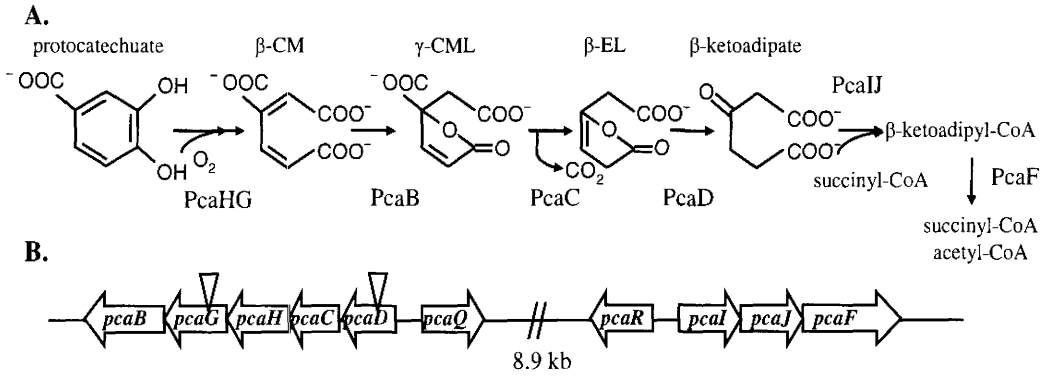


Figure 2.1 Protocatechuate catabolism in S. meliloti. (A) The protocatechuate branch of the β-ketoadipate pathway is involved in the catabolism of protocatechuate to intermediates which are funneled into the tricarboxylic acid cycle. β-CM: β-carboxy-cis, cis-muconate; γ-CML: γ-carboxymuconolactone; β-EL: β-ketoadipate enol-lactone. (B) Schematic depiction of the pca genes on the pSymB megaplasmid. Triangles indicate the location of transposon insertions in S. meliloti strains RmG879 (left triangle) and RmG867 (right triangle).

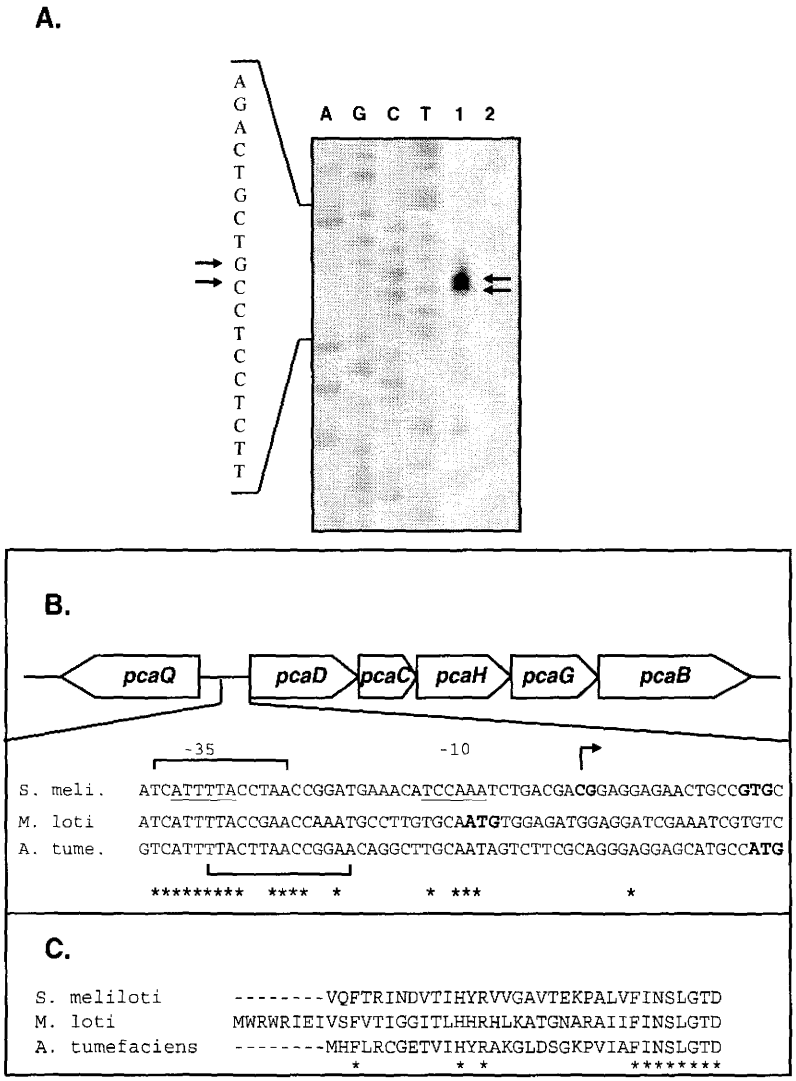


Figure 2.2 Analysis of the *pcaD* promoter in *S. meliloti*. (A) Primer extension reactions were performed using RNA isolated from *S. meliloti* strain Rm1021 grown in the presence (lane 1) and absence (lane 2) of 5 mM protocatechuate. Sequencing reactions were performed with the same primer used in primer extension reactions, and are shown to the left of the extension products. The arrowheads indicate the extension products and identify the corresponding nucleotides (right and left sides respectively). (B) Schematic depiction of the *pcaQ-pcaDCHGB* genes, including an alignment of the *S. meliloti pcaDp* region with sequence upstream of *pcaD* in *M. loti*, and *A. tumefaciens* (GenBank accession numbers NC_003078, NC_002678, and NC_003305, respectively). The two *S. meliloti pcaD* transcriptional start sites are identified (in bold) and the inferred -10 and -35 regions are underlined. Nucleotides conserved amongst all three species are indicated by an asterisk and predicted translational start sites are indicated (enlarged font). Two putative PcaQ binding sites present in *S. meliloti*, *A. tumefaciens*, and *M. loti* are indicated by square brackets. (C) Alignment of PcaD amino acid sequence (N-terminal) as annotated in *S. meliloti*, *M. loti*, and *A. tumefaciens* genome sequences. Invariant residues are indicated by an asterisk. Sequences were aligned using ClustalW (50).

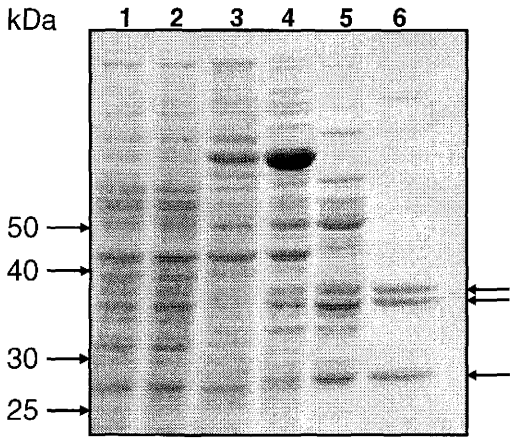
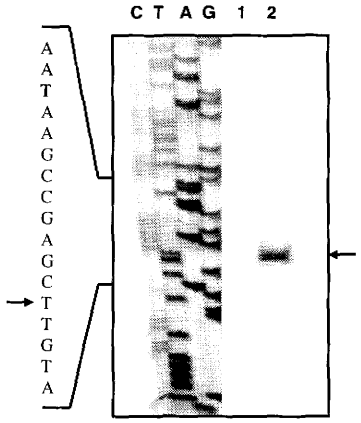


Figure 2.3 Purification of β-ketoadipate succinyl-CoA transferase in *S. meliloti*. *S. meliloti* Rm5004 carrying pTH1459 was grown in M9-minimal medium with 5 mM protocatechuate and 1 mM IPTG. Protein samples were subjected to SDS-PAGE in a 10% gel followed by staining with Coomassie brilliant blue. Lane 1: Crude cell lysate derived from uninduced culture. Lane 2: Crude cell lysate obtained from cells induced by 1 mM IPTG and 5 mM protocatechuate. Lane 3: Ammonium sulphate precipitate of induced cell lysate. Lane 4: Pooled eluate from phenyl sepharose column. Lane 5: Pooled eluate from Source 30Q (anion exchange) column. Lane 6: Pooled eluate from hydroxyapatite column. Molecular masses (in kilodaltons) are shown on the left. Arrowheads (right side) indicate the proteins identified by mass spectrometry. The 36 and 38 kDa proteins were identified as the α-subunits of a CoA-transferase encoded by gene *smb20587*; the 28 kDa protein was identified as the β-subunit of a CoA-transferase encoded by gene *smb20588*.

A.



B.

-35 -10 +1

TAAATTGACGCTTGCGAAGAACAGGCGCTTATTCGGCTCGAACATCACCGGAGGAAAACGATG

C.

Promoter sequences		Gene	Reference
-35	**-10**		
CTTGAC	ACTGATT-CGCGGAAGTG GGATTC	*incA1*	(18)
CTTGTC	TTGGGTC-AGCCTTGCCG GTATGT	*pckA*	(23)
CTTGAC	CAAATTC-CAGTAATAAG CAATTT	*ntrA*	(1)
CTTGAC	TTCGATC-GATGTTCGGG AGAATG	*hemA* P2	(17)
GTTGAC	CACTGAT-CGCTTTGAAG GAAGAA	*hemA* P1	(17)
CTTGAT	TCCATTAACTTCAGGGTT CTCTAA	*nodD1*	(7)
CTTGCG	CGCCAGC-GCAAGCCGCG CTAACA	*trpE*	(2)
ATTTTA	CCTAACCGGATGAAACAT CCAAAT	*pcaD*	This work
ATTGAC	GCTTGCG-AAGAACAGGC GCTTAT	*pcaI*	This work

Figure 2.4 Analysis of the *pcaI* promoter in *S. meliloti.* (A) RNA isolated from *S. meliloti* wild-type derivative strain Rm1021 was used in primer extension reactions. Lane 1: No extension products were observed using RNA obtained from cells grown in the absence of protocatechuate. Lane 2: A single extension product was observed using RNA derived of cells grown in the presence of protocatechuate. The arrowheads indicate the extension product and identify the corresponding nucleotide (right and left sides respectively). (B) The *pcaI* promoter region, including the inferred -10 and -35 regions (underlined). The transcriptional start site is identified (in bold, enlarged font) and the start codon is also indicated (enlarged font). (C) Comparison of the *pcaI* promoter with other previously determined promoters in *S. meliloti.*

CHAPTER THREE

Binding site determinants for the LysR-type transcriptional regulator PcaQ in the legume endosymbiont *Sinorhizobium meliloti*

Preface

This chapter describes a study of the transcriptional regulator PcaQ, as relating to the regulation of the *pcaDCHGB* operon in *S. meliloti*. Michelle Anstey created several genetic constructs used in this study, including the expression vector (pTH1979), used for the overexpression of *pcaQ*. Michelle also helped with the initial purification of PcaQ, and with defining the optimal conditions for the electrophoretic mobility shift assays. These experiments comprised Michelle's undergraduate senior thesis project, and were performed under my direct supervision. I repeated these initial experiments, and all data included in this chapter are my own, including the mobility shift assays (Figures 2, 3, and 5). As primary author, I wrote the manuscript submitted for publication, with editing by Turlough Finan. This chapter has been published in the Journal of Bacteriology, and is reprinted with permission from the American Society for Microbiology (license number 2082190622878).

3.1 Abstract

LysR-type transcriptional regulators represent one of the largest groups of prokaryotic regulators described to date. In the Gram-negative legume endosymbiont *Sinorhizobium meliloti*, enzymes involved in the protocatechuate branch of the β-ketoadipate pathway are encoded within the *pcaDCHGB* operon, which is subject to regulation by the LysR-type protein PcaQ. In this work, purified PcaQ is shown to bind strongly (K_D: 0.54 nM) to a region -78 to -45 upstream of the *pcaD* transcriptional start site. Within this region, we have defined a PcaQ-binding site of dyad symmetry that is required for regulation of *pcaD* expression *in vivo* and for binding of PcaQ *in vitro*. We also demonstrate that PcaQ participates in negative auto-regulation by monitoring expression of *pcaQ* via a transcriptional fusion to *lacZ*. Although *pcaQ* homologues are present in many α-proteobacteria, this work details the first reported purification of this regulator, as well as a characterization of its binding site, which is conserved in *Agrobacterium tumefaciens*, *Rhizobium leguminosarum*, *R. etli*, and *Mesorhizobium loti*.

3.2 Introduction

In a soil environment, plant-derived aromatic acids represent significant carbon and energy sources. The first step in the metabolism of these compounds involves their conversion into either protocatechuate or catechol, which are subsequently metabolized to tricarboxylic acid intermediates via the β-ketoadipate pathway (18). This metabolic pathway has been documented in many members of the family *Rhizobiaceae* (23, 35, 38-40), suggesting that the β-ketoadipate pathway is important for survival in these soil-dwelling microorganisms.

Sinorhizobium meliloti is a gram-negative, soil-dwelling bacterium that participates in a symbiotic relationship with the legume alfalfa through the establishment of nitrogen fixing root nodules. Enzymes involved in the protocatechuate branch of the β-ketoadipate pathway in *S. meliloti* are encoded within *pcaDCHGB* and *pcaIJF* operons, which are subject to regulation by products encoded by *pcaQ* and *pcaR*, respectively (23). The regulator encoded by *pcaQ* is a member of the LysR-type superfamily of transcriptional regulators (LTTRs) and PcaQ homologues are present in many species of α-proteobacteria (3, 4, 7, 23, 34, 36, 37).

LysR-type regulators comprise one of the largest groups of prokaryotic transcriptional regulators characterized to date; these proteins regulate a diverse range of regulons, including genes whose products are involved in nitrogen and carbon fixation, biofilm formation, oxidative stress response, bacterial virulence, and the catabolism of various compounds including aromatic acids (10, 16, 19, 22, 23, 28, 34, 46, 47, 51, 55, 57). LTTR proteins consist of a conserved helix-turn-helix DNA binding motif located at the N-terminal portion of the polypeptide, whereas the C-terminus includes an inducer-binding site. As a general rule, LTTRs act as transcriptional activators by inducing expression of a target gene(s) upon interaction with a coeffector molecule, although there are also reports of these proteins acting as repressors (9, 19). As transcriptional activators, LTTRs typically associate with two distinct binding sites (47). A recognition-binding site (RBS) is often positioned upstream of the target gene's promoter, and may

be sufficient to elicit regulator binding even in the absence of a coinducer. Interaction of the LTTR with an activation-binding site (ABS), located near the -35 regulatory region of the target gene, generally occurs in the presence of a coeffector and is required to induce target gene expression via interaction with RNA polymerase.

In this report, we describe data from electrophoretic mobility shift assays which indicate that PcaQ binds with high affinity to sequence upstream of the *pcaD* promoter. In DNAseI footprinting assays, a region protected by PcaQ was located -78 to -45 upstream of the *pcaD* transcriptional start site. Employing site directed mutagenesis, we demonstrate that PcaQ recognition and binding of the *pcaD* regulatory region involves a sequence of partial dyad symmetry (5' ATAACC-N_4-GGTTAA 3'), as determined by both *in vitro* binding assays and *in vivo* expression analyses. By measuring expression of a *pcaQ::lacZ* fusion in *S. meliloti*, we also show that PcaQ participates in negative auto-regulation, possibly through interaction with the same binding site necessary for regulation of *pcaD* expression.

3.3 Materials and Methods

Bacterial strains and culture conditions

The bacterial strains and plasmids used throughout this study are described in Table 3.1. *Escherichia coli* strains were grown aerobically at 37°C in LB broth; *S. meliloti* strains were grown aerobically at 30°C in LB broth supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ (LBmc) or M9-minimal medium (Difco). M9-minimal medium was supplemented with 1.0 mM $MgSO_4$, 0.25 mM $CaCl_2$, 1 μg/mL D-biotin, and 10 ng/mL $CoCl_2$. Carbon sources were added to M9-minimal medium as follows: 5 mM protocatechuate (Sigma-Aldrich) or (v/v) 0.5% glycerol. For *E. coli*, antibiotics were added at the following concentrations (μg/mL): ampicillin (Ap): 50; chloramphenicol (Cm): 20; gentamicin (Gm): 10. For *S. meliloti*, the following concentrations of

antibiotics were used (μg/mL): streptomycin (Sm): 200; spectinomycin (Sp): 200; gentamicin: 60; rifampicin (Rif): 20.

Overexpression and purification of PcaQ

Using *S. meliloti* Rm1021 genomic DNA as a template, *pcaQ* was PCR amplified using the following primers: (5'-GTGATACATATGATCGACGCTCGCGTTAAG-3'; 5'-ACTCGAGGGCCGTCCTCTTTGCTTCC-3'), and cloned into the expression vector pET-21a (Novagen) via *Nde*I and *Xho*I restriction sites, to create a C-terminal fusion with a His·Tag. The location of the hexa-histidine tag at the carboxyl terminus of the protein was chosen to minimize any influence the foreign tag may exert upon the ability of PcaQ to bind DNA, as has been demonstrated with other LTTRs (5, 30). DNA sequencing of the cloned regions confirmed the absence of mutations within the 937 bp *pcaQ* gene fragment, and the designated plasmid, pTH1979, was then transformed into *E. coli* BL21 (DE3) pLysS (Stratagene).

E. coli strain M924 (BL21 (DE3) pLysS (pTH1979)) was subcultured into 100 to 200 mL of pre-warmed LB broth (with appropriate antibiotics) to an optical density at 600 nm (OD_{600}) of approximately 0.2 and grown to an OD_{600} ~ 0.6 to 0.8. Expression of *pcaQ* was induced with 0.1 mM isopropyl-β-D-thiogalactoside (IPTG) for 3 hours at 37°C. Cells were then centrifuged for 20 minutes at 9,300 x *g* and pellets were stored at -20°C prior to lysis.

Pellets were thawed on ice, resuspended into 4 mL buffer (50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole; pH 8.0), and sonicated with a Sonifier model 350 (Branson Sonic Power Co.) for five 10s-bursts on ice. The lysate was incubated with 2 mL Ni-NTA agarose beads (Qiagen) for one hour at 4°C prior to loading the chromatography column (BIO-RAD). The column was washed with 20 mM to 100 mM imidazole in buffer (50 mM NaH_2PO_4 and 300 mM NaCl; pH 8.0) and protein was eluted with 250 mM imidazole.

For FPLC size exclusion chromatography, eluant from the Ni-NTA column was dialyzed into a column buffer (100 mM Tris-HCl, 300 mM NaCl, 1 mM DTT and 1 mM

EDTA; pH 8.0). Six hundred microlitres of sample was loaded onto the Sephadex G-200 size exclusion column and eluted at 0.2 mL/min. Fractions of volume 1 mL were collected and the purified protein was visualized via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Pooled fractions were dialyzed into a buffer (50% (v/v) glycerol, 20 mM Tris-HCl, 50 mM KCl; pH 8.0) and stored at -20°C.

The molecular mass of PcaQ was determined using a Sephadex G-200 size exclusion column calibrated using the following molecular mass standards: catalase (MW, 232 kDa), aldolase (158 kDa), albumin (67 kDa), ovalbumin (43 kDa), chymotrypsin A (25 kDa) and ribonuclease A (13 kDa). Blue Dextran 2000 was used to calculate the void volume of the column. Molecular mass standards were run in sets of three at a final volume of 600 μL.

Electrophoretic Mobility Shift Assays

PCR amplified probes and annealed synthetic oligonucleotides were purified via polyacrylamide or agarose gels prior to labeling. For K_D determination and experiments comparing PcaQ binding activity between wild-type and mutant binding sites, assays involved 194 bp PCR amplified products spanning the intergenic region, using primers: (102: 5' CGCTCTAGACAAATGTCTGCAGATGG 3') and (103: 5' ATTCTAGAGATAGTGAATCGTGACGTCG 3'), with either pTH2209 or a derivative serving as a template. Probes were 5' end-labeled using γ^{32}P[ATP] (Perkin Elmer) and T_4 polynucleotide kinase (New England Biolabs). Binding reactions (15 μL total) contained 10,000 c.p.m. of one of the labeled DNA probes, and 500 ng herring sperm non-specific competitor DNA, in a buffer containing 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 200 ng/mL BSA and (v/v) 4% glycerol. Purified PcaQ was added as specified, and the binding reaction was incubated on ice for 15 minutes and then incubated at room temperature for 25 minutes. Reactions were loaded onto a 6% non-denaturing polyacrylamide gel (200 bp probes) or 8% non-denaturing polyacrylamide gel (45 to 78 bp oligomers). The gels were run at 150 Volts for 5 minutes and 80 Volts at room temperature for the remaining time. Gels were dried

onto filter paper, exposed on a phosphoimager screen, and visualized. To determine K_D, mobility assays were performed using eight concentrations of purified PcaQ (ranging from 0 to approximately 25 nM) in triplicate, for a total of twenty-four assays; this experiment was performed twice.

DNAseI Footprinting Reactions

Primer 103 was 5' end-labeled using γ^{32}P[ATP] (Perkin Elmer) and T_4 polynucleotide kinase (New England Biolabs) and purified using QIAquick Nucleotide Removal Kit (QIAGEN). Labeled primer was incorporated into the probe via PCR amplification using pFL2211 as a template, and unlabeled primer 102. PCR product was resolved upon an 8% polyacrylamide gel and purified product was electro-eluted from the gel, and further purified using QIAquick PCR Purification Kit (QIAGEN). Binding reactions were performed using ~45,000 c.p.m, 100 ng polydI/dC (as non-specific competitor DNA), and binding buffer as described for electrophoretic mobility shift assays. 0.01 Units DNAseI (Invitrogen) was incubated in the presence or absence of 100 to 500 ng purified PcaQ (approximately 15 to 70 nM) for 2 minutes at room temperature (22°C). Reactions were stopped with the addition of 10 volumes (500 µL total) buffer PN (QIAGEN), and DNAseI digested DNA was immediately purified via QIAquick Nucleotide Removal Kit, according to manufacturer's directions. DNA was eluted using nuclease-free water (50 µL), and samples were concentrated via Eppendorf Vacufuge at 30°C for 20 minutes. Pellets were resuspended in loading buffer and samples were resolved upon an 8% polyacrylamide (7 M urea) sequencing gel. Sequencing reactions were generated using the primer 103 via the Sequenase Version 2 DNA Sequencing kit (USB), with plasmid pFL2211 as a template for sequencing reactions.

Site directed mutagenesis

Platinum *Pfx* DNA polymerase (Invitrogen) was used according to supplier's protocol. Briefly, reactions were set up as follows: 0.3 mM each dNTP (10 mM stock), 1 mM $MgSO_4$, 0.3 µM primer, 2X *Pfx* amplification buffer, 50 ng template DNA

(pTH2209), 2U Platinum *Pfx* polymerase. To provide a template for site directed mutagenesis experiments, plasmid pTH2209 was constructed as follows. Primers 102 and 103 were used to PCR amplify a 194 bp product spanning the intergenic region (-122 to +55), which was cloned into pUCP30T via *Xba*I. Following amplification, 20 U *Dpn*I (New England Biolabs) was added directly to the amplification mixture, which was incubated for 2 hours at 37°C. Product DNA was purified using QIAquick spin columns (Qiagen) and transformed into competent *E. coli* DH5α; transformants were selected for Gm^r and all plasmids were sequenced to verify the presence of site directed mutations (MOBIXlab, McMaster University, Hamilton, Ontario, Canada).

To construct the *gfpuv* reporter fusions, the pUCP30T derivatives (each bearing a site directed mutation) were used as template DNA for PCR amplification using primers (5' TGTCTGCAGATAGTGAATCGTGACGTCG 3') and (5' GTATCTAGACAGATGGCGAAACTTAACGC 3'). PCR products were cloned into pOT1 (2) via *Xba*I and *Pst*I sites using standard molecular biology techniques to yield a *pcaD*::*gfp* fusion. All plasmids were sequenced to confirm the presence of mutations, and the orientation of transcriptional fusion with *gfpuv*.

Construction of an S. meliloti pcaQ::Ω strain

An *S. meliloti pcaQ* mutant was required for reporter enzyme assays (GfpUV), however the only strain available encoded gentamicin resistance, and thus it was necessary to construct an additional strain which lacked resistance to this antibiotic. Accordingly, an Ω cassette encoding streptomycin/spectinomycin resistance from pHP45Ω (42) was introduced into a *Pst*I site located 35 bp downstream of the predicted translational start site of *pcaQ* as follows. The Ω cassette was PCR amplified and cloned into plasmid pTH1577 via *Pst*I, yielding pTH1958. pTH1577 has been previously described (23), and consists of an 893 bp fragment centered upon the *Pst*I site located within *pcaQ*, cloned into the suicide vector pJQ200 uc1 (43). pTH1958 was transferred via conjugation into rifampicin resistant *S. meliloti* strain Rm5000, and recombinants were selected by plating onto LB agar supplemented with rifampicin and gentamicin. A

single Gm^r colony was grown overnight in LBmc in the absence of antibiotic selection and the resulting culture was plated onto LB agar supplemented with 5% sucrose and spectinomycin. Sp^r/$sucrose^r$ colonies were patched onto LB agar plus gentamicin to confirm excision of the Gm^r suicide plasmid. The *pcaQ*::Ω allele was transferred from Rm5000 to RmP110 via transduction by selecting for Sp^r transductants. The resulting RmP110 *pcaQ*::Ω strain was designated RmP1676. As expected, RmP1676 exhibited a Pca^- phenotype when plated upon M9-minimal medium with protocatechuate as a sole carbon source.

GfpUV assays

Plasmids were transferred by conjugation from donor *E. coli* DH5α into *S. meliloti* recipient strains RmP110 and RmP1676 using a tri-parental mating with *E. coli* helper strain MT616. Streak purified transconjugants were used in GfpUV assays as follows. Overnight *S. meliloti* LBmc cultures were washed and sub-cultured into M9-minimal salts with carbon source and gentamicin as specified to achieve OD_{600} ~ 0.2 to 0.5. Cultures were incubated at 30°C for four to six hours before harvesting for enzyme assays. GfpUV fluorescence was quantified by dividing each emission output by its respective OD_{600}; all assays were performed in triplicate for each experiment, and experiments were performed a minimum of three times.

β-Galactosidase assays

Plasmids were transferred by conjugation into *S. meliloti* recipient strains RmG212 (Rm1021 *lac*) and RmP134 (Rm1021 *lac pcaQ*::Ω) (23) and transconjugant colonies were streak-purified prior to use in enzyme assays. β-galactosidase enzyme assays were performed as previously described (23) and enzyme activities were calculated based upon Miller (27).

3.4 Results

Overexpression and purification of PcaQ

The LysR-type transcriptional regulator PcaQ is known through genetic analyses to participate in the regulation of expression of the *pcaDCHGB* operons in *S. meliloti* and *A. tumefaciens* (23, 34). The *pcaQ* gene is transcribed divergently from the *pcaDCHGB* operon (Figure 3.1A), and we recently mapped the *pcaD* transcriptional start sites to G and C residues located 14 and 15 nucleotides upstream of the *pcaD* start codon (23). For simplicity however, nucleotide positions throughout this work are reported with regard to the upstream transcriptional start site identified (i.e., C residue located 15 nucleotides upstream of *pcaD*). To facilitate the analysis of the *pcaD* promoter, we wished to purify PcaQ through the use of a hexa-histidine tag located at the C-terminus of the protein (Figure 3.1B). Accordingly, *pcaQ* was overexpressed in *E. coli* and the purified protein eluted from a size exclusion chromatography column as a 149 kDa protein peak, indicating that this regulator exists in solution as a tetramer (Figure 3.1C). We have determined that *pcaQ·His* complements the regulatory phenotype of a *pcaQ*-minus strain of *S. meliloti in trans*, thus indicating that PcaQ·His is capable of activating transcription *in vivo* (data not shown).

Purified PcaQ binds upstream of the pcaD promoter

PcaQ regulates expression of the *pcaDCHGB* promoter in *S. meliloti* and *A. tumefaciens*, inducing expression of the operon in cells when grown in the presence of protocatechuate (23, 34). To determine whether PcaQ recognizes and binds the *pcaD* promoter region, electrophoretic mobility shift assays were performed using purified PcaQ. A 194 bp probe spanning the *pcaDQ* intergenic region (-122 to +55) was shown to bind PcaQ in the absence of a co-inducing metabolite (Figure 3.2). The apparent equilibrium dissociation constant (K_D) of PcaQ was determined as 0.54nM for the *pcaD* promoter region, indicating an interaction of high affinity even in the absence of a coeffector molecule. As a negative control, a probe consisting of the *pcaIJF* upstream

region was also tested for PcaQ binding; expression of these genes is not regulated by PcaQ and thus the upstream sequence should not be recognized by the regulator (23). As expected, the addition of up to 50 ng of PcaQ (24 nM) did not result in a shift of the *pcaIJF* promoter probe (data not shown), confirming that the interaction observed between PcaQ and the *pcaD* promoter region is specific.

DNAseI footprinting analysis of PcaQ binding

LysR-type transcriptional regulators are known to recognize sequences established upon a $TN_{11}A$ core motif (15, 47). Seven such motifs are present within the 94 bp *pcaDQ* intergenic region and it was therefore necessary to perform DNAseI footprinting experiments to identify specific sequences involved in PcaQ binding (Figure 3.3A). PcaQ protected nucleotides located upstream of the *pcaD* promoter, with a footprint located at approximately –78 to –45 bp, with respect to the *pcaD* transcriptional start site (Figure 3.3C). These results are in agreement with previously published descriptions of other LysR-type regulators, which often bind upstream of the promoters subject to their regulation (14, 16, 17, 19, 21, 22, 47, 49, 51). To confirm that the area protected by PcaQ was sufficient for PcaQ recognition and binding, mobility shift assays were performed using oligonucleotide probes including or excluding the protected region (Figure 3.3B). These assays confirmed that the region -87 to -45 (Figure 3.3B; lanes 1 to 4) is necessary and sufficient for PcaQ binding *in vitro*, consistent with results obtained from the DNAseI footprinting experiments. As well, a probe extending -83 to -6 (lanes 9 to 12) was shifted by the addition of purified regulator. In contrast, binding was not detected using a probe extending -51 to +6 (lanes 5 to 8) in the presence of up to 48 nM PcaQ, suggesting that the primary (or only) binding site is located -83 to -51 upstream of the *pcaD* transcriptional start site.

Identification of a motif that is required for PcaQ binding in vivo

It was previously demonstrated that an *A. tumefaciens pcaD::lacZ* transcriptional fusion exhibits protocatechuate-inducible expression in several related species, including

S. meliloti (37). Accordingly, it is thus likely that PcaQ binding sites are conserved between these species, and an alignment of the intergenic regions of *S. meliloti*, *R. leguminosarum*, *A. tumefaciens*, *M. loti*, and *R. etli* was performed to facilitate the identification of possible conserved binding sites within the footprinted region (Figure 3.4). Examination of the alignment reveals two distinct areas that have been conserved in all species; one of these encompasses and surrounds the -35 hexanucleotide promoter region associated with the *S. meliloti pcaD* gene and is located outside the region identified as being necessary and sufficient for PcaQ binding, as determined by our *in vitro* analyses. The second conserved region is comprised of two halves of an AT-rich sequence of partial dyad symmetry (5' ATAACC-N_4-GGTTAA 3') that includes a single '$TN_{11}A$' motif. This region is located upstream of the predicted -35 hexanucleotide region (-72 to -57), and falls within the region protected from DNAseI digestion.

The upstream region was targeted for site directed mutagenesis experiments to confirm whether the identified motif is required for PcaQ regulation *in vivo*. Highly conserved positions within the putative PcaQ binding site were systematically mutated through the introduction of single point mutations at each position (Figure 3.4; enclosed in box); A and T residues were replaced with G, whereas G residues were substituted with C. The wild-type and mutant binding sites were cloned as 184 bp fragments (-112 to +54) into the broad host range reporter plasmid pOT1 to generate *pcaD::gfpuv* transcriptional fusions. These plasmids were conjugated into wild-type and PcaQ-minus *S. meliloti* strains RmP110 and RmP1676, respectively, and GfpUV assays were performed to determine whether expression of *pcaD::gfpuv* was induced by growth in the presence of protocatechuate (Table 3.2).

Expression of the *pcaD* promoter (as determined by GfpUV specific activity) was induced in cells of wild-type *S. meliloti* strain RmP110 grown in the presence of protocatechuate (Table 3.2; pTH2276). Inducible expression of the *pcaD* promoter was reduced upon mutation of any one of the highly conserved nucleotides targeted for mutagenesis (as underlined) within the putative PcaQ binding site (5' <u>ATAA</u>CC-N_4-<u>GGTTAA</u> 3') (Figure 3.4). Single point mutations introduced at positions A(-72)G

(pTH2396) and A(-58)G (pTH2298) yielded particularly severe reductions in inducible expression, resulting in 87 and 90 percent decreases in activation respectively, as compared to the wild-type regulatory region. In contrast, mutations introduced into other positions (such as G(-62)C and A(-57)G) did not have such a severe impact upon the regulated expression of *pcaD::gfpuv*; in these instances, expression is induced by growth with protocatechuate at levels approaching that exhibited in the wild-type control. As expected, *pcaD::gfpuv* expression in all instances was not induced in a PcaQ-minus background, despite the addition of protocatechuate to growth medium (data not shown).

An additional '$TN_{11}A$' motif was identified -63 to -51 from the transcriptional start site of the *S. meliloti pcaD* promoter. The simultaneous introduction of two mutations within the motif (T(-63)G and A(-51)G; pTH2282) yielded a protocatechuate-inducible expression comparable to that observed in wild-type sequence, indicating that these nucleotides are not important with respect to PcaQ-mediated regulation. Consistent with these data is the lack of conservation of this region between species (Figure 3.4).

Identification of a motif that is required for PcaQ binding in vitro

We have shown *in vivo* that the introduction of mutations within the putative PcaQ binding site (5'ATAACC-N_4-GGTTAA 3') negatively affects the regulation of *pcaD* expression in *S. meliloti.* One mechanism by which this may occur is that mutations within this region impede PcaQ recognition and/or binding of the regulatory site; it is also possible that these mutations permit binding of the regulator, but inhibit a conformational change in PcaQ that is necessary to elicit a regulated response (i.e., interaction with RNA polymerase). To examine whether mutations within this site also affect PcaQ binding, mobility shift assays were performed to compare binding of PcaQ to wild-type and mutant regulatory sequences. Representative data are shown in Figure 3.5. Densitometry analyses indicate that in all cases, PcaQ binding is reduced at least 2-fold by the introduction of mutations within the putative binding sequence, as compared to the wild-type sequence (Table 3.3). This is most evident with respect to mutations at

positions A(-72)G and A(-58)G, where PcaQ binding to the mutant binding site is decreased more than 5-fold compared to wild-type.

PcaQ participates in negative auto-regulation

To determine whether PcaQ auto-regulates its own expression in *S. meliloti*, the *pcaDQ* intergenic region was cloned into the replicating plasmid pMP220 (50), with the *pcaQ* promoter in the same orientation as the promoterless reporter gene *lacZ*. This plasmid, pTH467, was then transferred via conjugation into RmG212 (Rm1021 *lac* mutant) and RmP134 (*pcaQ*::Ω derivative of RmG212) to examine expression of *pcaQ* (as measured by β-galactosidase activity). Reporter enzyme assays were performed using cells grown in 0.5% glycerol with and without 5 mM protocatechuate (Table 3.4). Expression of *pcaQ* was increased approximately 5-fold in the PcaQ-minus strain as compared to wild-type *S. meliloti*, indicating that *pcaQ* expression is repressed by its encoded product under the conditions tested.

3.5 Discussion

Despite a widespread occurrence in members of the class α-proteobacteria, this work describes the first purification of a member of the PcaQ family and the characterization of a PcaQ binding site. As a member of the superfamily of LysR-type transcriptional regulators, PcaQ exhibits many traits common to this group of proteins. FPLC size exclusion chromatography performed upon partially purified PcaQ (Figure 3.1) indicates that this protein likely exists as a tetramer in solution, as has been reported for other LTTRs (26, 29, 47, 48). In particular, the crystal structure of the full-length LysR-type regulator CbnR indicates that this protein exists as a tetramer formed by the association of a dimer of dimers, and it is believed that this corresponds to the biologically active form of this protein (29).

Electrophoretic mobility shift assays demonstrated that PcaQ binds to DNA fragments carrying the –87 to –51 nucleotide region upstream of the *pcaDCHGB* operon

(Figures 3.2 and 3.3), in the absence of any co-inducing metabolite. Other LTTRs have also been shown to bind DNA in the absence of any co-inducing molecule(s) (14, 41, 47, 49). The apparent dissociation constant for PcaQ under these conditions was determined to be K_D: 0.54 x 10^{-9} M, indicating an interaction of high affinity. This dissociation constant is comparable to that obtained from other LTTRs, with reported values ranging from at least 7.0 x 10^{-11} M (41) to 0.9 x 10^{-6} M (53).

DNAseI footprinting experiments revealed that PcaQ strongly protected a region -78 to -45 relative to the *pcaD* transcriptional start site. This location upstream of the -35 promoter region is consistent with that observed for several LysR-type proteins (14, 16, 19, 21, 22, 47, 49, 51). Examination of Figure 3.3A also revealed protected sequence spanning approximately -40 to –21. This footprint was more difficult to discern, however its presence was observed in multiple independent assays. In *A. tumefaciens* and *R. leguminosarum*, OccR and NodD have been demonstrated to bind DNA as tetramers, respectively (1, 12). The length of the protected sequence observed in our assays suggests that PcaQ may bind upstream of *pcaD* as a tetramer however we have not confirmed directly that this is the case. A probe encompassing -51 to +6 was not shifted by PcaQ in mobility shift assays (Figure 3.3B; lanes 5 to 8), suggesting that the region -40 to -21 is not sufficient to permit PcaQ binding. In contrast, the upstream region alone (-87 to -45) permitted PcaQ binding in the absence of inducer (Figure 3.3B; lanes 1 to 4).

In order to document whether conserved nucleotides within the region protected by PcaQ are essential for the regulation of *pcaD* expression, site directed mutagenesis experiments were performed. Expression of the *pcaD* promoter was then monitored in *S. meliloti in vivo* through the use of the reporter protein GfpUV. It is important to note that although the absolute levels of expression of each site directed mutant have been reduced with respect to wild-type sequence, most mutant sequences retained the ability of protocatechuate to up-regulate expression of the *pcaD::gfpuv* fusion, albeit to a lesser extent than that observed in wild-type sequence (Table 3.2). That this residual regulated response was nonetheless dependent upon PcaQ was evident when examining expression of each fusion in *pcaQ* mutant strain RmP1676; without exception, expression of each

fusion in this genetic background remained at a comparable low level that was unaltered by the presence of protocatechuate (data not shown). At present, we do not have an explanation to account for the observation that the basal level of *pcaD* expression (i.e., uninduced) in many mutants is decreased in comparison to wild-type levels.

Similar studies involving mutagenesis of LTTR binding sites have been performed and often these mutations result in reduced transcription (6, 17, 22, 49). For example, the introduction of point mutations within a putative MetR binding site reduced expression of a *metH::lacZ* fusion (6). However, gel mobility shift assays revealed that only a subset of these mutations likewise affected MetR binding to the regulatory site. In the case of PcaQ, our assays revealed that although all mutations affecting *pcaD::gfpuv* expression *in vivo* likewise reduced PcaQ binding ability (as determined by densitometry analysis; Table 3.3), only two mutations (A(-72)G and A(-58)G) reduced binding >5-fold. It is worth noting that these same mutations had the strongest negative impacts upon *pcaD::gfp* expression, correlating the ability of PcaQ to bind *in vitro* with transcription activation. As with Byerly et al., the A(-72)G and A(-58)G mutations correspond to positions located at the outer edge of the palindromic binding site (5'ATAACC-N_4-GGTTAA 3'), possibly reflecting the relative importance of outer positions with respect to LTTR binding.

Systematic mutagenesis of a binding site recognized by the LysR-type protein AphB in *Vibrio cholerae* revealed that the promoter proximal dyad arm may be more important than the distal arm for the activation of gene expression (22). Based upon our analyses, positions in both dyad arms of the PcaQ binding site contribute similarly with regard to the transcriptional activation of *pcaD* expression (Table 3.2). It is possible that this inconsistency may reflect subtle variations in the manner by which these LysR-type proteins regulate gene expression, however differences in experimental design (such as the type of nucleotide substitutions used in each study) may also account for this discrepancy.

In *A. tumefaciens*, β-ketoadipate pathway intermediates β-carboxy-*cis,cis*-muconate and γ-carboxymuconolactone act as inducing agents in the presence of PcaQ,

resulting in an induction of *pcaDCHGB* expression (34, 36). The identity of comparable co-inducing metabolites in *S. meliloti* is unknown however it is likely that the same two pathway intermediates serve a similar function in *S. meliloti* (23). β-carboxy-*cis,cis*-muconate and γ-carboxymuconolactone are unstable compounds (31, 32), and the synthesis of these metabolites requires the enzymatic activities of protocatechuate 3,4-dioxygenase (PcaHG) and 3-carboxymuconate cycloisomerase (PcaB), respectively. We attempted but were unable to obtain a DNAseI footprint with purified PcaQ in the presence of β-carboxy-*cis,cis*-muconate (data not shown). It is intriguing that multiple co-effector molecules have been identified to act in concert with PcaQ, as a recent report has been published describing the synergistic effect exerted upon BenM-mediated transcriptional activation by the metabolites benzoate and muconate (5).

In the absence of data including DNA binding assays performed with a co-effector, it is difficult to propose a comprehensive model of PcaQ regulation. Our results indicate that the region -78 to -45 upstream of *pcaD* in *S. meliloti* encompasses a high affinity binding site (or recognition-binding site; RBS), and a second region (-40 to –21) may include a secondary interaction site (or activation-binding site; ABS). The ability of PcaQ to bind a probe containing only the putative RBS (spanning -87 to -45), but not a probe containing solely the putative ABS (–51 to +6) is not inconsistent with this hypothesis. For example, OccR interacts with sequence flanking the -35 regulatory region associated with *occQ*; these binding sites are required for modulation of DNA bending in response to the ligand octopine, and do not contribute to the high affinity binding of this regulator (56). It is worth noting the presence of highly conserved sequence on both sides of the -35 region of *pcaD* (Figure 3.4); these conserved residues may in fact be part of the ABS.

The gene encoding *pcaQ* is adjacent to, and divergently transcribed from, the *pcaDCHGB* operon. This spatial organization is a common feature observed in LTTRs (20, 21, 47, 53) and it has been proposed that this permits a transcriptional coupling of the divergent promoters via DNA supercoiling introduced by the transcribing RNA polymerase, as modeled by the *ilvYC* operon in *E. coli* (44). As well, the organization of

pcaQ and *pcaDCHGB* as divergent transcriptional units raises the possibility of PcaQ influencing the expression of both sets of genes by occupying the same binding site(s). While many LTTRs repress self-expression (9, 22, 33, 36, 47, 49), this is not always the case as these proteins may also act as either activators (17, 19) or fail to influence the expression of their own gene (45, 47). In *S. meliloti,* PcaQ represses its own expression, as is reflected by a 5-fold increase in *pcaQ* expression in a PcaQ-minus strain (Table 3.4), which is within the range reported for other LTTRs in comparable studies (22, 47). To further examine the regulation of *pcaQ* expression, we attempted to identify the *pcaQ* transcriptional start site via primer extension using mRNA isolated from wild-type and PcaQ-minus strains of *S. meliloti*, however we were unable to obtain an extension product. Similar problems have been reported in mapping the start sites of other LTTRs (5, 11, 22), and this is likely due to the relatively low expression of the encoding gene. In the absence of an identified transcriptional start site, we nonetheless note that the close proximity of the PcaQ binding site to the predicted *pcaQ* translational start codon (separated by 6 nucleotides) suggests that binding of the LysR protein to this site may prevent efficient transcription of *pcaQ*. While effecting auto-repression, LTTRs often bind to sites located within their own coding sequence (22, 33, 54). We examined the nucleotide sequence of *pcaQ* for candidate binding sites located within the gene, but we were unable to identify any additional sites with strong similarity to the consensus PcaQ binding site. Intriguingly, a scan of the entire *S. meliloti* genome with a consensus PcaQ binding site results in two positive hits; one of these is located upstream of *pcaD* and corresponds to the binding site characterized in this study. The second site is located upstream of an ABC-type transport system whose expression is induced by protocatechuate (24). It seems likely that this transport system is dedicated to the uptake of protocatechuate, and we are pursuing studies to determine whether this system in fact transports protocatechuate and whether expression of this system is regulated by PcaQ.

3.6 References

1. **Akakura, R. and S. C. Winans**. 2002. Mutations in the *occQ* operator that decrease OccR-induced DNA bending do not cause constitutive promoter activity. J. Biol. Chem. **277:**15773-15780.

2. **Allaway, D., N. A. Schofield, M. E. Leonard, L. Gilardoni, T. M. Finan, and P. S. Poole**. 2001. Use of differential fluorescence induction and optical trapping to isolate environmentally induced genes. Environ. Microbiol. **3:**397-406.

3. **Buchan, A., L. S. Collier, E. L. Neidle, and M. A. Moran**. 2000. Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. Appl. Environ. Microbiol. **66:**4662-4672.

4. **Buchan, A., E. L. Neidle, and M. A. Moran**. 2004. Diverse organization of genes of the beta-ketoadipate pathway in members of the marine *Roseobacter* lineage. Appl. Environ. Microbiol. **70:**1658-1668.

5. **Bundy, B. M., L. S. Collier, T. R. Hoover, and E. L. Neidle**. 2002. Synergistic transcriptional activation by one regulatory protein in response to two metabolites. Proc. Natl. Acad. Sci. USA **99:**7693-7698.

6. **Byerly, K. A., M. L. Urbanowski, and G. V. Stauffer**. 1991. The MetR binding site in the *Salmonella typhimurium metH* gene: DNA sequence constraints on activation. J. Bacteriol. **173:**3547-3553.

7. **Contzen, M. and A. Stolz**. 2000. Characterization of the genes for two protocatechuate 3, 4-dioxygenases from the 4-sulfocatechol-degrading bacterium *Agrobacterium radiobacter* strain S2. J. Bacteriol. **182:**6123-6129.

8. **Cowie, A., J. Cheng, C. D. Sibley, Y. Fong, R. Zaheer, C. L. Patten, R. M. Morton, G. B. Golding, and T. M. Finan**. 2006. An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. Appl. Environ. Microbiol. **72:**7156-7167.

9. **Deghmane, A. E., D. Giorgini, L. Maigre, and M. K. Taha**. 2004. Analysis *in vitro* and *in vivo* of the transcriptional regulator CrgA of *Neisseria meningitidis* upon contact with target cells. Mol. Microbiol. **53:**917-927.

10. **Deghmane, A. E., S. Petit, A. Topilko, Y. Pereira, D. Giorgini, M. Larribe, and M. K. Taha**. 2000. Intimate adhesion of *Neisseria meningitidis* to human epithelial cells is under the control of the *crgA* gene, a novel LysR-type transcriptional regulator. EMBO J. **19:**1068-1078.

11. **Doty, S. L., M. Chang, and E. W. Nester**. 1993. The chromosomal virulence gene, *chvE*, of *Agrobacterium tumefaciens* is regulated by a LysR family member. J. Bacteriol. **175:**7880-7886.

12. **Feng, J., Q. Li, H. L. Hu, X. C. Chen, and G. F. Hong**. 2003. Inactivation of the *nod* box distal half-site allows tetrameric NodD to activate *nodA* transcription in an inducer-independent manner. Nucleic Acids Res. **31:**3143-3156.

13. **Finan, T. M., E. Hartweig, K. LeMieux, K. Bergman, G. C. Walker, and E. R. Signer**. 1984. General transduction in *Rhizobium meliloti*. J. Bacteriol. **159:**120-124.

14. **Fisher, R. F. and S. R. Long**. 1989. DNA footprint analysis of the transcriptional activator proteins NodD1 and NodD3 on inducible *nod* gene promoters. J. Bacteriol. **171:**5492-5502.

15. **Goethals, K., M. Van Montagu, and M. Holsters**. 1992. Conserved motifs in a divergent *nod* box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. Proc. Natl. Acad. Sci. USA **89:**1646-1650.

16. **Goller, C., X. Wang, Y. Itoh, and T. Romeo**. 2006. The cation-responsive protein NhaR of *Escherichia coli* activates *pgaABCD* transcription, required for production of the biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. J. Bacteriol. **188:**8022-8032.

17. **Grob, P., D. Kahn, and D. G. Guiney**. 1997. Mutational characterization of promoter regions recognized by the *Salmonella dublin* virulence plasmid regulatory protein SpvR. J. Bacteriol. **179:**5398-5406.

18. **Harwood, C. S. and R. E. Parales**. 1996. The beta-ketoadipate pathway and the biology of self-identity. Annu. Rev. Microbiol. **50:**553-590.

19. **Heroven, A. K. and P. Dersch**. 2006. RovM, a novel LysR-type regulator of the virulence activator gene *rovA*, controls cell invasion, virulence and motility of *Yersinia pseudotuberculosis*. Mol. Microbiol. **62:**1469-1483.

20. **Ieva, R., C. Alaimo, I. Delany, G. Spohn, R. Rappuoli, and V. Scarlato**. 2005. CrgA is an inducible LysR-type regulator of *Neisseria meningitidis*, acting both as a repressor and as an activator of gene transcription. J. Bacteriol. **187:**3421-3430.

21. **Jones, R. M., B. Britt-Compton, and P. A. Williams**. 2003. The naphthalene catabolic (*nag*) genes of *Ralstonia* sp. strain U2 are an operon that is regulated by NagR, a LysR-type transcriptional regulator. J. Bacteriol. **185:**5847-5853.

22. **Kovacikova, G. and K. Skorupski**. 2002. Binding site requirements of the virulence gene regulator AphB: differential affinities for the *Vibrio cholerae* classical and El Tor *tcpPH* promoters. Mol. Microbiol. **44:**533-547.

23. **MacLean, A. M., G. MacPherson, P. Aneja, and T. M. Finan**. 2006. Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. Appl. Environ. Microbiol. **72:**5403-5413.

24. **Mauchline, T. H., J. E. Fowler, A. K. East, A. L. Sartor, R. Zaheer, A. H. Hosie, P. S. Poole, and T. M. Finan**. 2006. Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proc. Natl. Acad. Sci. USA **103:**17933-17938.

25. **Meade, H. M., S. R. Long, G. B. Ruvkun, S. E. Brown, 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.

26. **Miller, B. E. and N. M. Kredich**. 1987. Purification of the *cysB* protein from *Salmonella typhimurium*. J. Biol. Chem. **262:**6006-6009.

27. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

28. **Mulligan, J. T. and S. R. Long**. 1985. Induction of *Rhizobium meliloti nodC* expression by plant exudate requires *nodD*. Proc. Natl. Acad. Sci. USA **82:**6609-6613.

29. **Muraoka, S., R. Okumura, N. Ogawa, T. Nonaka, K. Miyashita, and T. Senda**. 2003. Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. J. Mol. Biol. **328:**555-566.

30. **Ogawa, N., S. M. McFall, T. J. Klem, K. Miyashita, and A. M. Chakrabarty**. 1999. Transcriptional activation of the chlorocatechol degradative genes of *Ralstonia eutropha* NH9. J. Bacteriol. **181:**6697-6705.

31. **Ornston, L. N. and D. Parke**. 1977. The evolution of induction mechanisms in bacteria: insights derived from the study of the beta-ketoadipate pathway, p. 209-262. Academic Press, New York, NY.

32. **Ornston, L. N. and R. Y. Stanier**. 1966. The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida*. J. Biol. Chem. **241:**3776-3786.

33. **Ostrowski, J. and N. M. Kredich**. 1991. Negative autoregulation of *cysB* in *Salmonella typhimurium*: in vitro interactions of CysB protein with the *cysB* promoter. J. Bacteriol. **173:**2212-2218.

34. **Parke, D.** 1993. Positive regulation of phenolic catabolism in *Agrobacterium tumefaciens* by the *pcaQ* gene in response to beta-carboxy-*cis*,*cis*-muconate. J. Bacteriol. **175:**3529-3535.

35. **Parke, D.** 1995. Supraoperonic clustering of *pca* genes for catabolism of the phenolic compound protocatechuate in *Agrobacterium tumefaciens*. J. Bacteriol. **177:**3808-3817.

36. **Parke, D.** 1996. Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds, from *Agrobacterium tumefaciens*. J. Bacteriol. **178:**266-272.

37. **Parke, D.** 1996. Conservation of PcaQ, a transcriptional activator of *pca* genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. J. Bacteriol. **178:**3671-3675.

38. **Parke, D. and L. N. Ornston**. 1984. Nutritional diversity of *Rhizobiaceae* revealed by auxanography. J. Gen. Microbiol. **130:**1743-1750.

39. **Parke, D. and L. N. Ornston**. 1986. Enzymes of the beta-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium* spp. and constitutive in *Bradyrhizobium* spp. J. Bacteriol. **165:**288-292.

40. **Parke, D., F. Rynne, and A. Glenn**. 1991. Regulation of phenolic catabolism in *Rhizobium leguminosarum* biovar *trifolii*. J. Bacteriol. **173:**5546-5550.

41. **Parsek, M. R., R. W. Ye, P. Pun, and A. M. Chakrabarty**. 1994. Critical nucleotides in the interaction of a LysR-type regulator with its target promoter region. *catBC* promoter activation by CatR. J. Biol. Chem. **269:**11279-11284.

42. **Prentki, P. and H. M. Krisch**. 1984. *In vitro* insertional mutagenesis with a selectable DNA fragment. Gene **29:**303-313.

43. **Quandt, J. and M. F. Hynes**. 1993. Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. Gene **127:**15-21.

44. **Rhee, K. Y., M. Opel, E. Ito, S. Hung, S. M. Arfin, and G. W. Hatfield**. 1999. Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the *ilvYC* operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **96:**14294-14299.

45. **Rocha, E. R., G. Owens, Jr., and C. J. Smith**. 2000. The redox-sensitive transcriptional activator OxyR regulates the peroxide response regulon in the obligate anaerobe *Bacteroides fragilis*. J. Bacteriol. **182:**5059-5069.

46. **Schell, M. A.** 1985. Transcriptional control of the *nah* and *sal* hydrocarbon-degradation operons by the *nahR* gene product. Gene **36:**301-309.

47. **Schell, M. A.** 1993. Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. **47:**597-626.

48. **Schell, M. A., P. H. Brown, and S. Raju**. 1990. Use of saturation mutagenesis to localize probable functional domains in the NahR protein, a LysR-type transcription activator. J. Biol. Chem. **265:**3844-3850.

49. **Schell, M. A. and E. F. Poser**. 1989. Demonstration, characterization, and mutational analysis of NahR protein binding to *nah* and *sal* promoters. J. Bacteriol. **171:**837-846.

50. **Spaink, H. P., J. H. Robert, C. A. Okker, E. P. Wijffelman, and B. J. J. Lugtenberg**. 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Plant Mol. Biol. **9:**27-39.

51. **Tartaglia, L. A., G. Storz, and B. N. Ames**. 1989. Identification and molecular analysis of *oxyR*-regulated promoters important for the bacterial adaptation to oxidative stress. J. Mol. Biol. **210:**709-719.

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

53. **Tralau, T., J. Mampel, A. M. Cook, and J. Ruff**. 2003. Characterization of TsaR, an oxygen-sensitive LysR-type regulator for the degradation of *p*-toluenesulfonate in *Comamonas testosteroni* T-2. Appl. Environ. Microbiol. **69:**2298-2305.

54. **Urbanowski, M. L. and G. V. Stauffer**. 1989. Genetic and biochemical analysis of the MetR activator-binding site in the *metE metR* control region of *Salmonella typhimurium*. J. Bacteriol. **171:**5620-5629.

55. **van den Bergh, E. R., L. Dijkhuizen, and W. G. Meijer**. 1993. CbbR, a LysR-type transcriptional activator, is required for expression of the autotrophic CO_2 fixation enzymes of *Xanthobacter flavus*. J. Bacteriol. **175:**6097-6104.

56. **Wang, L. and S. C. Winans**. 1995. The sixty nucleotide OccR operator contains a subsite essential and sufficient for OccR binding and a second subsite required for ligand-responsive DNA bending. J. Mol. Biol. **253:**691-702.

57. **Wheelis, M. L. and L. N. Ornston**. 1972. Genetic control of enzyme induction in the beta-ketoadipate pathway of *Pseudomonas putida*: deletion mapping of *cat* mutations. J. Bacteriol. **109:**790-795.

58. **Yuan, Z. C., R. Zaheer, and T. M. Finan**. 2006. Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. J. Bacteriol. **188:**1089-1102.

Table 3.1 Bacterial strains and plasmids used in study.

Strain or plasmid	Relevant characteristics	Source or reference
S. meliloti		
Rm1021	Smr derivative of wild-type strain SU47	(25)
RmG212	Rm1021 lac; Smr	Strain collection
RmP110	Rm1021 with wild-type pstC; Smr	(58)
RmP134	RmG212 pcaQ::Ω; Smr, Gmr	(23)
RmP137	RmG212 (pTH467); Smr, Tcr	This work
RmP138	RmP134 (pTH467); Smr, Gmr, Tcr	This work
RmP1676	RmP110 pcaQ::Ω; Spr, Smr	This work
Rm5000	SU47 rif-5	(13)
Plasmid		
pUCP30T	Cloning vector; Gmr	Genbank accession no. U33752
pOT1	Broad host range gfpuv transcriptional reporter; Gmr	(2)
pTH467	434 bp EcoRI-XbaI PCR product encompassing pcaDQ intergenic region in pMP220 (pcaQ::lacZ); Tcr	This work
pTH1522	Reporter vector used in construction of S. meliloti reporter gene fusion library; Gmr	(8)
pFL2211	1.6 kb insert encompassing 781 bp pcaD and 740 bp pcaQ into pTH1522; clone obtained from S. meliloti reporter gene fusion library; Gmr	(8)
pTH2209	194 bp XbaI PCR product extending across pcaD / pcaQ intergenic region in pUCP30T; Gmr	This study
pTH2273	pTH2209 site directed mutagenesis (SDM); T(-63)†→G and A(-51)→G; Gmr	This study
pTH2276	184 bp PCR-amplified insert from pTH2209 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2282	184 bp PCR-amplified insert from pTH2273 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2294	pTH2209 SDM; A(-58)→G; Gmr	This study
pTH2295	pTH2209 SDM; A(-57)→G: Gmr	This study
pTH2298	184 bp PCR-amplified insert from pTH2294 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2299	184 bp PCR-amplified insert from pTH2295 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2336	pTH2209 SDM; T(-60)→G; Gmr	This study
pTH2337	184 bp PCR-amplified insert from pTH2336 into pOT1 to generate pcaD::gfp fusion; Gmr	This study

pTH2387	pTH2209 SDM; T(-59)→G; Gmr	This study
pTH2388	pTH2209 SDM; A(-72)→G; Gmr	This study
pTH2389	pTH2209 SDM; T(-71)→G; Gmr	This study
pTH2390	pTH2209 SDM; A(-70)→G; Gmr	This study
pTH2391	pTH2209 SDM; A(-69)→G; Gmr	This study
pTH2392	pTH2209 SDM; G(-61)→C; Gmr	This study
pTH2393	pTH2209 SDM; G(-62)→C; Gmr	This study
pTH2395	184 bp PCR-amplified insert from pTH2387 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2396	184 bp PCR-amplified insert from pTH2388 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2397	184 bp PCR-amplified insert from pTH2389 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2398	184 bp PCR-amplified insert from pTH2390 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2399	184 bp PCR-amplified insert from pTH2391 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2400	184 bp PCR-amplified insert from pTH2392 into pOT1 to generate pcaD::gfp fusion; Gmr	This study
pTH2401	184 bp PCR-amplified insert from pTH2393 into pOT1 to generate pcaD::gfp fusion; Gmr	This study

[†]Position with reference to *pcaD* transcriptional start site.

Table 3.2 Expression analysis of *pcaD::gfpuv* in *S. meliloti*.

Plasmid	Position of Mutation[b]	GfpUV Specific Activity[a] [SD] Uninduced	Induced	Fold Induction
pTH2276	None	1107 (80)	15748 (316)	14.2
pTH2396	A(-72)G	317 (36)	691 (53)	2.2
pTH2397	T(-71)G	202 (16)	598 (30)	3.0
pTH2398	A(-70)G	198 (26)	1821 (83)	9.2
pTH2399	A(-69)G	214 (10)	1507 (29)	7.1
pTH2282	T(-63)G; A(-51)G	981 (48)	14126 (665)	14.4
pTH2401	G(-62)C	326 (27)	3596 (19)	11.0
pTH2400	G(-61)C	223 (30)	1028 (18)	4.6
pTH2337	T(-60)G	264 (14)	720 (32)	2.7
pTH2395	T(-59)G	718 (41)	4195 (193)	5.8
pTH2298	A(-58)G	343 (20)	419 (21)	1.2
pTH2299	A(-57)G	551 (48)	4112 (32)	7.4
pOT1	None	208 (90)	158 (10)	0.8

[a] Shown are averages of three independent cultures of *S. meliloti* strains subcultured into M9 minimal medium with 0.5% glycerol ± 5 mM protocatechuate. SD, standard deviation

[b] Position indicated with respect to *pcaD* transcriptional start site.

Table 3.3 Quantification of PcaQ binding to the *pcaD* regulatory region *in vitro*.

Position of Mutation[a]	Percentage of Bound Probe[b] [SD]
None (wild-type)	84 (1)
A(-72)G	13 (<1)
T(-71)G	24 (6)
A(-70)G	57 (4)
A(-69)G	41 (8)
G(-62)C	60 (5)
G(-61)C	53 (7)
T(-60)G	33 (<1)
T(-59)G	53 (7)
A(-58)G	16 (4)
A(-57)G	64 (2)

[a] Position indicated with respect to *pcaD* transcriptional start site.
[b] Shown are averages obtained from two independent mobility shift assays. SD, standard deviation

Table 3.4 Expression of *pcaQ::lacZ* in *S. meliloti*.

Strain	Relevant genotype	Growth condition	β-galactosidase activity (Miller units [SD])
RmP137	Rm1021 *lac* (pTH467)	Glycerol	272 (8.4)
		Glycerol plus PCA[b]	359 (12.2)
RmP138	Rm1021 *lac pcaQ*::Ω (pTH467)	Glycerol	1391 (35.2)
		Glycerol plus PCA	1635 (18.4)

[a] Shown are averages of three independent cultures of *S. meliloti* strains subcultured into M9 minimal medium with 0.5% glycerol ± 5 mM protocatechuate. SD, standard deviation

[b] PCA, protocatechuate.

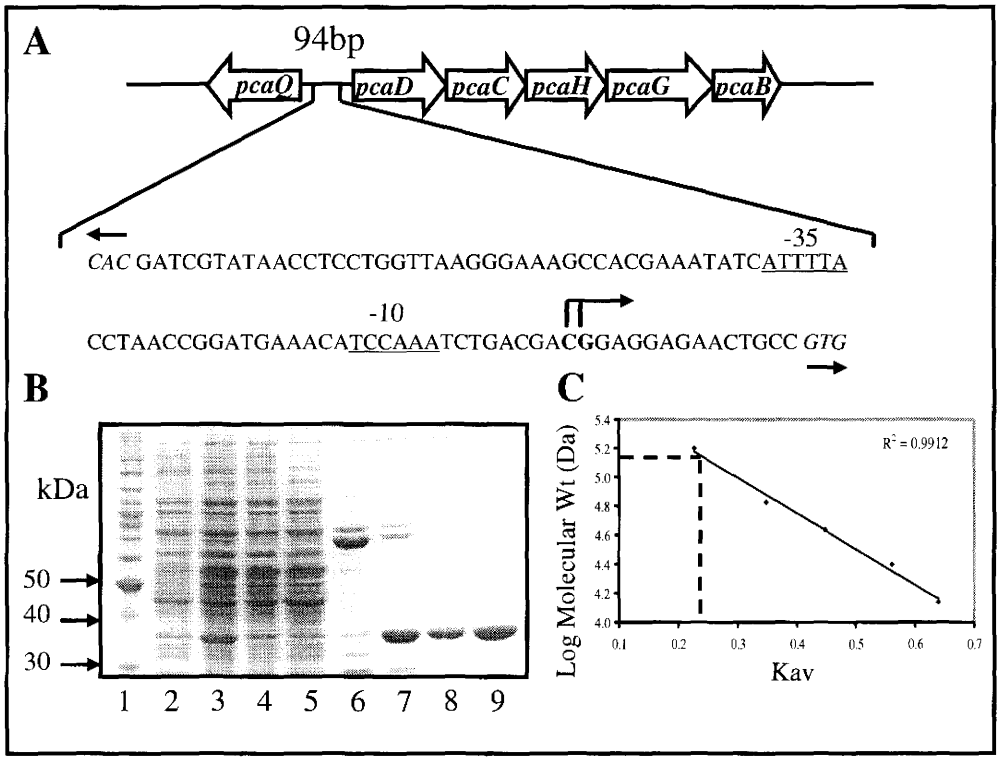


Figure 3.1. (A) Schematic depiction of the 94 bp intergenic region located between *pcaQ* and *pcaDCHGB* on the pSymB megaplasmid of *S. meliloti*. The predicted translational start codons of *pcaQ* and *pcaD* are italicized and the direction of translation is indicated by arrows. The two *pcaD* transcriptional start sites are identified in bold with the direction of transcription indicated by bent arrows; inferred -10 and -35 hexanucleotide regions associated with *pcaD* are underlined. (B) Overexpression and purification of *S. meliloti* PcaQ carrying a C-terminal His-tag. Protein samples were visualized by SDS-PAGE in a 10% polyacrylamide gel followed by staining with Coomassie brilliant blue. Lane 1, BenchMark protein ladder (Invitrogen); lane 2, crude cell lysate obtained from uninduced *E. coli* strain M924; lane 3, crude cell lysate obtained from *E. coli* strain M924, induced by 0.1mM IPTG; lane 4, flow through collected from Ni-NTA column; lane 5, eluate collected from 20mM imidazole wash; lane 6, eluate collected from 50mM imidazole wash; lane 7, eluate collected from 100mM imidazole wash; lane 8, eluate collected from 250mM imidazole wash; lane 9; pooled eluate obtained following FPLC size exclusion chromatography. (C) Standard curve used to estimate molecular mass of purified PcaQ, established using globular proteins as described in materials and methods. The K_{av} value obtained for PcaQ is 0.229; corresponding to an estimated molecular mass of PcaQ of 149 kDa.

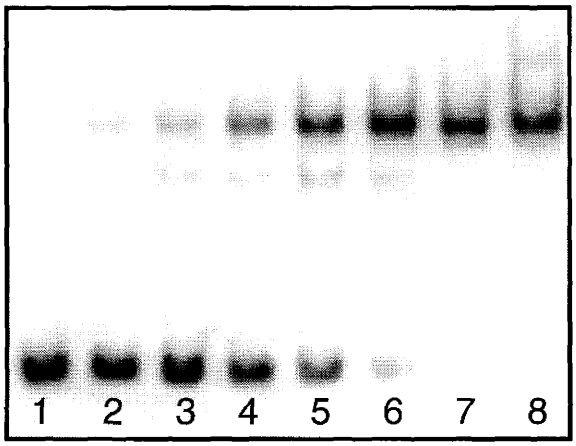


Figure 3.2. Electrophoretic mobility shift assay for PcaQ binding to the *S. meliloti* intergenic region. A 194 bp PCR amplified probe was end-labeled using γ^{32}P[ATP] and the resulting probe was incubated in the presence of increasing amounts of purified PcaQ prior to resolution upon a 6% non-denaturing polyacrylamide gel. Each assay contained 500ng herring sperm DNA as a non-specific competitor. Lanes 1 to 8 contain 0, 0.2, 0.5, 1.0, 2.5, 4.9, 12.3, and 24.7nM PcaQ.

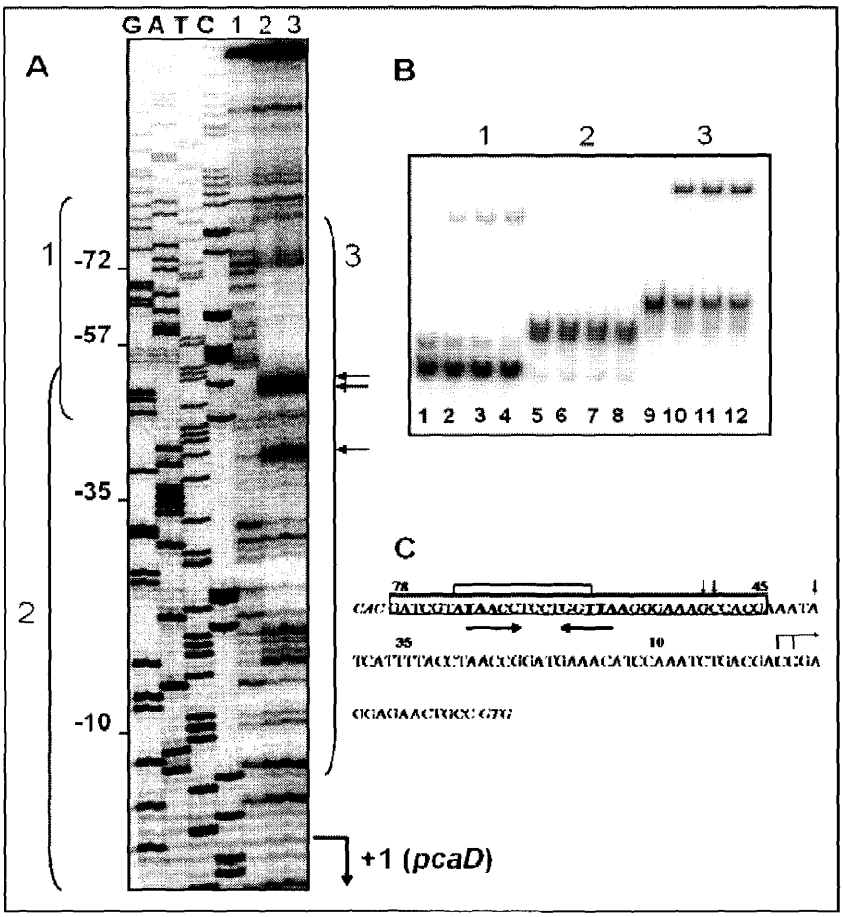


Figure 3.3. Identification of sequence involved in PcaQ binding within the *S. meliloti pcaDQ* intergenic region. (A) A 194 bp labeled probe was subjected to DNAseI digestion in the presence and absence of purified PcaQ. Lanes G, A, T, and C represent corresponding nucleotides as determined via sequencing reactions performed using pFL2211 as a template. Lane 1, probe digested with 0.01U DNAseI at room temperature for two minutes; lane 2, in the presence of 100ng PcaQ (14.4nM); lane 3, in the presence of 500ng PcaQ (72.2nM). Brackets 1, 2, and 3 correspond to probes used in subsequent mobility shift assays (panel B). Arrows located on the right indicate regions of hypersensitivity. (B) Electrophoretic mobility shift assays performed to confirm the location of a putative PcaQ binding site as determined by DNAseI footprinting experiment. Assays of group 1 (lanes 1 to 4), group 2 (lanes 5 to 8), and group 3 (lanes 9 to 12) were performed using probes corresponding to regions (-87 to -45), (-51 to +6), and (-83 to -6), as bracketed in panel A. Purified PcaQ was added to each group of assays in the following quantities: 0ng; 25ng (12nM); 50ng (24nM); and 100ng (48nM). (C) Sequence analysis of region protected by PcaQ from DNAseI digestion. Sequence upstream of *pcaD* including -35 and -10 hexanucleotide promoter regions (as underlined); *pcaD* transcriptional start sites (G and C) are indicated by an arrow bent in the direction of transcription; and the predicted translational start codons of *pcaD* and *pcaQ* are indicated by italics. The 34 nucleotide region protected by PcaQ from digestion is boxed in, whereas in bold are the nucleotides identified by alignment as conserved residues (Figure 4) upstream of *pcaD* in *A. tumefaciens*, *R. etli*, *M. loti*, and *R. leguminosarum*. The square bracket indicates a '$TN_{11}A$' motif within the conserved sequence and arrows above the sequence correspond to DNAseI hypersensitive sites identified during DNAseI footprinting experiments. Arrows below the sequence indicate inverted repeats within the PcaQ binding site.

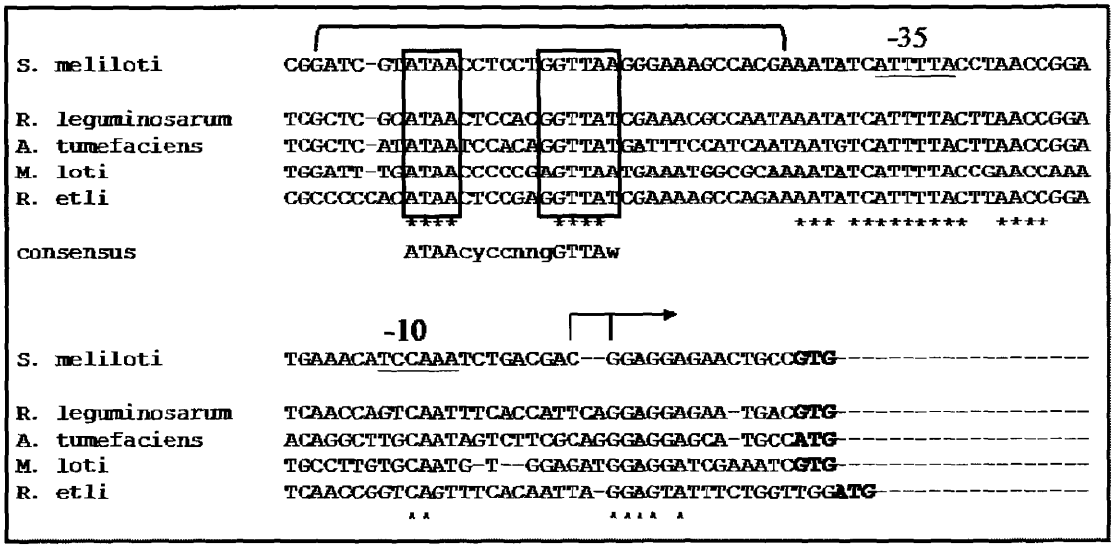
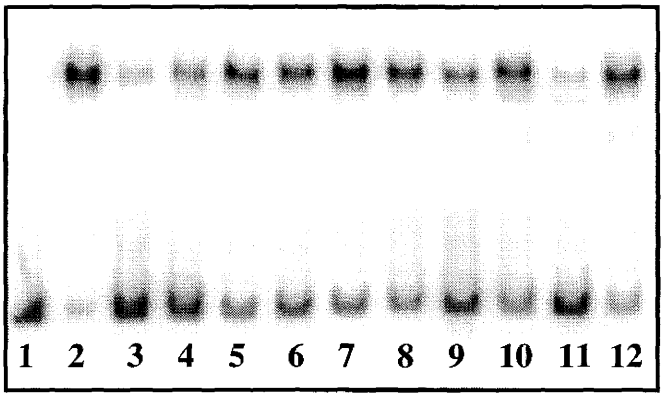


Figure 3.4. Alignment of *pcaDQ* intergenic regions from *S. meliloti*, *R. leguminosarum*, *A. tumefaciens*, *M. loti*, and *R. etli*. Residues conserved in all species are indicated by an asterisk and the bracket outlines the region protected from DNAseI digestion by PcaQ. Positions targeted for site directed mutagenesis in *S. meliloti* are enclosed in a box. Within the *S. meliloti* sequence, -35 and -10 hexanucleotides regions are underlined and the *pcaD* transcriptional start sites are indicated by arrows bent in the direction of transcription. A consensus sequence (5'ATAAcyccnngGTTAAw 3') was established based upon the alignment; uppercase letters indicate nucleotides conserved in all species and lowercase letters indicate nucleotides conserved in all but one species. The alignment was performed using ClustalW (52) (http://www.ebi.ac.uk/clustalw/).



Position		72	71	70	69	62	61	60	59	58	57
Nucleotide	WT	A	T	A	A	G	G	T	T	A	A
		↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
Substitution		G	G	G	G	C	C	G	G	G	G

Figure 3.5. Effect of site directed mutations within a putative PcaQ binding site upon ability of PcaQ to bind upstream of *pcaD* in *S. meliloti*. Electrophoretic mobility shift assays were performed using probes containing wild-type *pcaD* promoter region (lanes 1 and 2) or probes including single point mutations within a putative PcaQ binding site (lanes 3 to 12), as indicated by text below figure. Assays were performed using 2.5ng purified PcaQ (1.2nM), with the exception of a control in which PcaQ was not added (lane 1).

CHAPTER FOUR

Regulation of an ABC-type transport system by the protocatechuate LysR-protein PcaQ in *S. meliloti*

Allyson M. MacLean, Wilfried Haerty,

and Turlough M. Finan.

Preface

This chapter describes the identification and regulation of an ABC-type transporter in *S. meliloti* whose expression is induced in the presence of protocatechuate. During this study, Vladmir Jokic created certain genetic constructs and helped with the site directed mutagenesis experiments while working as an undergraduate student under my direct supervision. As well, Vladmir performed the Gfp assays to monitor the effect of mutagenesis upon gene expression. Wilfried Haerty wrote the computational script used to scan the genomes of Proteobacteria for PcaQ-binding motifs. With input from Turlough Finan, I have conceived of and designed all experiments described throughout this study, and (with the exception of the Gfp assays) I have performed these experiments myself. Finally, I have written this chapter in its entirety, with editing by Turlough Finan. Upon completion of protocatechuate uptake assays, this chapter will be submitted for publication to an appropriate journal.

4.1 Abstract

Lignin-derived aromatic acids offer soil-dwelling microorganisms a source of energy, and many soil inhabiting bacteria of the diverse class α–proteobacteria encode the ability to catabolize aromatic compounds via the protocatechuate branch of the β–ketoadipate pathway. *Sinorhizobium meliloti* may exist as a free-living soil saprophyte or as a highly differentiated endosymbiont within nitrogen-fixing root nodules of the leguminous plant alfalfa. This report describes the regulation of an ABC-type transport system that we infer is involved in the uptake of protocatechuate in *S. meliloti*. We demonstrate that expression of a gene encoding the periplasmic solute binding protein of the transport system (*smb20568*) is induced by growth with protocatechuate and is subject to regulation by the LysR-type protein PcaQ. Purified PcaQ is shown to bind sequence upstream of *smb20568* and we have identified a regulator binding site at positions -73 to -58 upstream of the *smb20568* transcriptional start site. Mutagenesis of conserved nucleotides within the putative regulatory element alters PcaQ-dependent regulation of *smb20568* expression *in vivo*, consistent with this sequence encoding a PcaQ-binding site. A consensus PcaQ-binding motif was generated based upon conserved binding sites positioned upstream of *pcaD* in α-proteobacteria. Using this consensus sequence, we identified PcaQ-binding motifs in members of α-, β-, and γ-proteobacteria, revealing an unexpected taxonomic and regulatory distribution for this LysR-type protein.

4.2 Introduction

The β-ketoadipate pathway is a key metabolic pathway through which the aromatic substrates protocatechuate and catechol are catabolized to yield succinate and acetyl-CoA (Harwood & Parales, 1996). Plant-derived aromatic acids such as *p*-hydroxybenzoate and protocatechuate (3,4-dihydroxybenzoate) offer soil-dwelling microorganisms a source of energy, and the protocatechuate branch of the β-ketoadipate pathway is present in many members of the family *Rhizobiaceae* (Parke & Ornston, 1984; Parke & Ornston, 1986; Parke, et al., 1991; Parke, 1995; Parke, 1996a; MacLean, et al., 2006). In a nutrient-limited environment such as soil, a competitive advantage may be afforded to species with the ability to efficiently scavenge organic compounds such as protocatechuate. Although genes and enzymes associated with the β-ketoadipate pathway have been the subject of many studies in α-proteobacteria (Parke & Ornston, 1984; Parke, et al., 1985; Parke & Ornston, 1986; Parke, et al., 1987; Parke, et al., 1991; Parke, 1995; Parke, 1996a; Parke, 1996b; Buchan, et al., 2000; Buchan, et al., 2001; Buchan, et al., 2004; MacLean, et al., 2006; MacLean, et al., 2008), transport systems involved in the uptake of protocatechuate and related compounds such as *p*-hydroxybenzoate have yet to be identified in this class of bacteria.

The Gram-negative legume microsymbiont *Sinorhizobium meliloti* encodes the protocatechuate branch of the β-ketoadipate pathway on the pSymB megaplasmid (MacLean, et al., 2006). Protocatechuate catabolism genes are organized in two operons (*pcaDCHGB* and *pcaIJF*), whose expression is regulated by the transcriptional regulators PcaQ and PcaR, respectively (MacLean, et al., 2006). PcaQ is a member of the LysR-family of regulatory proteins and we have previously described a PcaQ-binding motif that is located upstream of *pcaD* in *S. meliloti* and is conserved in related rhizobia (MacLean, et al., 2008). To date, this regulatory protein has only been described in α–proteobacteria, as modulating the expression of *pcaD* (Parke, 1993; Parke, 1995; Parke, 1996a; Parke, 1996b; MacLean, et al., 2006; MacLean, et al., 2008). We here report that PcaQ additionally regulates the expression of a protocatechuate-inducible ABC-type transport system in *S. meliloti* through interaction with a binding site –73 to –58 upstream

PhD Th

of a gene encoding the periplasmic solute binding protein of the transporter. Using an *in silico* approach, we have also identified conserved PcaQ-binding sites positioned upstream of ABC-type transport systems and genes encoding the enzymes protocatechuate 3,4-dioxygenase (*pcaHG*) and β-ketoadipate enol-lactone hydrolase (*pcaD*) in the genomes of α-, β-, and γ-proteobacteria.

4.3 Materials and Methods

Bacterial strains and growth conditions

All plasmids and bacterial strains used throughout this study are described in Table 4.1. *Escherichia coli* strains were grown aerobically at 37°C in LB broth. *S. meliloti* strains were grown aerobically at 30°C in LB broth supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ (LBmc) or in M9-minimal medium (Difco) supplemented with 1.0 mM $MgSO_4$, 0.25 mM $CaCl_2$, 1 μg/mL D-biotin, and 10 ng/mL $CoCl_2$. 0.5% (v/v) glycerol and 5 mM protocatechuate (Sigma-Aldrich) were added to M9-minimal medium as a source of carbon. For *E. coli* strains, the following antibiotic concentrations were used (μg/mL), chloramphenicol (Cm): 20; gentamicin (Gm): 10; ampicillin (Amp): 50; tetracycline (Tc): 10. For *S. meliloti*, streptomycin (Sm): 200; spectinomycin (Sp): 200; gentamicin: 60; rifampicin (Rif): 20; tetracycline: 5.

Construction of S. meliloti transporter mutants

Two *S. meliloti* transporter mutant strains were constructed for future use in protocatechuate-uptake assays as follows. In the first strain, expression of the gene encoding a periplasmic solute binding protein (*smb20568*) was disrupted via introduction of an in-frame deletion. Primers (5' TCGTCTAGAACGTGACGATGGTTCTGG 3') and (5' TCTAAGCTTCGATCGTCATCAGCACCTG 3') were used to amplify a 1,965 bp fragment which was cloned into pUCP30T via *Xba*I and *Hind*III, yielding pTH1948. A *Pst*I digest was performed upon pTH1948 to generate a 687 bp deletion within

smb20568 (in-frame deletion of 229 amino acids), to create plasmid pTH1949. The *S. meliloti* insert in pTH1949 (with the 687 bp deletion) was then subcloned into the suicide vector pJQ200 (Quandt & Hynes, 1993) in the following manner. pTH1949 was digested with *Xba*I and *Hind*III to liberate the *S. meliloti* insert DNA from the pUCP30T backbone. Klenow fragment of DNA polymerase I (New England Biolabs) was added to digested DNA to yield blunt-ended DNA fragments which were resolved and purified from the agarose gel and cloned into the suicide vector pJQ200 uc-1 via *Sma*I to create pTH1959. This plasmid was transferred into wild-type *S. meliloti* strain RmP110 and single cross-over recombinants were selected by plating upon LB agar supplemented with Gm. A single purified Gm^r transconjugant was grown overnight in LBmc in the absence of antibiotic selection. Aliquots of the overnight culture were plated upon LB agar in the presence of 5% sucrose to select for cells in which the integrated plasmid had recombined out of the genome. Sucrose resistant ($sucrose^r$) colonies were patched upon Gm to confirm the loss of the pJQ200 derivative plasmid. Ten $sucrose^r$ and Gm^s strains were screened via whole cell PCR amplification of *smb20568* for the presence of a 687 bp deletion; of these, five strains yielded products consistent with the incorporation of a deletion within this gene. Sequencing reactions were performed upon genomic DNA to confirm the genotype of each strain using primers complementary to sequence external to the region cloned into pJQ200. One of the strains was selected for use in this study and was designated as RmP1710.

A second transporter mutant strain was constructed through the insertion of an antibiotic cassette into *smb20787*. In addition to disrupting expression of *smb20787*, the presence of a cassette at this location in the operon will prevent expression of the three transporter genes located downstream of this gene (encoding one permease and two ATP binding cassettes) due to the polar nature of the mutation. A streptomycin/spectinomycin resistance cassette was cloned into a *Sac*II site within *smb20787* as follows. Primers (5' ATGCGGCCGCTGCATCGTTGGTTTGG 3') and (5' ATGCGGCCGCAATAGCCGGTGACG 3') were used to PCR amplify sequence spanning a *Sac*II site located 187 bp downstream of the predicted translational start site of

smb20787. The 1,149 bp amplified product was cloned into plasmid pTH1883 via a *Not*I restriction site, yielding pTH1908. pTH1883 is a derivative of pVO155 (Oke & Long, 1999) in which the reporter gene *gusA* has been deleted; this plasmid was selected because it is unable to replicate in *S. meliloti* and thus may be used as a means of recombining the antibiotic cassette into the *S. meliloti* genome. The streptomycin/spectinomycin resistance cassette from pHP45Ω (Prentki & Krisch, 1984) was PCR amplified and cloned into pTH1908 following digestion with *Sac*II to create pTH1917. This plasmid was transferred by conjugation into the Rif resistant *S. meliloti* strain Rm5000; recombinants were selected by plating upon LB agar supplemented with Rif and Sp. Transconjugants were patched upon LB agar plus Nm to screen for loss of pTH1917.

Southern hybridization was performed to confirm integration of the Ω cassette within *smb20787* as follows. Briefly, primers (5' CGAGATCGAGCGAGAGTACCAGC 3') and (5' GGAGAACGTGACGATGGTTCTGG 3') were used to amplify a probe corresponding to *smb20787*. *Sal*I digests were performed upon genomic DNA isolated from two Sp^r and Nm^s colonies and wild-type strain RmP110. In both putative *smb20787*::Ω mutants, an ~2kb shift (compared to wild-type) in one of the three hybridized *Sal*I fragments was noted, consistent with the incorporation of the 2.1kb cassette into the *Sac*II site and excision of pTH1917. One of the two mutant strains was selected and the *smb20787*::Ω allele was transferred from Rm5000 into RmP110 by transduction through selection of Sp^r transductants. This strain (RmP110 *smb20787*::Ω) was designated RmP1712.

Construction of reporter gene fusion plasmids

Plasmids pTH1913, pTH1914, and pTH1927 were created through the integration of plasmid pTH1508 with pFL1131, pFL2665, and pFL2211, respectively (see Table 4.1). pFL1131, pFL2665, and pFL2211 are derived from pTH1522, and are consequently unable to replicate in *S. meliloti*. Integration of these plasmids with the broad-host-range plasmid pTH1508 via recombination at *attP* and *attB* sites yields a single plasmid that

can replicate in *S. meliloti* and carries the intact transcriptional fusions associated with the pTH1522 derivatives.

The *S. meliloti* reporter fusion genome library (Cowie, et al., 2006) did not yield a suitable reporter gene fusion with the *pcaIJF* operon, and thus pTH1971 was constructed in the following manner. Sequence upstream of *pcaI* was PCR amplified as a 1,123 bp fragment using primers (5' ATCAGATCTGAGTTCGTCGACGATCTCC 3') and (5' TAGGTACCCAGGTTGATCGACATCACC 3'). This region was cloned into the transcription fusion reporter vector pTH1705 (Cowie, et al., 2006) via *Bgl*II and *Kpn*I to generate a *pcaI*::*gfp+/lacZ* fusion. The resulting plasmid (pTH1960) was integrated with pTH1508 to create the broad-host range replicating plasmid pTH1971.

To generate a reporter gene fusion with *smb20568* (pTH1972), sequence spanning the intergenic region of *smb20568* and *smb20569* was PCR amplified as a 627 bp fragment using primers (5' TAGGTACCTGCGATGACGAAACTGACG 3') and (5' ATCAGATCTGGTAGATGATCTCGATGTCG 3') and cloned into pTH1705 via *Bgl*II and *Kpn*I to create pTH1961. pTH1972 was then constructed through the integration of pTH1961 with pTH1508; pTH1972 thus carries a *smb20568*::*gfp+/lacZ* fusion.

These plasmids were transferred by triparental conjugation into *S. meliloti* strains RmP110 (wild-type), RmP1676 (RmP110 *pcaQ*::Ω), and RmP708 (RmP110 *pcaR*::Ω). RmP708 was constructed through the transduction of *pcaR*::Ω from RmK1014 (Rm1021 *pcaR*::Ω; MacLean, et al., 2006) into RmP110 via selection for resistance to spectinomycin.

Construction of pTH2454

As previously described, plasmid pTH1979 contains *pcaQ* cloned into the expression vector pET-21a (Novagen) to create a translational fusion with a hexahistidine tag at the carboxyl terminus of the *S. meliloti* protein (MacLean, et al., 2008). This plasmid was used as template DNA for the PCR amplification of *pcaQ·his* using primers (5' GTGAGATCTAAGAAGGAGATATACATATGATCGAC 3') and (5' AGAATTCGTTAGCAGCCGGATCTCAGTG 3'). The amplified product (*pcaQ·his*)

was cloned into the broad host range replicating plasmid pTH1227 via *Bgl*II and *Eco*RI to create plasmid pTH2454, in which expression of *pcaQ·his* is regulated by a P_{tac} promoter. After sequencing (to confirm the absence of mutations within the coding sequence), the plasmid was transferred via conjugation into *S. meliloti* strain RmP1811 (RmP110 *pcaQ*::Ω *smb20568*::*gfp+/lacZ*) in which expression of *pcaQ* has been disrupted through the integration of a spectinomycin/streptomycin antibiotic resistance cassette. RmP1811 also contains a transcriptional fusion (*smb20568*::*gfp+/lacZ*), created through the integration of plasmid pFL1131, that allows expression of *smb20568* to be monitored via GFP+ and β–galactosidase reporter enzyme assays.

Primer extension

Total RNA was isolated from an *S. meliloti* RmP110 culture grown aerobically at 30°C in LBmc ± 5 mM protocatechuate to an optical density at 600 nm ($O.D._{600}$) of 0.8. RNA extraction was performed using a hot phenol method as previously described (MacLellan, et al., 2005). Approximately 60 µg of total RNA was used in each extension reaction, as described (MacLellan, et al., 2005). Two primers were used to yield extension products: (5' CCAGAATGATCCTTCTCATATTTCCTCC 3') and (5' CGACGACTCCGACCTTGATCGTATCC 3'). Sequenase version 2.0 DNA sequencing kits (USB) were used for sequencing reactions, which were performed upon plasmid pFL1131. The same primers were used in both sequencing and primer extension reactions.

β-galactosidase assays

S. meliloti strains were grown in LBmc overnight and washed in 0.85% NaCl prior to sub-culturing into M9-minimal medium supplemented with a carbon source as indicated. Cells were incubated for a minimum of four hours with shaking at 30°C; aliquots of cultures were then obtained and used in enzyme assays. β-galactosidase assays were performed as previously described (MacLean, et al., 2006).

Electrophoretic mobility shift assays

A 246 bp probe spanning the *smb20568* and *smb20569* intergenic region (-149 to +87; *smb20568* transcriptional start site) was PCR amplified using primers (5' ATCTAGATACAGGCAGGAGCTGCTTCG 3') and (5' ATTCTGCAGCAGAATGATCCTTCTCATATTTCCT 3'). PCR products were purified using a polyacrylamide gel before labeling reactions were performed using γ^{32}P[ATP] (Perkin Elmer) and T_4 polynucleotide kinase (New England Biolabs). Assays were performed using purified PcaQ·His as previously described (MacLean, et al, 2008).

Site directed mutagenesis

Plasmid pTH2410 was used as a template for mutagenesis reactions and was constructed as follows. Primers (5' ATCTAGATACAGGCAGGAGCTGCTTCG 3') and (5' ATCTAGAGCCAGAATGATCCTTCTCATATTTCCT 3') were used to PCR amplify the *smb20568/smb20569* intergenic region, which was cloned into pUCP30T via *Xba*I. Site directed mutagenesis reactions were performed using Platinum *Pfx* DNA polymerase (Invitrogen) according to the supplier's protocol. After amplification, 20U *Dpn*I (New England Biolabs) was added directly to the reaction mixture, which was incubated at 37°C for 2 hours. Plasmid DNA was purified using QIAquick spin columns (Qiagen) and transformed into chemically competent *E. coli* DH5α. Transformants were selected via Gm^r and plasmids were sequenced to confirm the presence of site directed mutations by Mobixlab (McMaster University, Hamilton, Ontario, Canada).

For the purpose of constructing *gfpuv* reporter fusions, the pUCP30T derivative plasmids (each with a site directed mutation) were used as template DNA for the PCR amplification of the *smb20568/smb20569* intergenic region using primers as described for the amplification of the 246 bp probe utilized in the mobility shift assays. Each PCR product was cloned into the broad host range reporter plasmid pOT1 (Allaway, et al., 2001) via *Xba*I and *Pst*I to generate a *smb20568::gfpuv* fusion. All plasmids were sequenced by Mobixlab to confirm the presence of the appropriate mutation and the orientation of the gene fusion with *gfpuv*.

GfpUV assays

Assays were performed as previously described (MacLean, et al., 2008). Fluorescence was determined by dividing the emission output of each sample by its respective $O.D._{600}$.

Computational analyses

A consensus PcaQ-binding site was generated using conserved sequence upstream of *pcaD* in *Agrobacterium tumefaciens* C58, *Mesorhizobium loti* MAFF303099, *Rhizobium etli* CFN42, *R. leguminosarum* bv. *viciae* 3841, *Sinorhizobium meliloti* Rm1021, and *Sinorhizobium medicae* WSM419, in addition to the PcaQ-binding site upstream of *smb20568* in *S. meliloti*. The fully sequenced genomes and plasmids of all α-proteobacteria (217 genomes and plasmids, available as of December, 2008; NCBI Genome database), *Pseudomonas* (11), *Burkholderia* and *Ralstonia* (85), were scanned for the consensus PcaQ-binding motif (5' ATAAYY$(N)_x$ RRTTA 3'; x = 3, 4, or 5). The output data was screened manually and putative binding sites positioned upstream of genes associated with aromatic acid catabolic enzymes and/or LysR-type regulators were extracted for further analysis. Sequence alignments were performed using ClustalW2 (EMBL-EBI). The sequence logo was constructed using WebLogo (Crooks, et al., 2004).

4.4 Results

Identification of a transport system with protocatechuate-inducible expression

A large-scale expression analysis of *S. meliloti* transport systems led to the identification of genes encoding a putative ABC-type transport system whose expression was induced by growth in the presence of protocatechuate (Mauchline, et al., 2006). The five genes encoding this system (*smb20568-smb20784*) are located less than 10 kb from previously characterized *pca* genes encoded on the pSymB megaplasmid (MacLean, et al., 2006). In order to confirm these initial results, additional reporter gene fusions were constructed and β-galactosidase assays were performed upon *S. meliloti* strains carrying

the replicating reporter plasmids (Figure 4.1). Expression of a gene encoding the periplasmic solute binding protein (*smb20568*) was induced 6- to 8-fold by the addition of protocatechuate to growth medium (Figure 4.1; pTH1913, pTH1972). In contrast, the intergenic region between *smb20568* and *smb20787* was not sufficient to induce gene expression when carried upon a replicating plasmid (Figure 4.1; pTH1914). Similarly, expression of an adjacent ABC-type transport system (*smb20569-smb20571*) encoded upstream of and on the same strand as *smb20568-smb20784* was not affected by growth in the presence of the aromatic compound (data not shown). These data suggest that the protocatechuate-inducible expression of the transport system is solely dependent upon a regulatory region located upstream of *smb20568*. Reporter gene fusions to *pcaD* (pTH1927) and *pcaI* (pTH1971) were included as positive controls; as expected, expression of these genes (as determined by β-galactosidase activity) was induced 9-fold and 5-fold in the presence of protocatechuate, respectively.

Expression of the transport system is regulated by PcaQ

Genes encoding products relevant to the catabolism of protocatechuate in *S. meliloti* and *A. tumefaciens* are organized into two operons (*pcaDCHGB* and *pcaIJF*) whose expression is regulated by the transcriptional regulators PcaQ and PcaR (Parke, 1993; Parke, 1995; Parke, 1996a; MacLean, et al., 2006). In order to determine whether the protocatechuate-inducible expression of the transporter genes requires either of these transcriptional regulators, expression of *smb20568::gfp+/lacZ* was monitored in PcaQ- and PcaR-minus strains of *S. meliloti* (Table 4.2). In RmP708 (RmP110 *pcaR*::Ω), expression of *smb20568* (as determined by β-galactosidase activity) was induced 7-fold in the presence of protocatechuate, as was observed for the expression of this gene in a wild-type background (RmP110), suggesting that the IclR-type regulator PcaR is not required for the regulation of *smb20568* expression. Expression of *pcaI* is known to be regulated by PcaR (MacLean, et al., 2006) and expression of this gene was not induced in the PcaR-minus strain RmP708, as expected. Expression of the transporter gene was not induced by growth with protocatechuate in RmP1676 (RmP110 *pcaQ*::Ω), indicating that

the LysR-type regulator PcaQ may be involved in modulating *smb20568* expression. Likewise, *pcaD* expression was not up-regulated in RmP1676, consistent with previous studies in which PcaQ has been shown to regulate expression of this gene (MacLean, et al., 2006; MacLean, et al., 2008). PcaQ is not involved in the regulation of *pcaI* expression, and expression of this gene was induced by growth with protocatechuate in the PcaQ-minus strain (Table 4.2).

PcaQ·His activates transcription of smb20568 in vivo

We have reported the purification of PcaQ as a translational fusion protein with a C-terminal hexahistidine tag in a previous study (MacLean, et al., 2008). Although purified PcaQ·His was shown to bind a regulatory site upstream of *pcaD* with high affinity in that study, we did not demonstrate whether PcaQ·His is capable of activating gene expression *in vivo*. To address this, we introduced *pcaQ·his in trans* into an *S. meliloti* strain in which the native *pcaQ* was disrupted, and measured expression of *smb20568* via reporter enzyme assays (Table 4.3). Our results indicate that expression of *smb20568* was induced up to 6-fold in cells grown in the presence of protocatechuate when *pcaQ·his* was supplied *in trans* on plasmid pTH2454. In contrast, expression of the transport gene was not induced in a control strain carrying the empty plasmid (pTH1227).

Identification of smb20568 transcriptional start site

We wished to further examine the regulation of *smb20568*, and primer extension analysis was performed to identify the promoter associated with this gene. Total RNA was isolated from wild-type *S. meliloti* strain RmP110 grown in the presence and absence of protocatechuate and primer extension reactions were performed using two different primers. With both primers, at least two extension products were obtained using RNA isolated from *S. meliloti* grown with and without protocatechuate (Figure 4.2A). The larger of the two extension products corresponds to a transcriptional start site located 29 nucleotides upstream of the predicted *smb20568* translational start codon. In this instance, a greater product yield was obtained using RNA isolated from cells grown with

protocatechuate; this is consistent with an up-regulation of *smb20568* expression in the presence of protocatechuate. The smaller extension product corresponds to a transcriptional start site located within the predicted coding sequence of *smb20568* and is likely due to the premature termination of the extended product.

Examination of the *smb20568* promoter reveals little sequence similarity to a subset of promoters previously described in *S. meliloti* (MacLellan, et al., 2006). Nonetheless, alignment of the transport gene promoter with the *S. meliloti pcaD* promoter reveals areas of conserved sequence within the -35 and -10 regulatory regions (Figure 4.2C). These conserved sequences are also present upstream of *pcaD* in related species of rhizobia and agrobacteria.

Purified PcaQ binds upstream of smb20568

Our results demonstrate that the LysR-type protein encoded by *pcaQ* is required for the protocatechuate-inducible expression of *smb20568*. To examine whether PcaQ binds sequence in the promoter region of this gene, electrophoretic mobility shift assays were performed using purified PcaQ·His. Assays were performed using a 246 bp probe encompassing the intergenic region between *smb20568/smb20569* (extending –149 to +87; with respect to the transcriptional start site of *smb20568*). PcaQ was shown to bind labeled probe with high affinity in the absence of a co-inducing molecule (Figure 4.3). This result indicates that PcaQ may be directly involved in regulating expression of the transporter genes through interaction with a binding site located upstream of *smb20568*.

Identification of a PcaQ binding site upstream of smb20568

Examination of the *smb20568/smb20569* intergenic region reveals a potential PcaQ-binding site located –73 to –58 with respect to the transcriptional start site of *smb20568* that is very similar in position and nucleotide composition to a previously characterized PcaQ-binding site upstream of *pcaD* in *S. meliloti* (MacLean, et al., 2008), and conserved in *Sinorhizobium medicae, Agrobacterium tumefaciens, Mesorhizobium loti, Rhizobium etli,* and *R. leguminosarum* (Figure 4.4).

Site directed mutagenesis was employed to determine whether the putative PcaQ-binding site upstream of *smb20568* is required for the protocatechuate-inducible expression of the transport system. Six positions within the binding site were targeted for mutagenesis; nucleotides in four of these positions have been conserved in PcaQ-binding sites of upstream of *pcaD* in all six species (Figure 4.4; black arrows), and mutations introduced at any of these locations should affect *smb20568* expression based upon a similar analysis of *pcaD* regulation in *S. meliloti* (MacLean, et al., 2008). As in the earlier study, we replaced A and T residues with a guanosine at each of these positions. In another instance, a transversion mutation was introduced to examine the effect of substituting a purine within a position occupied solely by pyrimidines in all sequences (C(-69)A; pTH2426). Finally, we substituted an adenosine encoded beside the right-hand arm of each binding site with a guanosine (A(-62)G; pTH2424) or a cytosine (A(-62)C; pTH2425) to generate a site that more closely approximated the PcaQ-binding consensus sequence. Wild-type and mutant PcaQ-binding sites (positions -149 to +87) were cloned into the broad host range plasmid pOT1 to generate a transcriptional fusion between *smb20568* and a gene encoding green fluorescent protein (GfpUV). Expression of *smb20568*::*gfpuv* was measured in wild-type (RmP110) and PcaQ-minus (RmP1676) strains of *S. meliloti* cultured in the presence and absence of protocatechuate (Table 4.4).

Expression of *smb20568* (as determined by GfpUV specific activity; pTH2414) was induced greater than 10-fold in RmP110 grown with protocatechuate, consistent with previous data (Figure 4.5). The introduction of mutations within five of the six nucleotides targeted for mutagenesis (underlined) within the putative binding site (5' ATAACCGGGGGATTAT 3') resulted in a decreased level of expression in response to growth with protocatechuate. Interestingly, one of these mutations (C(-69)A; pTH2426) resulted in an increased level of *smb20568* expression in uninduced cells (versus wild-type) that was not influenced by addition of protocatechuate (i.e., constitutive expression). This particular mutation also gave rise to a similarly high expression level in a *pcaQ*-minus background (Figure 4.5).

The substitution of A(-62)C (pTH2425) did not affect the degree of induction observed in response to protocatechuate (>11-fold increase in expression) although the absolute level of expression in both uninduced and induced cells was decreased as compared to the wild-type control. Replacement of an A residue with G at this position (A(-62)G) doubled the level of *smb20568* expression (>20-fold) in response to the aromatic acid.

Prediction of putative PcaQ-binding sites in Proteobacteria

A consensus PcaQ-binding motif was generated using the experimentally verified binding sites upstream of *pcaD* (MacLean, et al., 2008) and *smb20568* (this work) in *S. meliloti*, and conserved sites upstream of *pcaD* in related rhizobia (Figure 4.4). Using the consensus sequence, we scanned the genomes of α-, β- (*Burkholderia* and *Ralstonia*), and γ-proteobacteria (*Pseudomonas*) for potential regulatory sites. The majority of genomes yielded between 0 to 3 hits to putative PcaQ-binding sites, although we observed an unusually large number of hits (>50 hits per genome) in members of the order *Rickettsiales* (*Rickettsia*, *Ehrlichia*, *Orientia*, *Wolbachia*, *Candidatus*) and the genus *Bartonella*, which do not encode *pca* genes such as *pcaQ*. The quality of putative PcaQ-binding sites was assessed based upon location of hit (within or between coding sequence), the number and position of mismatches within the sequence (as compared to the consensus sequence), and proximity to known target genes (relating to aromatic acid metabolism or genes encoding LysR-proteins). Using these criteria, we identified 38 putative PcaQ-binding sites within the genomes of all three classes of proteobacteria (Table 4.5). A sequence logo was constructed using the predicted PcaQ binding sites, and indicated a high degree of sequence conservation within the left and right-hand arms of the PcaQ binding site (positions -73 to -71; and -61 to -58, respectively). In contrast, the central 4 nucleotides (positions -67 to -64) do not appear to be conserved. We did not detect any hits within the coding sequence of aromatic acid catabolism genes, indicating that PcaQ may not bind sequence within target genes, as has been reported for other LysR-proteins (Chugani, et al., 1998). In the α-proteobacteria, conserved binding sites

were located upstream of either *pcaD* (encoding β-ketoadipate enol-lactone hydrolase) or an ABC-type transport system. In β– and γ–proteobacteria, the sites were positioned upstream of *pcaH* (encoding the β subunit of protocatechuate 3,4-dioxygenase), indicating that expression of the dioxygenase may be regulated by PcaQ in these bacteria.

4.5 Discussion

The genomes of rhizobia are particularly enriched in genes encoding transport systems, as befitting a group of organisms that inhabit a complex and variable environment such as soil (MacLean, et al., 2007). Mauchline and coworkers recently reported an expression analysis encompassing the entire complement of *S. meliloti* transport systems against a wide range of possible inducers and growth conditions (Mauchline, et al., 2006). Intriguingly, the expression of an ABC-type transport system (*smb20568-smb20784*) was highly induced by growth with protocatechuate. The proximity of the encoded transport system to protocatechuate catabolic genes suggested that this system may be involved in the transport of protocatechuate.

Our results confirm that expression of *smb20568* (and we infer the entire gene cluster) is induced by growth with protocatechuate and are consistent with the presence of a protocatechuate-responsive regulatory region(s) located upstream of *smb20568*. The LysR-type protein PcaQ participates in the regulation of *pcaD* expression in *S. meliloti* and *A. tumefaciens* (Parke, 1993; Parke, 1996a; MacLean, et al., 2006; MacLean, et al., 2008), and we report that expression of *smb20568* is likewise regulated by PcaQ. In a PcaQ-minus background, expression of *smb20568* was no longer induced by growth with protocatechuate (Table 4.2), and the introduction of *pcaQ-his in trans* (pTH2454) rescued the regulatory phenotype exhibited by this strain (Table 4.3). Previous studies have demonstrated that the addition of a histidine tag at the carboxyl terminus of a LysR protein did not compromise the ability of the regulator to effect transcriptional activation (Ogawa, et al., 1999; Bundy, et al., 2002), and we have demonstrated that purified

PcaQ·His is able to bind DNA with a high degree of specificity and affinity (MacLean, et al., 2008). However it was nonetheless important to determine whether PcaQ·His has the ability to activate transcription *in vivo* as the influence of a foreign tag may vary between regulators and the ability of PcaQ·His to initiate transcription is an important indicator of whether this fusion protein has an activity that is comparable to the wild-type protein.

We wished to examine the regulation of *smb20568* expression more closely, and we mapped the transcriptional start site of *smb20568* via primer extension analysis (Figure 4.2). The corresponding promoter sequence was highly similar to the *pcaD* promoter region, as previously identified in *S. meliloti* (MacLean, et al., 2006) and conserved in other α-proteobacteria. Similarity between the *pcaD* and *smb20568* intergenic sequences extends beyond the promoter regions to include a putative PcaQ-binding site -73 to -58 upstream of *smb20568* that matches the position and nucleotide composition of a characterized PcaQ binding site associated with *pcaD* in *S. meliloti* (MacLean, et al., 2008).

Evidence that PcaQ regulates expression of *smb20568* via interaction with this putative binding site is three-fold. Firstly, we have determined that the PcaQ-binding motif is conserved amongst many representatives of Proteobacteria, implying that this non-coding sequence imparts a functional role (Figure 4.6). Secondly, the addition of purified PcaQ to a radiolabeled probe encompassing the candidate PcaQ-binding site resulted in an upwards shift of the probe, consistent with the interaction of this protein with a binding site encoded therein (Figure 4.3). Thirdly, the introduction of mutations within the putative PcaQ-binding site disrupted the regulated expression of *smb20568 in vivo* (Table 4.4).

We have previously described the effect of introducing point mutations within a PcaQ binding site upstream of *pcaD* in *S. meliloti* (MacLean, et al., 2008) and the identification of a second putative binding site upstream of *smb20568* presented an opportunity to expand upon our earlier analyses. Each of the six positions targeted for mutagenesis in this study are relevant to the regulation of *smb20568* expression, as replacement of any of these conserved nucleotides resulted in an abnormal expression of

the transporter gene (Table 4.4 and Figure 4.5). In four of these positions (A(-73)G; A(-71)G; T(-61)G; A(-59)G), the fold-induction of *smb20568*::*gfpuv* expression in the presence of protocatechuate was reduced considerably (1.5 to 3.6-fold induction) from that observed with the wild-type sequence (10.9-fold induction). A comparison of a similar experiment involving the binding site upstream of *pcaD* in *S. meliloti* (MacLean, et al., 2008) indicates that our data correspond well with those of the previous study (Figure 4.7). As well, we note a general correlation between the degree of nucleotide conservation at a particular position and the severity of the regulatory phenotype upon mutagenesis. For example, the greatest effect upon the regulated expression of both *pcaD* and *smb20568* was observed upon mutagenesis of A(-59)G, a position that is invariant in all PcaQ-binding sites predicted in this study (Figure 4.6).

Examination of the predicted PcaQ-binding sites in α-, β-, and γ-protoebacteria reveals a highly conserved cytosine at a position adjacent to the left-arm of the site (Figure 4.6). We generated a transversion mutation at this position (C(-69)A) and observed that expression of *smb20568* was no longer induced by protocatechuate, but rather remained constitutive at a level approximately 2.5-fold greater than that observed in the uninduced wild-type control. This constitutive expression was independent of PcaQ as a comparable level of expression was also observed in RmP1676 (*pcaQ*::Ω) (Figure 4.5; pTH2426). Constitutive expression of *pcaD* has been reported in *A. tumefaciens*, however the mutation in that instance mapped to the -35 hexanucleotide region of the gene (Parke, 1996a). In this case, the mutation is located upstream of the *smb20568* promoter and the increased, constitutive, gene expression might be attributed to the generation of an UP element. UP elements consist of AT-rich sequence upstream of a promoter that interact with the C-terminal domain of the RNA polymerase α-subunit (Ross, et al., 1993). We note that the position of the mutation (-69) is farther upstream than might be expected to influence expression (typically –60 to –40) however UP elements may still exert an effect if positioned further upstream in an appropriate orientation relative to the -10 and -35 hexamers (Ross, et al., 1993; Gourse, et al., 2000). In any case, the transversion of C → A within the context of the AT-rich PcaQ-binding

site appears to have strengthened interactions with the RNA polymerase in a manner that is independent of PcaQ activation.

In another instance, the introduction of a point mutation within the PcaQ-binding site resulted in a stronger up-regulation of *smb20568* expression. The substitution of A(-62)G was performed to generate a site upstream of *smb20568* that more closely approximates the consensus PcaQ-binding site. The *smb20568* sequence represented only four of the 38 predicted binding sites to encode an A residue in lieu of G at this position (Figure 4.6), and previous analyses of the *S. meliloti pcaD*-associated binding site indicated that mutagenesis of the conserved G had a deleterious effect upon the regulation of *pcaD* expression (MacLean, et al., 2008). Hence, we hypothesized that an A(-62)G mutation might increase the protocatechuate-inducible expression of *smb20568*.

The site directed mutation (A(-62)G) created a regulatory sequence that permitted a greater than 20-fold induction of *smb20568* expression in the presence of protocatechuate (Table 4.4). In contrast, the wild-type regulatory binding site yielded an 11-fold increase in *smb20568* expression. Thus, the presence of an adenosine at this position (as encoded upstream of *smb20568*) resulted in a lower level of induction than that observed when a guanosine was encoded within the sequence (as encoded in the majority of binding sites). Of the four binding sites encoding adenosine at this position, three sites are associated with ABC-type transport systems which are located near *pca* genes and presumably mediate the uptake of protocatechuate (in *R. etli*, *S. meliloti*, *S. medicae*). In addition, we note that a transition mutation may have occurred in the *R. etli* motif to replace a thymidine (encoded in all other sites) with a cytosine, within the right-arm of the site (T(-60)C). A putative PcaQ-binding site is positioned upstream of another ABC-type transport gene cluster in *M. loti* and similarly contains a mutation which (based upon our analyses in *S. meliloti*) would significantly reduce the level of induced gene expression. In the *M. loti* site, a 1 nucleotide deletion likely occurred to replace an adenosine (encoded in all other predicted PcaQ-binding sites) with a cytosine, within the left-hand arm of the binding site (A(-73)C). Finally, binding sites located upstream of transport genes in *Brucella* species also contain a substitution (T(-61)A) within a

conserved position of the right-hand arm. We have demonstrated that mutations comparable to those in the *R. etli, M. loti,* and *Brucella* sp. sites (A(-73)G ; T(-61)G ; T(-60)G; upstream of *pcaD* and/or *smb20568* in *S. meliloti*) resulted in a significantly decreased level of gene expression under inducing conditions (ranging from 15 to 41% expression observed from the wild-type regulatory region). Of the 10 predicted PcaQ-binding sites positioned upstream of transport genes, 9 contain nucleotide substitutions that likely reduce the level of gene expression under inducing conditions (identified in red font; Figure 4.6). The preferential accumulation of mutations within this subset of PcaQ-binding sites may simply be a reflection of the non-essential nature of aromatic acid transport systems. Protocatechuate (and related compounds) is sufficiently hydrophobic to permit passive diffusion across the cell membrane. The functional significance of encoding transport systems with a decreased responsiveness to protocatechuate (in terms of gene expression) nonetheless imparts an interesting physiological consequence. Although utilized as an energy source by many prokaryotes, protocatechuate (and the metabolite β-carboxy-*cis, cis*-muconate) is toxic if allowed to accumulate to high intracellular concentrations. Possibly the presence of nucleotide substitutions within the PcaQ binding sites may reflect an adaptation to control the import of protocatechuate into the cell by limiting the expression of a relevant transport system. Quite simply, a high level of expression of an aromatic acid transport system may be maladaptive in an environment rich in protocatechuate and related compounds.

An interesting result from the *in silico* prediction of PcaQ-binding sites was the identification of such motifs upstream of *pcaHG* in *Pseudomonas*, *Ralstonia*, and *Burkholderia*. The genomes of these genera were selected for analysis based upon an annotation of the LysR-protein PcaQ in the genome of atleast one species from each group. While the presence of a PcaQ homologue in *Pseudomonas* has been noted (Overhage, et al., 1999; Jimenez, et al., 2002), there is currently no experimental evidence supporting the regulation of *pca* genes by PcaQ in these bacteria. Expression of *pcaHG* is known to be up-regulated by growth with protocatechuate in *Pseudomonas*, however the regulator involved has never been identified (Harwood & Parales, 1996;

Jimenez, et al., 2002). The presence of candidate PcaQ-binding sites indicates that expression of *pcaHG* is likely regulated by PcaQ, and offers the first direct evidence that PcaQ-mediated regulation may extend to species outside of the class α-proteobacteria. We note that the binding sites identified in β–proteobacteria atypically contain a 7 nucleotide spacer between the left and right-hand arms of the binding site; all other predicted sites encode 8 nucleotides (Table 4.5). Further study is required to determine what effect (if any) such a deletion might have upon the regulation of gene expression in *Ralstonia* and *Burkholderia*.

While the identification of a putative PcaQ-binding motif upstream of a gene may be suggestive of a mode of regulation, such a prediction does not constitute conclusive evidence of regulation in the absence of experimental data. This caveat is best exemplified by the detection of candidate PcaQ-binding sites in five species of *Brucella*. The PcaQ-binding motifs identified upstream of *pcaD* in *Brucella* species match the consensus sequence, however the target *pcaDCHGB* operons lack atleast one gene necessary for the catabolism of protocatechuate in three of the five species (Table 4.5). For example, *B. ovis* encodes frameshift mutations in the majority of (pseudo)genes relevant to the metabolism of protocatechuate and *p*-hydroxybenzoate (*pcaD, pcaH, pcaB, pcaJ, pobR*), and thus it is extremely unlikely that this species has the ability to degrade these compounds. The loss of a peripheral metabolic pathway (such as the β–ketoadipate pathway) in *B. ovis, B. canis*, and *B. abortus* reflects the lifestyle adopted by these facultative intracellular animal pathogens, as the ability to metabolize aromatic acids may no longer be relevant to the ecological success of these bacteria and their genomes may be undergoing contraction via gene decay. In contrast, *B. suis* and *B. melitensis* encode the intact β-ketoadipate pathway, and it has been proposed that the persistence of *B. suis* in contaminated soils may in part be due to the β–ketoadipate genes (Paulsen, et al., 2002). Regardless, the PcaQ-binding sites in (atleast) *B. ovis*, *B. canis* and *B. abortus* are unlikely to contribute to the functional regulation of gene expression in these species.

Within members of γ-proteobacteria, transport systems related to aromatic acid catabolism belong to the major facilitator superfamily (MFS) of transporter proteins (Harwood, et al., 1994; Collier, et al., 1997; Williams & Shaw, 1997; Leveau, et al., 1998; D'Argenio, et al., 1999; Chaudhry, et al., 2007). We have described the regulation of an ABC-type transport system that we infer is involved in the import of protocatechuate in the α–proteobacteria *S. meliloti*. The evidence indicating that *smb20568-smb20784* encode a system involved in the transport of protocatechuate is as follows: i) *smb20568-smb20784* are located in close proximity to known *pca* catabolic genes, and aromatic acid genes are often clustered together within supraoperons; ii) expression of *smb20568* is induced by growth in the presence of protocatechuate; iii) expression of *smb20568* is regulated by the LysR-type protein PcaQ, a regulator that also modulates expression of protocatechuate catabolism genes (*pcaDCHGB*). The recruitment of an ABC-type transport system for the uptake of an aromatic acid in *S. meliloti* likely reflects the relative abundance of this family of transport systems in the *S. meliloti* genome (Galibert, et al., 2001). We have identified additional ABC-type transport systems in *Sinorhizobium medicae*, *R. etli*, *R. leguminosarum*, *M. loti*, and *Brucella* species that are encoded near aromatic acid catabolism genes and include potential PcaQ-binding sites (Figure 4.6). In *R. etli* and *R. leguminosarum*, the transport genes are linked to a putative *p*-hydroxybenzoate hydroxylase gene (on the chromosome) and not to *pca* genes (located on plasmids p42e and pRL11, respectively) raising the possibility that *p*-hydroxybenzoate is also (or alternatively) recognized as a substrate by these systems.

Additional experiments are necessary to directly demonstrate that the system encoded by *smb20568-smb20784* participates in the uptake of protocatechuate. Particularly, transport assays utilizing radiolabeled protocatechuate are essential to monitor uptake in wild-type *S. meliloti* and a mutant strain(s) in which the genes encoding the putative uptake system have been disrupted. We have constructed two *S. meliloti* strains (as described in materials and methods) in which expression of the uptake genes is disrupted via an in-frame deletion of *smb20568* (RmP1710) or through the

integration of an antibiotic resistance cassette within *smb20787* (RmP1712). Transport and growth assays involving these strains are necessary and will be performed to conclusively determine whether *smb20568-smb20784* encode a novel aromatic acid uptake system.

4.6 References

Allaway, D., Schofield, N.A., Leonard, M.E., Gilardoni, L., Finan, T.M., & Poole, P.S. (2001). Use of differential fluorescence induction and optical trapping to isolate environmentally induced genes. Environmental Microbiology, **3**:397-406.

Buchan, A., Collier, L.S., Neidle, E.L., & Moran, M.A. (2000). Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. Applied and Environmental Microbiology, **66**:4662-4672.

Buchan, A., Neidle, E.L., & Moran, M.A. (2004). Diverse organization of genes of the β-ketoadipate pathway in members of the marine *Roseobacter* lineage. Applied and Environmental Microbiology, **70**:1658-1668.

Buchan, A., Neidle, E.L., & Moran, M.A. (2001). Diversity of the ring-cleaving dioxygenase gene *pcaH* in a salt marsh bacterial community. Applied and Environmental Microbiology, **67**:5801-5809.

Bundy, B.M., Collier, L.S., Hoover, T.R., & Neidle, E.L. (2002). Synergistic transcriptional activation by one regulatory protein in response to two metabolites. Proceedings of the National Academy of Sciences U.S.A., **99**:7693-7698.

Chaudhry, M.T., Huang, Y., Shen, X.H.., Poetsch, A., Jiang, C.Y., & Liu, S.J. (2007). Genome-wide investigation of aromatic acid transporters in *Corynebacterium glutamicum*. Microbiology, **153**:857-865.

Chugani, S.A.; Parsek,M.R.; Chakrabarty,A.M. (1998). Transcriptional repression mediated by LysR-type regulator CatR bound at multiple binding sites. Journal of Bacteriology, **180**:2367-2372.

Collier, L.S., Nichols, N.N., & Neidle, E.L. (1997). *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. Journal of Bacteriology, **179**:5943-5946.

Cowie, A., Cheng, J., Sibley, C.D., Fong, Y., Zaheer, R., Patten, C.L., Morton, R.M., Golding, G.B., & Finan, T.M. (2006). An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. Applied and Environmental Microbiology, **72**:7156-7167.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Research, **14**:1188-1190.

D'Argenio, D.A., Segura, A., Coco, W.M., Bunz, P.V., & Ornston, L.N. (1999). The physiological contribution of *Acinetobacter* PcaK, a transport system that acts upon protocatechuate, can be masked by the overlapping specificity of VanK. Journal of Bacteriology, **181**:3505-3515.

Friedman, A.M., Long, S.R., Brown, S.E., Buikema, W.J., & Ausubel, F.M. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene, **18**:289-296.

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.

Gourse, R.L., Ross, W., & Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the α-subunit of RNA polymerase in promoter recognition. Molecular Microbiology, **37**:687-695.

Harwood, C.S., & Parales, R.E. (1996). The β-ketoadipate pathway and the biology of self-identity. Annual Review of Microbiology, **50**:553-590.

Harwood, C.S., Nichols, N.N., Kim, M.K., Ditty, J.L., & Parales, R.E. (1994). Identification of the *pcaRKF* gene cluster from *Pseudomonas putida*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. Journal of Bacteriology, **176**:6479-6488.

Jiménez, J.I., Miñambres, B., García, J.L., Díaz, E. (2002). Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environmental Microbiology, **4**:824-841.

Leveau, J.H., Zehnder, A.J., & van der Meer, Jr. (1998). The *tfdK* gene product facilitates uptake of 2,4-dichlorophenoxyacetate by *Ralstonia eutropha* JMP134 (pJP4). Journal of Bacteriology, **180**:2237-2243.

MacLean, A.M., Anstey, M.I., & Finan, T.M. (2008). Binding site determinants for the LysR-type transcriptional regulator PcaQ in the legume endosymbiont *Sinorhizobium meliloti*. Journal of Bacteriology, **190**:1237-1246.

MacLean, A.M., Finan, T.M., & Sadowsky, M.J. (2007). Genomes of the symbiotic nitrogen-fixing bacteria of legumes. Plant Physiology, **144**:615-622.

MacLean, A.M., MacPherson, G., Aneja, P., & Finan, T.M. (2006). Characterization of the β-ketoadipate pathway in *Sinorhizobium meliloti*. Applied and Environmental Microbiology, **72**:5403-5413.

MacLellan, S.R., MacLean, A.M., & Finan, T.M. (2006). Promoter prediction in the rhizobia. Microbiology, **152**:1751-1763.

MacLellan, S.R., Smallbone, L.A., Sibley, C.D., & Finan, T.M. (2005). The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of α-proteobacterial plasmids. Molecular Microbiology, **55**:611-623.

Mauchline, T.H., Fowler, J.E., East, A.K., Sartor, A.L., Zaheer, R., Hosie, A.H., Poole, P.S., & Finan, T.M. (2006). Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proceedings of the National Academy of Sciences of U.S.A., **103**:17933-17938.

Ogawa, N., McFall, S.M., Klem, T.J., Miyashita, K., & Chakrabarty, A.M. (1999). Transcriptional activation of the chlorocatechol degradative genes of *Ralstonia eutropha* NH9. Journal of Bacteriology, **181**:6697-6705.

Oke, V., & Long, S.R. (1999). Bacterial genes induced within the nodule during the *Rhizobium*-legume symbiosis. Molecular Microbiology, **32**:837-849.

Overhage, J., Kresse, A.U., Priefert, H., Sommer, H., Krammer, G., Rabenhorst, J., Steinbüchel, A. (1999). Molecular characterization of the genes *pcaG* and *pcaH*, encoding protocatechuate 3,4-dioxygenase, which are essential for vanillin catabolism in *Pseudomonas* sp. strain HR199. Applied and Environmental Microbiology, **65**:951-960.

Parke, D. (1996a). Characterization of PcaQ, a LysR-type transcriptional activator required for catabolism of phenolic compounds, from *Agrobacterium tumefaciens*. Journal of Bacteriology, **178**:266-272.

Parke, D. (1996b). Conservation of PcaQ, a transcriptional activator of *pca* genes for catabolism of phenolic compounds, in *Agrobacterium tumefaciens* and *Rhizobium* species. Journal of Bacteriology, **178**:3671-3675.

Parke, D. (1995). Supraoperonic clustering of *pca* genes for catabolism of the phenolic compound protocatechuate in *Agrobacterium tumefaciens*. Journal of Bacteriology, **177**:3808-3817.

Parke, D. (1993). Positive regulation of phenolic catabolism in *Agrobacterium tumefaciens* by the *pcaQ* gene in response to β-carboxy-*cis,cis*-muconate. Journal of Bacteriology, **175**:3529-3535.

Parke, D., & Ornston, L.N. (1986). Enzymes of the β-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium* spp. and constitutive in *Bradyrhizobium* spp. Journal of Bacteriology, **165**:288-292.

Parke, D., & Ornston, L.N. (1984). Nutritional diversity of *Rhizobiaceae* revealed by auxanography. Journal of General Microbiology, **130**:1743-1750.

Parke, D., Ornston, L.N., & Nester, E.W. (1987). Chemotaxis to plant phenolic inducers of virulence genes is constitutively expressed in the absence of the Ti plasmid in *Agrobacterium tumefaciens*. Journal of Bacteriology, **169**:5336-5338.

Parke, D., Rivelli, M., & Ornston, L.N. (1985). Chemotaxis to aromatic and hydroaromatic acids: comparison of *Bradyrhizobium japonicum* and *Rhizobium trifolii*. Journal of Bacteriology, **163**:417-422.

Parke, D., Rynne, F., & Glenn, A. (1991). Regulation of phenolic catabolism in *Rhizobium leguminosarum* biovar *trifolii*. Journal of Bacteriology, **173**:5546-5550.

Paulsen, I.T., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J., Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O., Salzberg, S.L,. Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M., Fraser, C.M. (2002). The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proceedings of the National Academy of Sciences U.S.A. **99**:13148-13153.

Prentki, P., & Krisch, H.M. (1984). In vitro insertional mutagenesis with a selectable DNA fragment. Gene, **29**:303–313.

Quandt, J., & Hynes, M.F. (1993). Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. Gene, **127**:15-21.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., & Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the α-subunit of RNA polymerase. Science, **262**:1407-1413.

Williams, P.A., & Shaw, L.E. (1997). *mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow on exogenous *cis,cis*-muconate as the sole carbon source. Journal of Bacteriology, **179**:5935-5942.

Yuan, Z.C., Zaheer, R., & Finan, T.M. (2006). Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. Journal of Bacteriology, **188**:1089-1102.

Table 4.1 Strains and plasmids used in this study.

Strain or plasmid	Relevant characteristics	Source or reference
Strains		
E. coli		
DH5α	F⁻ Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *deoR recA1 endA1 hsdR17*(r_K^- m_K^-) *supE44* λ^- *thi-1 gyrA96 relA1*	Bethesda Research Laboratories, Inc.
S. meliloti		
RmP110	Rm1021 with wild-type *pstC*; Sm^r	Yuan, et al., 2006
RmP708	RmP110 *pcaR*::Ω; Sm^r, Sp^r	This study
RmP1676	RmP110 *pcaQ*::Ω; Sm^r, Sp^r	MacLean, et al., 2008
RmP1710	RmP110 Δ*smb20568*, 687 bp in-frame deletion created via pTH1959; Sm^r, Gm^s, $sucrose^r$	This study
RmP1712	RmP110 *smb20787*::Ω; Sm^r, Sp^r	This study
RmP1811	RmP110 *pcaQ*::Ω *smb20568*::*gfp*⁺/*lacZ*; Sm^r, Sp^r, Gm^r	This study
Plasmid		
pHP45Ω	pBR322 derivative with Ω element; Amp^r, Sp^r, Sm^r	Prentki & Krisch, 1984
pJQ200 uc1	Suicide vector with *sacB* to select for plasmid excision; Gm^r	Quandt & Hynes, 1993
pLAFR1	IncP cosmid cloning vector; Tc^r	Friedman, et al., 1982
pOT1	Broad host range *gfpuv* transcription reporter; Gm^r	Allaway, et al., 2001
pUCP30T	Cloning vector; Gm^r	Genbank accession no. U33752
pFL1131	2,619 bp fragment extending from within *smb20571* to *smb20568* (1581256 – 1583875 nts; pSymB) in pTH1522; Gm^r	Cowie, et al., 2006
pFL2211	1,616 bp fragment spanning *pcaD/Q* intergenic region (IG) (1592430 - 1594046) in pTH1522; Gm^r	Cowie, et al., 2006
pFL2665	1,623 bp fragment spanning *smb20568/smb20787* IG region (1579505 - 1581128) in pTH1522; Gm^r	Cowie, et al., 2006
pTH1227	Broad-host-range derivative of pFus1 with P_{tac} promoter inserted	J. Cheng and T.M. Finan,

	upstream of *gusA*; Tc^r	unpublished data
pTH1508	Broad-host-range plasmid used for integration with pTH1522 derivatives via *attP/attB*; Tc^r	J. Cheng and T.M. Finan, unpublished data
pTH1522	Plasmid used in construction of *S. meliloti* gene fusion library; *gfp*$^+$, *lacZ*, *rfp*, *gusA*; Gm^r	Cowie, et al., 2006
pTH1705	Derivative of pTH1522; transcription fusion reporter plasmid; may be integrated with pTH1508 via *attP* to form replicating plasmid; Gm^r	Cowie, et al., 2006
pTH1883	pVO155 derivative lacking *gusA*, suicide plasmid in *S. meliloti*; Amp^r, Kn^r	Lab collection
pTH1908	1,149 bp PCR amplified fragment spanning *smb20787* into pTH1883 via *Not*I; Amp^r, Kn^r	This study
pTH1913	Integration of pFL1131 with pTH1508 (*smb20568*::*gfp*$^+$/*lacZ*); Gm^r, Tc^r	This study
pTH1914	Integration of pFL2665 with pTH1508 (*smb20787*::*gfp*$^+$/*lacZ*); Gm^r, Tc^r	This study
pTH1917	ΩSm/Sp^r into pTH1908 via *Sac*II; Amp^r, Kn^r, Sm^r, Sp^r	This study
pTH1927	Integration of pFL2211 with pTH1508 (*pcaD*::*gfp*$^+$/*lacZ*); Gm^r, Tc^r	This study
pTH1948	1,965 bp PCR amplified fragment spanning *smb20568* into pUCP30T via *Xba*I/*Hind*III; Gm^r	This study
pTH1949	pTH1948 with a 687 bp deletion via *Pst*I; Gm^r	This study
pTH1959	*Xba*I/*Hind*III insert from pTH1949 into pJQ200 via SmaI; Gm^r	This study
pTH1960	1,123 bp PCR amplified region encompassing *pcaI* promoter into pTH1705 via *Bgl*II/*Kpn*I; Gm^r	This study
pTH1961	627 bp PCR amplified region encompassing *smb20568/smb20569* IG region into pTH1705 via *Bgl*II/*Kpn*I; Gm^r	This study
pTH1971	Integration of pTH1960 with pTH1508 (*pcaI*::*gfp*$^+$/*lacZ*); Gm^r, Tc^r	This study

pTH1972	Integration of pTH1961 with pTH1508 (*smb20568*::*gfp*$^{+}$/*lacZ*); Gm^r, Tc^r	This study
pTH1973	Integration of pTH1962 with pTH1508 (*smb20570*:: *gfp/lacZ*); Gm^r, Tc^r	This study
pTH2410	242 bp PCR product encompassing *smb20568/smb20569* IG into pUCP30T via *Xba*I; Gm^r	This study
pTH2414	246 bp PCR amplified insert from pTH2410 into pOT1 (*smb20568*::*gfpuv*); Gm^r	
pTH2424	246 bp PCR amplified insert from pTH2410 derivative with A(-62)G^{a} site directed mutation (SDM) into pOT1; Gm^r	This study
pTH2425	246 bp PCR amplified insert from pTH2410 derivative with A(-62)C SDM into pOT1; Gm^r	This study
pTH2426	246 bp PCR amplified insert from pTH2410 derivative with C(-69)A SDM into pOT1; Gm^r	This study
pTH2427	246 bp PCR amplified insert from pTH2410 derivative with A(-59)G SDM into pOT1; Gm^r	This study
pTH2428	246 bp PCR amplified insert from pTH2410 derivative with T(-61)G SDM into pOT1; Gm^r	This study
pTH2432	246 bp PCR amplified insert from pTH2410 derivative with A(-73)G SDM into pOT1; Gm^r	This study
pTH2433	246 bp PCR amplified insert from pTH2410 derivative with A(-71)G SDM into pOT1; Gm^r	This study
pTH2454	PCR amplified *pcaQ·his* into pTH1227 via *Bgl*II/*Eco*RI; Tc^r	This study

Table 4.2. Expression of *pca* genes in wild-type and regulator-minus strains of *S. meliloti*

Strain Genotype	Gene Fusion	β-galactosidase Activity[a] (Miller units) [SD] Uninduced	Induced	Fold Induction
RmP110 (pTH1927)	*pcaD*	264 (6)	2474 (134)	9
(pTH1971)	*pcaI*	220 (15)	1020 (65)	5
(pTH1972)	*smb20568*	353 (16)	2354 (8)	7
RmP110 *pcaR*::Ω (pTH1927)	*pcaD*	184 (6)	1416 (71)	8
(pTH1971)	*pcaI*	208 (25)	205 (4)	1
(pTH1972)	*smb20568*	132 (5)	956 (49)	7
RmP110 *pcaQ*::Ω (pTH1927)	*pcaD*	141 (7)	145 (12)	1
(pTH1971)	*pcaI*	75 (4)	371 (43)	5
(pTH1972)	*smb20568*	162 (8)	158 (11)	1

[a] Shown are averages of assays obtained using *S. meliloti* strains subcultured into M9 minimal medium with 0.5% glycerol in the absence (uninduced) and presence (induced) of 5 mM protocatechuate.
SD, standard deviation

Table 4.3. Expression of *smb20568* as regulated by PcaQ·His in *S. meliloti*

Strain	β-galactosidase Activity[a] (Miller units) [SD]		Fold Induction
	Uninduced	Induced	
RmP1811 (pTH1227)	117 (9)	116 (25)	1
RmP1811 (pTH2454)	115 (12)	696 (14)	6

[a] Shown are averages of assays obtained using *S. meliloti* strains subcultured into M9 minimal medium with 0.5% glycerol in the absence (uninduced) and presence (induced) of 5 mM protocatechuate. RmP1811, RmP110 *pcaQ*::Ω *smb20568*::*gfp*$^{+}$/*lacZ*
SD, standard deviation

Table 4.4. Analysis of expression of *smb20568::gfpuv* in *S. meliloti*.

Plasmid	Position of Mutation[b]	GfpUV Specific Activity[a] [SD]		Fold Induction [SD]
		Uninduced	Induced	
pTH2414	None	2,472 (280)	26,995 (1,839)	11[c]
pTH2432	A(-73)G	1,287 (279)	3,363 (311)	3
pTH2433	A(-71)G	1,152 (204)	3,153 (430)	3
pTH2426	C(-69)A	6,655 (1,911)	6,515 (939)	1
pTH2424	A(-62)G	2,390 (187)	51,708 (3,971)	22 (3)
pTH2425	A(-62)C	1,751 (94)	20,720 (1,691)	12 (2)
pTH2428	T(-61)G	1,305 (171)	4,729 (230)	4
pTH2427	A(-59)G	1,965 (490)	2,982 (443)	2

[a] Shown are averages of three independent experiments involving *S. meliloti* strains subcultured into M9 minimal medium with 0.5% glycerol ± 5 mM protocatechuate. In each experiment, assays are performed in triplicate. SD, standard deviation.

[b] Position indicated with respect to *smb20568* transcriptional start site.

[c] When not indicated, standard deviation is less than 1.

Table 4.5. Description of putative PcaQ-binding motifs in Proteobacteria

Species	Class	Nucleotide sequence	Position	Replicon
A. caulinodans ORS 571		ATAA CCCTCAGG TTAT	*pcaDHGB*	Chromosome
A. tumefaciens C58	α	ATAA TCCACAGG TTAT	*pcaDCHGB*	Linear chromsome
B. abortus S19	α	ATAA CCTGTGAG TTAT	*pcaDCHG*	Chromosome II
		ATAA CTGTTCGG ATAT	ABC-type transporter	Chromosome II
B. canis ATCC 23365	α	ATAA CCTGTGAG TTAT	*pcaDCGB*	Chromosome II
		ATAA CTGTTCGG ATAT	ABC-type transporter	Chromosome II
B. melitensis 16M	α	ATAA CCTGTGAG TTAT	*pcaDCHGB*	Chromosome II
		ATAA CTGTTCGG ATAT	ABC-type transporter	Chromosome II
B. ovis ATCC 25840	α	ATAA CCTGTGAG TTAT	*pcaCG*	Chromosome II
		ATAA CTGTTCGG ATAT	ABC-type transporter	Chromosome II
B. suis 1330	α	ATAA CCTGTGAG TTAT	*pcaDCHGB*[a]	Chromosome II
		ATAA CTGTTCGG ATAT	ABC-type transporter	Chromosome II
M. loti MAFF303099	α	ATAA CCCCCGAG TTAA	*pcaDCHGBpobA*	Chromosome
		CTAA CCCCAGGG TTAA	ABC-type transporter	Chromosome
Mesorhizobium sp. BNCI	α	ATAA CTTTGCGA TTAA	4-hydroxybenzoyl-CoA thioesterase	Chromosome
O. anthropi ATCC 49188	α	ATAA CCCTGTAG TTAT	*pcaDCHGB*	Chromosome II
R. etli CIAT	α	ATAA CTCCTAAG TTAT	*pcaDCHGB*	pA
R. etli CFN42	α	ATAA CCGGCAAA TCAT	ABC-type transporter	Chromosome
		ATAA CCTCGGAG TTAT	*pcaDCHGB*	p42e
R. leguminosarum bv. viciae 3841	α	ATAA TTGTACAG TTAT	ABC-type transporter	Chromosome
		ATAA CTCCACGG TTAT	*pcaDCHGB*	pRL11
R. leguminosarum bv. *trifolii* WSM2304	α	ATAA CTCCCAGG TTAT	*pcaDCHGB*	pRLG202
S. meliloti Rm1021	α	ATAA CCGGGGGA TTAT	*smb20568*	pSymB

		ATAA CCTCCTGG TTAA	*pcaDCHGB*	pSymB
S. medicae WSM419	α	ATAA CCGGGGAA TTAT	ABC-type transporter	pSMED01
		ATAA CTCCCTGG TTAA	*pcaDCHGB*	pSMED01
B. multivorans ATCC 17616	β	ATAA CCACCCG TTAT	*pcaHG*	Chromosome II
B. pseudomallei 1106	β	ATAA CAGTTAG TTAT	*pcaHG*	Chromosome II
B. vietnamiensis G4	β	ATAA CTCCCGG TTAT	*pcaHG*	Chromosome II
R. solanacearum MolK2	β	ATAA CATCCGG TTAT	*pcaHG*	Chromosome
P. aeruginosa PA7	γ	ATAA CCAAACGG TTAT	*pcaHG*	Chromosome
P. aeruginosa PAO1	γ	ATAA CCAAATGG TTAT	*pcaHG*	Chromosome
P. aeruginosa UCBPP-PA14		ATAA CCAAATGG TTAT	*pcaHG*	Chromosome
P. fluorescens Pf-5	γ	ATAA CCATTTGG TTAT	*pcaHG*	Chromosome
P. putida W619	γ	ATAA CCATTTGG TTAA	*pcaHG*	Chromosome
P. stutzeri A1501	γ	ATAA CCCCTGGG TTAT	*pcaHG*	Chromosome
P. syringae pv. *phaseolicola* 1448A	γ	ATAA CCATTTGG TTAT	*pcaHGKBC*	Chromosome
P. syringae pv. tomato str. DC3000	γ	ATAA CCAATTGG TTAT	*pcaHGKBC*	Chromosome

[a] The gene annotated as *pcaL* is incorrectly annotated, and encodes *pcaD*.

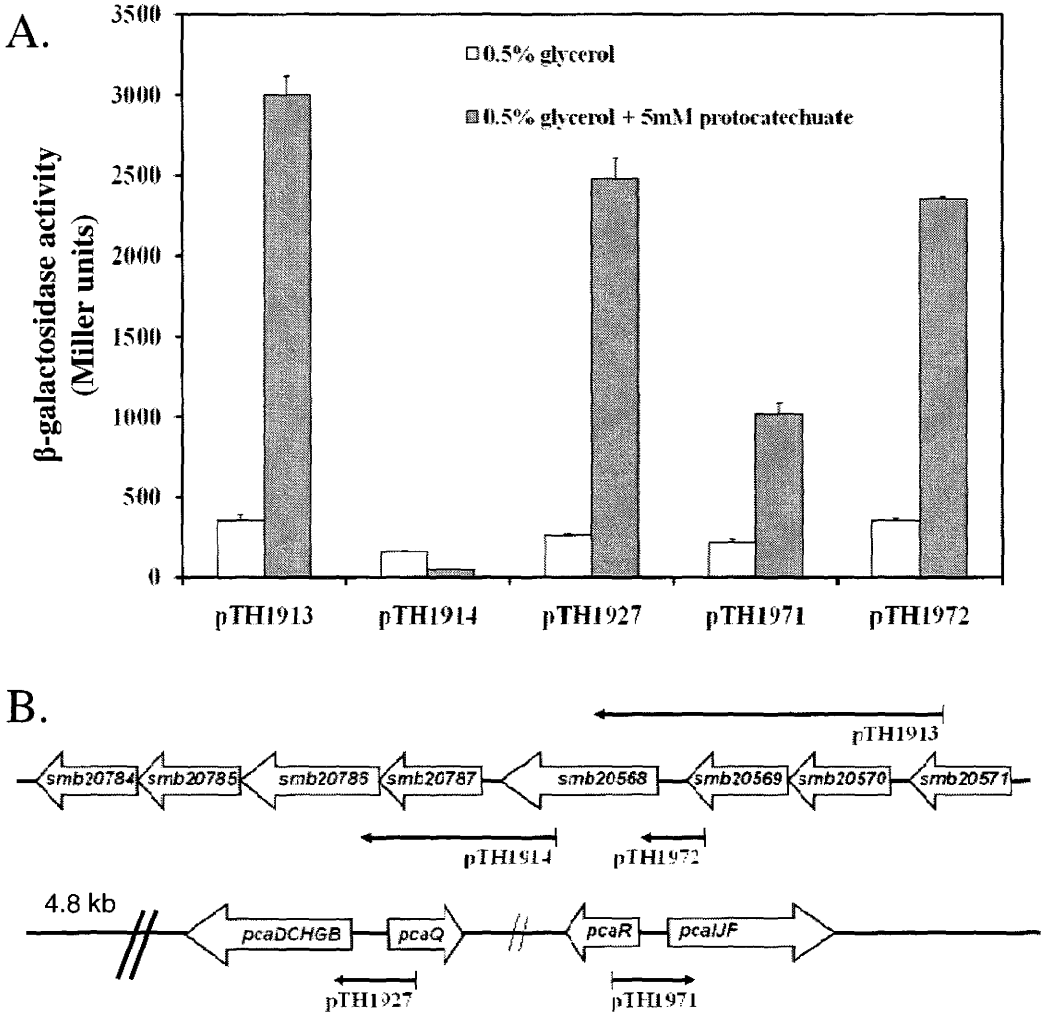


Figure 4.1. Expression and organization of genes encoding an ABC-type transport system we infer is involved in the uptake of protocatechuate in *S. meliloti*. (A) Expression of genes encoding the transport system as measured by β–galactosidase activity. Reporter enzyme assays were performed upon *S. meliloti* wild-type strain RmP110 carrying a plasmid with a transcriptional gene fusion between an *S. meliloti* insert and *gfp+/lacZ* as depicted in panel B. (B) Organization of genes encoding the putative protocatechuate uptake system (*smb20568-smb20784*) and metabolic enzymes (*pcaDCHGB* and *pcaIJF*) on the pSymB megaplasmid. Gene annotation as follows: *smb20784* and *smb20785*, ATP-binding protein; *smb20786* and *smb20787*, permease protein; *smb20568*, periplasmic solute binding protein; *smb20569*, ATP-binding protein; *smb20570*, periplasmic solute binding protein; *smb20571*, permease protein.

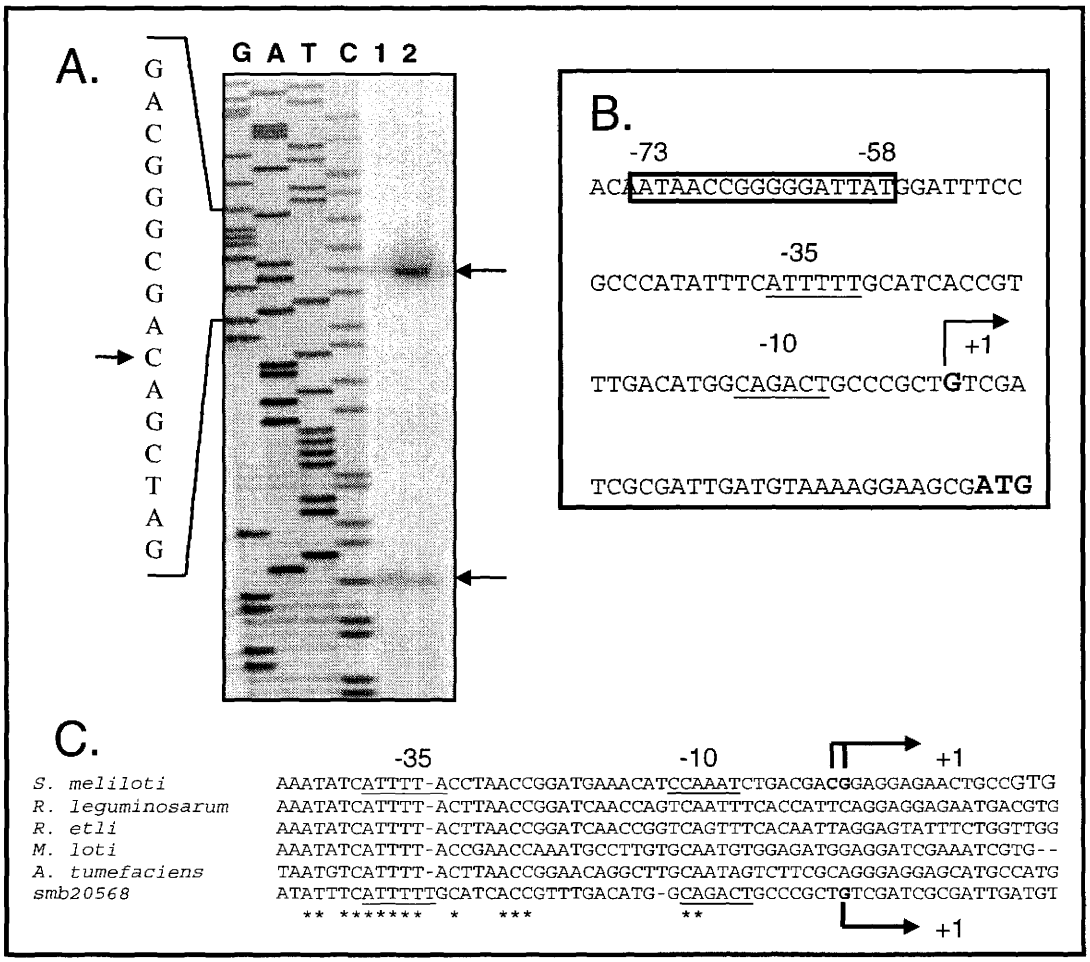


Figure 4.2. Identification and analysis of a promoter associated with an ABC-type transport system encoded by *smb20568-smb20784*. (A) Primer extension reactions were performed using mRNA isolated from *S. meliloti* wild-type strain RmP110 grown in the absence (lane 1) and presence (lane 2) of protocatechuate. Arrows (right) indicate the extension products obtained using two primers (results shown for one primer only) and an arrow (left) identifies the corresponding nucleotide. (B) Schematic depiction of the regulatory regions upstream of *smb20568*. Inferred -10 and -35 hexanucleotide regions are underlined and the transcriptional start site is identified by enlarged, bold, font with an arrow indicating the direction of transcription. A putative PcaQ binding is enclosed in a box and the translational start codon of *smb20568* is indicated (enlarged and bold font). (C) Alignment of sequence upstream of *pcaD* in *S. meliloti* (GenBank accession number, NP_438031), *Rhizobium leguminosarum* (YP_771122), *R. etli* (YP_472225), *Mesorhizobium loti* (NP_107573), and *A. tumefaciens* (AAK88901) with the promoter region of *smb20568*. Inferred -10 and -35 regions associated with *S. meliloti pcaD* and *smb20568* (MacLean, et al., 2006) are underlined; transcriptional start sites are indicated by arrows bent in the direction of transcription. Invariant nucleotides are indicated by an asterisk.

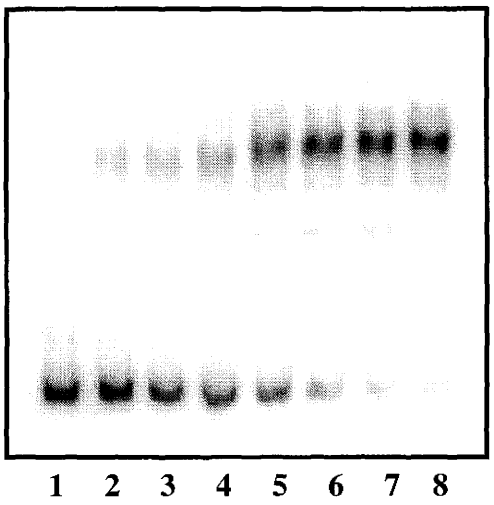


Figure 4.3. Electrophoretic mobility shift assay for PcaQ binding to the intergenic region upstream of *smb20568.* A 246 bp radiolabeled probe (extending −149 to +87; with respect to the transcriptional start site of *smb20568*) was incubated in the presence of increasing concentrations of purified PcaQ·His prior to resolution upon a nondenaturing polyacrylamide gel. Each binding reaction was performed in the presence of 500 ng herring sperm DNA as a nonspecific competitor. Lanes 1 to 8 contained 0, 0.2, 0.5, 1.2, 2.5, 4.9, 12.4, and 24.7 nM PcaQ, respectively.

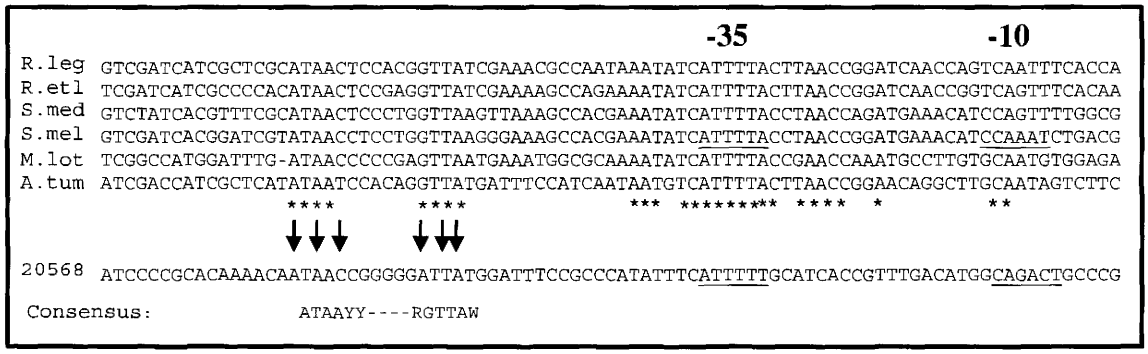
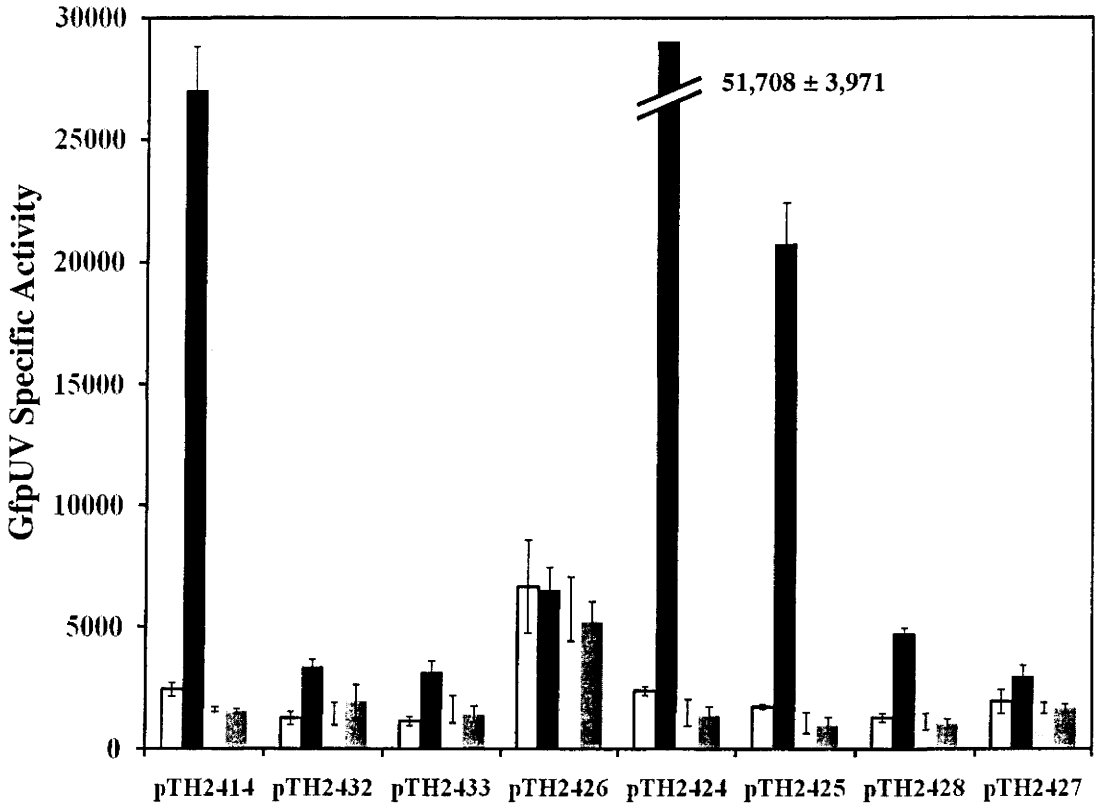


Figure 4.4. Identification of a putative PcaQ-binding site located upstream of *smb20568* in *S. meliloti*. Aligned are nucleotide sequences corresponding to sequence upstream of *pcaD* in *S. meliloti* (*S.mel*), *S. medicae* (*S.med*), *R. leguminosarum* (*R.leg*), *R. etli* (*R.etl*), *M. loti* (*M.lot*), and *A. tumefaciens* (*A.tum*) and the nucleotide sequence upstream of *smb20568*. The inferred -35 and -10 promoter regions associated with *pcaD* and *smb20568* in *S. meliloti* are underlined. Invariant nucleotides in *pcaD* sequences are indicated with an asterisk. Arrows indicate nucleotides targeted for site directed mutagenesis in this study: black arrows identify nucleotides that have been conserved in all sequences and grey arrows identify additional nucleotide that have been selected for mutagenesis.



Nucleotide	**WT**	**A**	**A**	**C**	**A**	**A**	**T**	**A**
Position		**73**	**71**	**69**	**62**	**62**	**61**	**59**
Substitution		**G**	**G**	**A**	**G**	**C**	**G**	**G**

Figure 4.5. Analysis of *smb20568*::*gfpuv* expression in *S. meliloti* wild-type and *pcaQ*::Ω strains. *S. meliloti* strains carrying plasmids as indicated were subcultured into M9-minimal medium with 0.5% glycerol ± 5 mM protocatechuate, and incubated at 30°C for four to six hours. Point mutations were introduced in a putative PcaQ-binding site as indicated and positions are given with respect to the transcriptional start site of *smb20568*. Shown is the mean expression as determined from three independent experiments; in each experiment, assays are performed in triplicate. Error bars reflect standard deviation observed between experiments. White bars, *S. meliloti* RmP110, uninduced; black bars, *S. meliloti* RmP110, induced; light grey bars, *S. meliloti* RmP1676 (RmP110 *pcaQ*::Ω), uninduced; dark grey bars, *S. meliloti* RmP1676, induced.

A.

B.canis transporter	TGCCGGAAAC	ATAA	CTGTTCGG	ATAT	CGGTTATAAGGATTTTTTCATTTTACATGACCA
B.ovis transporter	TGCCGGAAAC	ATAA	CTGTTCGG	ATAT	CGGTTATAAGGATTTTTTCATTTTACATGGCCA
B.suis transporter	TGCCGGAAAC	ATAA	CTGTTCGG	ATAT	CGGTTATAAGGATTTTTTCATTTTACATGACCA
B.abortus transporter	TGCCGGAAAC	ATAA	CTGTTCGG	ATAT	CGGTTATAAGGATTTTTTCATTTTACATGACCA
B.melitensis transp.	TGCCGGAAAC	ATAA	CTGTTCGG	ATAT	CGGTTATAAGGATTTTTTCATTTTACATGACCA
R.etli transporter	GCACAGAACC	ATAA	CCGGCAAA	TCAT	GTGTTTGGGGATTTATTTCATTTTACATCACCA
R.leg transporter	ACAC--AATC	ATAA	TTGTACAG	TTAT	GGATTACGGTACTTATTTCATTTTACATCACCA
B.canis pcaD	CACTTTT-CT	ATAA	CCTGTGAG	TTAT	GATTTCGGTCCTGAACATCATTTTACTTAACCA
B.melitensis pcaD	CACTTTT-CT	ATAA	CCTGTGAG	TTAT	GATTTCGGTCCTGAATATCATTTTACTTAACCA
B.suis pcaD	CACTTTT-CT	ATAA	CCTGTGAG	TTAT	GATTTCGGTCCTGAACATCATTTTACTTAACCA
B.abortus pcaD	CACTTTT-CT	ATAA	CCTGTGAG	TTAT	GATTTCGGTCCTGAATATCATTTTACTTAACCA
B.ovis pcaD	CACTTTT-CT	ATAA	CCTGTGAG	TTAT	GATTTCGGTCCTGGATATCATTTTACTTAACCA
O.anthropi pcaD	CACTTAT-CT	ATAA	CCCTGTAG	TTAT	GATTTTTGCCAATAATATCATTTTACTTAACCA
M.loti transporter	TTCGAGTCCC	CTAA	CCCCAGGG	TTAA	TGAAACACCGCAAAATATCATTTTACCGAACCA
M.loti pcaD	CATGGATTTG	ATAA	CCCCCGAG	TTAA	TGAAATGGCGCAAAATATCATTTTACCGAACCA
R.etli p42e pcaD	ATCGCCCCAC	ATAA	CTCCGAGG	TTAT	CGAAAAGCCAGAAAATATCATTTTACTTAACCG
R.leg pcaD	ATCGCTC-GC	ATAA	CTCCACGG	TTAT	CGAAACGCCAATAAATATCATTTTACTTAACCG
R.etli pA pcaD	ATCGACC-AC	ATAA	CTCCTAAG	TTAT	CGAAAAGCCAGAAAATATCATTTTACTTAACCG
A.tumefaciens pcaD	ATCGCTC-AT	ATAA	TCCACAGG	TTAT	GATTTCCATCAATAATGTCATTTTACTTAACCG
A.caulinodans pcaD	CAT---CCAC	ATAA	CCCTCAGG	TTAT	AGATCGTTCCGGAAATATCATTTTACGCGGGGC
S.meliloti pcaD	CACGGATCGT	ATAA	CCTCCTGG	TTAA	GGGAAAGCCACGAAATATCATTTTAACCTAACC
R.leg WSM2304 pcaD	CATCGCTCAC	ATAA	CTCCCAGG	TTAT	CGAAAGGCTGATAAATATCATTTTACTTAACCG
S.medicae pcaD	CACGTTTCGC	ATAA	CTCCCTGG	TTAA	GTTAAAGCCACGAAATATCATTTTACCTAACCA
P.syr pv phaseo. pcaH	AGTAT---GC	ATAA	CCATTTGG	TTAT	GGATAACC-GGTTTTATTTCAATTTTCTATAAG
P.syr pv tomato pcaH	AGTAT---GC	ATAA	CCAATTGG	TTAT	GGATGCAT-GGTTTTATTTCAGTTCTCTGTAAG
P.fluo Pf5 pcaH	GTTAT---CC	ATAA	CCATTTGG	TTAT	TGATTGAG-GGCGGCATTTCAATTTTCCCCGCT
R.solanacearum pcaH	AGGATTTTCC	ATAA	C-ATCCGG	TTAT	CGATGAATCGGCAAGAAGTCAATTTACTTCACT
P.aeru PAO1 pcaH	GATAAA--GC	ATAA	CCAAATGG	TTAT	GTATGGGGCCGGA-TAATTCACTGTCCCCGTCC
P.aeru UCBPP pcaH	GATAAA--GC	ATAA	CCAAATGG	TTAT	GTATATGGCTGGA-TAATTCACTGTCCCCGCCC
P.aeru PA7 pcaH	GATTTC--GC	ATAA	CCAAACGG	TTAT	GTGTCTGGCCGGA-TAATTCACTGTCCCCGTCC
P.stutzeri pcaH	CGCAAT--CC	ATAA	CCCCTGGG	TTAT	GGATCGGCCCGTT-CTTTTCAATTTCGGCCTGC
P.putida pcaH	ACTAAT--GC	ATAA	CCATTTGG	TTAA	GTAACGGTCGGCG-GAATTCAATATTCATCTGG
S.meliloti transp.	GCACAAAACA	ATAA	CCGGGGGA	TTAT	GGATTTCCGCCCA-TATTTCATTTTTGCATCAC
S.medicae transp.	TCCAAAAGCA	ATAA	CCGGGGAA	TTAT	GAATTTCCGCCCA-TATTTCATTTTTGCATCAC
Meso BNCI	CAATCCGCTT	ATAA	CTTTGCGA	TTAA	TGCCGTCGGCGTT-CAGTCCGGCCGCCGCAGAA

B.

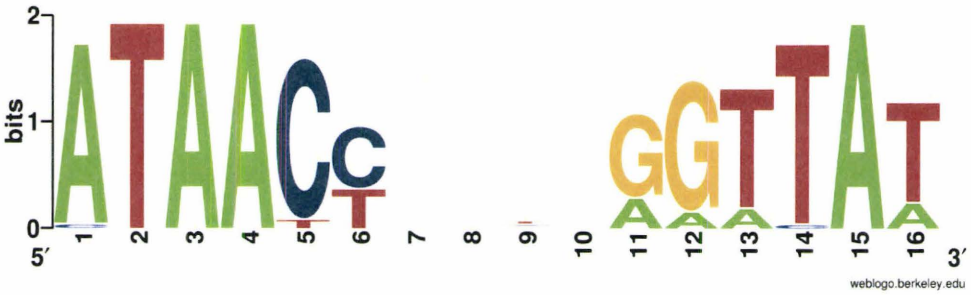


Figure 4.6. Analysis of the PcaQ-binding motif. (A) Alignment of predicted PcaQ-binding sites in members of α-, β-, and γ-proteobacteria. The right and left-arms of the binding sites are enclosed in a box. Red font indicates nucleotides within a binding site that are predicted to decrease the level of induced expression of a target gene. The -35 hexamers associated with the *pcaD* and *smb20568* promoters are underlined. (B) Sequence logo of the PcaQ-binding site. The logo was constructed using the experimentally verified binding sites upstream of *pcaD* and *smb20568* in *S. meliloti* and predicted binding sites conserved in α- and γ-proteobacteria. Predicted binding sites in β-proteobacteria were excluded from the analysis due to a 1 bp deletion within the motif.

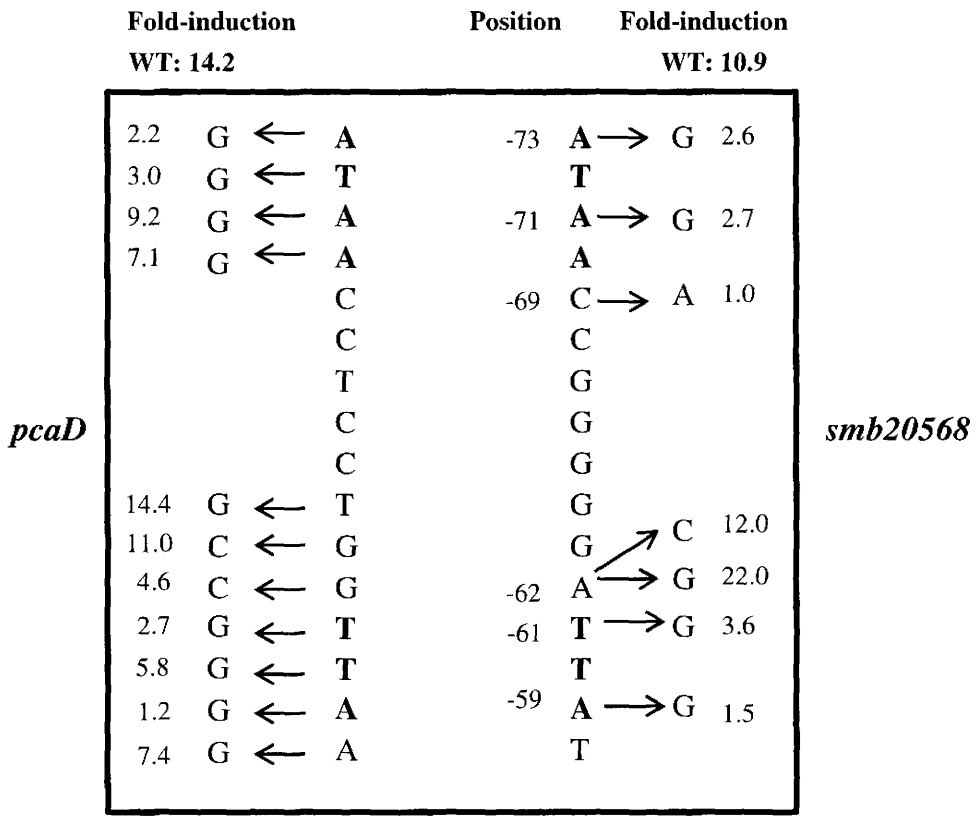


Figure 4.7. A comparison of a directed mutagenesis of PcaQ-binding sites upon the regulation of *pcaD* and *smb20568* expression in *S. meliloti*. Point mutations were introduced within PcaQ-binding sites located upstream of *pcaD* (left) and *smb20568* (right) as indicated. The effect of each mutation upon gene expression was determined using transcriptional fusions to the reporter protein GfpUV. Fold induction: average GfpUV specific activity in cells grown with protocatechuate (induced) divided by average activity in cells grown without protocatechuate (uninduced). Nucleotides that are highly conserved are in bold font. For simplicity, positions are indicated with repect to the *smb20568* transcript start site. Data relating to the *pcaD* site were obtained from a previous study (MacLean, et al., 2008) and are included here for the purpose of comparison.

CHAPTER FIVE

The legume endosymbiont *Sinorhizobium meliloti* encodes an ATP-binding cassette (ABC) transport system involved in the uptake of hydroxyproline.

Allyson M. MacLean, Catharine White, Jane Fowler, and

Turlough M. Finan

Preface

This chapter describes the identification of a transport system involved in the uptake of *trans*-4-hydroxy-L-proline in *S. meliloti*. The transport system was initially identified by Jane Fowler during the completion of a M.Sc. degree, and Jane performed the initial assays that demonstrated the hydroxyproline-inducible expression of the transporter genes. Alison Cowie constructed the regulator-minus strain of *S. meliloti* (RmP1406) and performed reporter enzyme assays to confirm the previous results obtained by Jane Fowler. Cathy White performed the primer extension experiment to map the transcript start site associated with *hypM* and provided valuable advice during revision of the manuscript. Under my direct supervision, Vladmir Jokic demonstrated that the hydroxyproline-inducibe TRAP-T system encoded by *smb20320-smb20322* was not essential to growth with hydroxyproline. I performed all uptake assays, as required to demonstrate that *hypMNPQ* mediate hydroxyproline transport. I performed the expression studies in nodules, using an *S. meliloti* strain I created for this purpose. Using the pLAFR1 clone bank, I isolated the cosmids capable of complementing the hydroxyproline-minus growth phenotype of RmF909. As well, I constructed the *S. meliloti* strain carrying an in-frame deletion encompassing *hypMNPQ* (RmP1114), using a plasmid (pTH2131) previously constructed by Jane Fowler. For consistency, I repeated all reporter enzyme assays and growth assays; data reported for these experiments is my own. Finally, as primary author, I wrote this chapter in its entirety, with editing by Turlough Finan and Cathy White. This manuscript has been submitted for possible publication to the journal Molecular Plant-Microbe Interactions.

5.1 Abstract

Hydroxyproline-rich proteins in plants offer a source of carbon and nitrogen to soil-dwelling microorganisms in the form of root exudates and decaying organic matter. This report describes an ABC-type transporter dedicated to the uptake of hydroxyproline in the legume endosymbiont *Sinorhizobium meliloti*. An *S. meliloti* strain lacking the transporter genes (Δ*hypMNPQ*) is unable to growth with or transport *trans*-4-hydroxy-L-proline when this compound is available as a sole source of carbon. Expression of *hypM* is shown to be up-regulated in the presence of *trans*-4-hydroxy-L-proline and *cis*-4-hydroxy-D-proline, as modulated by a repressor (HypR) of the GntR/FadR subfamily. Although alfalfa root nodules are comprised of hydroxyproline-rich proteins, we demonstrate that the transport system is not highly expressed in nodules, suggesting that bacteroids are not exposed to free hydroxyproline *in planta*. In addition to *hypMNPQ*, we report that *S. meliloti* encodes a second independent mechanism that enables transport of *trans*-4-hydroxy-L-proline. This secondary transport mechanism is induced in proline-grown cells and likely entails a system involved in L-proline uptake. Although this study represents the first genetic description of a prokaryotic hydroxyproline transport system, the ability to metabolize this amino acid may nonetheless contribute significantly towards the ecological success of plant-associated bacteria such as the rhizobia.

5.2 Introduction

Members of the order *Rhizobiales* (including *Sinorhizobium*, *Mesorhizobium*, *Rhizobium*, and *Bradyrhizobium*) participate in a symbiotic partnership with plant species within the legume family. The soil-dwelling and free-living bacteria are attracted by plant-derived flavonoids and isoflavonoids, which induce a set of bacterial genes (*nod*) involved in the synthesis of lipochitin oligosaccharides referred to as Nod factors (Cooper, 2007). Plant recognition of symbiotically relevant Nod factors triggers root hair deformation, cell division, and the production of an infection thread necessary for the invasion of the host plant with the endosymbiont (Geurts, et al., 2005). These events culminate in the development of root-borne nodules which house the nitrogen-fixing bacteria; the plant offers the rhizobia a source of carbon in the form of reduced photosynthate in exchange for a supply of fixed nitrogen.

Hydroxyproline (primarily in the form of *trans*-4-hydroxy-L-proline) commonly occurs in plant tissues in the form of hydroxyproline-containing proteins, including hydroxyproline-rich glycoproteins (HRGPs) that are a component of plant cell walls and contribute towards a structural support mechanism (Khashimova, et al., 2003). HRGPs consist of at least three related groups of glycosylated and hydroxyproline-containing proteins, including extensins, arabinogalactan-proteins, and proline/hydroxyproline-rich proteins. Hydroxyproline has been detected in the root nodules formed by leguminous plants (Cassab, et al., 1985; Benhamou, et al., 1991; Frueauf, et al., 2000) and hydroxyproline-rich extensin has been reported as the predominant structural protein of root cell walls and nodules of *Medicago truncatula* (Frueauf, et al., 2000). Brewin and coworkers reported a subclass of extensins (referred to as root nodule extensins) specific to legumes that are expressed in both uninfected root cells and nodules (Rathbun, et al., 2002). These hydroxyproline-rich extensins represent a significant component of the infection thread lumen in pea root hairs, and are secreted during even the earliest stages of rhizobial infection in *Vicia hirsuta* (Rae, et al., 1991; Rae, et al., 1992; Rathbun, et al., 2002).

The prevalence of hydroxyproline in plant tissue and secretion of hydroxyproline-rich proteins into the rhizosphere offer a potential source of nitrogen and carbon to plant-associated microorganisms (Knee, et al., 2001). Additionally, hydroxyproline has been detected in soil samples, humic acids, and decaying leaf litter (Griffith, et al., 1976; Sowden, et al., 1976; Morita & Sowden, 1981; Lahdesmaki & Phspanen, 1989). Hydroxyproline catabolism has been documented in several species of soil-dwelling bacteria, including *Pseudomonas putida* (Adams, 1959; Adams, 1973; Singh & Adams, 1965; Thacker, 1969; Gryder & Adams, 1969; Jayaraman & Radhakrishnan, 1965a; Jayaraman & Radhakrishnan, 1965b; Manoharan & Jayaraman, 1979; Manoharan, 1980). In this species, hydroxyproline metabolic enzymes and a dedicated uptake system are induced in cells grown in the presence of this compound, and these genes are likely regulated as a single transcriptional unit or share a common regulator (Gryder & Adams, 1969; Gryder & Adams, 1970). Hydroxyproline uptake kinetics and growth experiments with *P. putida* indicate that the affinity of the transport system for *trans*-4-hydroxy-L-proline is much greater than that for *cis*-4-hydroxy-D-proline (Gryder & Adams, 1969; Gryder & Adams, 1970). Data reported in previous studies, including competition experiments between L-proline and *trans*-4-hydroxy-L-proline, imply the possibility of a shared hydroxyproline/proline uptake system in *Pseudomonas* (Gryder & Adams, 1969; Gryder & Adams, 1970; Manoharan, 1980). However, the genes encoding a hydroxyproline-specific transport system have yet to be identified in any species.

In this study, we describe an ABC-type transport system (encoded by *hypMNPQ*) that is essential for growth with and transport of hydroxyproline in *Sinorhizobium meliloti*. We demonstrate that expression of the uptake system is induced by hydroxyproline, and is negatively regulated by a member of the GntR superfamily of transcriptional regulators (HypR). Although alfalfa nodules contain hydroxyproline-rich proteins such as extensin, we report that genes encoding the *S. meliloti* hydroxyproline uptake system are not highly expressed in either young or senescent root nodules. The addition of L-proline strongly reduces uptake of labeled *trans*-4-hydroxy-L-proline in transport assays, and growth with L-proline induces an uptake system that can transport

trans-4-hydroxy-L-proline in *S. meliloti*. Nonetheless, we establish that proline grown cells induce a separate transport system that is not specific to hydroxyproline and that cannot compensate for the deletion of *hypMNPQ* when hydroxyproline is available as a sole source of nitrogen or carbon.

5.3 Materials and Methods

Bacterial strains and growth conditions

Escherichia coli and *S. meliloti* strains used in this study are described in Table 5.1. *E. coli* strains were grown aerobically in LB broth or agar plates incubated at 37°C. *S. meliloti* strains were grown aerobically in LB broth supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ (LBmc) or in M9-minimal medium (Difco) at 30°C. M9-minimal medium was supplemented with 1.0 mM $MgSO_4$, 0.25 mM $CaCl_2$, 1 μg/mL D-biotin, and 10 ng/mL $CoCl_2$. Carbon sources were added to M9-minimal medium as follows: 10 mM glucose, 10 mM succinate, 0.5% (vol/vol) glycerol, 5 mM L-proline (hydroxyproline-free; Sigma-Aldrich), 5 mM *trans*-4-hydroxy-L-proline (Sigma-Aldrich), or 5 mM *cis*-4-hydroxy-D-proline (Sigma-Aldrich). When required, antibiotics were added to growth medium at the following concentrations for *S. meliloti* (in μg/mL): streptomycin (Sm), 200; spectinomycin (Sp), 200; gentamicin (Gm), 60; tetracycline (Tc), 5; rifampicin (Rif), 20.

Isolation of a cosmid capable of complementing a Hyp⁻ strain of S. meliloti

Spot matings with RmF909 as a recipient strain were performed using the *S. meliloti* pLAFR1 clone bank (Friedman, et al., 1982) and transconjugants were plated upon M9-minimal media with *trans*-4-hydroxy-L-proline as a sole carbon source and neomycin to counter-select the *E. coli* donor strain. Ten Hyp^+Nm^r *S. meliloti* colonies were patched onto LB agar containing tetracycline to confirm the presence of the cosmid. Spot matings with six of these Hyp^+ colonies were performed with *E. coli* recipient strain DH5α to facilitate cosmid DNA extraction. Cosmid DNA was isolated from the six

independent *E. coli* strains and DNA sequencing reactions were performed by Mobixlab (McMaster University, Hamilton, Canada) using primers complementary to the cosmid.

Construction of an in-frame deletion of the hydroxyproline uptake system

An in-frame deletion within the putative hydroxyproline transport system encoded by *hypMNPQ* was generated as follows. A 2,723 bp fragment spanning *hypM* and *hypQ* was PCR amplified using primers (5' AAGCGGCCGCAACAACGAACCGATCG 3') and (5' AAGCGGCCGCAGGCGAAAATCTGATCG 3') and cloned into the suicide vector pJQ200 uc1 (Quandt & Hynes, 1993) via *Not*I to yield pTH2130. This plasmid was then digested with *Sal*I to excise a 1,689 bp fragment encompassing *hypNP,* in addition to portions of *hypM* and *hypQ*. The linearized plasmid was purified from an agarose gel and incubated with T_4 DNA ligase to generate a plasmid carrying the *S. meliloti* sequence minus the ~1.7 kb *Sal*I fragment (pTH2131). The suicide plasmid pTH2131 was introduced into *S. meliloti* wild-type derivative strain RmP110 via conjugation using the *E. coli* helper strain MT616 (tri-parental mating). Transconjugants in which plasmid pTH2131 had integrated into the pSymB megaplasmid via homologous recombination were selected by plating onto LB agar supplemented with streptomycin plus gentamicin. A single *S. meliloti* Sm^rGm^r colony was purified (RmP1113) and inoculated into LBmc broth for incubation overnight in the absence of antibiotic selection. Aliquots of the culture were plated upon LB agar plus 5% sucrose to select for cells in which the suicide plasmid (encoding *sacB*) had excised from the genome; $sucrose^r$ colonies were screened for Gm^s to confirm loss of the plasmid. The genotypes of fifteen $sucrose^r/Gm^s$ colonies were screened via whole-cell PCR for the presence of a 1.7 kb deletion within *hypMNPQ*; of these, seven colonies yielded PCR products consistent with such a deletion. One of the seven strains was purified and designated as RmP1114. The primers utilized to identify the deletion mutants via whole-cell PCR correspond to sequence external to the region cloned into pJQ200.

To provide a means of complementing the in-frame deletion of the ABC-type transport system *in trans*, plasmid pTH2513 was constructed as follows. A region

encompassing the entire uptake system encoded by *hypMNPQ* was amplified by PCR using primers (5' ATTAAGCTTCGTCGCTTACCAGAACATGC 3') and (5' ATTATGCATGCAATTGCAGCCGTGAGG 3') and cloned via *Hind*III and *Nsi*I restriction sites into a modified version of the vector pJP2 (pTH1582). The 3,240 bp amplified fragment extends 241 bp upstream of the predicted translational start site of *hypM* and includes the intergenic region upstream of this gene, thus allowing expression of the operon to be determined by its native promoter. The replicating plasmid pTH2513 was transferred to *S. meliloti* strains by tri-parental conjugation; transconjugants were selected by plating upon LB agar supplemented with tetracycline and streptomycin.

Construction of an S. meliloti hypR::Ω strain

An Ω cassette specifying streptomycin and spectinomycin resistance from pHP45Ω (Prentki & Krisch, 1984) was introduced into a *Xho*I site located 345 bp downstream of the translational start site as follows. A 953 bp fragment encompassing the *Xho*I site was PCR amplified using *S. meliloti* strain Rm1021 genomic DNA as a template with primers (5' CCAAGCTTGCGATCTGTGCTGATTTCG 3') and (5' CCTAAGATCACAGAGCAATTTCG 3'); this fragment was cloned into plasmid pUCP30T to create pTH2216. The Ω cassette was PCR amplified using pHP45Ω as a template and the cassette was cloned into pTH2216 via *Xho*I restriction site to create pTH2217. This plasmid was transferred through conjugation into *S. meliloti* Rif^r strain Rm5000; Rif^r and Sp^r transconjugants were selected to isolate strains in which the plasmid (which is unable to replicate in *S. meliloti*) had integrated into the genome by homologous recombination. Rif^r/Sp^r colonies were screened for Gm^s to identify double recombinants in which the plasmid had simultaneously excised from the megaplasmid. To confirm the genotype of the strains, Southern hybridization was performed with genomic DNA isolated from five Rif^r/Sp^r and Gm^s strains digested with *Sac*II. Digested DNA was hybridized with a labeled probe corresponding to *hypR*; in each case, an ~2 kb increase was noted in the putative *hypR*::Ω strains (~6.9 kb) as compared to the wild-type control (~4.8 kb), consistent with the incorporation of the Ω cassette into the gene and

subsequent excision of the vector backbone. In contrast, DNA isolated from two control Rif^r/Sp^r and Gm^r strains (consistent with a single recombination event) identified two fragments (~5.6 and 6.5 kb) which hybridized with labeled probe. One of the five Rm5000 *hypR*::Ω strains was purified and designated as RmP1406. The *hypR*::Ω allele was transferred from RmP1406 into RmP110 via transduction by selecting for Sp^r transductants; this strain (RmP110 *hypR*::Ω) is referred to as RmP1724.

Reporter enzyme assays

β-glucuronidase and β-galactosidase enzyme assays were performed as previously described (MacLean, et al., 2006). *S. meliloti* strain RmFL7003 belongs to a reporter gene fusion library described by Cowie et al. (Cowie, et al., 2006). The reporter vector includes a variant of green fluorescent protein (GFP+) (Scholz, et al., 2000) and fluorescence was assayed using an excitation wavelength of 485 nm and emission wavelength of 510 nm. Assays for GFP+ were performed using overnight LBmc cultures of *S. meliloti* that were subcultured into M9-minimal medium supplemented with a carbon source as indicated and incubated at 30°C for four to six hours. As a means of quantifying fluorescence, the emission output was divided by its $O.D._{600}$ to calculate the specific activity.

Primer extension

RNA was isolated from *S. meliloti* wild-type strain RmP110 grown in M9-minimal medium with glycerol and ammonium chloride as sole sources of carbon and nitrogen (uninduced sample) or *trans*-4-hydroxy-L-proline as a source of nitrogen and carbon (induced sample). The cultures were grown for 12 hours with aeration at 30°C to an $O.D._{600}$ of approximately 0.4. Cells were harvested and total RNA extracted as previously described (MacLellan, et al., 2005). The oligonucleotide (5' CGTCAGAGCGGTGGCGGCAACGAGG 3') was end-labeled with γ^{32}P-dATP (Amersham) and T_4 polynucleotide kinase (Fermentas), and then purified using a QIAquick nucleotide removal column (Qiagen). For each primer extension, 2 x 10^5

c.p.m of the labeled primer was combined with approximately 20 μg of RNA, 20 nmoles each of all 4 dNTPs, and reverse transcriptase buffer (Invitrogen) in a total volume of 16 μL. Each reaction mixture was heated at 65°C for 15 minutes, and cooled to 50°C over approximately 30 minutes. 2 μL of 100 mM DTT and 1 μL of RNaseOUT (Invitrogen) were added to each reaction, which were then incubated at 42°C for 2 minutes before adding 1 μL of Superscript III reverse transcriptase (Invitrogen). Primer extensions were incubated for 50 minutes at 50°C, and an equal volume of 2x Stop buffer (USB, Sequenase version 2.0 kit) was added to each sample to terminate the reaction. DNA sequencing reactions were performed using Sequenase version 2.0 DNA sequencing kit (USB), with purified plasmid pFL7003 as a template and the same oligonucleotide used to prime the extension reactions. Extension products and sequencing reactions were separated by electrophoresis in an 8% polyacrylamide gel containing 7.7 M urea. Data was analyzed using a Storm820 phosphorimager (Amersham).

Hydroxyproline transport assays

S. meliloti strains were grown overnight in LBmc and subcultured (1:100) into M9-minimal medium supplemented with glycerol, L-proline, or *trans*-4-hydroxy-L-proline as sources of carbon. When indicated, proline and hydroxyproline were also added as a sole source of nitrogen through the omission of ammonium chloride, which is normally present in M9 growth medium. Cultures were typically harvested after 14 to 18 hours incubation at 30°C via centrifugation at 4°C ($O.D._{600}$ ~0.8 to 1.2). Cells were washed twice in cold M9-minimal medium, and resuspended to $O.D._{600}$ of 2.0 in M9-minimal medium lacking ammonium chloride (M9 – N); resuspended cells were kept on ice prior to use in transport assays. Assays were performed with 25 μL cells ($O.D._{600}$ 2.0) added to 450 μL M9 – N, and equilibrated at 30°C for 5 minutes in a water bath. Assays were initiated by the addition of 25 μL labeled substrate (*trans*-4-hydroxy-L-proline [^{3}H(G)]; 20 Ci/mmol; American Radiolabeled Chemicals, Inc.) added to a final concentration of 10 μM, cells were vortexed gently and incubated at 30°C for the duration of the assay. 100 – 200 μL aliquots of the assay mixture were removed at timed

intervals and rapidly filtered through 0.45 μm nitrocellulose filters (Millipore; catalogue number HAWP02500), which were presoaked in M9 – N. Filtered cells were immediately washed with approximately ten volumes of M9 – N. Counts per minute were determined from the dried filters using a Tri-Carb 2900TR Liquid Scintillation Analyzer (PerkinElmer). For competition assays, unlabeled competitor was added 30 seconds prior to the addition of labeled substrate. Assays were performed up to 1 min following addition of substrate and the rate of uptake was always linear within this period of time. Protein concentration was determined from sonicated cell lysates according to the method developed by Bradford (Bradford, 1976). Activity was also assessed using induced cells treated with toluene (100 μL toluene added to 2 mL cells, $O.D._{600}$ 2.0). Uptake was not observed in toluene treated cells, and this control was used to estimate background counts per minute, which was subtracted from values obtained in assays with untreated cells. All transport assays were performed in triplicate, and values reported represent the mean and standard deviation of the three assays. All reported data are representative of a minimum of two independent experiments.

Plant Assays

Alfalfa seeds (*Medicago sativa* cv Iroquois) were surface sterilized with 95% ethanol (5 minutes) followed by 2.5% sodium hypochlorite (20 minutes) and rinsed with distilled, autoclaved water for 1 hour. Seeds were germinated on water agar plates (1.5% agar) in the dark for two days, and seven seedlings were then transferred to nitrogen-free vermiculite and sand (1:1 (w/w)) and inoculated with approximately $10^7 – 10^8$ *S. meliloti* cells suspended in 10 mL sterile, distilled water. Plants were grown in a Conviron plant growth chamber with an 18 hour (21°C) day and 6 hour (17°C) night. Plants were watered as required using autoclaved, distilled water, and nodules were harvested for expression assays 4 to 12 weeks after inoculation and stored at -80°C.

β-glucuronidase activity was quantified in root nodules as follows. 5 – 10 nodules were placed into an Eppendorf tube that was pre-chilled on ice. 750 μL of cold MMS buffer (40 mM MOPS, 20 mM KOH, 2 mM $MgSO_4$, 0.3 M sucrose; pH 7) was

added and nodules were crushed with mini pestles. Plant tissue was pelleted via centrifugation at 400 × *g* for 2 min and 500 µL supernatant was transferred to a fresh tube on ice. SDS was added to a final concentration of 0.01% and samples were incubated on ice for five minutes. 100 µL lysate was used per assay, and each reaction included 890 µL buffer (50 mM sodium phosphate, 50 mM DTT, 1 mM EDTA; pH 7). Samples were equilibrated in a waterbath at 37°C for ten minutes, and the assay was initiated by the addition of 10 µL of 4-nitrophenyl β-D-glucuronide (35 mg/mL; Sigma-Aldrich). Reactions were terminated by the addition of 200 µL of reaction mixture to 700 µL of 0.46 M Na_2CO_3. Specific activity was calculated as: (Abs_{405} x 1000)(time (minutes) x mg protein)$^{-1}$. The protein concentration of lysates was determined by the method established by Bradford (Bradford, 1976). Histochemical staining to detect β-glucuronidase activity was performed based upon a method described by Boivin et al. (Boivin, et al., 1990). Five to ten nodules were obtained from plants inoculated with each *S. meliloti* strain. The fresh nodules were immediately mounted upon a specimen plate using crazy glue (Instant Krazy Glue, Elmer's Products Canada) and nodules were sliced using a vibrating blade microtome (LEICA VT1000) into 90 µm longitudinal sections. Nodule sections were transferred into staining buffer (200 mM sodium phosphate, 10 mM EDTA, 0.5 mM $K_3Fe(CN)_6$, 1.5 mM $K_4Fe(CN)_6$, pH 7.0) containing 0.08% (w/v) X-gluc (5-bromo-4-chloro-3-indolyl-β-D-glucuronide; Sigma-Aldrich). Sections were stained overnight (typically 12-18 hours) at 30°C and then washed twice in 200 mM sodium phosphate buffer (pH 7.0) for several hours. Prior to microscopic examination, sections were cleared in a solution of 1% sodium hypochlorite for five minutes. Images were captured using a Nikon TE2000 inverted microscope.

For the purpose of monitoring expression of *hypM* in alfalfa root nodules, *S. meliloti* strain RmP1886 was constructed as follows. Primers (5' ATCAGATCTCATGTCGATTCGACGTTTCCTGC 3') and (5' TAGGTACCATGATGTCGTCCAGCTTGTCG 3') were used to PCR amplify a 370 bp fragment encompassing the intergenic region upstream of *hypM*. This fragment was cloned into pTH1705 (Cowie, et al., 2006) via *Bgl*II/*Kpn*I to create pTH2494. The

plasmid was transferred via conjugation into *S. meliloti* strain RmP110; recombinants in which the plasmid integrated into the genome were isolated by plating transconjugants upon LB agar containing streptomycin and gentamicin.

S. meliloti strain RmP778 was included in all assays as a negative control to reflect the endogenous level of β-glucuronidase activity associated with a single copy of the reporter gene in the *S. meliloti* genome. To create RmP778, a promoter (associated with a gene unrelated to hydroxyproline metabolism; *smb20568*) was cloned in the antisense orientation with respect to *gusA* in pTH1705 via *BglII/KpnI*. This plasmid was transferred into RmP110, and a purified recombinant was designated as RmP778.

5.4 Results

Identification of a gene cluster essential for growth with hydroxyproline

S. meliloti strain RmF909 carries a large-scale deletion within the pSymB megaplasmid (Charles & Finan, 1991) that renders the strain unable to metabolize both *trans*-4-hydroxy-L-proline and *cis*-4-hydroxy-D-proline, whereas wild-type *S. meliloti* is able to utilize both of these compounds as sole carbon and nitrogen sources. To identify the genes required for hydroxyproline catabolism, we screened the *S. meliloti* pLAFR1 clone bank for a cosmid that was able to complement the hydroxyproline-minus phenotype this strain. We sequenced cosmid DNA originating from six independent RmF909 derivative strains of *S. meliloti* in which the inability of the parental strain to metabolize hydroxyproline was complemented by a cosmid clone; sequencing reactions performed upon each cosmid yielded identical *S. meliloti* DNA sequences (255129 – 275255 nts; pSymB), indicating that the cosmids were siblings. The ~20 kb region of *S. meliloti* DNA contained within the cosmid (pTH2439) includes sequence specifying a single ABC-type transport system (encoded by *hypMNPQ*) positioned upstream of several genes predicted to encode metabolic enzymes (Figure 5.1). To determine whether this putative uptake system is required for hydroxyproline transport, an *S. meliloti* strain was constructed carrying a 1.7 kb in-frame deletion (RmP1114; Figure 5.1C) which disrupted *hypM* and *hypQ* and removed the *hypN* and *hypP* genes.

Wild-type *S. meliloti* strain RmP110 grows readily upon *trans*-4-hydroxy-L-proline with a growth rate constant (μ) of 0.20 hour^{-1} and mean generation time of 3.4 hours when this compound is available as a sole carbon source (for *cis*-4-hydroxy-D-proline; μ = 0.22 hour^{-1} and g = 3.2 hours) (Figure 5.2). Although *S. meliloti* strain RmP1114 grows at a rate that is comparable to the wild-type strain RmP110 when subcultured with succinate (Figure 5.2A) or glycerol (data not shown) as carbon sources, the mutant strain exhibits a considerable lag in growth when either *trans*-4-hydroxy-L-proline or *cis*-4-hydroxy-D-proline are available as a sole source of carbon (Figures 5.2B and 5.2C). This phenotype was complemented when the genes encoding the transport system (*hypMNPQ*) were provided to RmP1114 *in trans* (pTH2513), indicating that the putative uptake system is required for wild-type growth with hydroxyproline. *S. meliloti* deletion strain RmF909 was unable to utilize *trans*-4-hydroxy-L-proline and *cis*-4-hydroxy-D-proline as a source of carbon, however growth was restored upon complementation with the cosmid pTH2439 (Figures 5.2B and 5.2C). We note that strains carrying the cosmid clone pTH2439 exhibited a markedly reduced growth lag compared to the wild-type RmP110 subcultured into either of the hydroxyproline-containing media (Figures 5.2B and 5.2C); this effect may be attributed to an increased copy number of hydroxyproline catabolic and uptake genes located upon the cosmid.

Expression of hypM is induced in the presence of hydroxyproline and is regulated by the GntR transcriptional regulator HypR

We examined expression of *hypM* (encoding a putative periplasmic solute binding protein; Figure 5.1B) in strain RmFL7003 carrying an integrated *hypM*::*gfp*$^{+}$/*lacZ* reporter gene fusion (Cowie, et al., 2006). Based upon reporter enzyme assays, expression of *hypM* was induced greater than 5-fold in the presence of *trans*-4-hydroxy-L-proline and *cis*-4-hydroxy-D-proline (Table 5.2). A gene encoding a putative transcriptional regulator belonging to the GntR family of proteins (*hypR*) is located ~5 kb upstream of the hydroxyproline-inducible ABC-type transport system (Figure 5.1B). To determine whether HypR was involved in regulating expression of the uptake system, we

constructed a *hypR* insertion mutant in which an Ω cassette was introduced 345 bp downstream of the predicted translational start site. Expression of *hypM::gfp*$^{+}$*/lacZ* was increased atleast 20-fold in the *hypR*::Ω strain RmP1838 (RmFL7003 *hypR*::Ω) as compared to RmFL7003 when grown with succinate or glucose as carbon sources (Table 5.2). In addition, *hypM::gfp*$^{+}$*/lacZ* expression in RmP1838 was no longer hydroxyproline-inducible; gene expression occurred at a comparable high level under all growth conditions tested. *hypR* is located upstream of a putative dihydrodipicolinate synthase gene (*smb20259*) and it is formally possible that the insertion of a cassette within *hypR* might exert polar effects upon the expression of this gene. However, the intergenic region between these genes is sufficiently large (121 bp) to encode an independent promoter associated with *smb20259* and it is unlikely that the regulatory phenotype exhibited by RmP1838 is the result of a polar effect upon the expression of a putative dihydrodipicolinate synthase gene. Our results thus indicate that *hypR* encodes a product that represses expression of *hypM* and that this repression is relieved by hydroxyproline.

Identification of the hypM transcriptional start site

Primer extension reactions were performed using an oligonucleotide complementary to sequence within the coding region of *hypM* (+27 to +51 nucleotides downstream of the annotated translational start site). One major extension product was obtained using RNA extracted from wild-type (RmP110) cells grown with *trans*-4-hydroxy-L-proline (induced), indicating a transcriptional start site located 38 nucleotides upstream of the predicted start codon associated with *hypM* (Figure 5.3A). The absence of a comparable extension product in reactions using RNA derived from glycerol-grown cells (uninduced) is consistent with the hydroxyproline-inducible expression of this gene. Examination of the inferred -35 and -10 hexanucleotide regions associated with this transcriptional start site reveals similarity to the consensus promoter sequence previously described in *S. meliloti* (MacLellan, et al., 2006) (Figure 5.3C). In fact, the experimentally verified promoter associated with *hypM* and identified in this work was

previously predicted based upon a search of the *S. meliloti* genome for sequences matching the consensus sequence (see Table 5.2; MacLellan, et al., 2006).

Hydroxyproline-inducible uptake requires hypMNPQ

To directly examine hydroxyproline uptake, we used ^{3}H-radiolabeled *trans*-4-hydroxy-L-proline. Cells grown in the presence of *trans*-4-hydroxy-L-proline as a sole nitrogen and carbon source demonstrated the greatest rate of uptake at 10.7 nmols/min/mg protein (Figure 5.4A). In contrast, cells grown with succinate (data not shown) or glycerol (1.2 nmols/min/mg protein) were unable to efficiently transport hydroxyproline indicating that the system involved is hydroxyproline-inducible. *S. meliloti* strain RmP1114 (Δ*hypMNPQ*) was unable to transport *trans*-4-hydroxy-L-proline when grown in the presence of hydroxyproline (less than 0.5 nmols/min/mg protein), and uptake was restored to wild-type levels upon introduction of the *hypMNPQ* genes *in trans* on the plasmid pTH2513 (Figure 5.4B). Thus *hypMNPQ* are required for the uptake of this compound in hydroxyproline-grown cells (Figure 5.4B).

Experiments performed using unlabeled competitor versus labeled *trans*-4-hydroxy-L-proline indicate that both *cis*-4-hydroxy-D-proline and L-proline compete with *trans*-4-hydroxy-L-proline for binding and/or uptake in *S. meliloti* hydroxyproline-grown cells (Table 5.3). Proline reduced the uptake of labeled *trans*-4-hydroxy-L-proline as effectively as unlabeled *trans*-4-hydroxy-L-proline; in contrast, the *cis* epimer was less effective as a competitor. Similar findings have been reported in *P. putida*, where L-proline also strongly reduces the transport of *trans*-4-hydroxy-L-proline (Gryder & Adams, 1970).

Hydroxyproline uptake in S. meliloti cells grown with L-proline

The effectiveness of L-proline as a competitor against *trans*-4-hydroxy-L-proline in uptake assays raised the possibility that proline and hydroxyproline may share a common transport system, as has been suggested in *P. aeruginosa* (Manoharan, 1980). To investigate whether proline induces expression of a hydroxyproline uptake

mechanism, transport assays were performed upon cells cultured in growth medium containing L-proline as a sole source of carbon and nitrogen (Figure 5.4C). Proline-grown *S. meliloti* RmP110 imports labeled *trans*-4-hydroxy-L-proline at approximately one-half the rate observed in *trans*-4-hydroxy-L-proline grown cells (3.4 versus 8.7 nmols/min/mg protein, respectively). This rate of uptake is significantly greater than that exhibited by uninduced control cells grown in M9-glycerol (< 1 nmol/min/mg protein). We also examined uptake of *trans*-4-hydroxy-L-proline in *S. meliloti* strain RmP1114 (Δ*hypMNPQ*) grown in the presence of *trans*-4-hydroxy-L-proline or L-proline; as in earlier experiments, RmP1114 did not transport labeled *trans*-4-hydroxy-L-proline when grown in M9-glycerol plus *trans*-4-hydroxy-L-proline (a growth condition which induces uptake in RmP110; Figure 5.4B). However, proline-grown RmP1114 transported *trans*-4-hydroxy-L-proline at a rate (3.0 nmols/min/mg protein) comparable to that observed in proline-grown wild-type cells. These results are consistent with the presence of a second mechanism for hydroxyproline uptake into *S. meliloti* that is induced by L-proline, and is therefore likely to entail a proline transport system.

The hydroxyproline uptake system is not highly expressed in root nodules

Expression of *hypM* was monitored in alfalfa nodules through the use of a transcriptional fusion to the reporter gene *gusA* in *S. meliloti* strain RmP1886 (Figure 5.1C). As a positive control, alfalfa seedlings were inoculated with strain RmFL5810, which contains a *gusA* fusion with *minD*; this gene is known to be expressed in nodules (Cheng, et al., 2007). As negative controls, we utilized wild-type strain RmP110 and a strain (RmP778) in which *gusA* is integrated into the *S. meliloti* genome in the antisense orientation to a promoter unrelated to hydroxyproline metabolism. RmP778 was included to reflect the level of β-glucuronidase activity associated with a single copy of the reporter gene in the *S. meliloti* genome.

β–glucuronidase activity was measured in nodule extracts obtained from alfalfa plants 30 to 84 days post inoculation (d.p.i.) (Table 5.4). Our results indicate a low level of *hypM* expression that is approximately 4-fold greater than that observed in nodules

inoculated with the negative controls RmP110 (no *gusA* fusion) or RmP778 (*gusA* in antisense orientation to *smb20568*). In contrast, nodules harvested from plants inoculated with the positive control (*minD*::*gusA*) exhibited a higher level of expression, consistent with previous data obtained regarding expression of this gene in nodules (Cheng, et al., 2007). To investigate whether expression was localized to particular regions of the nodule, we stained longitudinal sections of 30 and 66 d.p.i. nodules for β–glucuronidase activity (Figure 5.5). β–glucuronidase activity was observed primarily within the distal and central portions of the nodules, corresponding to the infection or prefixing zone II and nitrogen fixing zone III.

5.5 Discussion

The legume endosymbiont *S. meliloti* is able to utilize hydroxyproline as a carbon and nitrogen source. To identify loci involved in hydroxyproline metabolism and transport, we isolated a cosmid clone that was able to complement an *S. meliloti* deletion strain for growth at the expense of hydroxyproline. Sequencing reactions revealed that the cosmid contained ~ 20 kb of *S. meliloti* DNA (255129 - 275255 nts; pSymB), including a single ABC-type transport system encoded by genes *hypM* (periplasmic solute binding protein), *hypN* and *hypP* (transporter permease proteins), as well as *hypQ* (ATP-binding protein).

We have presented data to show that *S. meliloti hypMNPQ* encode an ABC-type transport system for uptake of *trans*-4-hydroxy-L-proline. Mutant strain RmP1114 carrying a deletion spanning the *hypMNPQ* genes failed to transport *trans*-4-hydroxy-L-proline and grew slowly with this compound as a carbon and nitrogen source (Figures 5.2 and 5.4). The growth and transport phenotypes exhibited by RmP1114 were complemented when *hypMNPQ* was supplied *in trans*, thus the in-frame deletion only disrupted expression of the uptake system and not expression of any downstream metabolic genes (Figure 5.1B).

Expression of *hypM*, and we infer the entire *hypMNPQ* gene cluster, was induced by growth in the presence of either *trans*-4-hydroxy-L-proline or *cis*-4-hydroxy-D-

proline (Table 5.2). Our uptake assays indicated that *cis*-4-hydroxy-D-proline is less effective in reducing the uptake of labeled *trans*-4-hydroxy-L-proline in hydroxyproline-grown *S. meliloti* than either cold *trans*-4-hydroxy-L-proline or L-proline. In *P. putida*, the affinity of the transport system for *trans*-4-hydroxy-L-proline is much greater than that for *cis*-4-hydroxy-D-proline (Gryder & Adams, 1969; Gryder & Adams, 1970). Likewise, *P. aeruginosa* PAO is impermeable to *cis*-4-hydroxy-D-proline and cannot grow at the expense of this compound, although the strain readily transports and metabolizes *trans*-4-hydroxy-L-proline (Manoharan, 1980). While we did not directly examine uptake of *cis*-4-hydroxy-D-proline, our data are consistent with the *S. meliloti* hydroxyproline transport system having a greater affinity for *trans*-4-hydroxy-L-proline than the *cis* epimer. Saprophytic or plant-associated bacteria such as *S. meliloti* and *Pseudomonas* are more likely to encounter *trans*-4-hydroxy-L-proline in a natural habitat as this isomer is the predominate form of hydroxyproline present in plant (and animal) tissue (Adams & Frank, 1980). Accordingly, the evolution of an uptake system with a greater affinity for *trans*-4-hydroxy-L-proline than *cis*-4-hydroxy-D-proline may reflect the relative abundance of each stereoisomer in the habitats occupied by these soil-dwelling species.

The finding that growth of *S. meliloti* with L-proline enables immediate and efficient uptake of *trans*-4-hydroxy-L-proline in both wild-type and the Δ*hypMNPQ* strain RmP1114 (Figure 5.4C) suggests that an L-proline uptake system can also transport hydroxyproline. The apparent inability of this unidentified uptake system to rescue growth of RmP1114 in M9-minimal medium with hydroxyproline, as well as the lack of *trans*-4-hydroxy-L-proline uptake observed in succinate and glycerol grown cells, emphasizes the requirement of induction via proline metabolism. In *S. meliloti*, proline is known to be transported into the cell by the ABC-type transport system Hut, however expression of this system is not induced by proline (Boncompagni, et al., 2000). The osmoprotectant proline betaine is transported by the ABC-type transporter Prb in *S. meliloti*, yet expression of this system is not induced by proline and there is no evidence to indicate that this system imports proline (Alloing, et al., 2006). In enteric bacteria

such as *E. coli*, a proline permease gene (*putP*) is divergently transcribed with respect to the proline dehydrogenase (*putA*) (Mogi, et al., 1986) however *putA* in *S. meliloti* is not associated with a permease (Soto, et al., 2000). Consequently, the identity of the proline-induced transport system remains unresolved.

Alfalfa and other legume nodules contain high amounts of hydroxyproline-rich proteins, however it is unclear to what degree (if any) nodule-inhabiting bacteroids encounter hydroxyproline in a form that is available for metabolism. Nodules induced by a *hypM*::*gusA* fusion yielded an average 4-fold greater reporter enzyme activity than that observed in the negative controls (Table 5.4). We examined expression in green (i.e., senescent) root nodules harvested as late as 84 dpi, however we did not observe an increase in specific activity in older nodules. Longitudinal sections of 4 week old nodules were stained for β–glucuronidase activity and we observed uniform staining throughout nodules from plants inoculated with RmP1886 (*hypM*::*gusA*). Particularly, the distal and central regions of the nodule (prefixing zone II and nitrogen fixing zone III) showed evidence of gene expression (Figure 5.5). An examination of 9 week old nodules revealed less intense staining that was primarily limited to zones II and III. Weak staining was occasionally observed in the saprophytic zone V of nodules (Timmers, et al., 2000) however we conclude that *hypM* was not highly expressed in this region of the nodule. In summary, our results indicate that the expression of *hypM* is a fraction of that exhibited by moderately expressed *minD* (60%) and highly expressed *nifH* (16%; data not shown). As well, the hydroxyproline uptake gene cluster is not required for nodulation or nitrogen fixation (data not shown). Accordingly, it is unlikely that the hydroxyproline uptake genes play an essential role in the plant-microbe symbiosis, nor is expression of the transport system strongly induced in either young or senescent root nodules.

Identification of the transcriptional start site associated with *hypM* (Figure 5.3A) reveals a hydroxyproline-inducible promoter exhibiting sequence similarity to that of consensus promoters previously described in *S. meliloti* and *Rhizobium etli* (MacLellan, et al., 2006; Ramirez-Romero, et al., 2006). In fact, the promoter identified upstream of *hypM* was accurately predicted by MacLellan et al., thus validating the promoter

prediction methodology employed by this study. The consensus sequence identified in *R. etli* is recognized by SigA (or σ^{70}) (Ramirez-Romero, et al., 2006). By extension, it may be expected that the promoter element upstream of *hypM* is similarly recognized by the *S. meliloti* housekeeping sigma factor RpoD, however expression of this gene is also repressed by the GntR regulator HypR (discussed below). The observation that *hypM* expression is regulated via repression (as opposed to activation) may in part underlie the sequence similarity between the *hypM* and consensus promoter sequences, as a high level of *hypM* expression under permissive conditions is likely dependent upon RNA polymerase σ^{70} recognition of the promoter element in the absence of any transcriptional activator. The promoter prediction methodology proposed by MacLellan et al. was established using (primarily) factor-independent promoters, and it is possible that such predictions may also be biased towards a subset of promoters subject to negative regulation.

HypR is a member of the GntR-type family of transcriptional regulators that was initially described by Haydon and Guest (Haydon & Guest, 1991). This comprises a group of proteins that participates in the regulation of a diverse range of processes including sporulation (Rigali, et al., 2006), plasmid maintenance and transfer (Kataoka, et al., 1994; Lee, et al., 2003), the acquisition of phosphorus (Gebhard & Cook, 2008), and utilization of sugars (Franco, et al., 2006), aromatic compounds (Morawski, et al., 2000; Gerischer, 2002), organic acids (Georgi, et al., 2008), amino acids (Wiethaus, et al., 2008), and fatty acids (Fujita, et al., 2007). GntR proteins have been classified into one of six subfamilies (FadR, MocR, HutC, YtrA, PlmA, AraR) based upon sequence divergence between members at the C-terminal effector binding and oligomerization domains (Rigali, et al., 2002; Lee, et al., 2003; Franco, et al., 2006). As a means of determining which subfamily the hydroxyproline regulator belongs to, we constructed a phylogenetic tree using the full-length amino acid sequences of proteins representing members of each subfamily (Figure 5.6A). Based upon these analyses, the regulator encoded by *hypR* clusters within the FadR subfamily of GntR proteins. Consistent with the classification of HypR, this subfamily comprises many members involved in the

regulation of amino acid metabolic genes in both Gram positive and negative species (Rigali, et al., 2002).

The FadR subfamily of proteins recognizes binding sites established upon a (tnTGn$_3$ACna) consensus motif (Rigali, et al., 2002). Examination of the promoter region upstream of *hypM* reveals a possible regulator binding site (5' TTTTGTATACTA 3') that approximates the FadR consensus motif. An examination of adjacent metabolic genes identified additional putative binding sites exhibiting marked sequence similarity with the *hypM* motif (Figure 5.6B). Further studies are required to determine the role of these sites and the putative metabolic genes in hydroxyproline catabolism.

We recently reported the identification of a TRAP-T uptake system whose expression was induced by hydroxyproline (Mauchline, et al., 2006). Disruption of the genes encoding the TRAP-T system does not affect cell growth upon hydroxyproline (data not shown) and we conclude that the TRAP-T system is not essential for hydroxyproline uptake. It is possible that expression of the system is induced by a metabolite generated during hydroxyproline catabolism, however this was not investigated further.

This report represents the first genetic identification of a hydroxyproline uptake system in prokaryotes, however it is likely that other soil-dwelling and plant-associated bacteria encode hydroxyproline uptake and catabolic genes. The ability to catabolize hydroxyproline may be particularly relevant to plant-associated bacteria such as the rhizobia due to the abundance of hydroxyproline-rich glycoproteins in plant tissues, including the nodules, seeds, and roots of legumes (Cassab, et al., 1985; Rae, et al., 1991; Wood, et al., 1991; Rae, et al., 1992; Wisniewski, et al., 2000; Frueauf, et al., 2000; Knee, et al., 2001; Rathbun, et al., 2002). Natural processes such as root nodule senescence may even result in an enrichment of hydroxyproline in soils where legumes predominate. The ability to catabolize hydroxyproline may thus confer a selective advantage to the saprophytic bacteria that colonize the plant rhizosphere, thereby contributing towards the ecological success of these organisms.

Acknowledgments

We thank W. Haerty for the construction of the phylogenetic tree GntR family proteins. This work was supported by grants to T.M.F. from the Natural Sciences and Engineering Council of Canada and from Genome Canada through the Ontario Genomics Institute and the Ontario Research and Development Challenge Fund.

5.6 References

Adams, E. (1959). Hydroxyproline metabolism. I. Conversion to alpha-ketoglutarate by extracts of *Pseudomonas*. Journal of Biological Chemistry, **234**:2073-2084.

Adams, E. (1973). The metabolism of hydroxyproline. Molecular and Cellular Biochemistry, **2**:109-119.

Adams, E., & Frank, L. (1980). Metabolism of proline and the hydroxyprolines. Annual Review of Biochemistry, **49**:1005-1061.

Alloing, G., Travers, I., Sagot, B., Le Rudulier, D., & Dupont, L. (2006). Proline betaine uptake in *Sinorhizobium meliloti*: Characterization of Prb, an opp-like ABC transporter regulated by both proline betaine and salinity stress. Journal of Bacteriology, **188**:6308-6317.

Benhamou, N., Lafontaine, P.J., Mazau, D., & Esquerre-Tugaye, M.T. (1991). Differential accumulation of hydroxyproline-rich glycoproteins in bean root nodule cells infected with a wild-type strain or a C_4-dicarboxylic acid mutant of *Rhizobium leguminosarum* bv. *phaseoli*. Planta, **184**:457-467.

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

Boncompagni, E., Dupont, L., Mignot, T., Osteras, M., Lambert, A., Poggi, M.C., & Le Rudulier, D. (2000). Characterization of a *Sinorhizobium meliloti* ATP-binding cassette histidine transporter also involved in betaine and proline uptake. Journal of Bacteriology, **182**:3717-3725.

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

Cassab, G.I., Nieto-Sotelo, J., Cooper, J.B., van Holst, G.J., & Varner, J.E. (1985). A Developmentally Regulated Hydroxyproline-Rich Glycoprotein from the Cell Walls of Soybean Seed Coats. Plant Physiology, **77**:532-535.

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

Cheng, J., Sibley, C.D., Zaheer, R., & Finan, T.M. (2007). A *Sinorhizobium meliloti minE* mutant has an altered morphology and exhibits defects in legume symbiosis. Microbiology, **153**:375-387.
Cooper, J.E. (2007). Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. Journal of Applied Microbiology, **103**:1355-1365.

Cowie, A., Cheng, J., Sibley, C.D., Fong, Y., Zaheer, R., Patten, C.L., Morton, R.M., Golding, G.B., & Finan, T.M. (2006). An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti.* Applied and Environmental Microbiology, **72**:7156-7167.

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

Finan, T.M., Hartweig, E., LeMieux, K., Bergman, K., Walker, G.C., & Signer, E.R. (1984). General transduction in *Rhizobium meliloti.* Journal of Bacteriology, **159**:120-124.

Franco, I.S., Mota, L.J., Soares, C.M., & Sa-Nogueira, I. (2006). Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. Journal of Bacteriology, **188**:3024-3036.

Friedman, A.M., Long, S.R., Brown, S.E., Buikema, W.J., & Ausubel, F.M. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene, **18**:289-296.

Frueauf, J.B., Dolata, M., Leykam, J.F., Lloyd, E.A., Gonzales, M., VandenBosch, K., & Kieliszewski, M.J. (2000). Peptides isolated from cell walls of *Medicago truncatula* nodules and uninfected root. Phytochemistry, **55**:429-438.

Fujita, Y., Matsuoka, H., & Hirooka, K. (2007). Regulation of fatty acid metabolism in bacteria. Molecular Microbiology, **66**:829-839.

Gebhard, S., & Cook, G.M. (2008). Differential regulation of high-affinity phosphate transport systems of *Mycobacterium smegmatis:* Identification of PhnF, a repressor of the *phnDCE* operon. Journal of Bacteriology, **190**:1335-1343.

Georgi, T., Engels, V., & Wendisch, V.F. (2008). Regulation of L-lactate utilization by the FadR-type regulator LldR of *Corynebacterium glutamicum.* Journal of Bacteriology, **190**:963-971.

Gerischer, U. 2002. Specific and global regulation of genes associated with the degradation of aromatic compounds in bacteria. Journal of Molecular Microbiology and Biotechnology, **4**:111-121.

Geurts, R., Fedorova, E., & Bisseling, T. (2005). Nod factor signaling genes and their function in the early stages of *Rhizobium* infection. Current Opinion in Plant Biology, **8**:346-352.

Griffith, S.M, Sowden, F.J., & Schnitzer, M. (1976). The alkaline hydrolysis of acid-resistant soil and humic acid residues. Soil Biology and Biochemistry, **8**:529-531.

Gryder, R.M, & Adams, E. (1969). Inducible degradation of hydroxyproline in *Pseudomonas putida*: pathway regulation and hydroxyproline uptake. Journal of Bacteriology, **97**:292-306.

Gryder, R.M, & Adams, E. (1970). Properties of the inducible hydroxyproline transport system of *Pseudomonas putida*. Journal of Bacteriology, **101**:948-958.

Haydon, D.J., & Guest, J.R. (1991). A new family of bacterial regulatory proteins. FEMS Microbiology Letters, **63**:291-295.

Jayaraman, K., & Radhakrishnan, A.N. (1965a). Bacterial metabolism of hydroxyproline. I. Metabolism of L-allohydroxyproline by a *Pseudomonas*. Indian Journal of Biochemistry, **2**:145-153.

Jayaraman, K., & Radhakrishnan, A.N. (1965b). Bacterial metabolism of hydroxyproline. II. Metabolism of L-hydroxyproline in *Achromobacter*. Indian Journal of Biochemistry, **2**:153-158.

Jones, D.T., Taylor, W.R., & Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences, **8**:275-282.

Kataoka, M., Kosono, S., Seki, T., & Yoshida, T. (1994). Regulation of the transfer genes of *Streptomyces* plasmid pSN22: in vivo and in vitro study of the interaction of TraR with promoter regions. Journal of Bacteriology, **176**:7291-7298.

Khashimova, Z.S. (2003). Hydroxyproline-containing plant proteins. Chemistry of Natural Compounds, **39**:229-236.

Knee, E.M., Gong, F.C., Gao, M., Teplitski, M., Jones, A.R., Foxworthy, A., Mort, A.J., & Bauer, W.D. (2001). Root Mucilage from Pea and Its Utilization by Rhizosphere Bacteria as a Sole Carbon Source. Molecular Plant-Microbe Interactions, **14**:775-784.

Lahdesmaki, P., & Phspanen, R. (1989). Changes in concentrations of free amino acids during humification of spruce and aspen leaf litter. Soil Biology and Biochemistry, **21**:975-978.

Lee, M.H., Scherer, M., Rigali, S., & Golden, J.W. (2003). PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena sp.* strain PCC 7120. Journal of Bacteriolgoy, **185**:4315-4325.

MacLean, A.M., MacPherson, G., Aneja, P., & Finan, T.M. (2006). Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. Applied and Environmental Microbiology, **72**:5403-5413.

MacLellan, S.R., Smallbone, L.A., Sibley, C.D., & Finan, T.M. (2005). The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of alpha-proteobacterial plasmids. Molecular Microbiology, **55**:611-623.

MacLellan, S.R., MacLean, A.M., & Finan, T.M. (2006). Promoter prediction in the rhizobia. Microbiology, **152**:1751-1763.

Manoharan, T.H. (1980). On the regulation of L-hydroxyproline dissimilatory pathway in *Pseudomonas aeruginosa* PAO. Journal of Biosciences, **2**:107-120.

Manoharan, T.H., & Jayaraman, K. (1979). Mapping of the Loci Involved in the Catabolic Oxidation of L-Hydroxyproline in *Pseudomonas aeruginosa* PAO. Molecular and General Genetics, **172**:99-105.

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

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

Mogi, T., Yamamoto, H., Nakao, T., Yamato, I., & Anraku, Y. (1986). Genetic and physical characterization of *putP*, the proline carrier gene of *Escherichia coli* K12. Molecular and General Genetics, **202**:35-41.

Morawski, B., Segura, A., & Ornston, L.N. (2000). Repression of *Acinetobacter* vanillate demethylase synthesis by VanR, a member of the GntR family of transcriptional regulators. FEMS Microbiology Letters, **187**:65-68.

Morita, H., & Sowden, F. (1981). Effect of decomposition on the distribution of amino compounds in the acid hydrolysates of organic soils. Soil Biology and Biochemistry, **13**:327-329.

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

Prentki, P., & Krisch, H.M. (1984). In vitro insertional mutagenesis with a selectable DNA fragment. Gene, **29**:303–313.

Quandt, J., & Hynes, M.F. (1993). Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. Gene, **15**:15-21.

Rae, A.L., Perotto, S., Knox, J.P., Kannenberg, E.L., & Brewin, N.J. (1991). Expression of extracellular glycoproteins in the uninfected cells of developing pea nodule tissue. Molecular Plant-Microbe Interactions, **4**:563-570.

Rae, A.L., Bonfante-Fasolo, P., & Brewin, N.J. (1992). Structure and growth of infection threads in the legume symbiosis with *Rhizobium leguminosarum*. The Plant Journal, **2**:385-395.

Ramirez-Romero, M.A., Masulis, I., Cevallos, M.A., Gonzalez, V., & Davila, G. (2006). The *Rhizobium etli* sigma70 (SigA) factor recognizes a lax consensus promoter. Nucleic Acids Research, **34**:1470-1480.

Rathbun, E.A, Naldrett, M.J, & Brewin, N.J. (2002). Identification of a family of extensin-like glycoproteins in the lumen of *rhizobium*-induced infection threads in pea root nodules. Molecular Plant-Microbe Interactions, **15**:350-359.

Rigali, S., Derouaux, A., Giannotta, F., & Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. The Journal of Biological Chemistry, **277**:12507-12515.

Rigali, S., Nothaft, H., Noens, E.E.E., Schlicht, M., Colson, S., Muller, M., Joris, B., Koerten, H.K., Hopwood, D.A., Titgemeyer, F., & van Wezel, G.P. (2006). The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. Molecular Microbiology, **61**:1237-1251.

Scholz, O., Thiel, A., Hillen, W., & Niederweis, M. (2000). Quantitative analysis of gene expression with an improved green fluorescent protein. p6. European Journal of Biochemistry, **267**:1565-1570.

Singh, R.M., & Adams, E. (1965). Enzymatic deamination of delta-1-pyrroline-4-hydroxy-2-carboxylate to 2,5-dioxovalerate (alpha-ketoglutaric semialdehyde). The Journal of Biological Chemistry, **240**:4344-4351.

Soto, M.J., Jimenez-Zurdo, J.I., van Dillewijn, P., & Toro, N. (2000). *Sinorhizobium meliloti putA* gene regulation: a new model within the family *Rhizobiaceae*. Journal of Bacteriology, **182**:1935-1941.

Sowden, F.J, Griffith, S.M, & Schnitzer, M. (1976). The distribution of nitrogen in some highly organic tropical volcanic soils. Soil Biology and Biochemistry, **8**:55-60.

Thacker, R.P. (1969). Conversion of L-hydroxyproline to glutamate by extracts of strains of *Pseudomonas convexa* and *Pseudomonas fluorescens*. Archiv fur Mikrobiologie, **64**:235-238.

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

Timmers, A.C., Soupene, E., Auriac, M., de Billy, F., Vasse, J., Boistard, P., & Truchet G. (2000). Saprophytic intracellular rhizobia in alfalfa nodules. Molecular Plant-Microbe Interactions, **13**:1204-1213.

Wiethaus, J., Schubert, B., Pfander, Y., Narberhaus, F., & Masepohl, B. (2008). The GntR-like regulator TauR activates expression of taurine utilization genes in *Rhodobacter capsulatus*. Journal of Bacteriology, **190**:487-493.

Wisniewski, J.P., Rathbun, E.A., Knox, J.P., & Brewin, N.J. (2000). Involvement of Diamine Oxidase and Peroxidase in Insolubilization of the Extracellular Matrix: Implications for Pea Nodule Initiation by *Rhizobium leguminosarum*. Molecular Plant-Microbe Interactions, **13**:413-420.

Wood, K.V., Stringham, K.J., Smith, D.L, Volenec, J.J, Hendershot, K.L, Jackson, K.A, Rich P.J., Yang W.J., & Rhodes D. (1991). Betaines of Alfalfa : Characterization by Fast Atom Bombardment and Desorption Chemical Ionization Mass Spectrometry. Plant Physiology, **96**:892-897.

Yuan, Z.C., Zaheer, R., & Finan, T.M. (2005). Phosphate limitation induces catalase expression in *Sinorhizobium meliloti*, *Pseudomonas aeruginosa* and *Agrobacterium tumefaciens*. Molecular Microbiology, **58**:877-894.

Yuan, Z.C., Zaheer, R., & Finan, T.M. (2006). Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. Journal of Bacteriology, **188**:1089-1102.

Table 5.1. Strains and plasmids used in this study.

Strain or plasmid	Relevant characteristic(s)	Source or reference
Strains		
S. meliloti		
Rm1021	Sm^r derivative of wild-type strain SU47	Meade et al. 1982
Rm5000	SU47 *rif-5*; Rf^r	Finan et al. 1984
RmF909	Rm1021 ΔΩ5085-5047::TnV; Sm^r Nm^r	Charles and Finan, 1991
RmFL7003	RmP110 *hypM*::*gfp*$^+$/*lacZ* via int. of pFL7003 (268250 – 268699 nts); Sm^r Gm^r	Cowie et al. 2006
RmP110	Rm1021 with wild-type *pstC*; Sm^r	Yuan et al. 2006
RmP319	Rm1021 *nifH*::*gusA*; Sm^r Gm^r	Cheng et al. 2007
RmP778	RmP110 with *gusA* in the antisense orientation to *smb20568* via integration of pTH1961; Sm^r Gm^r	This study
RmP1113	Integration of pTH2131 into RmP110 (single cross-over); $sucrose^s$ Sm^r Gm^r	This study
RmP1114	RmP110 Δ*hypMNPQ* via excision of pTH2131 to generate in-frame deletion; $sucrose^r$ Sm^r Gm^s	This study
RmP1406	Rm5000 *hypR*::Ω; Rf^r Sm^r Sp^r	This study
RmP1724	RmP110 *hypR*::Ω; Sm^r Sp^r	This study
RmP1838	RmFL7003 *hypR*::Ω; Sm^r Sp^r	This study
RmP1886	RmP110 *hypM*::*gusA* via int. of pTH2494; Sm^r Gm^r	This study
Plasmids		
pJP2	Replicating plasmid with *gusA* reporter gene; Tc^r	Prell et al. 2002
pJQ200 uc1	Suicide vector with *sacB* to select for excision of integrated plasmid; Gm^r	Quandt and Hynes, 1993
pFL7003	449 bp fragment encompassing the *smb20262/hypM* intergenic region into pTH1522; Gm^r	
pLAFR1	IncP cosmid cloning vector; Tc^r	Friedman et al. 1982
pUCP30T	Cloning vector, unable to replicate in *S. meliloti*; Gm^r	GenBank accession no. U33752
pTH1522	Reporter vector used in construction of *S. meliloti* reporter gene fusion library; Gm^r	Cowie et al. 2006
pTH1582	Modified version of pJP2; introduction of promoterless *gusA* from pFus1; Tc^r	Yuan et al. 2005
pTH1705	pTH1522 with expanded MCS; Gm^r	Cowie et al. 2006
pTH1961	627 bp PCR amplified region encompassing	A.M.MacLean and

	smb20568/smb20569 IG region into pTH1705 in the antisense orientation to *gusA*; Gm^r	T.M. Finan, unpublished data
pTH2130	2723 bp *Not*I PCR product encompassing *hypMNPQ* in suicide vector pJQ200; Gm^r	This study
pTH2131	pTH2130 with 1689 bp deletion via *Sal*I digest; Gm^r	This study
pTH2216	953 bp *Hind*III PCR product encompassing *hypR* in vector pUCP30T; Gm^r	This study
pTH2217	ΩSm/Sp^r from pHP45Ω into pTH2216 via *Xho*I; Gm^r Sm^r Sp^r	This study
pTH2439	pLAFR1 derivative complementing Hyp^- phenotype of RmF909; Tc^r	This study
pTH2494	370 bp PCR product encompassing *smb20262-hypM* intergenic region into pTH1705 via *Bgl*II/*Kpn*I; Gm^r	This study
pTH2513	3240 bp PCR product encompassing *hypMNPQ* into pTH1582 via *Nsi*I/*Hind*III; Tc^r	This study

Table 5.2. Expression of *hypM::gfp*$^{+}$*/lacZ* in *S. meliloti* wild-type and *hypR*::Ω backgrounds.

Growth condition	GFP+ Specific Activity [SD[a]]	
	Wild-type[b]	*hypR*::Ω[c]
Succinate	599 (45)	12,920 (145)
Glucose	292 (113)	18,111 (290)
Glycerol	311 (81)	22,521 (200)
L-proline	914 (101)	16,592 (442)
trans-4-hydroxy-L-proline	4431 (173)	17,132 (147)
cis-4-hydroxy-D-proline	5439 (117)	15,322 (95)

[a] SD, standard deviation.

[b] Wild-type, RmP110 *hypM*::*gfp*+*/lacZ*

[c] RmP1838, RmP110 *hypM*::*gfp*+*/lacZ, hypR*::Ω

Table 5.3. Effect of competitors on *trans*-4-hydroxy-L-proline uptake.

Type of competitor	Percent of labeled *trans*-4-hydroxy-L-proline uptake in presence of unlabeled competitor[a]		
	10 μM	50 μM	100 μM
trans-4-hydroxy-L-proline	42 (5)	15 (6)	9 (<1)
cis-4-hydroxy-D-proline	82 (3)	43 (3)	26 (1)
L-proline	40 (6)	12 (4)	5 (1)

[a] Results are reported as percentage of uptake (standard deviation) observed in the absence of competitor, which averaged 9.8 nmols/min/mg in two independent experiments. Uptake was assayed using hydroxyproline-grown cells and 10 μM labeled substrate was used in each assay.

Table 5.4. Expression of *hypM*::*gusA* in young and senescent alfalfa root nodules.

Strain	Gene fusion	Days post inoculation	β-glucuronidase activity[a] (SD)
RmP110	None	42	5 (<1)
		84	7 (<1)
RmP1886	*hypM*::*gusA*	30	41 (1)
		42	28 (1)
		52	27 (<1)
		66	30 (<1)
		84	28 (<1)
RmFL5810	*minD*::*gusA*	42	53 (3)
		84	68 (2)
RmP778[b]	*smb20568*::*lacZ*	42	7 (<1)
		84	8 (<1)

[a]Results are representative of two independent experiments using nodules obtained from a minimum of fourteen alfalfa plants inoculated per strain; β-glucuronidase activity (standard deviation) is the average of assays performed in triplicate.

[b] RmP778 contains a *gusA* fusion in the opposite orientation with respect to P_{20568} and serves as a negative control for endogenous β–glucuronidase activity.

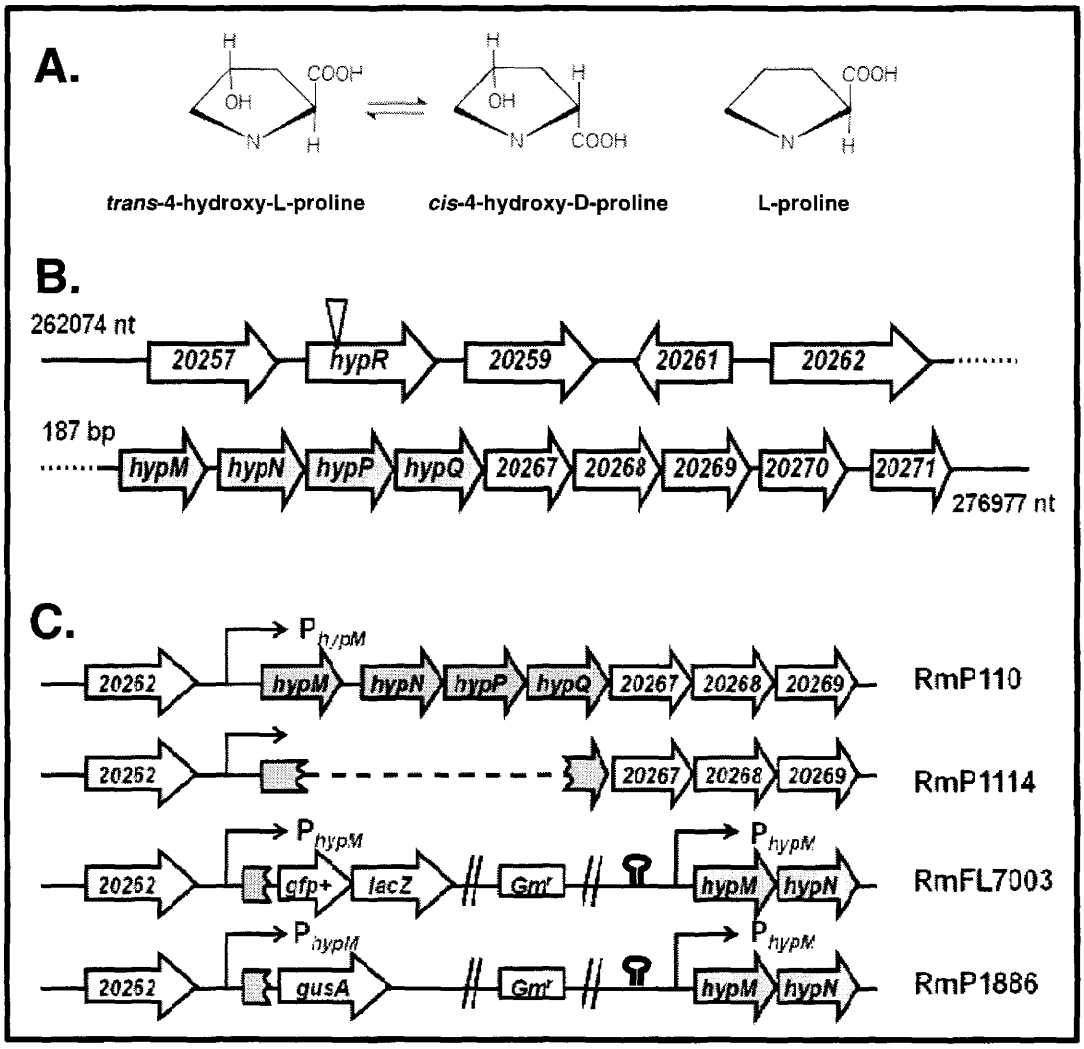


Figure 5.1. Hydroxyproline transport in *S. meliloti*. (A) Structure of two hydroxyproline isomers and L-proline. *S. meliloti* can utilize *trans*-4-hydroxy-L-proline and *cis*-4-hydroxy-D-proline as a sole source of carbon and nitrogen. (B) Schematic diagram of hydroxyproline transport genes encoded on the pSymB megaplasmid. *hypMNPQ* (shaded grey) encode an ABC-type transport system that is required for uptake of *trans*-4-hydroxy-L-proline in *S. meliloti* (this study). The inverted triangle indicates the insertion of an antibiotic resistance cassette within the GntR-type regulator gene (*hypR*). Gene annotation as follows: *smb20257*, adenylate cyclase; *hypR*, GntR-type transcriptional regulator; *smb20259*, dihydrodipicolinate synthase; *smb20261*, malate dehydrogenase; *smb20262*, semialdehyde dehydrogenase; *hypM*, periplasmic solute binding protein; *hypN* and *hypP*, permease protein; *hypQ*, ATP-binding protein; *smb20267*, D-amino acid oxidase; *smb20268* and *smb20270*, proline racemase; *smb20269* and *smb20271*, conserved hypothetical protein. (C) Schematic diagram of genetic constructs used in this study.

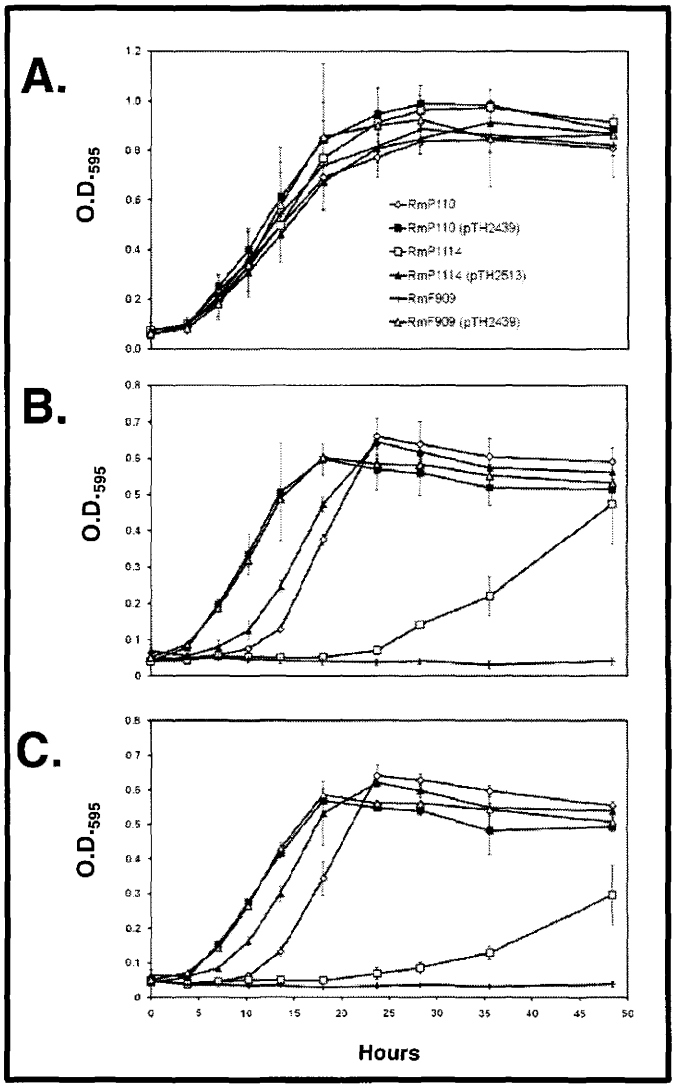


Figure 5.2. Growth of *S. meliloti* strains subcultured into M9-minimal medium with (A) 10 mM succinate; (B) 5 mM *trans*-4-hydroxy-L-proline; or (C) 5 mM *cis*-4-hydroxy-D-proline as sole carbon sources. Inoculum was from washed LBmc grown cells and cultures were incubated with shaking at 30°C for 48 hours; growth of each strain was assayed in triplicate cultures and shown is the average optical density at 595 nm ($O.D._{595}$) with error bars representing standard deviation. The growth rate constant (μ) and mean generation time (g) for wild-type *S. meliloti* in *trans*-4-hydroxy-L-proline (μ = 0.20 $hour^{-1}$; g = 3.4 hours) and *cis*-4-hydroxy-D-proline (μ = 0.22 $hour^{-1}$; g = 3.2 hours) were calculated during exponential growth (corresponding to t_0 = 10.2 and t = 18.1 hrs) using the following formulae. $\mu = ((\log_{10} N - \log_{10} N_0) \times 2.303) / (t - t_0)$ and $g = (\log_{10} N_t - \log_{10} N_0) / \log_{10} 2$, where N equals the number of cells.

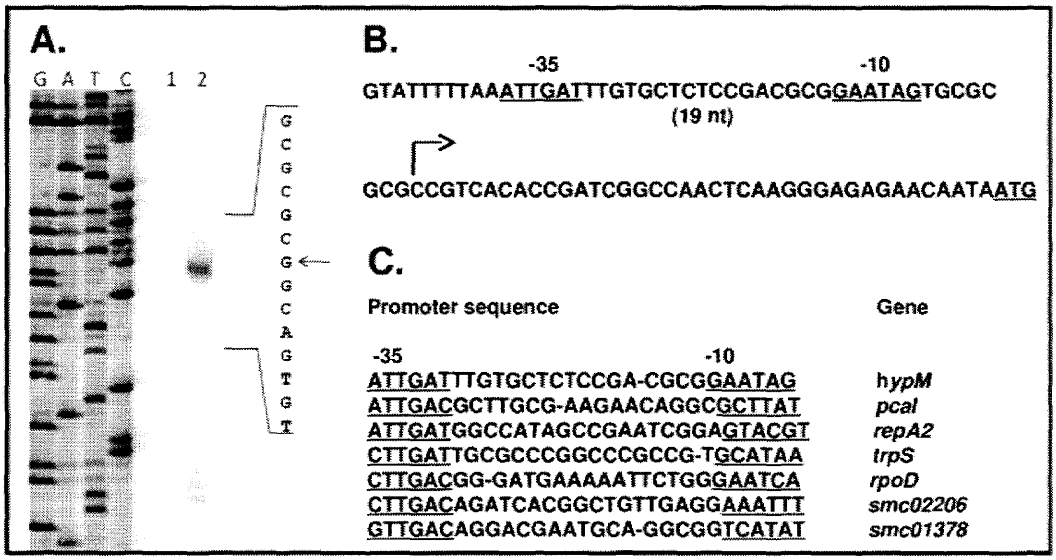


Figure 5.3. Analysis of the *hypM* promoter. (A) Primer extension reactions were performed using RNA isolated from *S. meliloti* wild-type strain RmP110 grown in M9-minimal medium with glycerol (lane 1) or *trans*-4-hydroxy-L-proline (lane 2) as a source of carbon (*trans*-4-hydroxy-L-proline was also added as a sole source of nitrogen). Sequencing reactions are included on the left, with the corresponding nucleotides indicated above each lane. A single major extension product was obtained using mRNA isolated from hydroxyproline-grown cells, corresponding to a start site located 38 nucleotides upstream of the annotated translational start codon associated with *hypM* (as indicated by the arrow). (B) The promoter region associated with *hypM*. The inferred -35 and -10 hexanucleotide regions are underlined and the transcriptional start site is indicated with an arrow bent in the direction of transcription. The translational start site of *hypM* is underlined. (C) The promoter of *hypM* is aligned with six promoters previously identified in *S. meliloti* (MacLellan, et al., 2006).

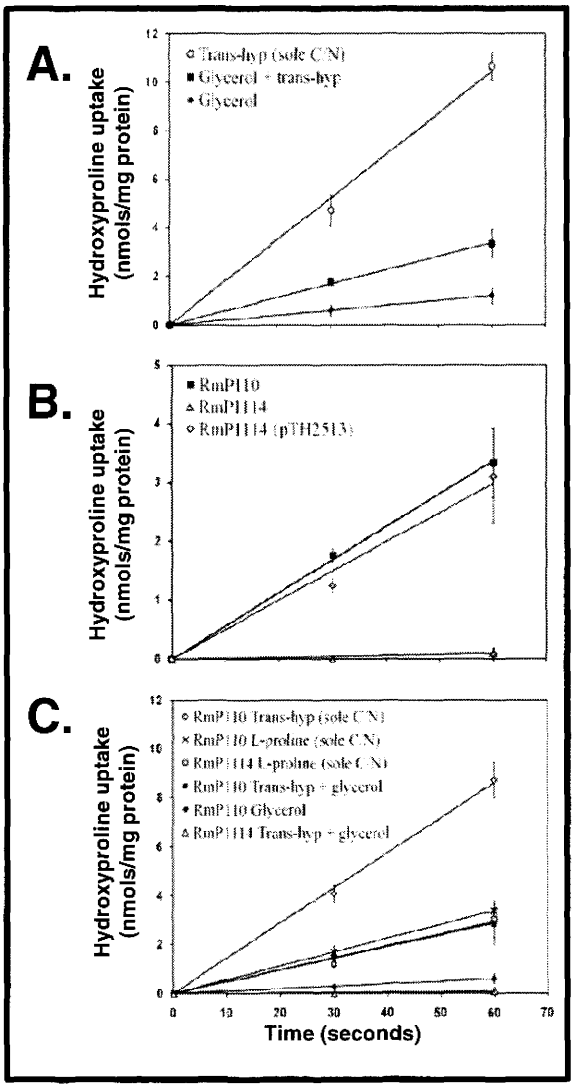


Figure 5.4. Uptake of labeled *trans*-4-hydroxy-L-proline (nmols/mg protein) in *S. meliloti*. (A) Hydroxyproline uptake observed in *S. meliloti* wild-type strain RmP110 grown in M9-minimal medium with carbon sources as indicated. (B) Hydroxyproline uptake observed in *S. meliloti* strains grown in M9-minimal medium with 0.5% glycerol + 5 mM *trans*-4-hydroxy-L-proline. (C) Hydroxyproline uptake observed in *S. meliloti* wild-type RmP110 and RmP1114 (*ΔhypMNPQ*) strains grown in M9-minimal medium with carbon and nitrogen sources as indicated. Assays were performed using 10 μM *trans*-4-hydroxy-L-proline [^{3}H(G)]. In all assays, ammonium chloride was present in growth medium as a source of nitrogen unless otherwise indicated. Trans-hyp, *trans*-4-hydroxy-L-proline. Sole C/N, *trans*-4-hydroxy-L-proline or L-proline is present as a sole source of carbon and nitrogen.

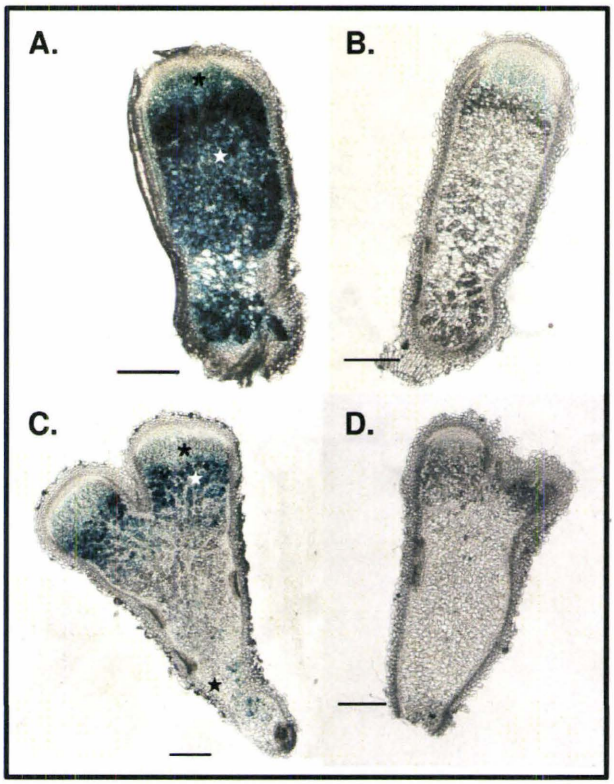


Figure 5.5. Histochemical staining for β–glucuronidase activity in alfalfa nodules. (A) 4 week old nodule inoculated with RmP1886 (*hypM*::*gusA*) shows evidence of staining throughout a longitudinal section, including the prefixing zone (zone II; black asterisk) and nitrogen fixing zone (zone III; white star). (B) 4 week old nodule inoculated with RmP778 (*gusA* in antisense orientation) shows little evidence of staining. (C) 9 week old nodule inoculated with RmP1886. Slight staining was occasionally observed in the saprophytic zone of the senescent nodule (zone V; black star). (D) 9 week old nodule inoculated with RmP778 shows no evidence of staining. Bars = 500 μm.

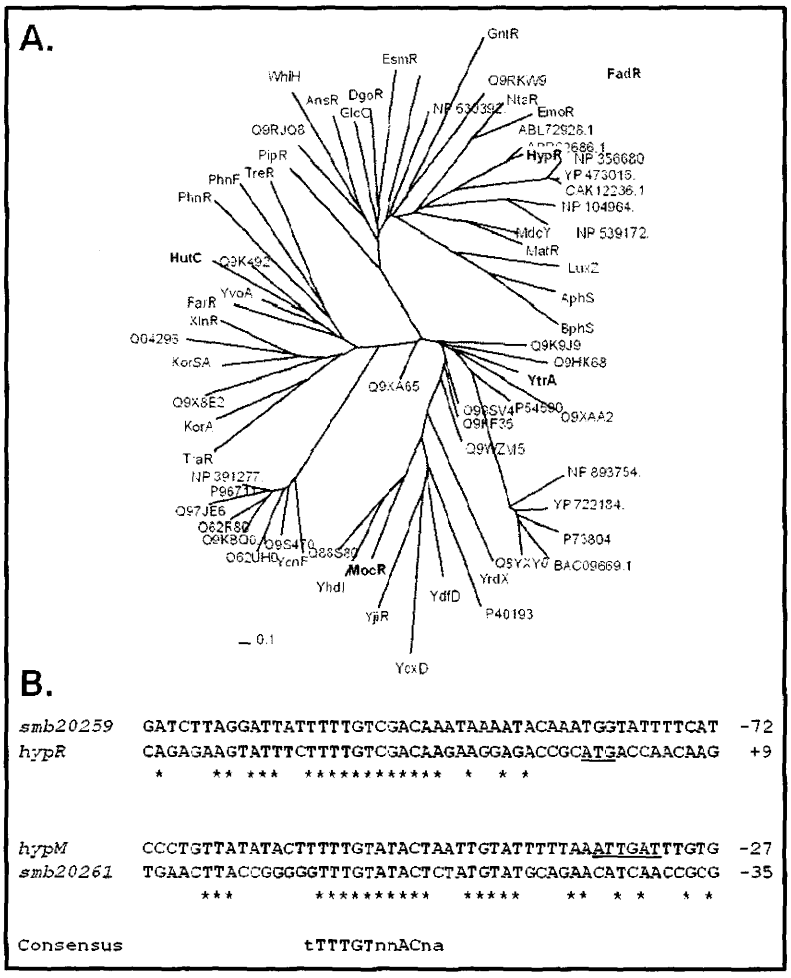


Figure 5.6. (A) Phylogenetic tree of GntR family member proteins. The unrooted tree was based on an alignment of the full-length amino acid sequences obtained for each protein. The multiple alignment was performed using ClustalW (Thompson et al. 1994) with manual adjustment of the aligned sequences. Distances between proteins were determined based upon the Jones-Taylor-Thornton matrix (Jones et al. 1992) using PRODIST (Felsenstein, 1989). (B) Alignment of putative HypR binding sites. The translational start site associated with *hypR* and -35 hexanucleotide region upstream of *hypM* are underlined. Nucleotide positions relative to the annotated start codons are indicated on the right, with the exception of *hypM*, in which the position is indicated with respect to the transcriptional start site. Invariant residues are indicated by an asterisk.

CHAPTER SIX

The growth of *Sinorhizobium meliloti* in bulk soil: A study of saprophytic competence

Allyson M. MacLean, Branka Poduška, and Turlough M. Finan.

Preface

This chapter describes the development and implementation of a soil microcosm system for the study of saprophytic competence in *S. meliloti*. The *S. meliloti* deletion strains were constructed based upon a method developed by Branka Poduska during her Ph.D. thesis, and the majority of the deletion strains used throughout this work were created by Branka. With input from Turlough Finan, all experiments were conceived and designed by myself. With some assistance from Vladmir Jokic (under my supervision), I developed and performed all experiments relating to the growth of *S. meliloti* in the soil microcosms. I have written this chapter in its entirety, with editing by Turlough Finan. This work is to be published as part of a study detailing the construction and phenotypic analysis of *S. meliloti* strains carrying deletions within pSymA and pSymB, with Branka Poduska as the primary author.

6.1 Abstract

Sinorhizobium meliloti is a Gram negative species of bacteria that participates in a symbiotic partnership with the agricultural legume alfalfa, and may also survive as a soil-borne population of free-living saprophytes in the absence of a host plant. In this report, we examined the growth of *S. meliloti* in sterile bulk soil as a means of identifying genes relevant to saprophytic fitness. We report that wild-type *S. meliloti* established a stable population ($\geq 10^8$ cells g^{-1} soil) when inoculated into bulk soil and that this population persisted for several weeks. Utilizing bulk soil microcosms, we assayed the growth of 40 strains of *S. meliloti* carrying large-scale deletions upon the pSymA or pSymB megaplasmid. Of these, two *S. meliloti* strains were identified as exhibiting a reduced ability to colonize bulk soil. RmP801 (Δ1255032 - 1308912 nts; pSymB) exhibited a 2- to 5-fold decrease in cell density when inoculated into sterile soil. RmP1815 (Δ1528150 – 1654191 nts; pSymB) attained a population density 5- to 25-fold lower than that observed by wild-type *S. meliloti*. Unexpectedly, a thiamine auxotroph was able colonize soil microcosms, suggesting that thiamine levels may not limit growth of *S. meliloti* in bulk soil. Although the majority of rhizobia in a natural environment persist as a population of soil-inhabiting microorganisms, this is the first study to systematically identify genes relevant to the survival and propagation of these bacteria in a bulk soil environment.

6.2 Introduction

Soil is a unique and dynamic microenvironment that is colonized by a richly diverse microbial community comprising bacteria, fungi, and archaea. As a growth substrate, soil constitutes a complex and heterogeneous mixture of nutrients and a scavenging bacterium must effectively compete against other microorganisms for access to energy sources. As a habitat, soil is a challenging environment and soil-dwelling microorganisms may encounter a range of stresses, including fluctuations in moisture content, nutrient availability, pH, salt concentration, and temperature (Aislabie, et al., 2006). Organisms may be exposed to chemical (heavy metals, pollutants, oxidative stress), physical (ultraviolet radiation, desiccation, freeze/thaw cycling), and biological stressors (predatory bacteria, protozoa, or bacteriophage, allelopathic interactions). Accordingly, soil-inhabiting bacteria have adapted active mechanisms to counteract such environmental threats and to enable efficient procurement of scarce energy and nutrient sources. In particular, members of the family *Rhizobiaceae* have evolved large genomes that encode a wide array of genes relevant to nutrient acquisition (i.e., transport, catabolic, and regulatory systems) and stress-related responses (MacLean, et al., 2007).

A few studies have identified genes relevant to the ecological performance of soil-dwelling bacteria (Rainey, 1999; Allaway, et al., 2001; Gal, et al., 2003; Rediers, et al., 2003; Silby & Levy, 2004; Ramos-Gonzalez, et al., 2005; Matilla, et al., 2007; Barr, et al., 2008). In vivo expression technology (IVET) has been employed to capture promoters with soil-inducible expression in *Pseudomonas* when inoculated into bulk soil and the rhizospheres of sugar beets, rice, and maize (Rainey, 1999; Gal, et al., 2003; Rediers, et al., 2003; Silby & Levy, 2004; Ramos-Gonzalez, et al., 2005) and *Rhizobium leguminosarum* inoculated into the pea rhizosphere (Barr, et al., 2008). As well, microarray analyses have been performed to examine the expression of *P. putida* in the maize rhizosphere (Matilla, et al., 2007). The implementation of IVET to the study of soil populations has provided valuable insights into environmental conditions as perceived by these bacteria, yet this method does have a few limitations (Rediers, et al., 2005). The identification of soil-inducible promoters may be biased towards highly

expressed genes, as influenced by the strength and method of selection. Transiently expressed genes may be lost during selection, even if the encoded products have important (albeit transitory) roles. As well, only a subset of genes is represented within fusion libraries and IVET studies thus do not constitute a comprehensive portrayal of environmental gene expression.

Sinorhizobium meliloti is a member of the ecologically diverse class α–proteobacteria, which includes obligate and facultative animal and plant pathogens and symbionts. *S. meliloti* is best characterized for its ability to form nitrogen-fixing root nodules upon alfalfa in the context of a symbiotic partnership with this legume. Nonetheless, *S. meliloti* and related rhizobia may alternatively adopt a free-living, saprophytic lifestyle that is independent of a host legume and thus these species may also comprise part of the soil microbial community (Denison & Kiers, 2004). While much emphasis has been placed upon elucidating the complex molecular and developmental dialogue that occurs between plant and microbe during symbiosis, comparatively little is known regarding the genetic basis of soil colonization. Such analyses are particularly pertinent to the development of commercial rhizobial strains, as these often fail to establish long term populations amidst competition from indigenous strains (Triplett & Sadowsky, 1992; Toro, 1996; Romdhane, et al., 2007). As well, the common agricultural practice of crop-rotation with non-leguminous plants such as cereals imposes additional selective pressure for ecological fitness in soil (Howieson, 1995). In this study, we examined the growth dynamics of *S. meliloti* strains inoculated into sterile, bulk soil. We report that wild-type *S. meliloti* strain RmP110 readily colonized bulk soil microcosms and established a stable population that persisted for the duration of the growth assay (i.e., 30 days). We describe two *S. meliloti* deletion mutants that exhibit a reduced ability to colonize bulk soil, indicating that loci within the deleted regions on pSymB must contribute towards saprophytic competence. Surprisingly, an *S. meliloti* thiamine auxotroph effectively colonized bulk soil ($\geq 10^8$ cells g^{-1} soil), indicating that thiamine levels may not limit the growth of this species even in the absence of a plant.

6.3 Materials and Methods

Bacterial strains and growth conditions

All bacterial strains and plasmids used throughout this study are described in Table 6.1. *Escherichia coli* were grown aerobically (with shaking) at 37°C in LB broth. *S. meliloti* strains were grown aerobically at 30°C in LB broth supplemented with 1.0 mM $MgSO_4$, 0.25 mM $CaCl_2$ (LBmc). For *E. coli*, antibiotics were added at the following concentrations (μg/mL): tetracyline (Tc): 10; chloramphenicol (Cm): 20; gentamicin (Gm): 10. For *S. meliloti*, the following concentrations of antibiotics were used (μg/mL): streptomycin (Sm): 200; neomycin: 200; gentamicin: 60; tetracycline (Tc): 5.

Preparation of sterilized soil

Soil was obtained from an agricultural field site located within a dairy farm near Guelph, Ontario, Canada (Figure A1; Appendix). The location of the farm[2], in close proximity to the water supply for the city of Guelph, precludes the application of pesticides, fertilizers, and herbicides onto the fields. On April 2, 2007, approximately 40 kg of soil was removed from a field cultivated with alfalfa and transported to McMaster University as three samples stored in clean buckets.

The soil was spread upon a clean tarp in a greenhouse (McMaster University) and material such as stones, sod, and sticks were removed manually. The soil was allowed to dry for 9 days and was passed through a sieve to remove all particles and fragments larger than 2 mm. The soil was then divided into samples of approximately 100 to 300 g, which were heat sealed in polyethylene freezer bags (FoodSaver; Jarden Corporation). Soil samples were subjected to γ-irradiation (using ^{60}Co as a source) under the supervision of Robert Pasuta (McMaster Institute for Applied Radiation Sciences) at the McMaster University Nuclear Reactor. Roughly one half of the samples were processed immediately for γ-irradiation, at a final dosage of 25.0 kGy (over a period of 54.3 hrs), while the remaining samples were stored at 4°C. Irradiated soil samples were then

[2] Address: Robert Jefferson, RR2 Lcd Royal City Mail, Guelph, ON, N1H 6H8.

screened for sterility via inoculation into a rich growth broth (LBmc broth). Following one to three days of incubation at 30°C (with shaking), the broth was examined for turbidity and 50 and 100 μL aliquots of the broth were plated upon Petri dishes of LB agar, which were incubated at 30°C for two days. Initial testing indicated the presence of multiple species of microorganisms within the irradiated soil. A subsequent attempt was made to sterilize all soil samples (including the previously irradiated samples) with γ–irradiation at a final dosage of 42.3 kGy however subsequent testing revealed the presence of *Deinococcus* in all irradiated samples. Accordingly, soil samples were autoclaved once (123°C; 17 psig; 20 minutes) immediately before use in growth assays. Samples are stored at -20°C (Alef & Nannipieri, 1995).

Chemical analysis of soil

Soil samples were analyzed by the University of Guelph Laboratory Services Agricultural and Food Laboratory (Guelph, Canada). Chemical analyses were performed upon samples of γ–irradiated soil and samples of γ–irradiated and autoclaved soil (Table 6.2).

Estimating soil moisture content

An aliquot soil was weighed and dried for a minimum of two days at ~60°C in an open container to facilitate the evaporation of water. The weight of the dried soil sample was then determined after the incubation period. The difference between the initial and final weights corresponds to the amount of water lost due to evaporation and thus gives an estimate of the amount of water present in the sample. This value was divided by the initial weight to calculate the percentage of water in the soil sample.

Microcosms

The physiochemical properties of the soil used in the growth assays are described in Table 6.2. Initial experiments involved the use of sterile 100 mL polystyrene tissue culture flasks with a vented cap (Sarstedt). In these experiments, 30 or 40 grams (dry weight) of γ–irradiated, autoclaved soil was transferred into the sterile flasks, which were

capped and incubated at room temperature in the dark. However, these early experiments were associated with a high rate of contamination, and thus the majority of growth assays were performed in soil microcosms as follows. 40 grams of γ–irradiated soil (dry weight) was added to 500 mL screw-capped glass bottles (Gibco) which were subsequently autoclaved (123°C; 17 psig; 20 minutes). After cooling to room temperature, microcosms were seeded with an inoculum of *S. meliloti* and moisture content was adjusted as specified (typically 20% (wt/vol)). In a trial experiment, water evaporation was estimated from six microcosms by recording the decrease in soil weight over time. Data obtained from the trial experiment yielded a mean average of 48 μL water evaporated per microcosm per day. Accordingly, moisture content was maintained over the duration of the growth assay through the addition of an equivalent volume of de-ionized, autoclaved water to each microcosm every five to seven days. Soil microcosms were incubated at room temperature (22°C) in the dark.

Inoculation of S. meliloti into soil microcosms

S. meliloti was grown overnight in LBmc and cells were washed three times with saline solution (0.85% NaCl) and once with de-ionized, autoclaved water (ddH_2O). Cells were diluted in ddH_2O to achieve an optical density at 600nm ($O.D._{600}$) of 1.0 (approximately 2.0×10^9 cells/mL) and a dilution series was established by aliquoting cells into ddH_2O. Typically, 1 mL of 10^{-2} dilution (approximately 2.0×10^7 cells) was mixed with 1380 μL ddH_2O and 40 g autoclaved soil samples were seeded with the suspended cell mixture, achieving roughly 5.0×10^5 cells g^{-1} soil with a final moisture content of 20%.

Enumeration of S. meliloti in soil microcosms

A minimum of two subsamples (each consisting of 0.5 g soil (dry weight)) were removed from each inoculated soil microcosm. Soil subsamples were weighed into sterile 2 mL Eppendorf tubes, and the soil was vigorously resuspended into 1 mL saline solution (0.85% NaCl) added to the tube. Samples were centrifuged at a low speed (100

x *g*, or 800 rpm; Eppendorf centrifuge 5415D) to pellet soil particles and debris (bacterial cells remain suspended in supernatant). Supernatant was transferred to a clean Eppendorf tube or 96-well microtiter plate and a dilution series was established by aliquoting supernatant into saline (20 µL supernatent diluted into 180 µL saline). Various dilutions of supernatant were plated upon LB agar (streptomycin was added to LB agar plates if indicated) and plates were incubated at 30°C for three days. Following incubation, the number and morphology of colonies per plate was recorded and total colony forming units per gram soil was calculated.

Identification of soil species via amplification of 16S rDNA

Several morphologically distinct colony types were isolated from soil following an initial round of γ–irradiation (total dosage, 25 kGy). Three unique colony types were purified by sequential streaking upon LB agar plates and incubation at 30°C. One representative colony from each purified strain was inoculated into LBmc broth and the culture was incubated at 30°C for two to three days. Bacterial cultures were harvested and genomic DNA was extracted as follows. 5 mL culture was centrifuged and pelleted cells were washed three times in saline (0.85% NaCl) and resuspended in 1 mL $T_{10}E_{25}$ (10 mM Tris-HCl, 25 mM EDTA; pH 8). 25 µL Proteinase K (10 mg/mL), 50 µL 25% SDS, and 125 µL 5 M NaCl was added to the mixture, which was incubated at 65°C for four hours. Lysate was allowed to cool to room temperature and was extracted three times in phenol:chloroform (1:1 (v/v)) and once with 100% chloroform. The aqueous phase was transferred to a clean 15 mL Falcon tube and genomic DNA was precipitated with 2 M ammonium chloride (final concentration) and 2 volumes of 100% ethanol. Precipitated DNA was washed twice with 70% ethanol and dried DNA was dissolved into 50 µL buffer $T_{10}E_1$ (10 mM Tris-HCl, 1 mM EDTA; pH 8). Universal primers (5'CTYAAAKRAATTGRCGG RRRSSC '3) and (5'CGGGCGGTGTGTRCAARRSSC 3') were used to PCR amplify the 16S rRNA gene from genomic DNA samples (Rivas et al., 2004). As a control, *S. meliloti* genomic DNA was extracted alongside the three

unidentified bacterial cultures and PCR amplification was also performed upon *S. meliloti* 16S rDNA.

Construction of S. meliloti deletion strains

Unless specified, the *S. meliloti* deletions strains used in this study were previously constructed by Branka Poduška (B. Poduška and T. M. Finan, manuscript in preparation). Briefly, deletions were generated using a FRT/Flp recombination system for site-specific excision of targeted regions within pSymA or pSymB. As a means of introducing FRT (Flp recombinase target) sites, two plasmids (derivatives of pTH1522 (Cowie, et al., 2006) and pTH1937 (B. Poduška and T. M. Finan, manuscript in preparation); each carrying a single FRT site) were sequentially integrated into the megaplasmid via homologous recombination. Subsequently, a replicating plasmid encoding Flp recombinase was transferred through conjugation into the recombinant *S. meliloti* strains. Upon expression, Flp recombinase catalyzes site-specific recombination through interaction with the FRT sites, resulting in the excision of intervening DNA (Andrews, et al., 1985). When appropriate, putative deletion strains are screened for the loss of antibiotic resistance markers associated with the integrated plasmids (i.e., Gm^r and Nm^r). The genotype of each strain was confirmed using PCR analysis.

Complementation of RmP1815 with pLAFR1 cosmid pT8

pT8 was originally isolated from a pLAFR1 clone bank of *S. meliloti* Rm1021 genomic DNA on the basis of its ability to complement an *S. meliloti* thiamine auxotroph (Finan et al., 1986). To confirm the presence of *thiCOGE* within pT8, DNA sequencing was performed by MOBIX (McMaster University, Hamilton, Canada) using primers complementary to the cosmid (5' CCTCGATCAGCTCTTGCACTCG 3') and (5' GCAGGTGCTGGCATCGACATTCAGC 3') (Vanbleu, et al., 2004). Based upon DNA sequence analysis, pT8 contains 24 kb of *S. meliloti* sequence (corresponding to 1620000 – 1643981 nts; pSymB) that encompasses *thiCOGE*. pT8 was transferred into RmP1815

via conjugation and transconjugants were selected upon LB agar plus tetracycline and streptomycin.

6.4 Results

Sterilization of soil via γ–irradiation and autoclaving

One hundred to 300 gram soil samples (total mass; approximately 20 kg) were subjected to γ–irradiation (total dosage, 25 kGy) in an attempt to achieve soil sterilization. To examine the effectiveness of the sterilization procedure, samples of the irradiated soil were inoculated into a rich growth broth, and aliquots of the broth were plated upon LB agar (Trevors, 1996). Multiple colonies were evident following incubation of the agar plates and a minimum of five morphologically distinct colony types were observed within a total of three plates. We wished to identify a subset of the contaminating microorganisms, and three colony types were purified via sequential transfer upon LB agar. Genomic DNA was isolated from a bacterial culture inoculated with one representative from each colony type and the nucleotide sequence of the PCR amplified 16S rRNA gene (475 nts) was determined (Appendix). A BLASTN search of the NCBI database reveals that the sequences most closely match those corresponding to the genera *Chelatococcus*, *Rhizobium*, and *Deinococcus*.

The isolation of viable bacteria from irradiated soil samples indicated that the dosage of γ–irradiation was insufficient to effectively sterilize the soil. Samples were subjected to a second round of γ–irradiation (total dosage, 42.3 kGy) however a single colony type was subsequently isolated from these soil samples. Sequencing of 16S rRNA gene identified this microorganism as belonging to the genus *Deinococcus*.

The presence of a viable population of *Deinococcus* (<100 cell g^{-1} soil) necessitated an alternative method of soil sterilization. Prior to the adoption of autoclaving as a standard method of soil sterilization, we compared the growth of *S. meliloti* wild-type derivative strain RmP110 in autoclaved and non-autoclaved soils (Figure 6.1). Microcosms containing γ–irradiated soil (± autoclaving) were inoculated

with approximately 10^5 cells g^{-1} soil and cell density was measured over a period of up to 20 days. The *S. meliloti* soil population increased to $>10^8$ cells g^{-1} soil within seven days post-inoculation, as observed in autoclaved and non-autoclaved soil samples. Examination of growth curves obtained from both soil treatments indicates that autoclaved soil supports the growth of an *S. meliloti* population to a level and at a rate that is equivalent to or exceeds that observed in non-autoclaved soils. Consequently, subsequent experiments were performed using γ–irradiated and autoclaved soil.

Influence of soil moisture content upon growth of S. meliloti

S. meliloti strain RmP110 was inoculated into sterile soil microcosms with varying percent soil moisture content to examine the influence of this abiotic variable upon cell survival and population growth (Figure 6.2). In all three conditions tested (percent moisture content (vol/wt); 20 to 30%), the population of *S. meliloti* RmP110 increased from an initial 10^5 cells g^{-1} soil to achieve a final density greater than 10^8 cells g^{-1} soil. A comparison of growth curves indicates that cell density was lowest in samples obtained from microcosms with a moisture content of 30%. Soils with moisture contents of 20 to 25% yielded similar growth curves, and subsequent experiments were performed upon microcosms with a moisture content adjusted to 20%. This value is comparable to those used in other studies which typically range from 15 to 30% (Lowendorf, et al., 1981; Heynen, et al., 1988; Kinkle, et al., 1993; Selbitschka, et al., 1995).

Growth of wild-type S. meliloti in a bulk soil microcosm

Having established an appropriate protocol to monitor the growth of *S. meliloti* in sterile bulk soil, we assayed the growth of wild-type *S. meliloti* in a soil microcosm for 30 days. *S. meliloti* strain RmP110 was inoculated into soil to a density of 10^3 cells g^{-1} soil and the population increased to approximately 5×10^8 cells g^{-1} soil within seven days, at which point a stable population was maintained for the remaining three weeks (representative data are shown in Figure 6.3). These assays indicate that *S. meliloti* wild-type strain RmP110 has the ability to persist as a viable population in bulk soil for one

month under the conditions tested. These growth data are consistent with previous reports detailing the persistence of *R. leguminosarum* and *S. meliloti* populations assayed in sterile soils to examine the effects of soil pH (Lowendorf, et al., 1981), protozoa predation (Heynen, et al., 1988), and other factors upon cell survival (Cebolla, et al., 1993; Selbitschka, et al., 1995).

Identification of loci within pSymB that contribute towards saprophytic competence in S. meliloti

We wished to identify genes encoding products essential to the survival and growth of *S. meliloti* in soil. As a means of achieving this objective, we screened *S. meliloti* strains carrying large-scale deletions within the pSymA or pSymB megaplasmids for growth in a sterile soil microcosm. The *S. meliloti* strains were generated as part of the PhD research project of Branka Poduška, through the use of a Flp/FRT recombination system in which regions within either megaplasmid were systematically deleted (Branka Poduška and T. M. Finan, unpublished data) and the size of each deletion ranged from 22 to 346 kb (average, 54 kb). A total of 40 deletion strains were assayed for growth in bulk soil and the deletions associated with these strains comprise a significant coverage of each megaplasmid (pSymA, Δ1,339,708 bp (or 99%); pSymB, Δ1,551,371 bp (or 92%)). *S. meliloti* strain RmP110 was included in all growth assays as a positive control, and microcosms were inoculated individually with *S. meliloti* strains to approximately 10^3 cells g^{-1} soil (Figures 6.4 – 6.6).

The majority of *S. meliloti* deletion strains yielded growth curves in soil that were similar to the wild-type control, indicating that the loss of genes encoded with the deleted regions did not adversely impact the ability of these strains to survive and grow in soil. Of the forty strains examined, two strains reproducibly demonstrated a reduced fitness in soil when compared to the wild-type control (Figure 6.5, ΔB116; Figure 6.6, ΔB123). These strains carry non-overlapping deletions within the pSymB megaplasmid, and in both cases the reduced fitness was reflected by a decrease in total colony forming units observed following inoculation into soil.

S. meliloti strain RmP801 (ΔB116; Δ1255032 - 1308912 nts; pSymB) exhibits a subtle growth phenotype when assayed in sterile soil, as reflected by a 2- to 4-fold decrease in final cell density as compared to a wild-type control (Table 6.3). When inoculated to an initial density of 10^3 cells g^{-1} soil, RmP801 increases up to 10^8 cells g^{-1} soil as observed with the wild-type RmP110. However the population of the mutant strain remains stable at a level that is approximately 2-fold lower than that observed for RmP110, as demonstrated in three independent experiments (representative data are shown in Figure 6.5B). RmP801 is characterized by a 54 kb deletion within pSymB that includes 53 genes, including genes predicted to encode products relevant to oxidative stress response and nucleoside catabolism (Table 6.4).

S. meliloti strain RmP1815 carries a 126 kb deletion (ΔB123) within pSymB that encompasses 118 genes, including the thiamine biosynthetic gene cluster *thiCOGE*. When inoculated into bulk soil, the population of RmP1815 initially increases at a rate comparable to that observed in the wild-type control however population growth levels off at a final cell density of approximately 10^7 cells g^{-1} soil (Figure 6.6). In contrast, the population of RmP110 exceeds 10^8 cells g^{-1} soil after one week incubation under identical conditions. The severity of the phenotype exhibited by RmP1815 increases over the duration of the growth assay, with an average 5-fold decrease in cell density (as compared with RmP110) observed within the first seven days and a greater than 25-fold reduction observed after two weeks incubation.

Thiamine auxotrophy is not linked to the reduced saprophytic competence of RmP1815

The thiamine biosynthetic gene cluster is encompassed by the deletion within RmP1815, making this strain a thiamine auxotroph. To determine whether the thiamine auxotrophy was responsible for the phenotype exhibited by RmP1815 in soil, the *thiCOGE* gene cluster was introduced into RmP1815 *in trans* via a pLAFR1 cosmid (pT8; Finan, et al., 1986). The transconjugant *S. meliloti* strain, alongside appropriate control strains, was inoculated into soil and growth was monitored for a total of 17 days

(Figure 6.7A). Consistent with earlier experiments, RmP1815 increased from an initial population of 10^3 cells g^{-1} soil to reach an average cell density of 10^7 cells g^{-1} soil whereupon the population remained constant at a level as much as 20-fold lower than that observed in microcosms inoculated with wild-type *S. meliloti*. RmP1815 (pT8) yielded a growth curve comparable to that of RmP1815, indicating that the presence of *thiCOGE in trans* did not complement the growth phenotype exhibited by the deletion strain in bulk soil (Figure 6.7A). The presence of the Tc^r cosmid in soil-borne RmP1815 (and RmP110) was monitored during the course of the assay by patching soil isolates upon LB agar plus tetracycline. This screening indicated that the cosmid was maintained within the majority of the soil population. The failure of the *thiCOGE* gene cluster to rescue the phenotype exhibited by RmP1815 in soil indicated that the loss of the *thi* genes was unlikely to (solely) account for this phenotype.

We constructed a series of *S. meliloti* strains carrying subdeletions within the 126 kb region encompassed by ΔB123 to narrow down the number of candidate genes whose deletion might be associated with the reduced saprophytic competence of RmP1815 (Figure 6.8). Six additional strains were screened for growth in sterile soil; of these, one strain demonstrated a reduced fitness consistent with that exhibited by RmP1815 (Figure 6.7B). This strain, RmP798, contains a 46 kb deletion that does not encompass the *thiCOGE* gene cluster. We note that a strain (RmP1884) in which the *thi* locus was deleted readily colonized the sterile soil microcosms (Table 6.5).

6.5 Discussion

Members of the genera *Sinorhizobium, Mesorhizobium, Rhizobium,* and *Bradyrhizobium* (collectively referred to as rhizobia) are well known for their ability to participate in a symbiotic partnership with a leguminous host. However, the lifestyle afforded these bacteria is not limited to that of a plant-inhabiting microsymbiont and the vast majority of rhizobia exist in the form of free-living and soil-dwelling organisms (Denison & Kiers, 2004; Sullivan, et al., 1996). In fact, nonsymbiotic strains of rhizobia (lacking genes required to fix nitrogen) may outnumber their symbiotic counterparts by

as much as 40 to 1 within the rhizosphere of a host plant (Segovia, et al., 1991). While the cultivation of a host plant may encourage the establishment of a rhizobial soil population, the long-term residence of rhizobia within field soils has been documented years subsequent the cultivation of a legume. For example, *Rhizobium leguminosarum* biovars *viciae* and *trifolii* have been detected at 10^4 to 10^5 nodulating cells g^{-1} soil in an agricultural field with no cultivation of legumes in twenty years, as determined by most probable number studies based upon plant infection tests (Hirsch, 1996). This field study reveals that rhizobia may readily colonize soils in the absence of host legume cultivation and emphasizes that these bacteria are well adapted to a saprophytic lifestyle. Surprisingly, few studies have attempted to identify the genes that play a role in the survival of rhizobia in soil. This study was undertaken to examine the growth dynamics of *S. meliloti* in a bulk soil environment and to identify genes encoded upon the megaplasmids pSymA and pSymB that play a major role in saprophytic competence.

Utilizing a rich loam soil harvested from an alfalfa field, we instituted a protocol that allowed the enumeration of *S. meliloti* in a sterile soil microcosm. We demonstrate that wild-type *S. meliloti* will attain a density of 10^8 cells g^{-1} soil when inoculated into sterile soil and that a stable population will persist at this level for a minimum of several weeks (Figure 6.3). In order to estimate the percentage of cells present in the supernatant upon resuspension of soil samples with saline (and thus the percentage of cells enumerated), we determined the percent recovery of cells from soil within two hours of inoculation. Our results indicate approximately 11% of the *S. meliloti* population (percent recovery 6 to 18%; average 11% ± 4, N = 7 samples) is recovered (data not shown), and thus up to 90% soil-borne bacteria remain adhered to the soil and are not enumerated. The efficiency with which bacterial cells adhere to soil particles varies between different species. For example, 94% of the total population of *Pseudomonas fluorescens* Pf0-1 remained associated with soil particles following three successive washes with saline, as compared to an estimated 12% of an *Escherichia coli* population under identical conditions (DeFlaun, et al., 1990). As well, the growth phase and nutrient status of cells may affect the efficiency of adherence, with cells harvested during the

exponential phase or from minimal medium adhering more efficiently to soil (DeFlaun, et al., 1990). The molecular basis of attachement is not well understood, however there is evidence that flagella and/or extracellular polysaccharides may be involved in mediating a firm adhesion of cells to soil (Balkwill & Casida, 1979; DeFlaun, et al., 1990).

The majority of deletion strains examined effectively colonized bulk soil microcosms under the conditions employed in this study and attained a final cell density comparable to that of the wild-type strain (approximately 10^8 cells g^{-1} soil) (Figures 6.4 to 6.6). The observation that large-scale gene loss (in most cases) does not negatively influence the growth of *S. meliloti* in soil likely reflects the extensive array of genes encoded in the genomes of rhizobia that are relevant to the saprophytic lifestyle (MacLean, et al., 2007). It is worth noting that we assayed the growth of strains in the context of individually inoculated microcosms, thus populations were allowed to establish in the absence of biotic factors such as intra- and interspecies competition. While the study of clonal soil populations greatly facilitates analyses, it is not representative of a real world environment in which a soil microorganism comprises part of a complex and dynamic microbial community. Accordingly, it is not reasonable to extrapolate the successful colonization of a sterile soil setting to an equivalent success in the presence of competing (and predating) microorganisms.

Our screening of *S. meliloti* deletion strains led to the identification of two strains exhibiting a reduced fitness in sterile bulk soil. The observation that both strains carry non-overlapping deletions upon the pSymB megaplasmid is interesting. pSymB is particularly enriched in solute transport systems and catabolic pathways, leading to the proposal that this replicon acts as a major contributor towards the saprophytic fitness of *S. meliloti* (Finan, et al., 2001).

S. meliloti strain RmP801 (ΔB116) reproducibly demonstrated an average 2-fold decrease in cell density (as compared to wild-type) when added as a sole inoculant in bulk soil (Table 6.3). The strain carries a 54 kb deletion that encompasses 53 annotated open reading frames, many of which do not have an assigned function (23 conserved or hypothetical proteins; Table 6.4). Of the 30 genes with an assigned function, almost one

half encode either putative transcriptional regulators (4) or transport systems (8). One gene located within the deleted region (*smb20860*) encodes a protein with strong similarity (68% identity) to a non-heme chloroperoxidase purified from *Pseudomonas pyrrocinia* (now *Burkholderia pyrrocinia*) (Wiesner, et al., 1988). The *B. pyrrocinia* enzyme catalyzes the chlorination of indole to 7-chloroindole, and is likely involved in the production of the chlorinated antibiotic pyrrolnitrin (Wiesner, et al., 1986; Wiesner, et al., 1988). *smb20860* also exhibits sequence similarity (48% identity) to a paralogous *S. meliloti* chloroperoxidase (*smc01944*) that is involved in neutralizing exogenous hydroperoxides via secretion into the periplasm and external environment (Barloy-Hubler, et al., 2004). Consistent with a role in oxidative stress response, expression of *smb20860* is up-regulated (> 3-fold) in response to exposure to 10 mM H_2O_2 (Barloy-Hubler, et al., 2004). Previous studies with *Pseudomonas* have demonstrated that the expression of genes associated with oxidative stress is up-regulated in soil (Rainey, 1999; Gal et al., 2003). Possibly, *smb20860* encodes a protein involved in the neutralization of reactive oxygen species encountered in soil, and loss of this gene renders a cell more susceptible to oxidative damage, thereby resulting in a decreased overall fitness.

An examination of the region deleted in RmP801 also reveals a few genes which may encode products relevant to the catabolism of purines in *S. meliloti*. Nucleosides accumulate in soil (Phillips, et al., 1997), as resulting from microbial biomass turnover and the natural decay of organic matter. Nucleosides may thus afford a source of nitrogen (and carbon) to soil-dwelling microorganisms, and expression of a putative nucleoside uptake system is induced in soil-borne *P. fluorescens* strain Pf0-1 (Silby & Levy, 2004). *S. meliloti* encodes xanthine- and allantoin-inducible uptake systems on the chromosome (Mauchline, et al., 2006), suggesting that *S. meliloti* may also salvage or catabolise these purine derivatives.

S. meliloti gene *smb20872* shows sequence similarity to a thyroid-hormone binding receptor in vertebrates (transthyretine), and thus has been annotated as a transthyretin-like protein. This protein also exhibits similarity (49% identity) to an enzyme (PucM; 5-hydroxyisourate hydrolase) involved in the catabolism of purines as an

energy and nitrogen source in *Bacillus subtilis* (Schultz, et al., 2001; Jung, et al., 2006). Consistent with this classification, we note in Smb20872 the presence of conserved amino acids specific to 5-hydroxyisourate hydrolase (Jung, et al, 2006). Expression of *pucM* in *B. subtilis* is up-regulated during nitrogen-limiting conditions as a means of catalyzing the conversion of uric acid to allantoin, which is subsequently metabolized to ammonia and carbon dioxide (Schultz, et al., 2001). The oxidation of uric acid to allantoin also requires the decarboxylation of 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU), as mediated by OHCU decarboxylase (Ramazzina, et al., 2006). Interestingly, *smb20874* (annotated as a conserved hypothetical protein) contains a conserved domain associated with OHCU decarboxylase (NCBI Conserved Domains; E value: 8e-55). Thus it may be that *smb20872* and *smb20874* encode two of three enzymes required for the metabolism of uric acid to allantoin (Figure 6.9). Allantoin is further metabolized in *B. subtilis* to ureidoglycolic acid, which is subsequently converted to glyoxylate, ammonia, and CO_2 via ureidoglycolate hydrolase. *smb20873* is predicted to encode ureidoglycolate hydrolase, further linking the *smb20872-smb20874* gene cluster to purine catabolism. If nucleosides constitute an important nutrient source, the growth phenotype exhibited by RmP801 may be due to the loss of genes (*smb20872-smb20874*) encoding products required for purine catabolism.

A second *S. meliloti* strain also exhibited a reduced fitness in bulk soil, as reflected by a 5 to 25-fold decrease in soil cell density when compared to a wild-type control (Figure 6.6A). RmP1815 (ΔB123) carries a 126 kb deletion upon pSymB, and we conducted an additional round of growth assays utilizing strains carrying smaller deletions within this region to further define loci associated with this phenotype (Figure 6.7B). We determined that *S. meliloti* strain RmP798 grows poorly when inoculated into sterile soil (yielding a growth curve comparable to RmP1815), and thus the 46 kb deletion carried by this strain must encompass genes relevant to saprophytic competence. The deletion associated with RmP798 spans 41 annotated open reading frames, and includes 13 conserved or hypothetical genes and 2 transposase insertion sequences (Table

6.6). In addition, we note the presence of a phosphonate metabolic gene cluster (*phnGHIJKL*) within the deleted region.

Rhizobia can utilize a variety of phosphonates as a source of phosphorus, including herbicides such as glyphosate (Liu, et al., 1991). *S. meliloti* encodes a C-P lyase that catalyzes the cleavage of the chemically stable carbon-phosphorus bond (Parker, et al., 1999). The C-P lyase comprises a membrane-associated multiple subunit assembly encoded by *phnGHIJKL*, and deletion of this gene cluster in *S. meliloti* renders a strain that is unable to process a wide range of phosphonates (Parker, et al., 1999). Although the availability of phosphorus in most soils is growth limiting, phosphonates accumulate in soil as resulting from anthropogenic (phosphonate-based fungicides and herbicides) and natural sources (primarily microbial and fungal) (Tate & Newman, 1982; Turner, et al., 2003; Koukol, et al., 2008). The loss of the *phnGHIJKL* gene cluster in strain RmP1815 would prevent the degradation of most phosphonates (as dependent upon C-P lyase activity) and may thereby considerably reduce the phosphorus resources available to this mutant in a soil microcosm.

In order to effectively monitor the growth of *S. meliloti* in a soil microcosm, we required an efficient method to sterilize soil while minimizing any physical or chemical changes that might occur as a result of the sterilization procedure. γ–irradiation is an effective method to achieve soil sterilization with minimal impact upon soil properties (Alef & Nannipieri, 1995; Lotrario, et al., 1995; McNamara, et al., 2003), however the heterogeneous nature of soil and complexity of soil-dwelling communities make it difficult to estimate an appropriate dosage. As well, larger samples of soil are considerably more difficult to effectively sterilize, in part due to sample shielding from γ–irradiation (McNamara, et al., 2003). We initially selected a dosage level (25 kGy) that is commonly utilized to sterilize medical equipment, foodstuffs, and that has previously been shown to sterilize soil samples (McNamara, et al., 2003). However, multiple colony types were isolated from our soil samples following γ–irradiation, suggesting that sterilization was not effective. As a means of identifying a subset of the contaminating microflora, 16S rDNA sequences associated with three colony types were determined.

BLASTN analysis revealed that the 16S rDNA sequences most closely correspond to soil-inhabiting bacteria of the genera *Rhizobium*, *Deinococcus*, and *Chelatococcus*. This result implies that the contaminating microorganisms are likely due to the incomplete eradication of the microbial soil community as opposed to originating from within the laboratory setting. A higher dosage of γ–irradiation (42 kGy) was sufficient to eliminate all but one microbial species in soil samples (as determined by growth upon LB agar) which was subsequently identified as a member of *Deinococcus* sp.

Deinococci are Gram positive bacteria that typically inhabit rich soils and have also been isolated from sewage sludge and animal feeds (Ito, et al., 1983). This genus is extremely resistant to ionizing radiation due an unusually efficient DNA repair mechanism that may have evolved as an adaptation to survive desiccation (Mattimore & Battista, 1996). Although the population of *Deinococcus* in our samples is small (less than 100 cells g^{-1} soil), these bacteria grow readily in a soil microcosm and we found that they will establish a population of $>10^8$ cells g^{-1} soil within one week (data not shown). We examined the feasibility of amending soil with streptomycin as a means of eliminating *Deinococcus* from our samples (*S. meliloti* RmP110 is resistant to streptomycin) and initial testing indicated that *Deinococcus* is sensitive to the antibiotic when grown upon LB agar supplemented with streptomycin at 5 μg mL^{-1}. However the addition streptomycin to soil (up to 5 mg g^{-1} soil) was ineffective at completely eliminating the *Deinococcus* population. The discrepancy between streptomycin sensitivities in soil versus solid growth medium is likely due to the adsorption of the antibiotic upon soil particles and/or the difficulty of achieving a homogenous distribution of the antibiotic (added as an aqueous solution) within the microcosm.

The presence of a viable population of *Deinococcus* necessitated the inclusion of an autoclaving step to achieve soil sterilization. Autoclaving has been demonstrated to alter both the chemical and physical properties of soil (Jenneman, et al., 1986; Lotrario, et al., 1995). Particularly, autoclaving reduces the surface area of soils by causing soil aggregation, which may alter the adsorption of organic compounds (Lotrario, et al., 1995). Autoclaving has also been demonstrated to alter the levels of chloride, aluminum,

potassium, and silica on the surfaces of sandstone, resulting in an overall increased negative charge associated with the surface of the rock particles (Jenneman, et al., 1986). Finally, autoclaving alters soil chemistry, in part by increasing the concentration of ammonium (Alef & Nannipieri, 1995) and also by destroying heat-labile molecules and organic compounds. Analyses of the physiochemical properties of soil samples ± autoclaving reveal differences in carbon (total, inorganic, and organic) and nitrogen (ammonium, nitrate, and nitrite) concentrations (Table 6.2). However, our results revealed that wild-type *S. meliloti* grew well in autoclaved soils and that growth may even exceed that observed in non-autoclaved soils (Figure 6.1). The latter observation may be attributed to the presence of a competing population of *Deinococcus* inhabiting non-autoclaved soils: after twelve days incubation, the population of *Deinococcus* (estimated at 6.8×10^8 cells g^{-1} soil) was roughly equivalent to that of *S. meliloti* (data not shown). It is important to consider that we should not assume autoclaved and non-autoclaved soils to have equivalent physiochemical properties based upon these results. Possibly, differences in soil chemistry may be manifest in the expression profile of soil-borne *S. meliloti*. This caveat must be considered as relevant to the future analyses of *S. meliloti* gene expression in autoclaved soils.

Thiamine is a water-soluble vitamin that is exuded from plant roots into the surrounding rhizosphere (West, 1939). In addition, many strains of rhizobia encode genes relevant to the *de novo* synthesis or salvage of this vitamin, which is a cofactor essential to the function of carbohydrate and amino acid metabolic enzymes (Finan, et al., 1986; Miranda-Rios, et al., 1997; Karunakaran, et al., 2006; Taboada, et al., 2008). We initially hypothesized that the poor growth of *S. meliloti* strain RmP1815 in bulk soil was linked to thiamine auxotrophy, as this strain lacks a gene cluster necessary for *de novo* thiamine biosynthesis (*thiCOGE*). However, introduction of the *S. meliloti* gene cluster (*thiCOGE*) *in trans* failed to rescue the growth phenotype exhibited by RmP1815 in soil (Figure 6.7A and Table 6.5). Most importantly, the reduced saprophytic competence exhibited by RmP1815 was shown to be linked to the deletion of a 46 kb region that does not include the *thi* genes (i.e., RmP798 is a thiamine prototroph). For these reasons, we

concluded that the poor colonization of RmP1815 in bulk soil was unrelated to its status as a thiamine auxotroph. We observed that *S. meliloti* strains RmP1882 and RmP1884 (respectively carrying 16 and 22 kb deletions encompassing the *thi* gene cluster; Figure 6.8) grew readily in a soil microcosm (Figure 6.7B and Table 6.5). This result is surprising, as prior studies have indicated that thiamine levels may be limiting in the alfalfa, pea, and tomato rhizospheres (Simons, et al., 1996; Streit, et al., 1996; Allaway, et al., 2001). As well, a *P. fluorescens* thiamine auxotroph exhibited a reduced ability to colonize the tomato rhizosphere when inoculate alone or in competition with the parental strain (Simons, et al., 1996), indicating that thiamine auxotrophy may impair ecological fitness. Presumably, autoclaving soil would further decrease the availability of this heat-labile vitamin in our bulk soil samples. Nonetheless, although we did not quantify the concentration of thiamine, our data indicate that levels of this vitamine are sufficient in autoclaved bulk soil to permit the growth and propagation of two *S. meliloti* thiamine auxotrophs (RmP1882 and RmP1884) to a level that is comparable to that of the wild-type strain.

Saprophytic competence is a multidimensional and complex phenotype that influences the ecological success of soil-dwelling microogranisms. The successful nodulation of a host legume by rhizobia is necessarily preceded by and dependent upon a successful colonization of the plant rhizosphere, and thus the saprophytic fitness of a rhizobial strain contributes directly towards its nodulation competitiveness. Studies have shown that commercial rhizobial strains typically do not persist in soils and are replaced by indigenous strains which are better adapted to survive in the local soil environment (Triplett & Sadowsky, 1992; Toro, 1996; Romdhane, et al., 2007). Consequently, the characterization of genes relevant to saprophytic competence is essential towards the improvement of rhizobial strains that are able to successfully establish long-term viable soil populations. This study represents an important contribution towards the identification of such genes in *S. meliloti*, however additional studies are required. Particularly, the identity of gene(s) associated with the poor growth of *S. meliloti* strains

RmP801 and RmP798 in soil remains to be elucidated and we are currently pursuing this objective.

6.6 References

Aislabie, J.M., Chhour, K.L., Saul, D.J., Miyauchi, S., Ayton, J., Paetzold, R.F., Balks, M.R. (2006). Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. Soil Biology and Biochemistry, **38**:3041-3056.

Alef, K., & Nannipieri, P. (1995). *Methods in Applied Soil Microbiology and Biochemistry.* San Diego, California: Academic Press.

Allaway, D., Schofield, N.A., Leonard, M.E., Gilardoni, L., Finan, T.M., & Poole, P.S. (2001). Use of differential fluorescence induction and optical trapping to isolate environmentally induced genes. Environmental Microbiology, **3**:397-406.

Andrews, B.J., Proteau, G.A., Beatty, L.G., Sadowski, P.D. (1985). The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. Cell, **40**:795-803.

Balkwill, D.L., & Casida, L.E. (1979). Attachment to autoclaved soil of bacterial cells from pure cultures of soil isolates. Applied and Environmental Microbiology, **37**:1031-1037.

Barloy-Hubler, F., Cheron, A., Hellegouarch, A., & Galibert, F. (2004). Smc01944, a secreted peroxidase induced by oxidative stresses in *Sinorhizobium meliloti* 1021. Microbiology, **150**:657-664.

Barr, M., East, A.K., Leonard, M., Mauchline, T.H., & Poole, P.S. (2008). In vivo expression technology (IVET) selection of genes of *Rhizobium leguminosarum* biovar *viciae* A34 expressed in the rhizosphere. FEMS Microbiology Letters, **282**:219-227.

Cebolla, A., Ruiz-Berraquero, F., & Palomares, A.J. (1993). Stable tagging of *Rhizobium meliloti* with the firefly luciferase gene for environmental monitoring. Applied and Environmental Microbiology, **59**:2511-2519.

Cowie, A., Cheng, J., Sibley, C.D., Fong, Y., Zaheer, R., Patten, C.L., Morton, R.M., Golding, G.B., & Finan, T.M. (2006). An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. Applied and Environmental Microbiology, **72**:7156-7167.

DeFlaun, M.F., Tanzer, A.S., McAteer, A.L., Marshall, B., Levy, S.B. (1990). Development of an adhesion assay and characterization of an adhesion-deficient mutant of *Pseudomonas fluorescens*. Applied and Environmental Microbiology, **56**:112-119.

Denison, R.F., & Kiers, E.T. (2004). Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. FEMS Microbiology Letters, **237**:187-193.

Finan, T.M., Kunkel, B., De Vos, G.F., & Signer, E.R. (1986). Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. Journal of Bacteriology, **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., Puhler, A. (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*. Proceedings of the National Academy of Sciences U.S.A, **98**:9889-9894.

Gal, M., Preston, G.M., Massey, R.C., Spiers, A.J., & Rainey, P.B. (2003). Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. Molecular Ecology, **12**:3109-3121.

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.

Heynen, C.E., van Elsas, J.D., Kuikman, P.J., & van Veen, J.A. (1988). Dynamics of *Rhizobium leguminosarum* biovar *trifolii* introduced into soil; the effect of bentonite clay on predation by protozoa. Soil Biology and Biochemistry, **20**:483-488.

Hirsch, P.R. (1996). Population dynamics of indigenous and genetically modified rhizobia in the field. New Phytologist, **133**:159-171.

Hirsch, P.R., & Spokes, J.D. (1994). Survival and dispersion of genetically modified rhizobia in the field and genetic interactions with native strains. FEMS Microbiology Ecology, **15**:147-160.

Howieson, J.G. (1995). Rhizobial persistance and its role in the development of sustainable agricultural systems in Mediterranean environments. Soil Biology and Biochemistry, **27**:603-610.

Huang, J., Su, Z., & Xu, Y. (2005). The evolution of microbial phosphonate degradative pathways. Journal of Molecular Evolution, **61**:682-690.

Ito, H., Watanabe, H., Takehisa, M., & Iizuka, H. (1983). Isolation and identification of radiation-resistant cocci belonging to the genus *Deinococcus* from sewage sludges and animal feeds. Agricultural and Biological Chemistry, **47**:1239-1247.

Jenneman, G.E., McInerney, M.J., Crocker, M.E., & Knapp, R.M. (1986). Effect of sterilization by dry heat or autoclaving on bacterial penetration through Berea sandstone. Applied and Environmental Microbiology, **51**:39-43.

Jung, D.K., Lee, Y., Park, S.G., Park, B.C., Kim, G.H., & Rhee, S. (2006). Structural and functional analysis of PucM, a hydrolase in the ureide pathway and a member of the transthyretin-related protein family. Proceedings of the National Academy of Sciences U.S.A, **103**:9790-9795.

Karunakaran, R., Ebert, K., Harvey, S., Leonard, M.E., Ramachandran, V., & Poole, P. (2006). Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. Journal of Bacteriology, **188**:6661-6668.

Kinkle, B.K., Sadowsky, M.J., Schmidt, E.L., & Koskinen, W.C. (1993). Plasmids pJP4 and r68.45 can be transferred between populations of *Bradyrhizobia* in nonsterile soil. Applied and Environmental Microbiology, **59**:1762-1766.

Koukol, O., Novak, F., & Hrabal, R. (2008). Composition of the organic phosphorus fraction in basidiocarps of saprotrophic and mycorrhizal fungi. Soil Biology and Biochemistry, **40**:2464-2467.

Liu, C.M., McLean, P.A., Sookdeo, C.C., & Cannon, F.C. (1991). Degradation of the herbicide glyphosate by members of the family *Rhizobiaceae*. Applied and Environmental Microbiology, **57**:1799-1804.

Lotrario, J.B., Stuart, B.J., Lam, T., Arands, R.R., O'Connor, O.A., & Kosson, D.S. (1995). Effects of sterilization methods on the physical characteristics of soil: implications for sorption isotherm analyses. Bulletin of Environmental Contamination and Toxicology, **54**:668-675.

Lowendorf, H.S., Baya, A.M., & Alexander, M. (1981). Survival of *Rhizobium* in Acid Soils. Applied and Environmental Microbiology, **42**:951-957.

MacLean, A.M., Finan, T.M., & Sadowsky, M.J. (2007). Genomes of the symbiotic nitrogen-fixing bacteria of legumes. Plant Physiology, **144**:615-622.

Matilla, M.A., Espinosa-Urgel, M., Rodriguez-Herva, J.J., Ramos, J.L., & Ramos-Gonzalez, M.I. (2007). Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. Genome Biology, **8**:R179.

Mattimore, V., & Battista, J.R. (1996). Radioresistance of *Deinococcus radiodurans*: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. Journal of Bacteriology, **178**:633-637.

Mauchline, T.H., Fowler, J.E., East, A.K., Sartor, A.L., Zaheer, R., Hosie, A.H., Poole, P.S., & Finan, T.M. (2006). Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proceedings of the National Academy of Sciences of U.S.A., **103**:17933-17938.

McNamara, N.P., Black, H.I.J., Beresford, N.A., & Parekh, N.R. (2003). Effects of acute gamma irradiation on chemical, physical and biological properties of soils. Applied Soil Ecology, **24**:117-132.

Miranda-Rios, J., Morera, C., Taboada, H., Davalos, A., Encarnacion, S., Mora, J., Soberon, M. (1997). Expression of thiamin biosynthetic genes (*thiCOGE*) and production of symbiotic terminal oxidase *cbb3* in *Rhizobium etli*. Journal of Bacteriology, **179**:6887-6893.

Parker, G.F., Higgins, T.P., Hawkes, T., & Robson, R.L. (1999). *Rhizobium* (*Sinorhizobium*) *meliloti phn* genes: characterization and identification of their protein products. Journal of Bacteriology, **181**:389-395.

Phillips, D.A., Joseph, C.M., & Hirsch, P.R. (1997). Occurrence of flavonoids and nucleosides in agricultural soils. Applied and Environmental Microbiology, **63**:4573-4577.

Rainey, P.B. (1999). Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. Environmental Microbiology, **1**:243-257.

Ramazzina, I., Folli, C., Secchi, A., Berni, R., & Percudani, R. (2006). Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes. Nature Chemical Biology, **2**:144-148.

Ramos-Gonzalez, M.I., Campos, M.J., & Ramos, J.L. (2005). Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: in vivo expression technology capture and identification of root-activated promoters. Journal of Bacteriology, **187**:4033-4041.

Rediers, H., Bonnecarrere, V., Rainey, P.B., Hamonts, K., Vanderleyden, J., & De Mot, R. (2003). Development and application of a *dapB*-based in vivo expression technology system to study colonization of rice by the endophytic nitrogen-fixing bacterium *Pseudomonas stutzeri* A15. Applied and Environmental Microbiology, **69**:6864-6874.

Rediers, H., Rainey, P.B., Vanderleyden, J., & De Mot, R. (2005). Unraveling the secret lives of bacteria: use of in vivo expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. Microbiology and Molecular Biology Reviews, **69**:217-261.

Rivas, R., Velazquez, E., Zurdo-Pineiro, J.L., Mateos, P.F., & Martinez Molina, E. (2004). Identification of microorganisms by PCR amplification and sequencing of a universal amplified ribosomal region present in both prokaryotes and eukaryotes. Journal of Microbiological Methods, **56**:413-426.

Romdhane, S.B., Tajini, F., Trabelsi, M., Aouani, M.E., & Mhamdi, R. (2007). Competition for nodule formation between introduced strains of *Mesorhizobium ciceri* and the native populations of rhizobia nodulating chickpea (*Cicer arietinum*) in Tunisia. World Journal of Microbiology and Biotechnology, **23**:1195-1201.

Schultz, A.C., Nygaard, P., & Saxild, H.H. (2001). Functional analysis of 14 genes that constitute the purine catabolic pathway in *Bacillus subtilis* and evidence for a novel regulon controlled by the PucR transcription activator. Journal of Bacteriology, **183**:3293-3302.

Segovia, L., Pinero, D., Palacios, R., & Martinez-Romero, E. (1991). Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. Applied and Environmental Microbiology, **57**:426-433.

Selbitschka, W., Dresing, U., Hagen, M., Niemann, S., & Puhler, A. (1995). A biological containment system for *Rhizobium meliloti* based on the use of recombination-deficient (*recA*-) strains. FEMS Microbiology Ecology, **16**:223-232.

Silby, M.W., & Levy, S.B. (2004). Use of in vivo expression technology to identify genes important in growth and survival of *Pseudomonas fluorescens* Pf0-1 in soil: discovery of expressed sequences with novel genetic organization. Journal of Bacteriology, **186**:7411-7419.

Simons, M., van der Bij, A.J., Brand, I., de Weger, L.A., Wijffelman, C.A. & Lugtenberg, B.J. (1996). Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting *Pseudomonas* bacteria. Molecular Plant-Microbe Interactions, **9**:600-607.

Streit, W.R., Joseph, C.M., & Phillips, D.A. (1996). Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021. Molecular Plant-Microbe Interactions, **9**:330-338.

Taboada, H., Encarnacion, S., Vargas, M. C., Mora, Y., Miranda-Rios, J., Soberon, M., Mora, J. (2008). Thiamine limitation determines the transition from aerobic to fermentative-like metabolism in *Rhizobium etli* CE3. FEMS Microbiology Letters, **279**:48-55.

Tate, K.R., & Newman, R.H. (1982). Phosphorus fractions of a climosequence of soils in New Zealand tussock grassland. Soil Biology and Biochemistry, **14**:191-196.

Toro, A. (1996). Nodulation competitiveness in the Rhizobium-legume symbiosis. World Journal of Microbiology and Biotechnology, **12**:157-162.

Trevors, J.T. (1996). Sterilization and inhibition of microbial activity in soil. Journal of Microbiological Methods, **26**:53-59.

Triplett, E.W., & Sadowsky, M.J. (1992). Genetics of competition for nodulation of legumes. Annual Review of Microbiology, **46**:399-428.

Turner, B.L., Mahieu, N., & Condron, L.M. (2003). The phosphorus composition of temperate pasture soils determined by NaOH–EDTA extraction and solution ^{31}P NMRspectroscopy. Organic Geochemistry, **34**:1199-1210.

Vanbleu, E., Marchal, K., & Vanderleyden, J. (2004). Genetic and physical map of the pLAFR1 vector. DNA Sequence, **15**:225-227.

West, P.M. (1939). Excretion of thiamin and biotin by the roots of higher plants. Nature, **144**:1050-1051.

Wiesner, W., van Pee, K.H., & Lingens, F. (1986). Detection of a new chloroperoxidase in *Pseudomonas pyrrocinia*. FEBS Letters, **209**:321-324.

Wiesner, W., van Pee, K.H., & Lingens, F. (1988). Purification and characterization of a novel bacterial non-heme chloroperoxidase from *Pseudomonas pyrrocinia*. The Journal of Biological Chemistry, **263**:13725-13732.

Table 6.1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	Source or reference
Strains		
S. meliloti		
Rm1021	Sm^r derivative of wild-type strain SU47	Meade, et al., 1982
RmP110	Rm1021 with wild-type *pstC*; Sm^r	Yuan, et al., 2006
RmP790	RmP110 Δ(1129758 – 1170466 nts; pSymB) corresponds to ΔB108; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan, unpublished data
RmP791	RmP110 Δ(1091104 – 1131168 nts; pSymB) corresponds to ΔB107; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP798	RmP110 Δ(1528150 – 1573735 nts; pSymB) corresponds to ΔB122; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP799	RmP110 Δ(1169073 – 1207052 nts; pSymB) corresponds to ΔB109; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP801	RmP110 Δ(1255032 – 1308912 nts; pSymB) corresponds to ΔB116; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP803	RmP110 Δ(1652558 – 1679723 nts; pSymB) corresponds to ΔB124; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP806	RmP110 Δ(1528150 – 1654191 nts; pSymB) corresponds to ΔB123; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP808	RmP110 Δ(869642 – 1092289 nts; pSymB) corresponds to ΔB106; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP811	RmP110 Δ(1322226 – 1529711 nts; pSymB) corresponds to ΔB118; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP823	RmP110 Δ(1204770 – 1226491 nts; pSymB) corresponds to ΔB112; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP874	RmP110 Δ(741497 – 870505 nts; pSymB) corresponds to ΔB139; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP878	RmP110 Δ(121311 – 467160 nts; pSymB) corresponds to	B. Poduška and T. M.

	ΔB142; Sm^r Tc^r	Finan
RmP884	RmP110 Δ(635019 – 764540 nts; pSymB) corresponds to ΔB148; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP928	RmP110 Δ(9549 – 48842 nts; pSymA) corresponds to ΔA102; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP930	RmP110 Δ(47717 – 92124 nts; pSymA) corresponds to ΔA103; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP932	RmP110 Δ(90324 – 125128 nts; pSymA) corresponds to ΔA104; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP934	RmP110 Δ(123788 – 186200 nts; pSymA) corresponds to ΔA105; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP936	RmP110 Δ(184519 – 251809 nts; pSymA) corresponds to ΔA106; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP941	RmP110 Δ(458916 – 507338 nts; pSymA) corresponds to ΔA118; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP943	RmP110 Δ(505335 – 577241 nts; pSymA) corresponds to ΔA119; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP945	RmP110 Δ(575671 – 624863 nts; pSymA) corresponds to ΔA120; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP947	RmP110 Δ(623673 – 678150 nts; pSymA) corresponds to ΔA121; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP949	RmP110 Δ(677157 – 727921 nts; pSymA) corresponds to ΔA122; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP951	RmP110 Δ(726673 – 775476 nts; pSymA) corresponds to ΔA123; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP953	RmP110 Δ(774293 – 830143 nts; pSymA) corresponds to ΔA124; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP955	RmP110 Δ(828417 – 881169 nts; pSymA) corresponds to ΔA125; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP957	RmP110 Δ(924516 – 1064644 nts; pSymA) corresponds to	B. Poduška and T. M.

	ΔA128; Sm^r Tc^r	Finan
RmP959	RmP110 Δ(1063642 – 1123504 nts; pSymA) corresponds to ΔA129; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP961	RmP110 Δ(1122176 – 1173730 nts; pSymA) corresponds to ΔA130; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP963	RmP110 Δ(1283082 – 1349931 nts; pSymA) corresponds to ΔA133; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP965	RmP110 Δ(1231998 – 1284751 nts; pSymA) corresponds to ΔA132; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP967	RmP110 Δ(1173115 – 1232916 nts; pSymA) corresponds to ΔA131; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP979	RmP110 Δ(250917 – 313654 nts; pSymA) corresponds to ΔA109; Sm^r Nm^r Gm^r Tc^r	B. Poduška and T. M. Finan
RmP987	RmP110 Δ(294016 – 459668 nts; pSymA) corresponds to ΔA114; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP1055	RmP110 Δ(1679725 – 50791 nts; pSymB) corresponds to ΔB161; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP1431	RmP110 Δ(483404 – 652707 nts; pSymB) corresponds to ΔB163; Sm^r Tc^r	B. Poduška and T. M. Finan
RmP1815	RmP110 Δ(1528150 – 1654191 nts; pSymB) corresponds to ΔB123; Sm^r Nm^r Gm^r	B. Poduška and T. M. Finan
RmP1880	RmP110 Δ(1606297 – 1620237 nts; pSymB); Sm^r Nm^r Gm^r Tc^r	This study
RmP1881	RmP110 Δ(1572422 – 1577234 nts; pSymB); Sm^r Nm^r Gm^r Tc^r	This study
RmP1882	RmP110 Δ(1618529 – 1634067 nts; pSymB); Sm^r Nm^r Gm^r Tc^r	This study
RmP1883	RmP110 Δ(1572422 – 1607703 nts; pSymB); Sm^r Nm^r Gm^r Tc^r	This study
RmP1884	RmP110 Δ(1632500 – 1654191 nts; pSymB); Sm^r Nm^r Gm^r Tc^r	This study
pLAFR1	IncP cosmid cloning vector; Tc^r	Friedman, et al., 1982
pT8	pLAFR1 derivative carrying ~24 kb *S. meliloti* insert (1620000 – 1643981 nts; pSymB), including *thiCOGE*; Tc^r	(Finan, et al., 1986)

Table 6.2. Physiochemical properties of soil used in this study.

Sample type	Not autoclaved[a]	Autoclaved
Total C	4.94 % dry	4.20 % dry
Inorganic C	1.53 % dry	1.38 % dry
Organic C	3.41 % dry	2.82 % dry
pH	7.6	7.5
NH_4-N	46.2 mg/kg dry	51.8 mg/kg dry
NO_3-N	2.16 mg/kg dry	1.22 mg/kg dry
NO_2-N	0.042 mg/kg	0.094 mg/kg
N	0.35 %	0.30 %
P	146 mg/kg dry	127 mg/kg dry
Mg	494 mg/kg dry	NT[b]
K	383 mg/kg dry	NT
S	0.05 % dry	0.05 % dry
Ca	2713 mg/kg dry	NT
Fe	36.46 mg/kg dry	NT
Mn	172.5 mg/kg dry	NT
Zn	24.79 mg/kg dry	NT
Percent soil moisture	20.64 %	20.65 %
Organic matter	5.9 % dry	5.3 % dry
Very coarse sand	0.8 % w/w	0.5 % w/w
Coarse sand	4.2 % w/w	3.8 % w/w
Medium sand	7.0 % w/w	6.7 % w/w
Fine sand	16.6 % w/w	18.0 % w/w
Very fine sand	21.2 % w/w	22.1 % w/w
Sand	49.8 % w/w	51.1 % w/w
Silt	35.4 % w/w	34.3 % w/w
Clay	14.8 % w/w	14.6 % w/w
Texture	Loam	Loam
Organic matter	5.9 % dry	5.3 % dry
Gravel	0.0 % w/w	0.0 % w/w

[a] Analysis performed following γ–irradiation of soil samples by the University of Guelph Laboratory Services Agricultural and Food Laboratory.
[b] NT, not tested.

Table 6.3. Growth of *S. meliloti* strains in sterile soil microcosms.

	Soil population[a] (cells g^{-1} soil)							
Day	0	1	2	7	13	28	34	39
RmP110	2.9×10^2	3.2×10^3	3.9×10^5	4.1×10^8	4.8×10^8	1.0×10^8	8.4×10^7	1.2×10^8
RmP801 (ΔB116)	1.9×10^2	1.9×10^3	2.2×10^5	4.5×10^8	3.1×10^8	2.4×10^7	3.6×10^7	4.4×10^7
Fold decrease		2	2	1	2	4	2	3
Day	0	1	2	5	7	9	13	17
RmP110	2.9×10^2	3.2×10^3	3.9×10^5	2.4×10^8	4.1×10^8	4.2×10^8	4.8×10^8	1.5×10^8
RmP1815 (ΔB123)	2.4×10^2	5.5×10^2	8.6×10^4	5.8×10^7	8.2×10^7	3.4×10^7	1.8×10^7	5.6×10^6
Fold decrease		6	5	4	5	12	28	27

[a] Shown is average soil population of four samples obtained from two microcosms. Data is also depicted graphically in Figures 6.5 and 6.6. Fold decrease, RmP110 cells g^{-1} soil divided by RmP801 cells g^{-1} soil or RmP1815 cells g^{-1} soil, as indicated.

Table 6.4. Description of genes deleted in *S. meliloti* strain RmP801 (ΔB116).

Gene	Gene Annotation	Gene	Gene Annotation
smb20851	Transcriptional regulator, SorC family	*smb20876*	Hypothetical protein
smb20852	Sugar kinase	*smb20877*	Oxidoreductase, possibly D-amino acid oxidase
smb20853	Sugar-alcohol dehydrogenase	*smb20878*	Conserved hypothetical protein
smb20854	Sugar uptake ABC transporter permease	*smb20879*	Putative protein required for attachment to host cells
smb20855	ABC transporter, ATP-binding protein	*rhlE2*	ATP-dependent RNA helicase
smb20856	ABC transporter, periplasmic solute-binding protein, may transport deoxyribose	*smb20881*	Conserved hypothetical exported protein
smb20857	Glucose-6-phosphate isomerase	*smb20882*	Putative heavy metal efflux pump
smb20858	L-lactate dehydrogenase	*smb20883*	Hypothetical protein
smb20859	Transcriptional regulator, AraC family protein	*smb20884*	Hypothetical protein
smb21677	Hypothetical exported protein signal peptide	*smb20885*	Conserved hypothetical membrane protein
smb20860	Probable non-heme chloroperoxidase	*smb20886*	Conserved hypothetical protein
smb20861	Dehydrogenase, FAD flavoprotein	*smb20887*	Conserved hypothetical protein
smb20862	Hypothetical protein	*smb20888*	Conserved hypothetical protein
smb20863	Mechanosensitive ion channel	*smb20889*	Hypothetical glycine-rich protein
smb20864	Conserved hypothetical exported protein	*araF*	Dihydroxy-acid dehydratase
smb20865	Conserved hypothetical protein	*araE*	2-oxoglutarate semialdehyde dehydrogenase
smb20866	Conserved hypothetical protein	*araD*	Conserved hypothetical protein
smb20867	Hypothetical protein	*araC*	ABC transporter, permease
smb20868	Two-component sensor histidine kinase	*araB*	ABC transporter, ATP-binding protein
smb20869	Two-component response regulator	*araA*	ABC transporter, periplasmic solute-binding protein
smb20870	Transcriptional regulator, LysR family protein	*gpbR*	transcriptional regulator, LysR family
smb20871	3-oxoacyl-[acyl-carrier-protein] reductase	*smb20897*	Hypothetical protein
smb20872	Conserved hypothetical protein, transthyretin-like protein	*smb20898*	Hypothetical protein
allA	Ureidoglycolate hydrolase	*smb21678*	Hypothetical protein
smb20874	Conserved hypothetical protein	*idhA*	Myo-inositol dehydrogenase protein
smb20875	Conserved hypothetical protein	*smb20900*	Diguanylate cyclase/phosphodiesterase

As annotated, http://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi (November, 2008)

Table 6.5. Growth of *S. meliloti* strains inoculated into a sterile soil microcosm.

	Soil population[a] (cells g^{-1} soil)												
Days	0	1	3	5	7	9	12	15	17	21	25	28	32
RmP110	3.9×10^2	5.9×10^3	6.3×10^6	1.8×10^8	2.1×10^8	2.5×10^8	2.0×10^8	1.6×10^8	1.6×10^8	7.4×10^7	1.2×10^8	1.4×10^8	2.1×10^8
RmP1815 (ΔB123)	2.4×10^2	1.8×10^3	3.0×10^6	5.5×10^7	6.6×10^7	5.6×10^7	4.6×10^7	2.6×10^7	1.7×10^7	1.0×10^7	3.8×10^6	5.4×10^6	4.4×10^6
RmP798 (ΔB122)	2.0×10^2	1.0×10^3	9.5×10^5	5.2×10^7	7.0×10^7	7.8×10^7	7.0×10^7	2.9×10^7	2.2×10^7	9.4×10^6	4.2×10^6	5.2×10^6	5.0×10^6
RmP1884	3.5×10^1	1.4×10^2	2.7×10^4	1.3×10^7	8.2×10^7	1.6×10^8	1.5×10^8	1.4×10^8	1.6×10^8	8.1×10^7	1.2×10^8	8.1×10^7	1.5×10^8
Fold decrease													
RmP1815		3	2	3	3	4	4	6	9	7	31	25	47
RmP798		6	7	3	3	3	3	5	7	8	28	27	41

[a] Shown is average soil population of four samples obtained from two microcosms. Data is also depicted graphically in Figure 6.7B. Fold decrease, RmP110 cells g^{-1} soil divided by RmP1815 cells g^{-1} soil or RmP798 cells g^{-1} soil, as indicated.

Table 6.6. Description of genes deleted in *S. meliloti* strain RmP798 (ΔB122).

Gene	Gene Annotation	Gene	Gene Annotation
smb20723	Iron ABC transporter periplasmic solute-binding protein	*smb20758*	Transcriptional regulator, GntR or ArsR family protein
smb20724	Conserved hypothetical exported protein	*phnG*	Probable protein PhnG
smb20725	Hypothetical membrane protein	*phnH*	Probable protein PhnH
smb20726	Conserved hypothetical membrane protein	*phnI*	Uncharacterized enzyme of phosphonate metabolism
smb20727	Hypothetical protein	*phnJ*	Uncharacterized enzyme of phosphonate metabolism
smb20728	Hypothetical protein	*phnK*	Phosphonate ABC transporter ATP-binding protein
smb20729	Conserved hypothetical protein	*phnL*	Phosphonate ABC transporter ATP-binding protein
smb22016	Hypothetical protein	*smb20765*	Chloramphenicol O-acetyltransferase
glnII	Glutamate-ammonia ligase	TRm22	Transposase of insertion sequence ISRm22 protein
gstI	Glutamine synthetase translation inhibitor	*Dak*	Dihydroxyacetone (glycerone) kinase
smb20747	Hypothetical enzyme, haloacid dehalogenaseepoxide hydrolase family	*smb20768*	Hypothetical protein
pssF	Glycosyltransferase	*smb20769*	Conserved hypothetical membrane protein
uxuB	Mannitol dehydrogenase	*smb20770*	Conserved hypothetical membrane protein
smb20750	Gluconate 5-dehydrogenase	*smb20771*	Conserved hypothetical protein
smb20751	3-hydroxyisobutyrate dehydrogenase	*pdxA2*	Pyridoxal phosphate biosynthetic protein PdxA
smb20752	Enoyl-CoA hydratase	*smb20773*	Transcriptional regulator, GntR family protein
smb20753	Acyl-CoA dehydrogenase	*smb20774*	Hypothetical protein
smb20754	Transcriptional regulator	*smb20775*	Hypothetical exported protein, TonB-dependent receptor protein
pccB	Propionyl-CoA carboxylase beta chain	*cyaK*	Adenylate cyclase
pccA	Propionyl-CoA carboxylase alpha chain	Trm19	Transposase of insertion sequence ISRm19 protein
bhbA	Methylmalonyl-CoA mutase		

As annotated, http://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi (November, 2008)

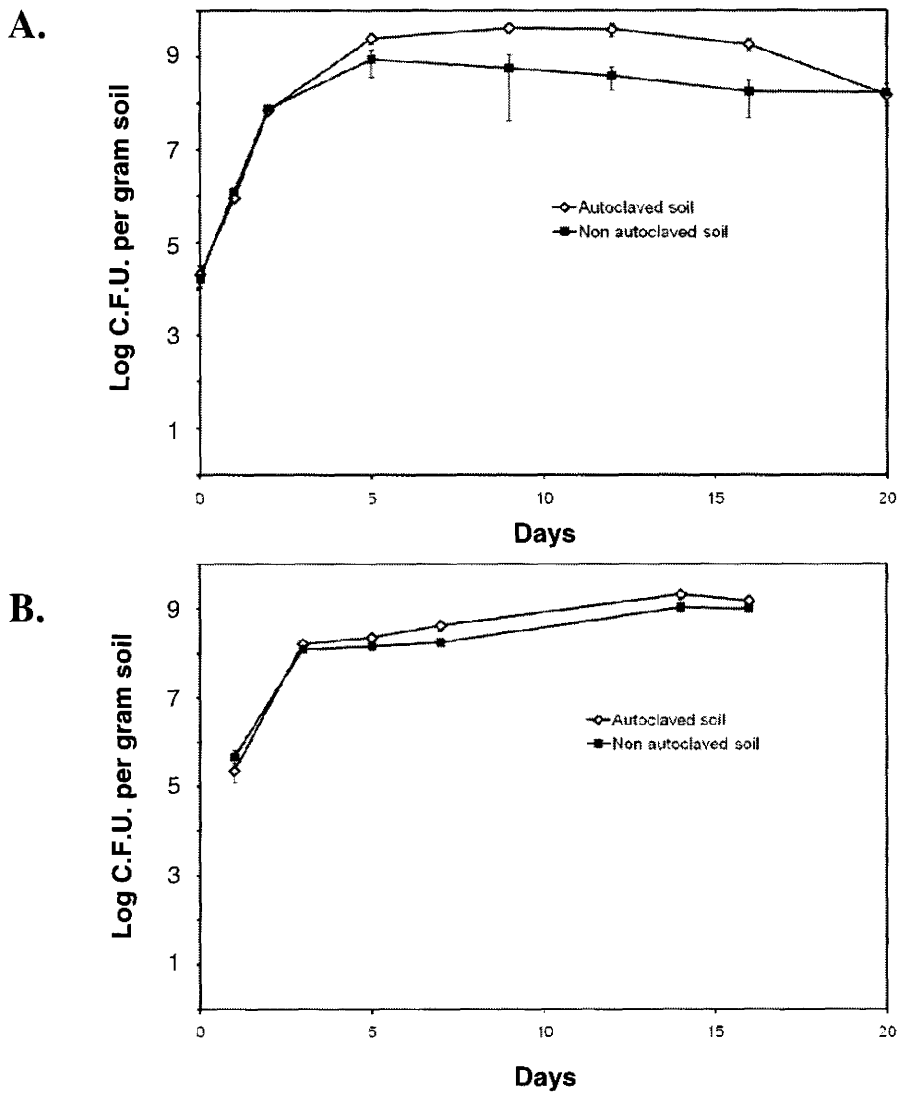


Figure 6.1. Growth of *S. meliloti* wild-type strain RmP110 in a soil microcosm. Growth was assayed using 40 grams (dry weight) γ–irradiated soil samples ± autoclaving once (123°C; 17 psig; 20 minutes). Shown is the average cell density and standard deviation estimated using two subsamples obtained from each of three replicate soil microcosms. (A) and (B) represent data obtained from two independent assays. Colony forming units were estimated by plating soil suspensions upon LB agar (autoclaved soil samples) or LB agar plus streptomycin (non autoclaved soil samples) to counterselect *Deinococcus*. Soil moisture was adjusted to 20%.

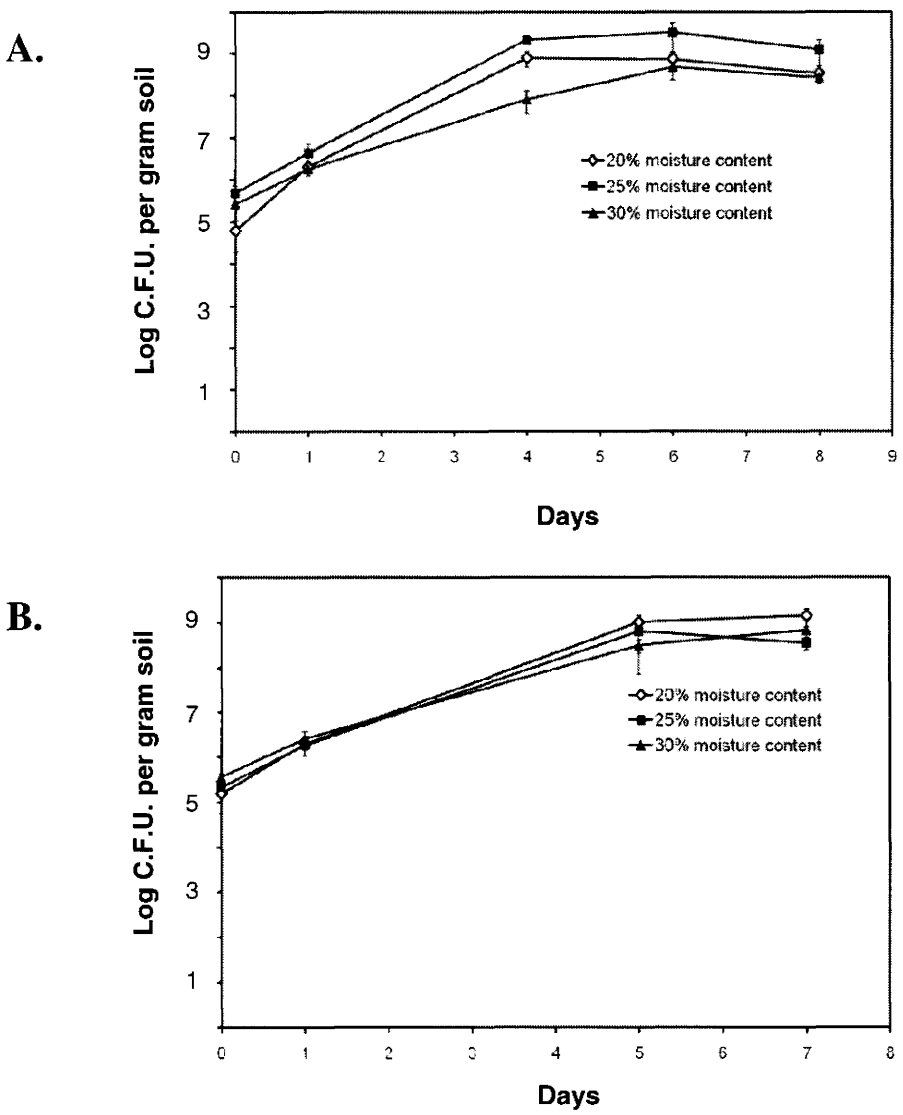


Figure 6.2. Influence of soil moisture content upon growth of *S. meliloti* wild-type strain RmP110 in a soil microcosm. Growth was assayed using (A) 30 grams and (B) 40 grams (dry weight) γ–irradiated soil. Shown is the average cell density and standard deviation estimated using two subsamples obtained from each of two replicate soil microcosms. Colony forming units were estimated by plating soil suspensions upon LB agar plus streptomycin to counterselect *Deinococcus* (this assay was performed prior to the adoption of autoclaving as a method of sterilization). Soil moisture was adjusted as indicated through the addition of autoclaved and deionized water.

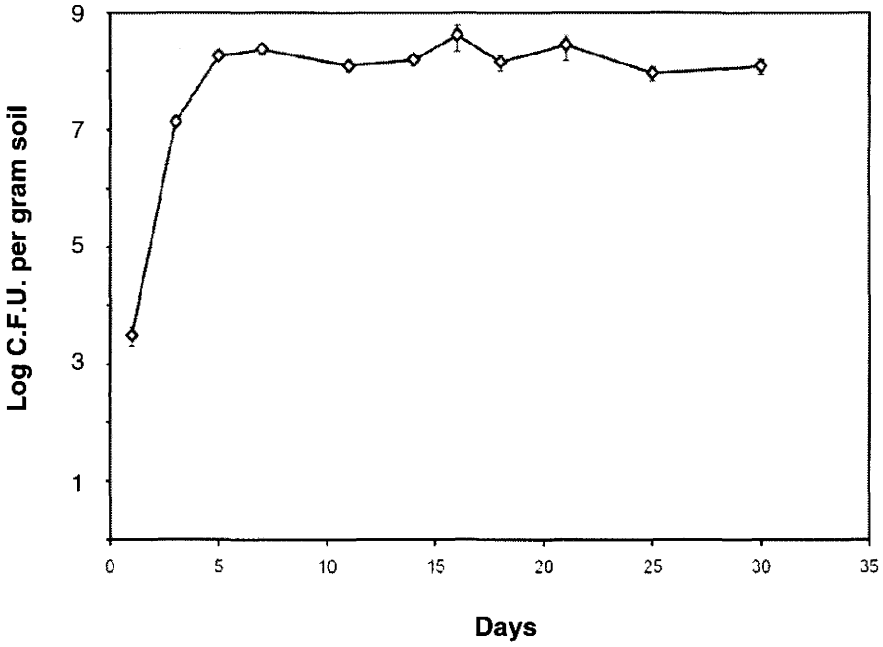


Figure 6.3. Growth of *S. meliloti* wild-type strain RmP110 in a sterile soil microcosm. Growth was assayed using 40 grams (dry weight) γ–irradiated and autoclaved soil. *S. meliloti* was inoculated to approximately 5 x 10^3 cells g^{-1} soil. Shown is the average cell density and standard deviation estimated using two subsamples obtained from each of six replicate soil microcosms. Colony forming units were estimated by plating soil suspensions upon LB agar. Soil moisture was adjusted to 20% and microcosms were incubated in the dark at 22°C for the duration of the assay.

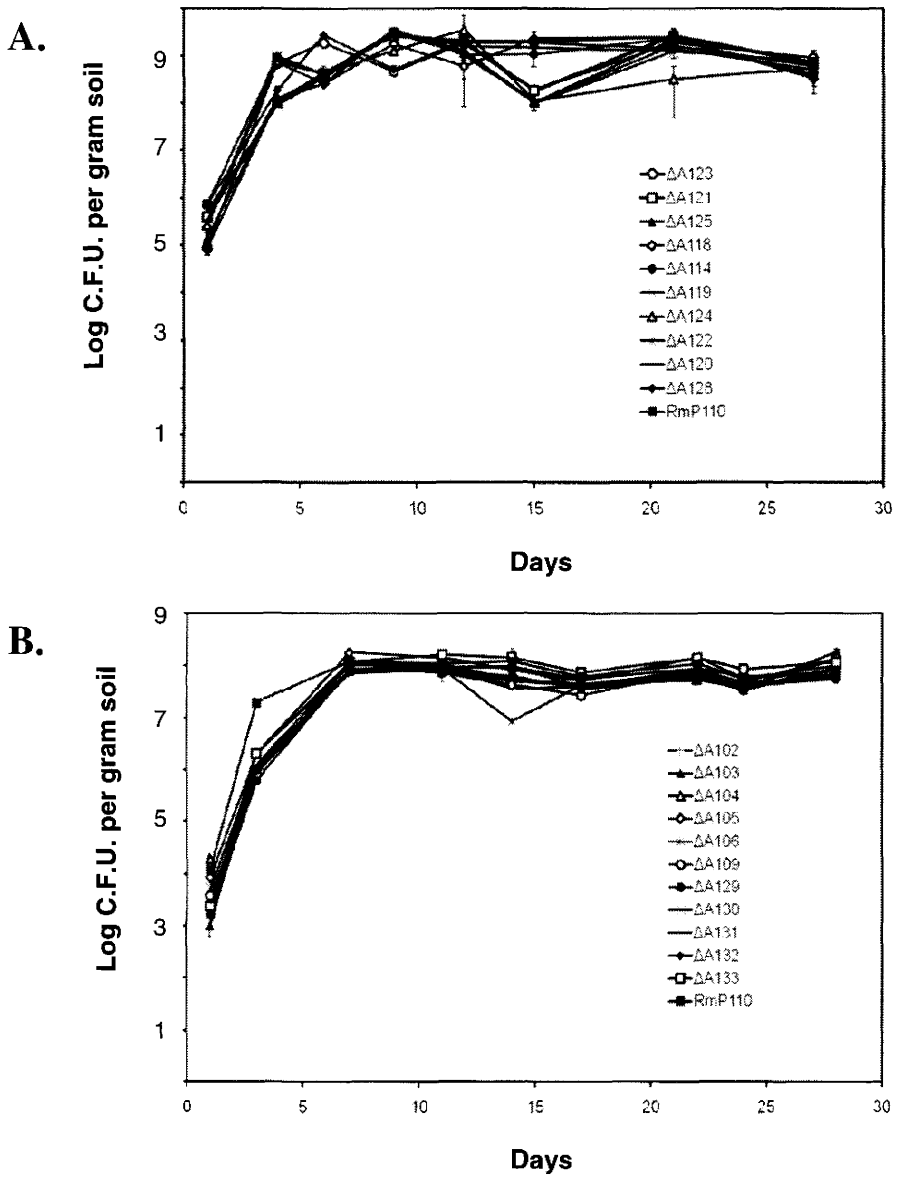


Figure 6.4. Growth of *S. meliloti* pSymA deletion strains in a sterile soil microcosm. Growth was assayed using 40 grams (dry weight) γ–irradiated and autoclaved soil. *S. meliloti* strains were inoculated to approximately 5 x 10^3 cells g^{-1} soil. Shown is the average cell density and standard deviation estimated using two subsamples obtained from one soil microcosm. Colony forming units were estimated by plating soil suspensions upon LB agar. Soil moisture was adjusted to 20% and microcosms were incubated in the dark at 22°C for the duration of the assay.

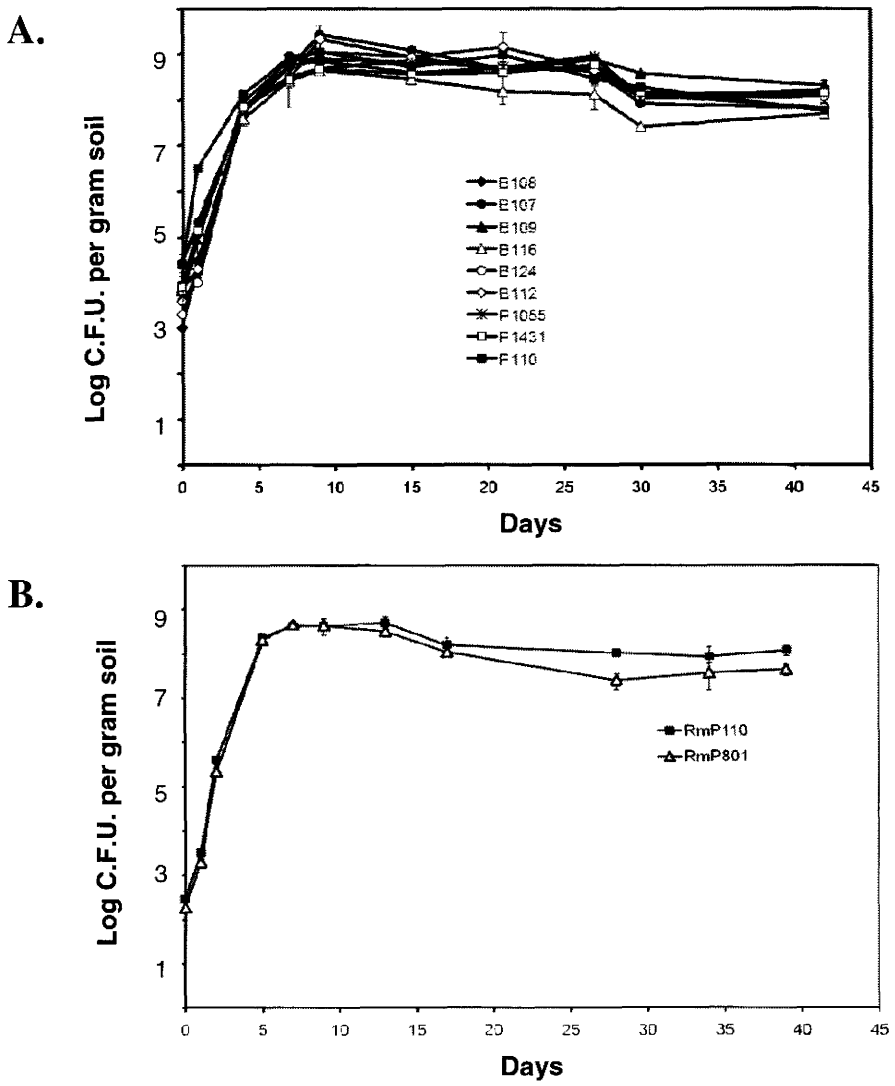


Figure 6.5. Growth of *S. meliloti* pSymB deletion strains in a sterile soil microcosm. Growth was assayed using 40 grams (dry weight) γ–irradiated and autoclaved soil. *S. meliloti* strains were inoculated to approximately 5 x 10^3 cells g^{-1} soil. (A) Screening of pSymB deletion strains for growth. Shown is the average cell density and standard deviation estimated using two subsamples obtained from one soil microcosm. (B) Growth of *S. meliloti* strain RmP801 (RmP110 ΔB116) versus RmP110. Shown is the average cell density and standard deviation estimated using two subsamples from two soil microcosms. Colony forming units were estimated by plating soil suspensions upon LB agar. Soil moisture was adjusted to 20% and microcosms were incubated at 22°C.

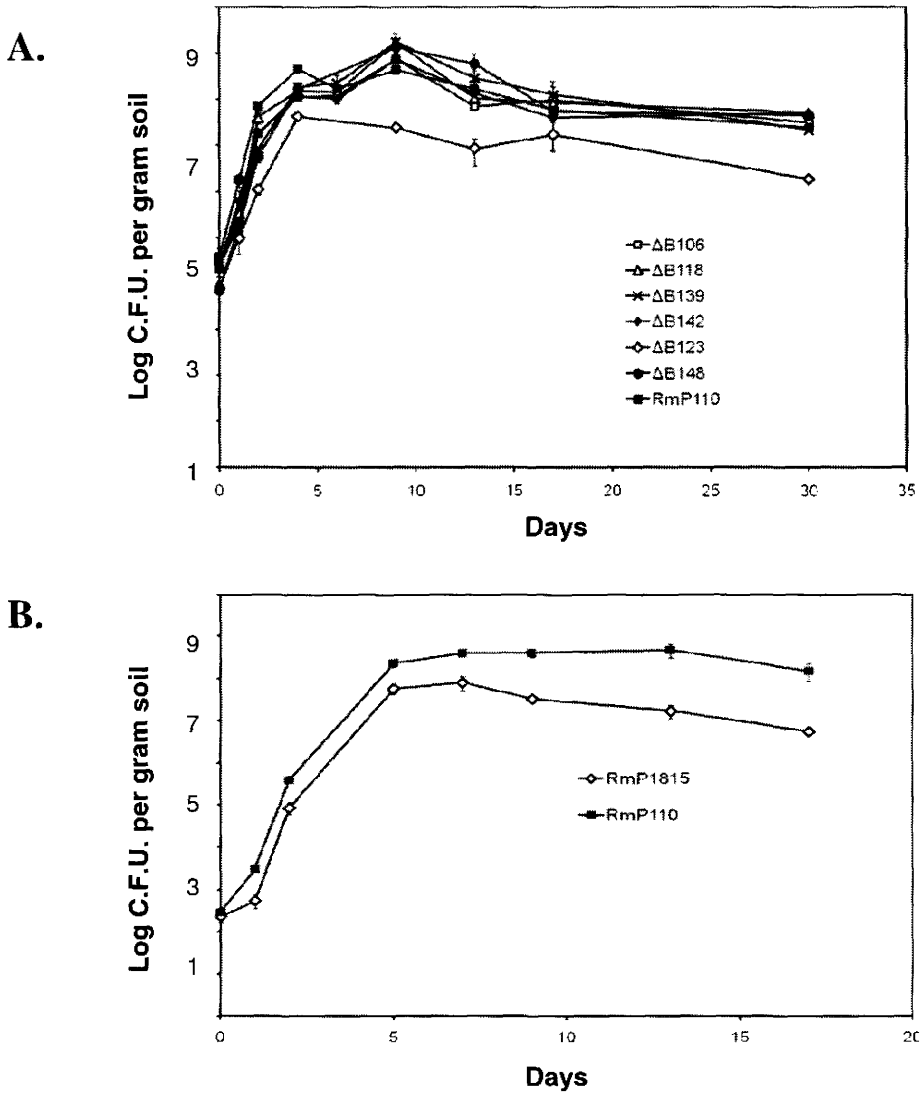


Figure 6.6. Growth of *S. meliloti* pSymB deletion strains in a sterile soil microcosm. Growth was assayed using 40 grams (dry weight) γ–irradiated and autoclaved soil. *S. meliloti* strains were inoculated to approximately 5 x 10^3 cells g^{-1} soil. (A) Screening of pSymB deletion strains for growth. Shown is the average cell density and standard deviation estimated using two subsamples obtained from one soil microcosm. (B) Growth of *S. meliloti* strain RmP1815 (RmP110 ΔB123) versus RmP110. Shown is the average cell density and standard deviation estimated using two subsamples from two soil microcosms. Colony forming units were estimated by plating soil suspensions upon LB agar. Soil moisture was adjusted to 20% and microcosms were incubated at 22°C.

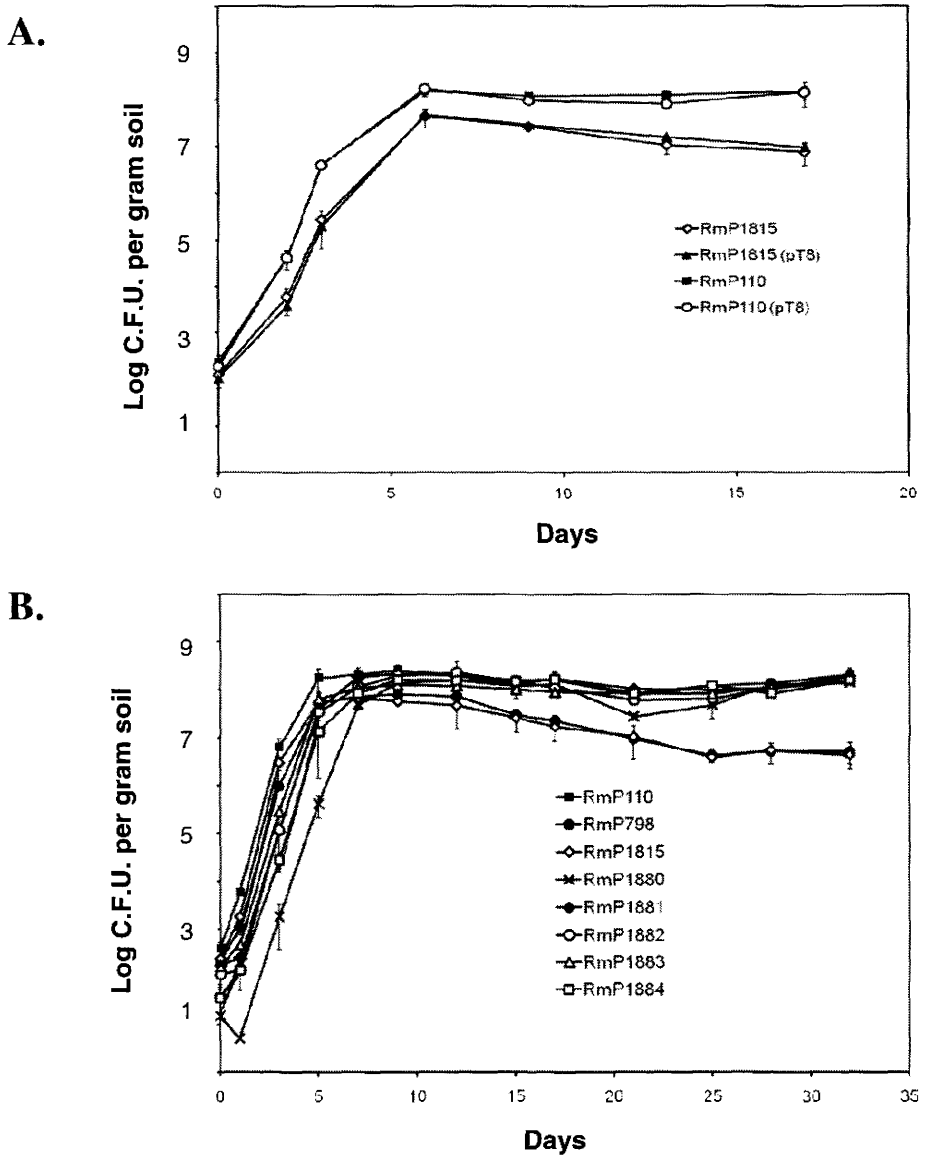


Figure 6.7. Growth of *S. melilot* strains in sterile soil. Growth was assayed using 40 grams (dry weight) γ–irradiated and autoclaved soil. *S. meliloti* strains were inoculated to approximately 5 x 10^3 cells g^{-1} soil. (A) Growth assay of *S. meliloti* thiamine auxotroph RmP1815 (RmP110 ΔB123) in soil. Cosmid pT8 carries the *thiCOGE* gene cluster. (B) Growth of *S. meliloti* strains carrying subdeletions within the boundaries of ΔB123, as depicted in Figure 6.9. Shown is the average cell density and standard deviation estimated using two subsamples from two soil microcosms. Colony forming units were estimated by plating soil suspensions upon LB agar. Soil moisture was adjusted to 20% and microcosms were incubated in the dark at 22°C.

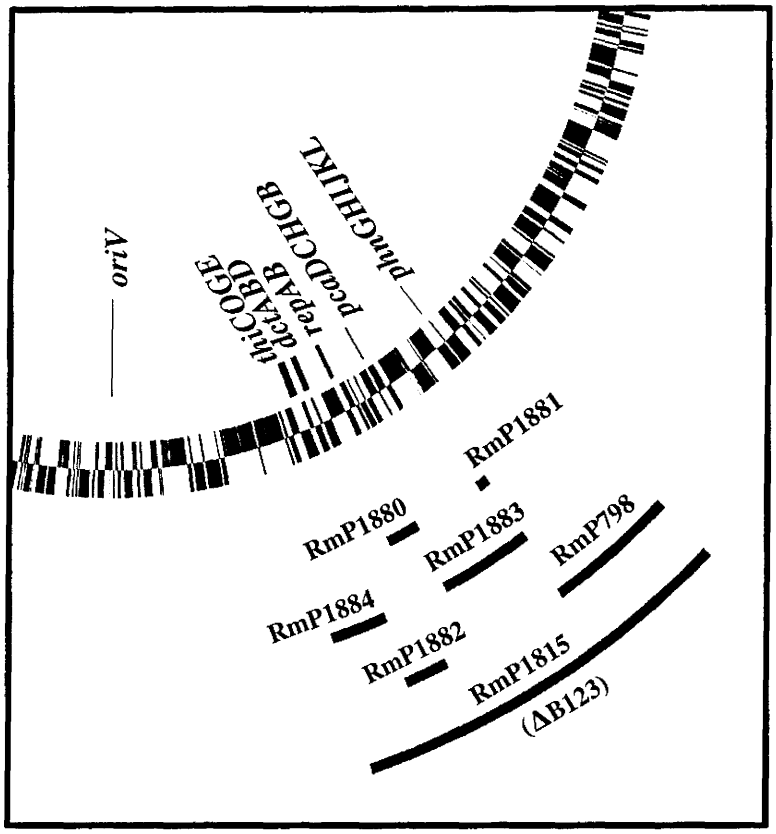


Figure 6.8. Schematic depiction of deletions generated in *S. meliloti* strains. ΔB123 corresponds to a 126 kb deletion within the pSymB megaplasmid that confers a phenotype of reduced saprophytic competence in sterile, bulk soil. Six *S. meliloti* strains were constructed carrying deletions within the boundaries of ΔB123, as indicated. Representative genes located within the deleted regions are indicated. Circleplot created by and used with permission of Dr. R.A. Morton.

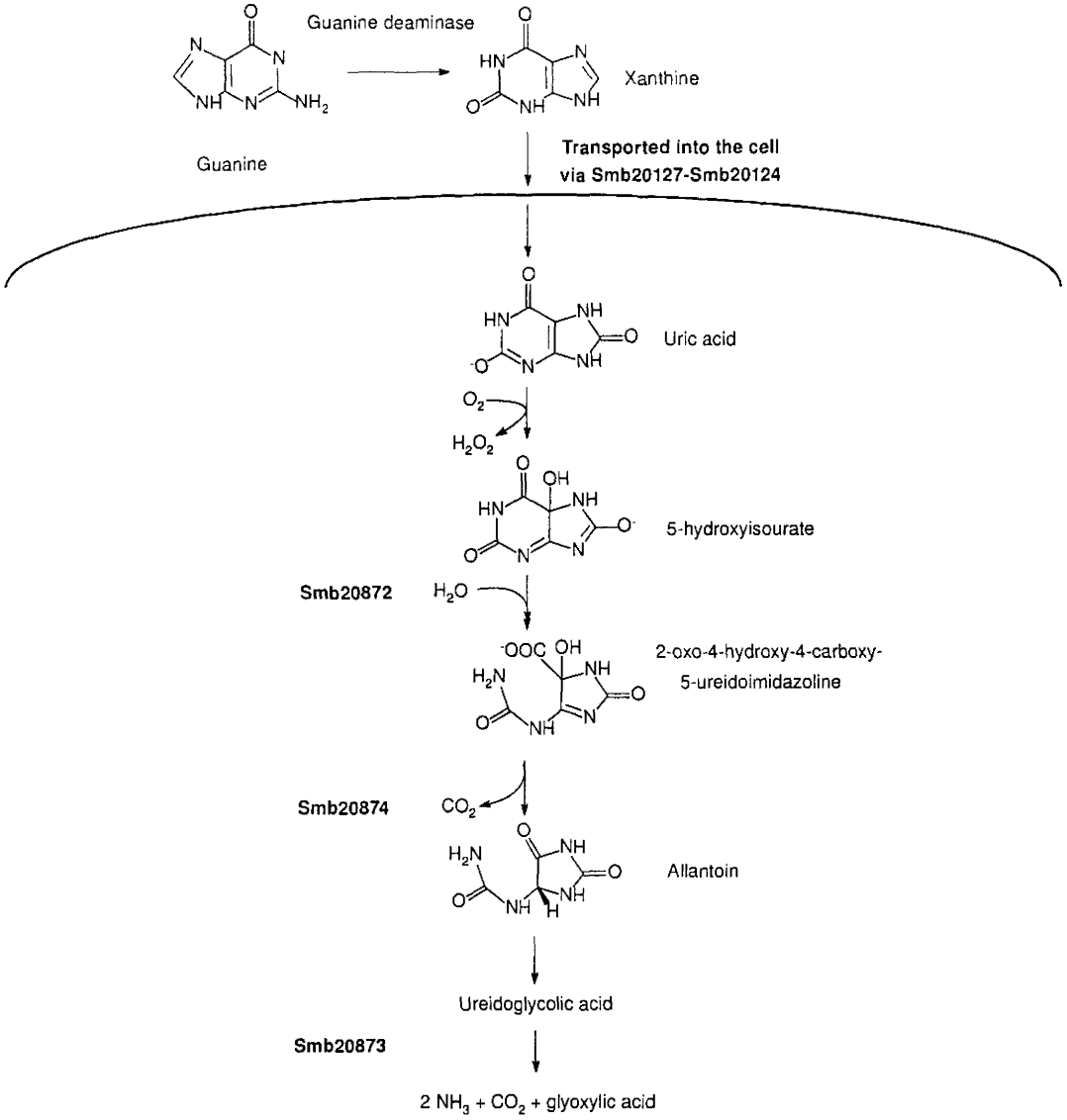


Figure 6.9. A putative purine catabolic pathway in *S. meliloti*. Expression of *smb20127-smb20124* is induced by xanthine (Mauchline, et al., 2006) and may encode a xanthine uptake system. *smb20872* encodes a protein with similarity to a 5-hydroxyisourate hydrolase (PucM), an enzyme which catalyzes the conversion of 5-hydroxyisourate to 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) in *Bacillus subtilis* (Schultz, et al., 2001; Jung, et al., 2006). *smb20874* encodes a protein containing an OHCU decarboxylase domain, an enzyme which converts OHCU to allantoin. *smb20873* is annotated as a ureidoglycolate hydrolase and may catalyze the catabolism of ureidoglycolic acid to ammonia, carbon dioxide, and glyoxylic acid.

APPENDIX

Figure A1. Harvesting and preparation of soil. (A) Dairy farm near Guelph, Ontario, Canada. The farm was owned and operated by Mr. Robert Jefferson, RR2 Lcd Royal City Mail, Guelph, ON, N1H 6H8.

(B) Soil was obtained from an alfalfa field in April, 2007.

APPENDIX

(C) Soil was sampled from the plough layer (0 to 25 cm) using shovels; samples were stored in clean buckets for immediate transport to McMaster University.

(D) Approximately 40 kg soil was spread upon a clean tarp in a greenhouse (McMaster University). Sticks, stones, and sod were removed manually from soil. Soil was covered with the tarp and allowed to air dry for 9 days.

APPENDIX

(E) Dried soil was sieved to remove particles greater than 2 mm

. (F) Soil prior to sieving (right) and after sieving (left).

APPENDIX

16S rDNA nucleotide sequences of 3 strains isolated from γ–irradiated soil (dosage, 25 kGy). Sequences were determined from PCR amplified products obtained using universal primers (5'CTYAAAKRAATTGRCGGRRRSSC '3) and (5'CGGGCGGTGTGTRCAARRSSC 3'). Best hits and E values are indicated as obtained following BLASTN analysis performed in October, 2008.

Best hit: *Deinococcus* sp. CC-FR2-10 (95% sequence identity) E value: 0.0

GACNGNGGTTNNATTCGAAGCNNCGCNAAAAACCTTACCNGTCTTGACATCCATGGAAC
CCCTGAGAGANNNGGGGGTGCCCTTCGGGGAACCATGANACAGGTGCTGCATGGCTGTC
GTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTACCTTT
AGTTGTCAGCATTCGGTTGGACACTCTAGAGGGACTGCCTATGAAAGTAGGAGGAAGGC
GGGGATGACGTCTAGTCAGCATGGTCCTTACNACCTGGGCTACACACGTGCTACAATGG
ATGGGACAACGCGCTGCCAGCCTGCGAAGGTGCGCGAATCGCTGAAACCCATCCCCAGT
TCAGATCGGAGTCTGCAACTCGACTCCGTGAAGTTGGAATCGCTAGTAATCGCAGGTCA
ACATACTGCGGTGAATACGTTCCCGGCCCTTGCACACACCGNCCGN

Best hits: *Rhizobium* sp. XJTC01 (98 % sequence identity) E value: 0.0

GACNGTGGTTNAATTNNNNCNACGCGCANAACCTTACCNCCCTTGACATCCCGATCGCG
GATACGAGAGATCGTATCCTTCAGTTCGGCTGGATCGGAGACAGGTGCTGCATGGCTGT
CGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCGCCCT
TAGTTGCCAGCATTCAGTTGGGCACTCTAAGGGGACTGCCGGTGATAAGCCGAGAGGAA
GGTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCTACAA
TGGTGGTGACAGTGGGCAGCGAGACCGCGAGGTCGAGCTAATCTCCAAAAGCCATCTCA
GTTCGGATTGCACTCTGCAACTCGAGTGCATGAAGTTGGAATCGCTAGTAATCGCAGAT
CANCATGCTGCGGTGAATACGTTCCCGGCCCTTGCACACACCGCCCG

Best hit: *Chelatococcus asaccharovorans* strain SAFR-017 (94% sequence identity)
E value: 0.0

CGNNGACNGNGGNNNNNTCGAANCNACGCGCAAAACCTTACCNCCTTTGACATGTCCGG
TTTGGATCCTGGAGACAGGTTCCTTCAGTTCGGCTGGCCGGAACACAGGTGCTGCATGG
CTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTCG
CCCCTAGTTGCCATCATTTGGTTGGGCACTCTAGGGGGACTGCCGGTGATAAGCCGAGA
GGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTACGGGCTGGGCTACACACGTGCT
ACAATGGCGGTGACAATGGGCAGCTAACCCGCGAGGGCATGCTAATCCCAAAAAGCCGT
CTCAGTTNNGANTGCACTCTGCAACTCGAGTGCATGAAGGTGGAATCNCTAGTNATCGT
GGATCANAANGCCNCCGTGAATACGTTCCNGGCCCTTGNACACACCGNCNG

APPENDIX

Protocatechuate offers a unique energy source, as this aromatic may serve as a source of carbon, but is also toxic to *S. meliloti* (and many other prokaryotes) at a concentration greater than five millimolar. Gord MacPherson isolated a catalase mutant (*S. meliloti* strain RmG878 contains a transposon insertion within *katA*) via Tn5 mutagenesis that was unable to grow in the presence of protocatechuate. This phenotype was not alleviated by the addition of an alternate carbon or energy source, and thus the poor growth of the *S. meliloti* mutant must be attributable to a toxic effect specific to protocatechuate. This work was extended by Ann Kim (under my supervision) as a senior undergraduate thesis project, who examined the phenotype in liquid growth medium.

The observation that a catalase mutant is unable to tolerate protocatechuate is intriguing, as polyphenols and aromatic acids such as protocatechuate undergo auto-oxidation, thereby leading to the generation of the reactive oxygen species hydrogen peroxide (H_2O_2). *S. meliloti* and related rhizobia colonize the flavonoid-rich environment of the rhizosphere, and catalases may contribute to the survival of these species in this habitat. Particularly, a catalase encoded by *R. etli* CFN42 (KatG) has been shown to be essential for the survival of this bacterium upon exposure to a crude extract obtained from the seed coat of bean (García-de los Santos, et al., 2008). The auto-oxidation associated with the polyphenol-rich extract produced measurable levels of H_2O_2, which increased more than 2-fold in the presence of *R. etli* (up to 1,300 μM H_2O_2 produced after a 16 hour incubation; García-de los Santos, et al., 2008). The failure of the *R. etli* catalase mutant to grow in the presence of the polyphenols emphasizes that cells may be subject to oxidative stress upon exposure to an aromatic such as protocatechuate. Thus, the phenotype exhibited by RmG878 when grown with protocatechuate may be due to the production of H_2O_2 that accompanies the auto-oxidation of this organic acid: in effect, the catalase mutant is unable to neutralize the accumulating H_2O_2, and does not survive.

During the course of my studies, I screened a small subset of the *S. meliloti* gene fusion library for growth upon M9-minimal medium supplemented with protocatechuate, protocatechuate and glycerol, or glycerol as a sole source(s) of carbon. Interestingly, two strains (*S. meliloti* RmFL1245 and RmFL1307), were unable to grow when plated upon M9-protocatechuate ± glycerol, although these strains grew upon M9-glycerol. Additionally, *S. meliloti* strain RmFL1178 was identified as exhibiting a comparable phenotype (lack of growth in the presence of protocatechuate) in an independent study intiated to examine the expression profiles of *S. meliloti* gene fusion strains subcultured into a variety of growth media. These initial observations were confirmed via growth curves performed using each strain subcultured into liquid growth medium supplemented with the appropriate carbon source.

S. meliloti strain RmFL1178 carries a reporter gene fusion within *phaA1C1D1E1F1G1*, thereby generating a null allele of the *pha* operon. The *pha* genes encode a potassium efflux system that is required for the formation of nitrogen-fixing root nodules, and is involved in an adaptive response to pH (Putnoky, et al., 1998). These genes encode transmembrane proteins, and it has been proposed that the system comprises a K+/H+ antiport that enables the maintenance of pH homeostasis via efflux of K+ (with a corresponding influx of H+) upon alkalinization of the cell (Putnoky, et al., 1998). The pK_a of protocatechuate is 4.48, and thus the majority of protocatechuate present in M9-minimal growth medium (pH 6.8) is in the form of the deprotonated base. Possibly, transport of protocatechuate into the cell results in the alkalinization of the internal milieu, which may be counteracted in a wild-type strain through the exchange of K+ and H+ via the Pha system. Accordingly, RmFL1178 (a *pha* mutant) may be unable to respond appropriately to this alkalinization, and thus fails to survive upon exposure to protocatechuate.

S. meliloti strain RmFL1245 contains a reporter gene fusion integrated within a gene predicted to encode an FAD-dependent dehydrogenase (*sma1414*). Expression of *smb1414* does not appear to be induced by growth with protocatechuate (based upon microarray analysis), and it is unclear why this strain exhibits a Pca^- phenotype.

RmFL1307 carries a gene fusion within the *rhbABCDEF* cluster, located upon pSymA. The integration of the library report vector disrupts expression of *rhbB*, a gene that encodes L-2,4-diaminobutyrate decarboxylase, an enzyme which is required for *rh*izobactin 1021 *b*iosynthesis (Lynch, et al., 2001). Rhizobactin is a citrate-based hydroxamate siderophore synthesized by *S. meliloti* strain Rm1021 under conditions of iron starvation, as regulated by Fur and RhrA (Persmark, et al., 1993; Lynch, et al., 2001). During periods of iron stress, the siderophore is secreted into the environment, and is transported into the cell via a specific outer membrane receptor protein (RhtA) upon chelation of iron. It is worth noting that protocatechuate (and related aromatics such as catechol) chelate iron, and this compound has even been incorporated into the iron siderophore Petrobactin synthesized by *Bacillus anthracis* (Garner, et al., 2004; Pfleger, et al., 2008). Accordingly, the availability of iron is likely limited in minimal growth medium supplemented with the iron-chelating protocatechuate, and the failure of RmFL1307 to grow under such conditions implies that Rhizobactin 1021 plays an important role in the acquisition of iron in this environment.

References

García-de los Santos, A., Lopez, E., Cubillas, C.A., Noel, K.D., Brom, S., & Romero, D. (2008). Requirement of a plasmid-encoded catalase for survival of *Rhizobium etli* CFN42 in a polyphenol-rich environment. Applied and Environmental Microbiology, **74**:2398-2403.

Garner, B.L., Arceneaux, J.E.L., & Byers, B.R. (2004). Temperature control of a 3,4-dihydroxybenzoate (Protocatechuate)-based siderophore in *Bacillus anthracis*. Current Microbiology, **49**:89-94.

Lynch, D., O'Brien, J., Welch, T., Clarke, P., Cuiv, P.O., Crosa, J.H., & O'Connell, M. (2001). Genetic organization of the region encoding regulation, biosynthesis, and transport of Rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. Journal of Bacteriology, **183**:2576-2585.

Persmark, M., Pittman, P., Buyer, J.S., Schwyn, B., Gill, P.R., & Neilands, J.B. (1993). Isolation and structure of Rhizobactin 1021, a siderophore from the alfalfa symbiont *Rhizobium meliloti* 1021. Journal of the American Chemical Society, **115**:3950-3956.

Pfleger, B.F., Kim, Y., Nusca, T.D., Maltseva, N., Lee, J.Y., Rath, C.M., Scaglione, J.B., Janes, B.K., Anderson, E.C., Bergman, N.H., Hanna, P.C., Joachimiak, A., & Sherman, D.H. (2008). Structural and functional analysis of AsbF: origin of the stealth 3,4-dihydroxybenzoic acid subunit for petrobactin biosynthesis. Proceedings of the National Academy of Sciences U.S.A. **105**:17133-17138.

Putnoky, P., Kereszt, A., Nakamura, T., Endre, G., Grosskopf, E., Kiss, P., & Kondorosi, A. (1998). The *pha* gene cluster of *Rhizobium meliloti* involved in pH adaptation and symbiosis encodes a novel type of K super(+) efflux system. Molecular Microbiology, **28**:1091-1101.

CONCLUSION

Sinorhizobium meliloti and related rhizobia are generally regarded as legume endosymbionts that also inhabit soil, however this characterization is merely reflective of their agricultural importance to humankind and does not accurately reflect the observation that the majority of these microorganisms may survive in the natural environment as soil saprophytes. In actuality, an ancestral saprophytic lifestyle certainly predated that of the modern plant symbiont, and effective symbiosis is preceded by and entirely dependent upon the successful colonization of the legume rhizosphere. *S. meliloti* represents an apposite organism for the study of saprophytic competence in soil bacteria as the effective nodulation of agricultural crops is contingent upon the ecological success of rhizobia in fields.

Soil comprises a rich and complex growth substrate, offering to inhabiting microorganisms a vast array of organic nutrients such as aromatic and amino acids, polysaccharides, nucleosides, vitamins, and essential elements. The challenges inherent in colonizing soil are underscored by the intrinsically heterogeneous nature of this habitat; environmental conditions may differ substantially within a relatively small area, as influenced by fluctuations in osmotic potential, nutrient availability, and pH. The environmental heterogeneity of soil simultaneously influences and is shaped by the complexity of soil microbial communities, which comprise an intricate network of competing, predating, and syntrophic interactions that link various microorganisms.

The dynamic character of soil necessitates a broad array of catabolic, transport, and stress-response mechanisms, and the genomes of soil bacteria have expanded by gene duplication and horizontal transfer to accommodate this life style. In this work, we examined the systems employed by *S. meliloti* for the acquisition and metabolism of two plant-associated compounds: the modified amino acid hydroxyproline and the aromatic acid protocatechuate. These compounds are each representative of a broader class of organic molecules available for catabolism by soil heterotrophs. Amino acids such as hydroxyproline serve as a source of carbon and nitrogen, and thus offer an important

resource in soils where nitrogen may be limited. Lignin-derived aromatic acids may comprise the largest pool of inert biological carbon in soils and humic acids, and as a key substrate for the β–ketoadipate pathway, protocatechuate is uniquely positioned to represent the catabolism of a wide range of aromatic monomers. In the final chapter of this thesis, we examined the growth of *S. meliloti* deletion strains in sterile, bulk soil. These experiments were initiated to adopt a more comprehensive and holistic approach to the study of saprophytic fitness than was permitted by a focused study of individual metabolic systems.

As is true of many α–proteobacteria, *S. meliloti* encodes a multipartite genome that includes a single circular chromosome and two megaplasmids. While the chromosome carries many genes encoding essential housekeeping functions relevant to the growth of a heterotrophic bacterium, the megaplasmids encode genes that allow *S. meliloti* to occupy a highly specialized niche as a legume endosymbiont (primarily pSymA) or to adopt the lifestyle of a free-living saprophyte in soil. It has been proposed that pSymB acts as a major contributor regarding saprophytic competence in *S. meliloti,* and the presence of genes relevant to the acquisition and metabolism of hydroxyproline and protocatechuate (both of which originate from the degradation of plant tissue in soil) are consistent with this hypothesis. Additionally, we describe two strains of *S. meliloti* that exhibit a reduced saprophytic competence as the result of carrying distinct lesions within the pSymB megaplasmid. Thus the observations documented in this thesis support the role of pSymB as an important factor contributing towards the competitive success of *S. meliloti* as a soil saprophyte.

The broad taxonomic distribution of the β-ketoadipate pathway within soil bacteria emphasizes the importance of this peripheral metabolic pathway to the survival of soil saprophytes. Yet we note that the loss of the protocatechuate transport and catabolic apparatus in an *S. meliloti* strain did not negatively affect the ability of this strain to colonize a sterile soil microcosm. Nor did the deletion of a region encompassing the hydroxyproline gene clusters adversely impact the growth of *S. meliloti* in bulk soil. While this work detailed a genetic characterization of the protocatechuate and

hydroxyproline systems, the ecological importance of these systems in an environmental setting has not been equally assessed. However, it is probable that the contribution of any particular nutrient source to the growth of a soil microbe will vary considerably between soils, and the observations of this thesis suggest that *S. meliloti* is a generalist with respect to the acquisition and utilization of nutrients in soil. For example, whereas the legume-inhabiting bacteroid primarily utilizes succinate as a source of carbon and energy, the free-living soil bacterium may exploit a variety of carbon sources, including protocatechuate and hydroxyproline.

The observation that large-scale deletions encompassing multiple transport and catabolic systems rarely translated into a phenotype of decreased saprophytic fitness in soil was surprising. However, saprophytic competence is a complex and multifaceted phenotype that is influenced by the environmental conditions inherent in the study. The soil microcosm experiments described in this work were predicated upon a rich, loam soil, which was harvested from an alfalfa field. While not selected for this quality, the richness of soil undoubtedly influenced the nature and severity of mutations required to reduce the survival and propagation of *S. meliloti* in soil. The use of a nutrient poor substrate (such as quartz sand) in lieu of the soil employed in this work may yield an increased number of *S. meliloti* 'saprophytic' mutants, however it is questionable whether such a system is in any way representative of a real world habitat. More appropriate would be the screening of *S. meliloti* deletion strains in competition against a wild-type strain (intraspecies competition) or against another soil species (interspecies competition), using the microcosms as described in this work. It should not be concluded from this study that the majority of genes encoded by pSymA and pSymB do not contribute towards the ecological fitness of *S. meliloti* in soil. Rather, this study underscores the nutritional versatility of *S. meliloti,* as the loss of certain metabolic pathways was often compensated by alternative pathways within the metabolic network. In effect, the richness of any soil is only relevant to the study of an organism that can exploit the available resources.