THE UNIVERSITY OF CALGARY

Characterization of *Rhizobium leguminosarum* genes homologous to chemotaxis chemoreceptors

by

Christopher K. Yost

A DISSERTATION
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

CALGARY, ALBERTA
JANUARY, 1998

Abstract

Methyl-accepting chemotaxis proteins (MCPs) play important roles in the chemotactic response of many bacteria. Oligonucleotide primers designed to amplify the conserved signalling domain of *mcps* by PCR were used to identify *mcp* genes in *Rhizobium leguminosarum*. Using a PCR derived probe created from these primers a genomic library of *R. leguminosarum* VF39SM was screened. Consequently, at least 5 putative *mcp* genes (termed *mcp-1* to *mcp-5*) were identified and isolated from the library. One of these putative genes (*mcp-2*) is located on one of the indigenous plasmids of VF39SM. A further 15 cosmids showing homology to a *mcp-3* probe were also isolated from a genomic library, three of which were found to be plasmid localized. The complete DNA sequences of the putative *mcp* genes *mcp-1*, *mcp-2*, and *mcp-3* were obtained. The three genes code for proteins with the characteristic domains of typical MCPs. Partial DNA sequences of two other putative *mcp* genes (*mcp-4*, and *mcp-5*) were highly homologous to sequences from the methylation domains of known MCPs. Mutants defective in either *mcp-1*, *mcp-2*, *mcp-3*, or *mcp-4* were created using insertional mutagenesis strategies. On swarm plates, mutation of *mcp-1* resulted in impairment of chemotaxis to a wide range of carbon sources, while phenotypes for the other three mutants have yet to be elucidated. The *mcp*-1, *mcp-2*, *mcp-3*, and *mcp-4* mutants were tested for loss of nodulation competitiveness. When co-inoculated with wild-type the *mcp-1* and *mcp-2* mutants did not form equal numbers of nodules relative to the wild-type. The *mcp-3* and *mcp-4* mutant were just as competitive as the wild-type. This study suggests that, in *R. leguminosarum*, *mcp* -like genes are present, and that some, at least play a role in early steps in the plant-microbe interaction. In addition to mutational studies, promoter-less *lacZ* fusions were created with *mcp-3* to study *mcp* gene expression in VF39SM. The regulatory analysis of *mcp-3* indicated that this gene is down-regulated in bacteroids and

that the down-regulation is not coupled to the *nif* gene activators, *fnr*, *fixK* or *fixL*. Further studies indicated that succinate does not act as a trigger to induce *mcp-3* down-regulation.

Acknowledgements

I thank my supervisor, Michael Hynes, for providing an ideal setting to pursue my graduate work. My research endeavours have benefited greatly by his support, encouragement and scientific knowledge. I sincerely believe that I could not have had a better supervisor, and I am not just saying that because he endorsed lab ski days... I thank Melanie Berndt and Patrice Rochepeau for supplying the VF39SM genomic library that was used in my thesis work. Additionally, I thank Gerrit Voordouw for supplying the *dcrA* primers used initially to detect *mcps* in VF39SM. A general thank you to the many lab co-workers who have contributed positively to "life in the lab" during my thesis studies. A particular thank you is warranted to Tanya Noel for her help in the final preparation of this thesis. I am grateful to my wife, Catherine, for her helpful-hand in the lab from time to time and her genuine interest in my work. She has given me so much during my PhD studies, and for this I am extremely thankful. Lastly, I am indebted to my family for their support, both financial and emotional, during my academic pursuits. I am fortunate to have such a supportive family.

Table of contents

List of tables

List of figures

Abbreviations and symbols

aa: Amino Acid
Amp: Ampicillin
bp: Base Pair
Cm: Chloramphenicol
CSPD:3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'-chloro)tricyclo[$^{3.3\ 1\ 1}$3,7]decan}-4-yl)phenyl phosphate
DIG: Digoxigenin
Gm: Gentimicin
HCD: Highly Conserved Signalling Domain
Kb: Kilobase Pairs
Kda: Kilodaltons
Km: Kanamycin
MCP: Methyl Accepting Chemotaxis Protein
MCS: Multiple Cloning Site
min: Minute
NBT: Nitroblue Tetrazolium Salt
Nm: Neomycin
ORF: Open Reading Frame
PAGE: Polyacrylamide Gel Electrophoresis
PCR: Polymerase Chain Reaction
SDS: Sodium Dodecyl Sulfate
Sm: Streptomycin
Sp: Spectinomycin
Tc: Tetracycline
Tn: Transposon
V/cm: Volts per centimeter
VMM: Vincent's Minimal Medium
w/v: Weight per Volume
X-gal: 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside
X-phos: 5-bromo-4-chloro-3-indolyl-b-D-phosphopyranoside

1.0 Introduction

Using a process known as chemotaxis, motile bacteria have the ability to monitor their environment, integrate the information, and swim towards attractants and away from repellents. Chemotaxis, in bacteria, was first reported by Pfeffer in 1884. In this early study, Pfeffer was able to demonstrate that motile bacteria will swim towards a capillary tube filled with attractant. Pfeffer's experiments did not explain the mechanisms involved in chemotaxis however the pioneering work of Pfeffer did set the stage for future studies on bacterial chemotaxis. In 1969 an important work by Julius Adler on *Escherichia coli* chemotaxis was published. By studying the chemotactic behaviour of *E. coli* metabolic mutants Adler (1969) surmised that chemotaxis to certain attractants does not require metabolism of the attractants. A mutant unable to metabolize galactose still responds chemotactically to galactose at levels similar to wild-type levels (Adler, 1969). Additional evidence that metabolism and chemotaxis can be uncoupled was provided by isolating mutants which can still metabolize galactose but are no longer attracted to galactose. Before this study it was generally believed that chemotaxis was the result of bacteria sensing the energy produced from the metabolism of an attractant. Adler's work (1969) contradicted this belief and showed that bacteria can detect attractants without metabolizing them and therefore he is considered the first person to document the existence of chemoreceptors in bacteria.

The initial work by Adler in 1969 led to the subsequent extensive molecular characterization of chemotaxis in *E. coli* (reviewed by Manson, 1992; Stock *et al.*, 1991; Armitage, 1992; and Parkinson, 1993). In fact, to date, the chemotactic signalling pathway of *E. coli* is the most understood signal transduction pathway. However, recent research into the chemotactic behaviour of other bacteria, both eubacteria and archaebacteria, indicate that there can be significant deviations from the signalling pathway found in *E. coli*, therefore

there is still much research to be conducted in the area of bacterial chemotaxis. In addition, although the signalling pathway is well understood in *E. coli* and a few other bacteria, the biological significance of chemotaxis is only beginning to be resolved and merits further investigation.

1.1 Molecular Events in Chemotaxis.

As mentioned earlier, the chemotaxis signalling pathway was first elucidated in *E. coli*. Subsequently, the *E. coli* model has been viewed as the paradigm for chemotactic signalling. This section reviews the *E. coli* model.

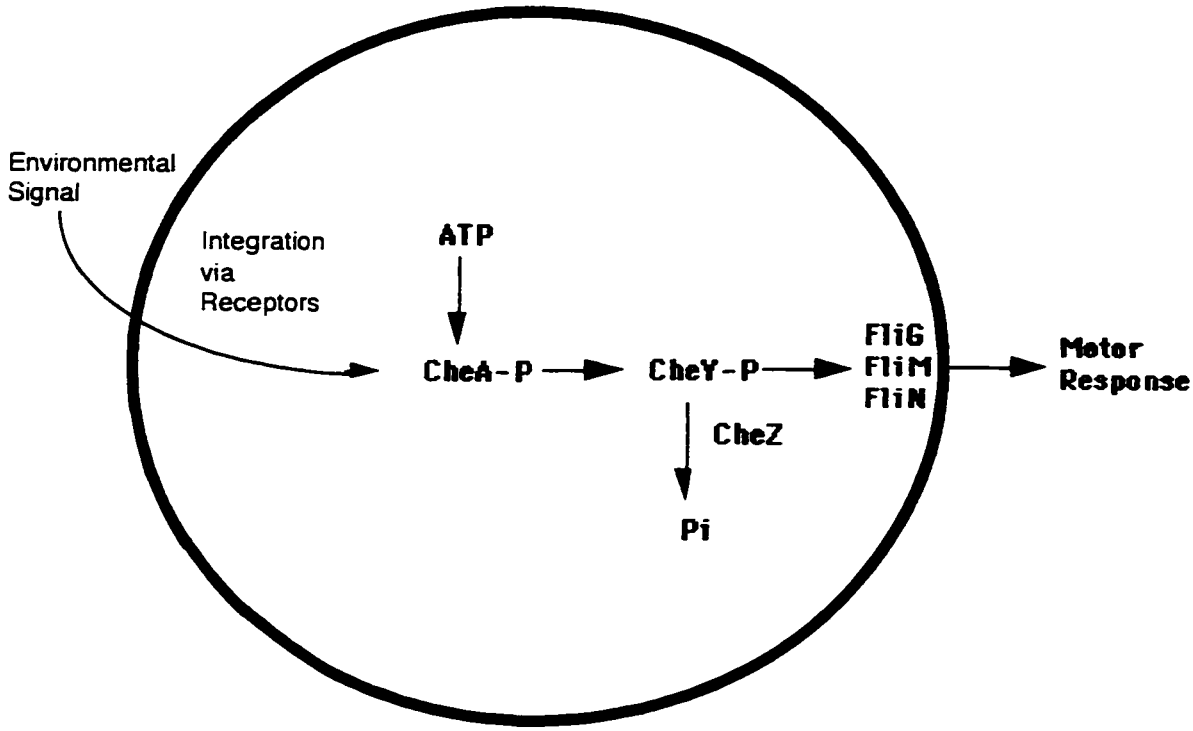
E. coli is peritrichously flagellated and propels itself by the rotation of its flagella (Silverman and Simon, 1974a). Counter-clockwise rotation results in co-ordinated flagellar rotation creating a bundle of flagella which propel the bacterium forward; a phenomenon termed smooth swimming (Larsen *et al.*, 1974). Conversely, clockwise rotation disturbs the flagellar bundle resulting in independent flagellar rotation and random reorientation of the cell; a phenomenon termed tumbling (Larsen *et al.*, 1974). How does the direction of flagellar rotation relate to chemotaxis? Components of the chemotactic signalling pathway control flagellar rotation. The genes in the chemotaxis operon of *E. coli* have been individually mutated and the effects on flagellar rotation noted. This data has played a major role in elucidating the chemotactic signalling pathway. Several pieces of experimental data indicate that a protein termed CheY interacts with the flagellar motor causing it to rotate clockwise (Parkinson, 1978; Parkinson *et al.*, 1983; Clegg and Koshland, 1984; Ravid *et al.*, 1986; Kuo and Koshland, 1987). I) Mutants in *cheY* exhibit constant counter-clockwise rotation (Parkinson, 1978). It should be noted that when there is no external signal acting on the motor it will rotate the flagellum counter-clockwise (Ravid *et al.*, 1986). II) Overproduction of CheY results in a bias to clockwise flagellar rotation, consequently cells over-expressing CheY tumble

with greater frequency (Clegg and Koshland, 1984; Kuo and Koshland, 1987). III) Flagellar motors in cell envelopes devoid of cytoplasm but reconstituted with CheY rotate in a clockwise manner (Ravid *et al.*, 1986). What protein transmits signals to CheY and how is this signal transmitted? Experimental evidence indicates that a protein named CheA modulates CheY by transferring phosphate groups to CheY (Hess *et al.*, 1987, 1988a, 1988b; Oosawa *et al.*, 1988; Wylie *et al.*, 1988; Borkovic *et al.*, 1989). CheA is a histidine kinase which can undergo autophosphorylation and consequently phosphorylate CheY (Hess *et al.*, 1987, 1988a; Stock *et al.*, 1988; Sanders *et al.*, 1989). Mutants in *cheA* rotate their flagella with a counter-clockwise bias, supporting the conclusion that CheA functions to stimulate CheY phosphorylation (Conley *et al.*, 1989). Another protein interacting with CheY is CheZ. CheZ acts as an antagonist to CheY, addition of CheZ accelerates the dephosphorylation of CheY-P (Hess *et al.*, 1988b). As predicted, when CheZ is absent, such as with *cheZ* mutants, cells tumble more frequently (Parkinson, 1978). Presumably, CheZ functions to help keep the signalling pathway balanced by preventing a build-up of CheY-P and therefore maintain chemotactic responsiveness. In fact, *cheZ* mutants show chemotactic responses that are abnormally extended in duration (Segall *et al.*, 1985).

The chemotactic signalling pathway described above is summarized in figure 1-1. In an environment lacking a stimulus, the rate limiting autophosphorylation of CheA results in intermediary levels of CheY-P (Stock *et al.*, 1991). Consequently, the motor will be experiencing periods of binding to CheY-P as well as unbound periods. This results in an alternation between clockwise (tumbling) and counter-clockwise (smooth swimming) rotation. The final behaviour is a random three dimensional walk (Berg and Brown, 1972). When an attractant is added, autophosphorylation of CheA is suppressed and consequently CheY is not phosphorylated (Borkovich and Simon, 1990). Subsequently there is a bias to counter-clockwise rotation and smooth

Figure 1-1. This schematic illustrates the backbone of the chemotactic signal transduction pathway. The key proteins involved in controlling the rotational phase of the flagellar motor are: CheA, CheY and CheZ (see text for details). Phosphorylation of a protein is denoted by -P. For simplicity, signal receptors have not been shown. However, many receptors transmit their signal to CheA themselves or indirectly through other proteins.

Figure 1-1 Chemotactic signalling pathway of *E. coli* :



swimming (Stock *et al.*, 1991). The cell will then migrate up the concentration gradient of attractant.

1.1.2 Methyl-Accepting Chemotaxis Proteins as Signal Transducers in Chemotaxis.

The previous section described the proteins involved in controlling flagellar rotation. This section describes a group of proteins which are involved in sensing chemoattractants and repellents. Methyl-accepting chemotaxis proteins (MCPs) sense a wide range of attractants and repellents (Hazelbauer *et al.*, 1990). In *E. coli* four MCPs have been cloned: Trg, Tar, Tap, and Tsr (Boyd *et al.*, 1983; Krikos *et al.*, 1983; Russo and Koshland, 1983; Bollinger *et al.*, 1984). MCPs can interact with an attractant directly; such as Tar which binds aspartate (Foster *et al.*, 1985). MCPs may also interact with the attractant indirectly, via binding proteins (Hazelbauer *et al.*, 1990). For example, Tar interacts with maltose via a specific periplasmic binding protein (Kossmann *et al.*, 1988). MCPs will also respond to repellents (Yamamoto *et al.*, 1990). The interactions between repellents and transducers are relatively non-specific; repellents do not recognize stereospecific sites on the transducer molecule (Hazelbauer *et al.*, 1990).

MCPs send their signalling information to the flagellar motor using the signalling components described in the previous section. However, an MCP can not interact with CheA directly, it requires an intermediary protein. CheW is a coupling protein which links MCPs to CheA (Liu and Parkinson, 1989; Gegner *et al.*, 1992). Therefore a ternary complex of MCP/CheW/CheA is necessary for signalling (Liu and Parkinson, 1989; Gegner *et al.*, 1992). An MCP with bound attractant will suppress CheA autophosphorylation (Borkovich and Simon, 1990), presumably by a conformational change in the MCP signalling domain that inhibits the interaction of ATP with the catalytic domains of CheA (Morrison,

and Parkison, 1997). By suppressing CheA phosphorylation the cell will adopt a smooth swimming behaviour, moving towards the attractant (Borkovich and Simon, 1990). Removal of the attractant results in the return to intermediary levels of CheA-P (Borkovich and Simon, 1990), upon which, the cell will resume its three dimensional walk.

Kort and colleagues (1975) were the first to show that a cell membrane protein involved in chemotaxis was methylated and hence called this protein methyl-accepting chemotaxis protein. CheR (referred to earlier as CheX) and CheB are responsible for the methylation and demethylation of MCPs, respectively (Springer and Koshland, 1977; Stock and Koshland, 1978; Hayashi *et al.*, 1979). Methylation is used for sensory adaptation (Springer *et al.*, 1979). If adaptation did not exist a cell would always be in an excited state (smooth swimming) in the presence of attractant. However, if a cell is left at a constant attractant concentration, after an initial excitatory response, the cell will adapt to the new attractant concentration and revert back to its pre-stimulatory random three-dimensional walk (Springer *et al.*, 1979). This observed adaptation is the result of MCP methylation. The sites for methylation occur in two domains in the C-terminal region (Stock *et al.*, 1991). Four, or more, glutamates are de/methylated in these domains (Stock *et al.*, 1991). Hazelbauer *et al.* (1990) have suggested that altered levels of methylation could cause conformational changes in the C-terminus of a MCP protein which would make the signalling domain more or less accessible to CheW/CheA.

Figure 1-2 is a schematic summarizing the role of MCPs in the chemotactic signalling backbone while figure 1-3 provides a diagrammatic representation of an MCP to illustrate the functional domains that are typical of most MCPs. The amino acid sequences of the two methylation domains and the signalling domain are highly conserved amongst bacterial species (Stock *et al.*, 1991).

Figure 1-2. The *E. coli* MCP Tar is used as an example in this schematic. Tar binds the chemoattractant aspartate. Proteins involved specifically in MCP signalling are: an MCP, CheW, CheR, and CheB. Phosphorylation of a protein is denoted by -P. Adomet is S-adenosylmethionine. CheW couples MCPs to CheA. CheR is a methyltransferase, while CheB is an esterase. CheR activity is unregulated, however ligand bound MCPs are better substrates than unliganded receptors. On the other hand, CheB activity is regulated by CheA. CheA-P will phosphorlyate CheB. CheB-P is the active form of the enzyme. CheR and CheB are involved in sensory adaptation (see text).

Figure 1-2 The role of an MCP in the *E. coli* chemotactic signalling pathway:

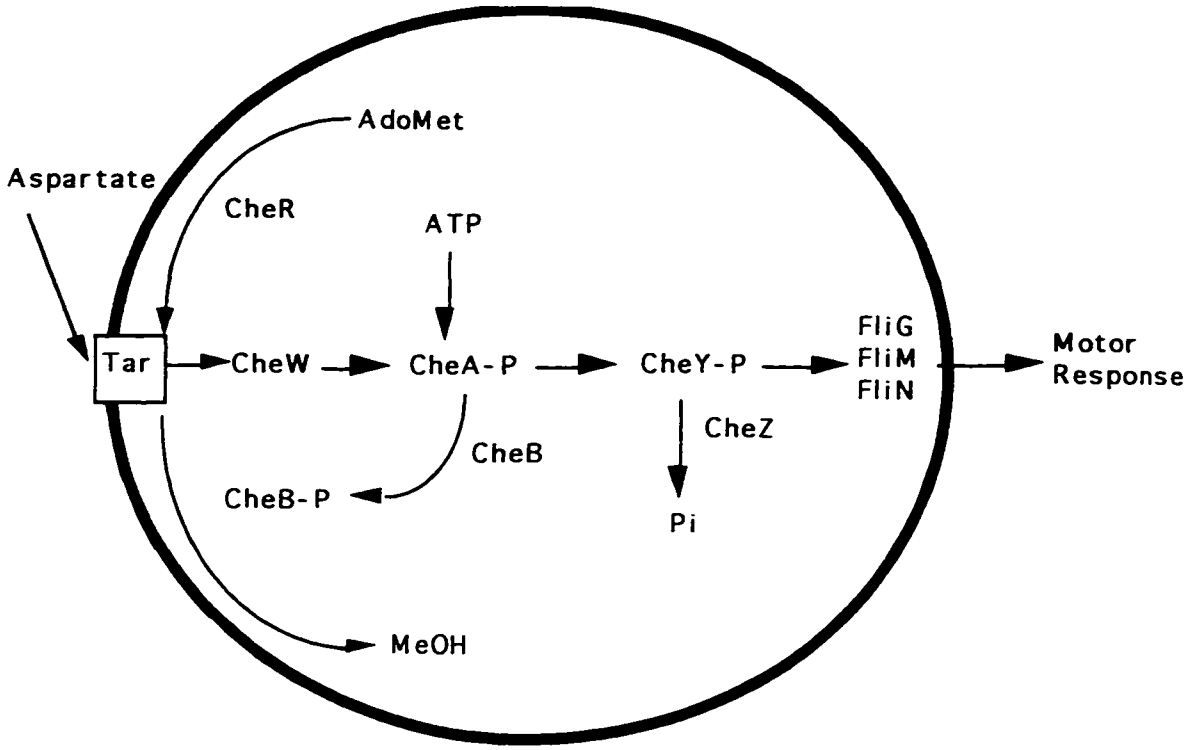
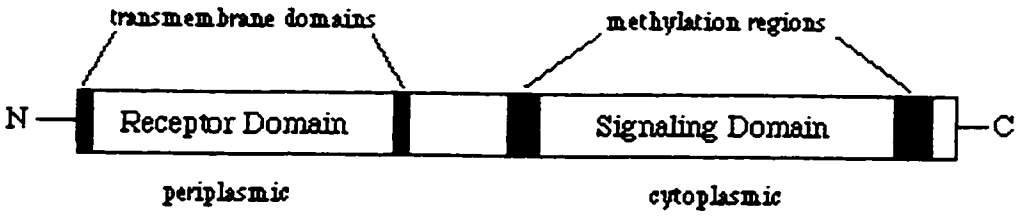


Figure 1-3. A classical MCP will have all the functional domains shown above. Receptors can vary in size, from 57.4 Kda for Tsr (Boyd *et al.*, 1983) to the 70 Kda McpA protein of *C. crescentus* (Alley *et al.*, 1992). Size variations usually stem from differences in the receptor domains. In Gram negative bacteria, the receptor domain is generally located in the periplasm due to the two flanking transmembrane domains. The area of highest homology between bacterial MCPs is found in the methylation and signalling domains.

Figure 1-3 Schematic of a classical MCP protein:



1.1.3 Methylation-Independent chemotactic signalling pathways in *E. coli* .

1.1.3.1 Chemotactic responses to phosphotransferase sugars.

E. coli is attracted to sugars such as glucose, mannitol and sorbitol (Adler and Epstein, 1974). These sugars are transported into the cell via the phosphotransferase-system (PTS) (for a general review of the PTS system see Postma *et al.*, 1985; Reizer *et al.*, 1988). Each sugar is transported by a sugar specific transport complex, known as EIIABC (Lengeler *et al.*, 1992). A number of different EIIABC complexes exist, each transporting a respective PTS sugar. A PTS sugar is phosphorylated during transport into the cell. The phosphate group originates from phosphoenol pyruvate and is transferred to the sugar via a phosphorylation cascade involving proteins, EI and HPr. Whereas many different EIIABC complexes exist, there are only single forms of EI and HPr. EI accepts a phosphate from phosphoenol pyruvate and subsequently transfers this phosphate to HPr. HPr-P donates its phosphate to the EIIABC complex which then phosphorylates the sugar during transport into the cytoplasm.

MCPs do not exist for sensing these sugars, instead, chemotaxis is dependent upon a functioning PTS system. Transport of a PTS sugar itself does not trigger a chemotactic response. Pecher and colleagues (1983) demonstrated this by showing that uptake of a sugar-phosphate does not initiate chemotaxis. In fact, the phosphorylation events occurring during transport are crucial for proper PTS sugar chemotaxis (described in Lengeler and Vogler, 1989). Mutants unable to express EI can bind and occasionally transport sugar substrates into the cell, however these sugars are not phosphorylated and the cell cannot respond chemotactically to them. Additionally, metabolism of a PTS-sugar is not necessary for proper chemotactic behaviour. Analogs of PTS-sugars that can

be transported and phosphorylated but not metabolized still trigger a chemotactic response (Lengeler *et al.*, 1981).

Is PTS chemotaxis linked to the chemotactic signalling pathway used by MCPs? Rowsell *et al.* (1995) used genetic reconstitution with an *E. coli* strain devoid of the chemotaxis operon (*cheA, cheW, cheY, cheZ, cheR, cheB*) to show that PTS-sugar chemotaxis requires CheA, CheW, and CheY for proper function. Therefore it is likely that the signals for PTS-sugar chemotaxis as well as MCP dependent chemotaxis converge at CheA. Lux and colleagues (1995) have further elucidated the link between PTS-chemotactic signalling and CheA. Using purified proteins and *in vitro* assays Lux *et al.* (1995) have shown that unphosphorylated EI can inhibit CheA phosphorylation. Using this result and experimental data describing phosphoryl transfer rates between EI and HPr, Lux and colleagues (1995) have proposed the following model: Transport of a PTS-sugar results in the rapid dephosphorylation of EI by HPr, the subsequent build up of unphosphorylated EI inhibits CheA phosphorylation stopping the flow of phosphates to CheY. This signalling phenomenon would result in a smooth swimming behaviour and movement up the concentration gradient.

1.1.3.2 Energy taxis and chemotaxis towards oxygen.

Recently, positive chemotaxis towards glycerol was observed in *E. coli* (Zhulin *et al.*, 1997), this contrasts earlier reports of glycerol being a repellent for *E. coli* (Oosawa and Imae, 1983). Both reports are correct, at high concentrations glycerol is a repellent detected in a classic MCP-dependent manner (Oosawa and Imae, 1984), while at low concentrations (1mM) glycerol acts as an attractant and methylation of an MCP is not required for proper taxis (Zhulin *et al.*, 1997). Notably, to function as a chemoattractant, glycerol must be metabolized. *E. coli* mutants able to transport glycerol but unable to metabolize

it cannot exhibit positive glycerol taxis (Zhulin *et al.*, 1997). Therefore glycerol is an example of energy taxis in *E. coli*.

Another example of energy taxis is the chemotactic response of *E. coli* to oxygen, reported by Shioi *et al.* (1988). Further elucidation of the areotaxis signalling pathway was provided by Rebbapragada and collegues (1997), who characterized a gene known as *aer* and its involvement in aerotaxis. The gene product of *aer* has high amino acid similarity to the signalling domain of MCP proteins, but does not contain predicted methylation regions (Rebbapragada *et al.*, 1997), however it is still referred to as a MCP chemotransducer. The N-terminus of Aer contains a putative FAD-binding site similar to NifL, a protein which regulates gene expression in response to redox changes (Rebbapragada *et al.*, 1997). It appears that Aer elicits a chemotactic response to oxygen by sensing intracellular energy levels rather than detecting oxygen itself as an environmental signal (Rebbapragada *et al.*, 1997). Interestingly, Aer may also play a role in glycerol taxis (Rebbapragada *et al.*, 1997), indicating that Aer may have a key signalling role in energy taxis.

1.2 Deviations from the *E. coli* chemotactic signalling paradigm.

The signalling pathway elucidated in *E. coli* is common to other enteric bacteria such as *Salmonella typhimurium*. However, the relatively recent surge of interest into the chemotactic signalling pathway of bacteria other than enterics has revealed that although the general elements of the signalling pathway are conserved there can be significant deviations. DNA sequencing of the chemotaxis operons from members of the α proteobacteria has revealed the presence of multiple *cheY* genes. In *E. coli* the chemotactic signalling genes are present as single copies and, excluding three MCP transducers (*trg*, *tsr*, and *aer*), reside in two operons. The chemotactic signalling operon of *Sinorhizobium meliloti* (Greck *et al.*, 1995), *Rhodobacter sphaeroides* (Ward *et*

al., 1995a), and *Agrobacterium tumefaciens* (E. Wright, personal communication) contain two functional and distinct *cheY* genes. The presence of two functional yet different *cheY* genes suggests that both play specific roles in chemotaxis and therefore add a new level of complexity not seen in the *E. coli* signalling pathway. The purpose of two *cheY* genes may be explained by the swimming behaviour of members of the a group of proteobacteria. Unlike *E. coli*, *Sinorhizobium meliloti* cannot change the direction of its flagellar rotation (Götz and Schmitt, 1987). Instead, *Sinorhizobium meliloti* exclusively rotate their peritrichous flagella in a clockwise direction resulting in smooth swimming. Tumbling is not observed in *S. meliloti* and an alternate tactic behaviour has been postulated for reorientation. *S. meliloti* intermittently varies its rotary speed asynchronously (Sourjik and Schmitt, 1996; R. Schmitt, T. Pitta and V. Sourjik, unpublished). Consequently, different rotary speeds of individual flagella could disrupt the flagellar bundle and cause uneven thrust to the cell (Sourjik and Schmitt, 1996) thereby resulting in a random walk similar to that observed in *E. coli*, but with sharp turns rather than abrupt changes in direction (Götz and Schmitt, 1987). How does *S. meliloti* modulate its speed of flagellar rotation? Sourjik and Schmitt (1996) have provided a possible model for this tactic behaviour This model includes an explanation for the need of two separate *cheY* genes. Studies using mutant strains deleted in either *cheY1* or *cheY2* have indicated that both CheY proteins are necessary for the full tactic response (Sourjik and Schmitt 1996). CheY2 is the main response regulator and when phosphorylated it acts to slow flagellar rotation (Sourjik and Schmitt 1996). CheY1 does not participate in altering flagellar rotation speed, but does interfere in smooth swimming (Sourjik and Schmitt 1996). It appears that the main role of CheY1 is to modulate CheY2 by competing for phosphorylation from CheA (Sourjik and Schmitt 1996). Since a *cheZ* gene has not been found in *S. meliloti* the role of CheY1 as a competitor for phosphorylation seems appropriate. This model may not include all members of the *Rhizobiaceae* family as some *Rhizobium* spp are bipolarly flagellated and their asynchronous flagellar rotation could not play a role in the tactic response of these bacteria. In

Rhizobium leguminosarum , a peritrichously flagellated bacterium, the swimming behaviour is uncharacterized. Therefore it is unknown if the model proposed by Sourjik and Schmitt (1996) includes other members of the *Rhizobiaceae* family.

Analysis of the chemotactic operon from *Sinorhizobium meliloti* has revealed another deviation from the *E. coli* operon. At the 5' end of the operon are two open reading frames (ORFs) that have homology to the HCD domain of *E. coli mcp* genes but lack transmembrane domains and contain no methylation regions (Greck *et al.*, 1995). These observations indicate the possible presence of novel chemotactic transducers which are presumably localized in the cytoplasm. The role of the two genes in chemotaxis is unknown since no phenotypes have been assigned to either gene.

E. coli, *Salmonella typhimurium* and *Enterobacter* have small families of MCPs, totalling 4 to 5. Identification of MCPs from other bacteria indicate that not all bacteria are limited to small MCP families. In the archaeon, *Halobacterium salinarium*, 13 MCPs have been identified (Zhang *et al.*, 1996). Zhang and colleagues (1996) have classified this family of proteins as MCPs based on DNA sequence data. Each gene's predicted amino acid sequence shows extensive homology to the signalling domain of Tsr. In addition, each gene contains putative methylation regions (Zhang *et al.*, 1996). *H. salinarium* MCPs can be grouped into 3 subfamilies based on structural differences; contrasting *E. coli* where MCPs are almost all identical in their structure, the only exception is Aer. The 3 subfamilies are described as: i) traditional MCPs with two transmembrane regions, a periplasmic domain, and a cytoplasmic domain, ii) MCPs with two or more membrane spanning regions, no periplasmic domain, and a cytoplasmic domain, or iii) MCPs with no membrane spanning regions, residing entirely within the cytoplasm (Zhang *et al.*, 1996). Large families of MCPs are not exclusive to the archaea. A large family of MCPs has been identified in *Desulfovibrio vulgaris* (Deckers and Voordouw, 1994). Partial

DNA sequencing of these genes revealed predicted amino acid sequence with high homology to the HCD domain of known MCPs. The explanation for large MCP families in certain bacteria is only speculative at the moment. A clearer picture may be drawn once phenotypes for each putative MCP are elucidated.

The role of MCPs in the chemotactic behaviour of certain bacteria is quite unclear. *Rhodobacter sphaeroides* is one such bacterium. In *E. coli*, glutamate chemotaxis is MCP-dependent, whereas in *R. sphaerodies* glutamate taxis is believed to be MCP independent (Sockett *et al.*, 1987). Jacobs *et al.* (1995) may have an explanation for the contrasting chemotactic behaviours. A glutamate transport-deficient mutant of *R. sphaerodies* is chemotactically impaired towards glutamate. This mutant of *R. sphaerodies* could bind glutamate at similar levels to wild type but it did not respond chemotactically to glutamate nor could it grow using glutamate as the sole carbon source (Jacobs *et al.*, 1995). However, glutamate chemotaxis was restored by complementing this mutant with the *gltP* gene of *E. coli* (Jacobs *et al.*, 1995). *gltP* codes for the H^+-linked glutamate carrier of *E. coli* (Deguchi *et al.*, 1989). These experiments indicate that glutamate taxis in *R. sphaerodies* requires the uptake of glutamate into the cell (Jacobs *et al.*, 1995), and support the earlier suggestion by Sockett *et al* (1987), and Armitage *et al.* (1990) that intracellular metabolism is necessary for chemotaxis in *R. sphaerodies*. Sockett and colleagues (1987) suggested that *R. sphaerodies* chemotaxis was methylation independent. However, the recent cloning of an *mcp* gene from *R. sphaerodies* (Ward *et al.*, 1995b) complicates the chemotaxis model for *R. sphaerodies*.

Given the diverse environments that bacteria inhabit it is not surprising to find diversity in chemotactic signalling pathways. The large breadth of knowledge about chemotactic signalling pathways will help us to study chemotaxis in the context of life-cycle and habitat of a particular bacterium. Investigating the biological significance of chemotaxis should provide new and interesting insights into the way bacteria interact within their environments.

1.3 The *Rhizobiaceae* Family and chemotaxis.

Members of the family *Rhizobiaceae* include *Agrobacterium* spp., *Bradyrhizobium* spp, *Sinorhizobium* spp., and *Rhizobium* spp. These bacteria are motile Gram negative flagellated rods that are found inhabiting soil environments Often, these bacteria will be found in association with plants. Members from the genus *Agrobacterium* are pathogenic to numerous plants causing diseases such as crown gall disease (Nester *et al.*, 1984), whereas members of the genera *Azorhizobium*, *Sinorhizobium*, *Bradyrhizobium*, and *Rhizobium* can form symbiotic relationships with leguminous plants.

The biological significance of chemotaxis has not been fully understood in the *Rhizobiaceae* family. Elucidating the role chemotaxis plays in the life of *Rhizobiaceae* should offer new insights into their complex life-cycle.

1.3.1 The Role of Chemotaxis in the Life-Cycle of *Agrobacterium tumefaciens*.

Agrobacterium infects a plant host by entering through wound sites on the plant. Interestingly, phenolic compounds secreted by the wounded plant tissue activate the virulence genes in *A. tumefaciens* (Stachel *et al.*, 1985; Bolton *et al.*, 1986). The majority of these virulence genes are carried on a large plasmid known as the Ti plasmid (Van Larebeke *et al.*, 1974; Holsters *et al.*, 1980). Strains cured of the Ti plasmid are avirulent (Van Larebeke *et al.*, 1974). Activation of virulence genes by phenolic compounds is accomplished through the two component regulatory system *virA/virG* (Stachel and Zambryski, 1986). VirA is a phenolic sensor and VirG is a positive transcriptional regulator that binds to the *vir* DNA box and activates transcription of the virulence genes. Acetosyringone appears to be a key phenolic in regulating virulence gene activation. In addition to phenolics, *vir* genes can be activated by some sugars

through a signalling pathway involving *chvE*. ChvE is a periplasmic sugar binding protein that responds to galactose, glucose, and several other sugars. In conjunction with VirA it up-regulates *vir* gene expression in the presence of these sugars (Cangelosi *et al.*, 1990). Interestingly, ChvE is also necessary for proper chemotaxis to these sugars (Cangelosi *et al.*, 1990).

A. tumefaciens chemotaxis has been implicated as a virulence factor for this organism (Hawes and Smith, 1989). The assumption that chemotaxis plays a biologically significant role in the life-cycle of *A. tumefaciens* has sustained a research effort into elucidating the chemotactic signalling pathway of this organism.

In addition to activating virulence genes, plant released phenolics act as chemoattractants for *A. tumefaciens* (Ashby *et al.*, 1987; Parke *et al.*, 1987; Ashby *et al.*, 1988). Ashby *et al.* (1987, 1988) have further suggested that the chemotaxis to one of these phenolics, acetosyringone, is dependent on genes residing on the Ti plasmid. Shaw *et al.* (1988) and Palmer and Shaw (1992) have used *virA* and *virG* mutant strains of *A. tumefaciens* to implicate the *virA* and *virG* gene products in acetosyringone chemotaxis. Palmer and Shaw (1992) have suggested that VirG has a higher affinity for a component of the chemotactic signalling pathway than for the *vir* DNA box. The peak concentration for chemoattraction to acetosyringone is 10^{-7}M, 500-fold lower than the maximal concentration for inducing virulence genes (Shaw, 1991). Therefore, at low levels of acetosyringone the small amounts of VirG-P produced would interact preferentially with the chemotaxis machinery, resulting in movement up the acetosyringone concentration gradient. As the acetosyringone concentration increases there would be an abundance of VirGP and the extra molecules would bind to the *vir* DNA box allowing virulence gene transcription to begin. This model still requires much investigation; the component of the chemotactic machinery that interacts with VirG is yet to be identified. As well, other issues must be resolved such as the inability of other

researchers to see a chemotactic response to acetosyringone in *A. tumefaciens* (Parke *et al.*, 1987).

1.3.2 The Role of Chemotaxis in the Free-living and Symbiotic State of *Rhizobium*.

The symbiotic relationship between root-nodule forming bacteria (collectively known as rhizobia) and leguminous plants has been the focus of intensive world-wide research efforts. Population growth is causing increased pressure on food production and consequently there is an increased need for fixed nitrogen fertilizer. As there is a growing concern for the environment and agricultural sustainability research into biological nitrogen fixation is becoming increasingly more important.

Symbiotic nitrogen fixation occurs in specialized structures known as nodules. Nodules result from a rhizobial infection of legume roots. The infection process is complex, involving multiple gene products and co-ordinated signalling events between the bacterium and plant host. Genes necessary for nodulation (*nod* genes) have been identified in rhizobia (Rossen *et al.*, 1984; Torok *et al.*, 1984; Downie *et al.*, 1985; Egelhoff *et al.*, 1985; Debelle and Sharma, 1986; Schofield and Watson, 1986). The characterization of one particular gene, *nodD*, has shown that the product of *nodD* is a regulatory protein that, in the presence of legume exudates, activates the transcription of other *nod* genes (Innes *et al.*, 1985; Mulligan and Long, 1985; Rossen *et al.*, 1985; Shearman *et al.*, 1986; Burn *et al.*, 1987). The compounds present in plant exudates which activate nodulation genes have been identified as flavonoids (Peters *et al.*, 1986; Redmond *et al.*, 1986). Flavonoid secretion helps to determine host specificity as only flavonoids secreted by the host plant will activate *nod* gene expression in the respective rhizobial sp. (reviewed by Fellay *et al.*, 1995). The expression of *nod* genes results in the production and secretion of Nod factors (Lerouge *et*

al., 1990; reviewed by Downie, 1994). Since the isolation of the first Nod factor by Lerouge *et al.* (1990), a number of Nod factors have been isolated and characterized. Structural analysis indicates that all Nod factors have a common acylated chito-oligosaccharide backbone, and that variations are due to the different substiuents added to the backbone (Downie, 1994). Like flavonoids, changes in Nod factor structure help to determine host specificity; a host plant will only respond to Nod factors secreted by its symbiont (reviewed by Fellay *et al.*, 1995). Nod factors trigger morphological changes within the plant root, such as root hair curling, which are necessary for the bacteria to infect the plant and continue the nodulation process (Spaink *et al.*, 1991; Truchet *et al.*, 1991; Stokkermans and Peters, 1994). (The plant infection process is very complex, for a detailed explanation of the process refer to Kannenberg and Brewin, 1994). Rhizobia attach to root hair surface, and enter the root through the formation of an infection thread (Vesper and Bauer, 1985; Turgeon and Bauer, 1982). The bacteria are confined to the infection thread (Napoli and Hubbell, 1975) until they reach the inner cortex of plant cells (Patel and Yang, 1981; Turgeon and Bauer, 1985). Here the bacteria are engulfed by the plasma membrane of a plant cell (Tu, 1975) (peribacteroid membrane), the bacteria multiply and divide; generally, after division each bacterium is surrounded by its own peribacteroid membrane (Robertson and Lyttleton, 1984; Roth and Stacey, 1989). After several rounds of division the bacteria differentiate into bacteroids and reside in structures termed symbiosomes (Patel and Yang, 1981). The terminally differentiated bacteroid reduces atmospheric nitrogen for the plant.

It has frequently been suggested that chemotaxis and motility must play an important role in the interaction of rhizobia with their legume hosts (Currier and Strobel, 1977; Ames & Bergman, 1981; Gulash *et al.*, 1984; Munoz Aguilar *et al.*, 1988; Caetano-Anollés *et al.*, 1988a,b; Bauer and Caetano-Anollés, 1990; Caetano-Anollés *et al.*, 1992; and reviewed by Bauer, 1991; Vande Broek & Vanderleyden, 1995). As a result there has been a sustained research effort

into rhizobial chemotaxis, however only now are we beginning to understand rhizobial chemotaxis at the molecular level.

During nodulation, rhizobia are known to attach to the plant root and in some instances, rhizobia attach at the tip of a developing root hair during infection (Kijne *et al.*, 1990). Researchers have wondered if rhizobia find sites of infection randomly or are bacteria directed to the proper infection site by responding to chemoattractants secreted by the plant. Gulash and colleagues (1984) have demonstrated that *Rhizobium* cells are attracted to the root tip surface. A portion of this attraction is not species specific and is likely due to plant exudates containing various sugar and metabolite chemoattractants. However, researchers have discovered that rhizobia are chemoattracted to flavonoids (Armitage *et al.*, 1988; Caetano-Anollés *et al.*, 1988b; Munoz Aguilar *et al.*, 1988; and Dharmatilake and Bauer, 1992). This chemotaxis can be species specific (Munoz Aguilar *et al.*, 1988). It is interesting to note that chemotaxis towards flavonoids can occur at a concentration ten fold lower than that needed for *nod* gene induction (Munoz Aguilar *et al.*, 1988). Therefore, a situation may occur where rhizobia are attracted to the proper infection site by plant secreted flavonoids. Once they are drawn to the root tip the flavonoid concentration is sufficiently high to induce *nod* genes and the nodulation process can begin. In fact, field trials have shown that non-motile mutant strains as well as general non-chemotactic mutants are less competitive than motile, chemotactically competent wild type strains at forming nodules (Ames and Bergman, 1981; Caetano-Anollés *et al.*, 1988; Bauer and Caetano-Anollés, 1990). When the mutants were co-inoculated with the wild type parent strain nodules from the mutant strain were under-represented.

The molecular characterization of chemotaxis and motility in the nodulation process is limited. For instance, the signalling pathway used for flavonoid taxis is unknown. However, some research has provided insight into flavonoid taxis. NodD, the gene product that senses flavonoids and acts as a transcriptional

activator of the other *nod* genes, is not involved in flavonoid chemotaxis (Dharmatilake and Bauer, 1992). Therefore, although nod gene induction and flavonoid chemotaxis have similar biochemical specificity, the signalling pathways used for each biological event are different. This is in contrast to the case of *A. tumefaciens* where *virA/virG* participates in both virulence gene regulation and chemotaxis to plant wound secreted phenolics (Shaw *et al.*, 1988; and Palmer and Shaw, 1992). The genes involved in nodulation and nitrogen fixation are localized to large plasmids known as sym plasmids (Johnston *et al.*, 1978; Nuti *et al.*, 1979; Dénarié *et al.*, 1981; Rosenberg *et al.*, 1981). Although *nodD* does not play a role in flavonoid chemotaxis it appears that some functions of flavonoid chemotaxis may be encoded on sym plasmids (Armitage *et al.*, 1988). Flavonoid taxis was studied in *R. leguminosarum* 8401 pRL1JI and its sym plasmid cured derivative strain 8401. Both strains were attracted to increasing concentrations of flavonoids, but, 8401 showed a significant decrease in its chemotactic response to the flavonoids apigenin and naringenin (Armitage *et al.*, 1988). These results indicate that although genes on the sym plasmid are not needed for a chemotactic response to flavonoids there may be some unidentified genes that enhance flavonoid taxis (Armitage *et al.*, 1988)

In the free-living state, *Rhizobium* can be found in the rhizosphere. The rhizosphere is defined as the area adjacent to plant roots. Presumably, competition for metabolites in the rhizosphere environment is fierce. *Rhizobium spp.* are capable of utilizing a wide variety of carbon metabolites (Parke and Ornston, 1984; Boivin *et al.*, 1991; Baldani *et al.* 1992). This metabolic diversity is beneficial to the rhizobial cell since it can quickly change its carbon source when one becomes too limiting. The role of chemotaxis in the rhizosphere is largely unknown. It is tempting to hypothesize that a chemotactic strain will be more apt at surviving the highly competitive rhizosphere environment than a non-chemotactic strain. The molecular signalling events involved in rhizosphere chemotaxis are not well understood, and require further study.

1.3.2.1 The Role of Methyl Accepting Chemotaxis Proteins in *Rhizobium* Chemotaxis.

The existence of MCPs in a wide range of bacterial genera indicates their importance in chemotaxis. Genes potentially coding for MCPs have been identified in rhizobia although their role in chemotaxis is still unknown. Earlier, the existence of MCPs in rhizobia was questionable. Conflicting reports using *in vivo* methylation studies raised doubt for the existence of *mcp* genes in rhizobia. Robinson and Bauer (1993) observed no increase or decrease in protein methylation after *Sinorhizobium meliloti* cells were incubated with attractants such as L-amino acids or D-mannitol. However, Armitage *et al.* (1988) demonstrated an increase in protein methylation after *Rhizobium leguminosarum* cells were incubated with the chemoattractant L-serine. The latter experiment was in agreement with a study conducted by Morgan *et al.* (1993). They were able to show increases in protein methylation after stimulation with known chemoattractants such as serine, as well as provide western blots of rhizobia strains that tested positive for MCPs after probing with *E. coli* anti-Trg (Morgan *et al.*, 1993). Subsequent to this research *mcp* homologous genes have been cloned from different rhizobia. Greck and colleagues (1995) cloned and sequenced the *che* operon of *Sinorhizobium meliloti*, identifying a putative MCP gene. The putative MCP lacks the characteristic transmembrane regions suggesting that it might be located entirely in the cytoplasm. In addition, this MCP appears to lack the characteristic methylation domains. Typical *mcp*–like genes have also been identified. Brito *et al.* (1996) have cloned and sequenced a *mcp* gene from *Rhizobium leguminosarum* bv. viciae UPM791. This *mcp* gene is localized on a nodulation plasmid adjacent to hydrogenase (*hup*) genes. Additional genes with homology to MCPs have been identified on the sym plasmid of *Rhizobium* species NGR234 (it should be noted that NGR234 is almost certainly a *Sinorhizobium* species) (Freiberg *et al.*, 1997). In each case no role for the *mcp*- like genes in chemotaxis has been shown.

1.4 Gene regulation within the nodule.

A rhizobial cell which differentiates into a bacteroid experiences numerous cellular changes. The most readily apparent change is the appearance of nitrogenase activity and the subsequent fixation of atmospheric nitrogen. Many of the other changes occurring during bacteroid differentiation are associated with the cell surface: bacteroids have weakened cell walls (Van Brussel *et al.*, 1977), there is a reduction in extracellular polysaccharide synthesis (Tully and Terry, 1985), changes in membrane-associated LPS occur (Brewin *et al.*, 1986; VandenBosch *et al.*, 1989; Sindhu *et al.*, 1990), and some surface proteins are down regulated (de Maagd *et al.*, 1989). Some of these changes can be induced in free-living cells by altering environmental conditions. For example growing cells in microaerobic conditions will trigger *nif* gene expression (David *et al.*, 1988). Some of the cell surface changes are also regulated by oxygen, as well as pH (Kannenberg and Brewin, 1989). Although oxygen concentration and pH play significant roles as signalling molecules not all the cellular changes are associated with these environmental cues (de Maagd *et al.*, 1994). Therefore there must be other, as yet unidentified, environmental signals involved in triggering the genetic events associated with bacteroid differentiation. Consequently, it stands to reason there should be novel unidentified signalling components that respond to these environmental cues. Research is ongoing to identify genes that are differentially regulated during transition from the free-living to bacteroid state. By studying such genes new signalling pathways involved in bacteroid differentiation may be elucidated (de Maagd *et al.*, 1994).

1.4.1 Expression of motility genes within the nodule.

Motility is an energetically costly phenotype (Macnab, 1990). Motility related proteins constitute a large percentage of total cellular protein (Macnab, 1990).

Consequently, it is not surprising that flagellar genes and chemotaxis genes are only expressed when needed. For example, in *E. coli* flagellar gene expression is under catabolite repression (Silverman and Simon, 1974). This seems logical as the bacterium would not need motility and chemotaxis to search for nutrients if it already has adequate nutrients available. Other conditions can occur in nature where motility is not essential. For example, in *Rhizobium* a non-motile mutant can form a functional nodule (Ames and Bergman, 1981). Therefore, although motility may be an asset in the stages before root infection it is not needed once rhizobia have entered the plant tissue. Since motility is not needed it is not surprising to learn that flagellar genes are not expressed in the bacteroid state. Using *lac Z* reporter fusions to the *fla* genes of *Sinorhizobium meliloti* it has been shown that bacteroids do not express flagella genes (K. Bergman, personal communication). It appears that motility genes are still expressed in the infection thread (K. Bergman, personal communication). Subsequently, flagellar and most probably chemotaxis genes are shut off when the rhizobial cells differentiate into bacteroids. Genes involved in motility appear to be expressed under the same regulatory hierarchy (Macnab, 1990). Consequently, it seems reasonable to assume that, in *Rhizobium*, chemotaxis genes may be regulated by the same means as flagellar genes. The mechanism behind this negative regulation is unknown. Studying *mcp* gene expression during the nodulation process may reveal novel signalling pathways and identify new signals involved in triggering bacteroid differentiation.

2.0 Research Objectives

When this research was initiated the existence of *mcp* genes in *Rhizobium* had not been clearly established. One objective of this research was to identify, isolate and characterize *mcp* genes from *Rhizobium leguminosarum* bv. viciae strain VF39SM. Characterization of cloned *mcp* genes included DNA sequencing and insertional mutagenesis of selected *mcp* genes. Phenotypic characterization of the mutants included swarm assays to various chemoattractants, as well as competition assays. To observe if mutations in MCPs decreased competitive fitness, plant tests were conducted assaying for efficiency of nodulation. Nodulation efficiency can be considered a method for detecting *Rhizobium* competitiveness in the rhizosphere. By observing MCP mutants in the rhizosphere more information regarding the ecological role of chemotaxis in the soil can be obtained. The results of these studies will contribute to information regarding the role of MCPs in *Rhizobium* chemotaxis.

In addition to chemotactic experiments, the regulation of *Rhizobium mcp* gene expression was investigated. *lacZ* gene fusions to selected *mcp* genes, cloned earlier in this study, were used to determine what signals regulate *mcp* gene expression. Specifically, the possibility of catabolite repression of *mcp* genes during free-living conditions was investigated. Investigations also determined if *mcp* genes are down regulated during nodulation. These studies will aid in identifying novel gene regulation and signalling pathways in *Rhizobium*.

3.0 Materials and Methods

3.1 Bacterial Strains, Plasmids and Media.

The bacterial strains and plasmids used in this study are listed in table 3-1. *R. leguminosarum* strains were grown on TY medium (Beringer, 1974) at 30 °C. *E. coli* strains were grown on LB media (Sambrook *et al.*, 1989) at 37 °C. Yeast extract swarm medium (YES) was composed of 0.01 % yeast extract, 1mM $MgSO_4$, and 0.3 % (w/v) agar. The chemotactic response of *R. leguminosarum* to specific carbon sources was assayed using swarm plates comprised of Vincent's minimal medium (Vincent, 1970) with 0.15 % agarose and the potential chemoattractant as the sole carbon source (1 mM final concentration). Carbon sources were purchased from Sigma-Aldrich Canada (Oakville, Ontario), all other chemicals were obtained from BDH (Poole, England). When necessary, *Rhizobium* strains were cultured in media containing antibiotics at the following concentrations:neomycin, 100 $\mu g\ ml^{-1}$, spectinomycin, 500 $\mu g\ ml^{-1}$, tetracycline 5 $\mu g\ ml^{-1}$, streptomycin 500 $\mu g\ ml^{-1}$, and gentamycin, 30 $\mu g\ ml^{-1}$. Antibiotic concentrations used when culturing *E. coli* strains were as follows: ampicillin, 100 $\mu g\ ml^{-1}$, kanamycin, 50 $\mu g\ ml^{-1}$, tetracycline 10 $\mu g\ ml^{-1}$, streptomycin 500 $\mu g\ ml^{-1}$, gentamicin 15 $\mu g\ ml^{-1}$ and spectinomycin 100 $\mu g\ ml^{-1}$.

3.2 Visualization of *Rhizobium* plasmids.

When necessary the plasmids of VF39SM were visualized on agarose gels using a modified Eckhardt technique (Eckhardt, 1978) described by Hynes *et al.* (1985), as modified by Hynes & McGregor (1990).

Table 3-1. Bacterial strains and plasmids:

Strain or Plasmid	Relevant Characteristic(s)	Source or Reference
E. coli:		
DH5α	*endA1*, *hsdR17*, *supE44*, *thi-1*, *recA1*, *gyrA96*, *relA1*, Δ(*argF-lacZYA*), U169, ϕ 80d*lacZ* ΔM15	Gibco BRL.
HB101	F- *hsdS20* (r_B- m_B-) *leu supE44 ara14 galK2 lacY1 proA2 rpsl20 syl-5 mtl-1 recA13 mcrB*	Sambrook *et al.*, 1989
C2110	Naladixic acid resistant, *his*, *rha*, *polA1*	Kent, U of Wisconsin, Madison, Wisconsin
MT616	mobilizer strain	Finan *et al.*, 1985
RP8611	Δ(*tsr*) $_{7028}$ Δ(*tar-tap*) $_{5201}$ zbd::Tn5 Δ(*trg*) $_{100}$, *leuB6*, *his4*, *rpsL136*, (*thi1*, *ara14*, *lacY1*, *mtl1*, *xyl5*, *tonA31*, *tsx78*)	Parkinson, U of Utah, Salt Lake City, Utah
S17-1	Sp^r. RP4 *tra* region, mobilizer strain	Simon *et al.*, 1983.
R. leguminosarum		
VF39SM	biovar *viciae*, Sm^r	Priefer, 1989
LRS39201	VF39SM cured of pRleVF39b	Hynes & McGregor, 1990
LRS39301	VF39SM cured of pRleVF39c	Hynes & McGregor, 1990
LRS39401	VF39SM cured of pRleVF39d	Hynes & McGregor, 1990
LRS39501	VF39SM cured of pRleVF39e	Hynes & McGregor, 1990
LRS39601	VF39SM cured of pRleVF39f	Hynes & McGregor, 1990
TP6	VF39SM *fnrN*::Gm^r mutant	Patschkowski *et al.*, 1996
TP9	VF39SM *fixL*::Gm^r mutant	Patschkowski *et al.*, 1996
TP11	VF39SM *fixK* deletion mutant	Patschkowski *et al.*, 1996
VF-MCP1$^-$	VF39SM, *mcp-1* ::ΩTc, Tc^r	this work
VF-747-	VF39SM, *mcp-1* :: ΩNm, Nm^r	this work
VF-MCP2$^-$	VF39SM, *mcp-2* ::ΩSp, Sp^r	this work
VF-MCP3$^-$	VF39SM, *mcp-3* ::ΩNm, Nm^r	this work
VF-MCP4$^-$	VF39SM, *mcp-4* ::ΩSp, Sp^r	this work

VF-MCP2⁻/3⁻	VF39SM, *mcp-2* :: ΩSp, *mcp-3* :: ΩNm, Sp^r, Nm^r	this work
VF-MCPF4	VF39SM, *mcp-3* ::*lacZ*	this work
Plasmids		
pBSIISk+	cloning vector, amp^r	Gibco BRL
puc4	cloning vector, amp^r	Pharmacia
pJQ200mp18	suicide vector with *sacB* system, Gm^r	Quandt & Hynes, 1993
pJQ200SK	suicide vector with *sacB* system, Gm^r	Quandt & Hynes, 1993
pJQ200uc1	suicide vector with *sacB* system, Gm^r	Quandt & Hynes, 1993
pRK7813	broad host range cloning vector, Tc^r	Jones & Gutterson, 1987
pRK600	conjugation helper plasmid, Cm^r	Finan *et al.*, 1985
pRK602	Cm^r plasmid carrying Tn5 (Km^r/Nm^r)	Finan *et al.*, 1985
pBSIISK+	cloning vector,ampicillin resistant	Stratagene
pTn3PAR	Tn3 transposon carrying *par* locus, *tnpA* inactive, kanamycin resistance and ampicillin resistance encoded.	Weinstein *et al.*, 1992
pSShe	supplies *tnpA* activity in *trans*. Chlorophenicol resistant	Weinstein *et al.*, 1992
pZ1918	promoterless *lac* Z gene flanked by inverted repeats of the puc19 MCS, ampicillin resistant	Schweizer, 1993
p1918	a vector with a MCS composed of inverted repeats of the puc19 MCS, ampicillin resistant	Schweizer, 1993
pJQ200::Tc	ΩTc inserted into pJQ200uc1, ΩTc flanked by *Not*I sites, ampicillin resistant, spectinomycin resistant.	this work
p1918::Sp	Ωsp cassette cloned into p1918 via a *Bam* HI site	this work
MCP1.B	8 kb VF39SM *Bam* HI fragment cloned into pBSIISK + Contained a fragment of the *mcp1* gene.	this work
VGL-747	VF39SM cosmid clone in pRK7813. Cosmid contains the entire *mcp1* gene	this work
MCP1.E	1.8 kb VF39SM *Eco* RI fragment cloned into pBSIISK +	this work
MCP2.B	7.5 kb VF39SM *Bam* HI fragment cloned into pBSIISK+	this work
MCP2.E	3.6 kb VF39SM *Eco* RI fragment cloned into pBSIISK+	this work
MCP3.B	7 kb VF39SM *Bam* HI fragment cloned into pBSIISK +	this work
MCP3.P	2.5 kb VF39SM *Pst* I fragment cloned into pBSIISK +	this work

MCP4.B	5 kb VF39SM *Bam* HI fragment cloned into pBSIISK +	this work
MCP4.C	1.7 kb VF39SM *Cla* I fragment cloned into pBSIISK +	this work
MCP5.B	6.7 kb VF39SM *Bam* HI fragment cloned into pBSIISK +	this work
MCP5.PL	1.2 kb VF39SM *Pst* I fragment cloned into pBSIISK +	this work
MCP1ΩTc	ΩTc cassette cloned into internal *Not* I site of MCP1.E. Contained in pJQ200mp18.	this work
MCP2ΩSp	ΩSp cassette cloned into the internal *Xho* I site of MCP2.E. Contained in pJQ200SK.	this work
MCP3ΩNm	ΩNm cassette cloned into the internal *Hind* III site of MCP3.P. Contained in pJQ200SK.	this work
MCP4ΩSp	ΩSp cassette cloned into the internal *Sma* I site of MCP4.C. Contained in pJQ200SK.	this work
MCP-F1	a *mcp-3::lacZ* fusion cloned into pRK7813 containing the *par* locus	this work
MCP-F2	a *mcp-3::lacZ* fusion in which the *lacZ* gene is in the opposite orientation to the *mcp-3* promoter. Construct has been cloned into pRK7813 containing the *par* locus	this work
FusMCP2	pRK7813::*mcp-2* ::*lacZ*	this work

3.3 Bacterial matings.

Bacterial cultures were grown overnight in 5 ml of culture medium. Equal volumes (usually 0.75 ml) of donor and recipient were added to a microfuge tube. When a helper plasmid was necessary triparental matings were conducted using the strain MT616 carrying pRK600. The cells were pelleted by centrifugation (10 000 rpm for 5 minutes). After centrifugation the supernatant was removed and the cells were resuspended in 50 µl of sterile distilled water. The 50 µl drop of cell suspension was placed on a TY agar plate and incubated overnight at 28 °C. Following overnight incubation, the cells were scraped off the plate using a 200 µl pipette tip and resuspended in a microfuge tube containing 500 µl of water; 100 µl was removed, plated on the appropriate selection media, and incubated for 1 to 3 days.

3.4 Recombinant DNA techniques and molecular methodologies.

Isolation of plasmid DNA was done using the alkaline lysis procedure outlined in Sambrook *et al.* (1989). Genomic DNA was prepared using a protocol described by Meade *et al.* (1982).

All restriction endonucleases and modifying enzymes were purchased from Gibco BRL (Burlington, Ontario), and used according to the manufacturer's specifications. Preparation of competent cells and transformations were performed as described in Sambrook *et al.* (1989).

Agarose was purchased from Gibco BRL (Burlington, Ontario) and agarose gels were routinely electrophoresised (6 V/cm) in TBE buffer (Sambrook *et al.*, 1989) using a final agarose concentration of 0.9 % w/v. DNA fragments were isolated

from agarose gels using Prep-A-Gene glass matrix (BioRad, Hercules, California); following the manufacturer's protocol.

Gels used for Southern blotting were treated as follows: The DNA in the gel was depurinated by bathing the gel in 0.25M HCl for 12 min. The DNA was then transferred onto nylon membrane supplied by Boehringer Mannheim Canada (Laval, Québec) using a protocol described by Sambrook *et al.* (1989). The Southern blots were hybridized to non-radioactive DIG labelled DNA probes using reagents and protocols purchased from Boehringer Mannheim Canada (Laval, Québec). Hybridizations were performed overnight at 67 °C followed by washes of 2x SSC, 0.1 % SDS at room temperature, and 0.1x SSC, 0.1 % SDS at 67 °C. All washes were done in duplicate. Hybridization signals were detected via either chemiluminescence (using CSPD substrate) or colour detection (using X-phos/NBT), following a protocol supplied by Boehringer Mannheim Canada (Laval, Québec).

3.4.1 DNA Sequencing and Analysis.

DNA was prepared for sequencing using the alkaline lysis procedure outlined in Sambrook *et al.* (1989); an additional phenol/chloroform and chloroform extraction were added to the procedure. DNA sequencing was performed using an Applied Biosystems Inc. automated sequencer operated by The University Core DNA Services (The University of Calgary, Calgary, Alberta). Nested deletions created with *Exo* III nuclease were used to obtain the complete DNA sequences of *mcp-1*, *mcp-2* and *mcp-3*. Enzymes and protocols for the *Exo* III deletions were supplied by Pharmacia (Uppsala, Sweden). When necessary, primers ordered from Gibco BRL (Burlington, Ontario) were used to fill in gaps between contiguous stretches of DNA sequence. Primers were also used to sequence the opposite DNA strand. The primers were designed using the computer program Oligo (National Biosciences Inc., Minineapolis, U.S.A.).

Partial DNA sequences of *mcp-4*, and *mcp-5* were obtained from subclones cloned into pBSIISK+ and sequenced using T7 and T3 primers.

DNA sequence was analysed using DNAsis (Hitachi Software Engineering Co. Ltd.). Sequence homology searches were performed using the "BLAST" sequence alignment program (Altschul *et al.*, 1990).

3.5 Cloning of *mcp* genes in *R. leguminosarum* VF39SM.

Primers designed from the DNA sequence of *dcrA*, a *mcp*-like gene from *Desulfovibrio vulgaris*, (Dolla *et al.*, 1992) were supplied by G. Voordouw and were used to amplify VF39SM genomic DNA by PCR. The primer sequences used were: 5'-CATGGTCTTCTCGGCAGCTTGCG and 5'-GAATCGCCGACCAGACCAACC. PCR cycling conditions were: 1 minute at 94 °C, 30 seconds at 60 °C and 1 minute at 72 °C for 30 cycles. The amplified DNA fragment was labelled with DIG during the PCR amplification using a protocol specified by Boehringer Mannheim Canada (Laval, Québec). The DIG labelled product was gel purified using band interception, as described by Sambrook *et al.* (1989). After gel purification it was used as a probe (termed PCR probe) to identify *mcp* genes from a partial genomic library of VF39SM.

The partial genomic library was constructed by digesting approximately 5 μg of VF39SM genomic DNA with *Bam*HI, to completion. The digested DNA was run on a 0.8% agarose gel and the smear of cut DNA ranging in size from 5 kb to 12 kb was excised from the gel. The DNA fragments were purified from the gel using the Prep a gene technique described earlier. The fragments were ligated into pBSII SK+ digested with *Bam* HI. The ligation products were transformed into DH5α. The transformation mixture was then plated on LB, ampicillin plates containing X-gal at a final concentration of 0.04% v/v.

A putative *mcp* gene was isolated from the first round of screening of the library. The identity of this putative *mcp* gene was verified by sequencing the region which hybridized to the PCR probe. Upon verification, this *mcp* subclone was labelled via random primer labelling (Boehringer Mannheim Canada, Laval, Québec) and was used to re-probe the partial genomic library. This second round of screening revealed four previously unidentified putative *mcp* genes. Each fragment was subcloned and sequenced to identify regions of homology to *mcp* genes.

3.5.1 Identification of additional putative *mcp* genes in VF39SM.

C1 5'-AGGCGGACCGAGCAGCAGGC and C2 5'-CTTGATTTCCTTTGCCGC were oligonucleotides designed from the DNA sequence of *mcp-3*. C1 and C2 were used as primers to create a new probe from VF39SM genomic DNA amplified by PCR. PCR cycling parameters were those previously described. This probe (termed C12) was used later to probe a complete genomic library of VF39SM.

The complete genomic library was prepared by M. Berndt and P. Rochepeau (University of Calgary, Calgary, Alberta). Total genomic DNA from VF39SM was partially digested with *Sau*3AI to give a maximum number of digested fragments approximately 30 kb to 40 kb in length. The fragments were ligated to the dephosphorylated cosmid pRK7813. The ligation mixture was packaged *in vitro* using the Boehringer Mannheim Canada (Laval, Québec) packaging kit according to the manufacturer's recommendation. Titration of the genomic library yielded approximately 2500 CFU. A thousand independent tetracycline resistant clones were subjected to alkaline lysis minipreps and cosmid DNA was verified on 0.8% agarose gels. Cosmid DNA from each individually purified clone was then dot-blotted on a positively charged membrane (Boehringer Mannheim Canada, Laval, Québec) using a vacuum blotter. Three

membranes were used to blot the entire library of clones. Two membranes contained 400 clones each while the third membrane contained 200 clones. The membranes were dried for 30 min. at 120 °C and were then ready for the hybridization procedures.

3.6 Identification of putative MCP proteins from VF39SM.

An antibody supplied by M. Alam (University of Hawaii, Honolulu, Hawaii), created to detect MCP proteins in *H. salinarium* (Zhang *et al.*, 1996), was used to probe western blots containing whole cell lysate from VF39SM. The antibody used was raised against a 23 amino acid synthetic peptide, representing the highest conserved portion of the signalling domain amongst eubacterial MCPs (Zhang *et al.*, 1996). This antibody has been termed HC23 (Zhang *et al.*, 1996).

Whole cell lysates of VF39SM was prepared as follows: VF39SM was grown to stationary phase in 10 ml of TY broth. The culture was passed through a french press. Protein concentration of the resulting lysate was obtained using a Biorad protein assay, and the procedure outlined by Biorad (Hercules, California). When necessary, cell lysates were concentrated by spinning 1 ml of the lysate in a speed-vac for 4 hrs, after which another Biorad protein assay was performed.

10 μg of protein from the cell lysates was loaded onto a 8% acrylamide (0.1 % bisacrylamide) SDS PAGE gel and electrophoresed. The loading dye used to load the samples onto the gel is that described by Sambrook *et al.* (1989). After electrophoresis the gel was western blotted onto nitrocellulose paper purchased from Biorad (Hercules, California) using a semi-dry transfer apparatus, and protocol supplied by Pharmacia (Upusula, Sweeden). After transfer the nitrocellulose filter was stained using Ponceaus stain to confirm the transfer of proteins to the filter was successful. The blot was then probed with

the HC23 antibody using a chemiluminescence detection system supplied by Gibco BRL (Burlington, Ontario). The protocol outlined by Gibco BRL (Burlington, Ontario) was used in the detection procedure.

3.7 Insertional mutagenesis of VF39SM *mcp* genes.

mcp genes were mutated by inserting antibiotic resistance cassettes into the open reading frames (ORFs) of the cloned *mcp* genes. The antibiotic resistance cassettes used were those developed by Prentki & Krisch (1984) and Fellay *et al.* (1987). The gene replacement in VF39SM was performed using the suicide vector, pJQ200Sk+ created by Quandt & Hynes (1993). Appendix 2 provides a diagrammatic representation of the gene replacement strategy used, following the method developed by Quandt and Hynes (1993). Southern blotting was used to verify that the wild type gene was replaced by the mutated gene.

3.8 Phenotypic analysis of the *mcp* mutant strains.

3.8.1 Microscopic analysis.

A phase contrast microscope and 40X objective were used to confirm that the *mcp* mutants were motile and to observe their swimming behaviour. To prepare a wet mount for viewing, 10 µl of cells grown overnight in TY were placed on a slide and sealed with a coverslip. The cells were then visualized under the phase contrast microscope for motility.

3.8.2 Swarm plate analysis.

The swarming behaviour of the *mcp* mutants was investigated by inoculating the cultures in VMM swarm media. Inoculation was done using a three step technique: I) A fresh culture was streaked onto a TY plate and incubated for 2 days. II) A stab was taken from a single colony and inoculated into a YES swarm plate. The plate was incubated for 2 days. III) A stab inoculum was taken from the distal edge of the swarming colony and was stabbed into a VMM swarm plate. Inoculated swarm plates were incubated at 28 °C for 3 days. Following incubation the diameter of the swarming culture was measured and compared to a wild type control.

3.8.3 Nodulation competition experiments.

Pisum sativum cv. Trapper were surface sterilized by washing the seeds in 50% bleach for 5 minutes, followed by a second wash in 70 % ethanol. Following these two washes the seeds were rinsed 3 times in sterile distilled water. The seeds were germinated by placing them on water agar plates (12.5 g agar for 1 litre of distilled water) and incubating them at room temperature in the dark for 3 days. Seedlings were transferred to modified magenta jars which were designed to resemble Leonard Jars (Vincent, 1970). Appendix 3 is a schematic representation of the modified magenta jar used in this study. The peas were grown in a vermiculite substrate.

Once the peas were transferred to the magenta jars the seedlings were inoculated with various VF39SM strains. For competition experiments wild-type was co-inoculated with a particular *mcp* mutant in a one to one ratio. The ratios were confirmed by performing viable plate counts on the inoculum. The inoculated peas were then grown for 5 weeks, after which the nodules were

harvested, surface sterilized and then crushed. To eliminate potentially misleading results due to artificial conditions inherent in the nodulation assay care was taken to harvest nodules near the crown that were initiated before the roots became container bound. Nodules were surface sterilized by washing them in a 20% solution of bleach for 5 minutes, preceded by a 5 minute wash in 70% ethanol. The nodules were then rinsed twice in sterile distilled water. Surface sterilized nodules were placed individually in microfuge tubes containing 50 μl of sterile distilled H_2O and crushed using inoculating sticks. 5 μl of the macerate was plated onto the appropriate selective media to distinguish which strain had formed the nodule; the wild-type or the *mcp* mutant strain. For each competition experiment 100 nodules were crushed and plated, with the exception of the VF-MCP4$^-$ competition experiment where 63 nodules were crushed and plated.

3.9 Assaying for β-galactosidase activity in strains carrying *mcp* gene/*lacZ* fusions.

Promoterless *lacZ* gene cassettes were ligated to *mcp-1*, *mcp-2*, and *mcp-3*, downstream of their respective promoters. The constructs were placed either on the broad host range plasmid pRK7813 for mobilization into VF39SM and other derivative strains or into the VF39SM genome by homologous recombination.

β–galactosidase activity was quantitated using the protocol described by Miller (1972). *Rhizobium* cultures to be assayed were grown for 36 to 48 hours (late log phase) in TY broth or VMM broth supplemented with a desired carbon source. To study *mcp* gene expression in the bacteroid, fusion strains were allowed to nodulate Trapper peas using the methods described in the previous section. To assay for β-galactosidase activity from the bacteroids a procedure

adapted from Wang *et al.* (1989) was used. Briefly, 10 nodules were picked from a 4 week old nodulated plant. The nodules were placed in a microfuge tube kept on ice and washed with a ice cold solution containing 0.25 M mannitol and 0.05 M Tris-HCl pH 7.6. Following the wash step, the nodules were suspended in 500 µl of ice cold 0.25 M mannitol and 0.05 M Tris-HCl pH 7.6. The nodules were then crushed using a microfuge mortar and the plant debris was allowed to settle to the bottom of the tube. 100 µl of the resultant bacteroid suspension was removed and used in a β-galactosidase assay. The assay procedure described by Miller (1972) was used once again.

3.10 Protocol for identifying regulatory mutants in *mcp* gene expression.

A strain of VF39SM which carries an *mcp-3*::*lacZ* fusion on its chromosome was created by mating pJQ200::*mcp-3*::*lacZ* into VF39SM (see appendix 2 for protocol) and then selecting for single recombinants using Gm as the selectable marker. To obtain a double recombinant, thereby removing the pJQ200 vector from the VF39SM chromosome, a single recombinant was grown to late log phase and 100 µl of the culture was plated onto VMM plates containing 5% sucrose. Resultant colonies were screened for Gm sensitivity and β-galactosidase expression. A colony which was Gm^s but maintained β-galactosidase activity was saved for subsequent use and termed VF-MCPF4.

VF-MCPF4 was subjected to Tn5 mutagenesis by mating pRK602 into the strain. This plasmid carries the wild-type Tn5. Resultant mutants were selected for on plates containing Sm, Nm. To screen for mutants that no longer expressed the *mcp3*::*lacZ* the individual colonies were transferred, by sterile toothpick, to VMM plates containing X-gal. Colonies which were white rather than blue were selected as potential *mcp-3* regulatory mutants. To separate

mutants which were white due to Tn5 disruption of the *lacZ* and those that are bona-fide regulatory mutants the mutants were complemented with the plasmid MCP-F1. If the mutant expressed β-galactosidase activity after complementation it was assumed that the mutation was caused by a Tn5 insertion in the *lacZ* gene rather than a disruption of a regulatory gene.

4.0 Results

4.1 Cloning of putative *mcp* genes from *R. leguminosarum* VF39SM.

PCR reactions using VF39SM genomic DNA and the *dcrA* primers (Dolla *et al.*, 1992) resulted in the amplification of a single fragment approximately 350 bp in size (see figure 4-1). The 350 bp fragment was used to probe a Southern blot of VF39SM genomic DNA, digested with a variety of restriction endonucleases. The probe hybridized to numerous DNA fragments of *Bam* HI digested VF39SM genomic DNA, ranging in size from 5 kb to 12 kb (see figure 4-2). The next step was to clone the DNA fragments that hybridized to the PCR probe. Consequently, to enrich for fragments of 5 to 12 kb in size, the size-biased genomic library of VF39SM described in Materials and Methods was created. To identify clones containing VF39SM DNA inserts the library transformants were plated on plates containing X-gal (0.04% v/v). White colonies were presumed to contain inserts. Four hundred such colonies were included in the library. Five putative MCP clones were subsequently isolated from this library; termed *mcp-1* to *mcp-5*. The restriction maps of each clone are provided in figure 4-3.

Figure 4-1. A 1.2% agarose gel showing the PCR products formed after amplification of VF39SM genomic DNA with the *dcrA* primers. Lane 1 is Lambda digested with *Hind* III, lane 2 is VF39SM genomic DNA PCR amplified using *dcrA* primers, lane 3 is VF39SM genomic DNA amplified using nucleotides containing DIG-dUTPs. This DNA fragment was subsequently gel purified and used a probe.

Figure 4-1 PCR amplification of VF39SM genomic DNA with *dcrA* primers:

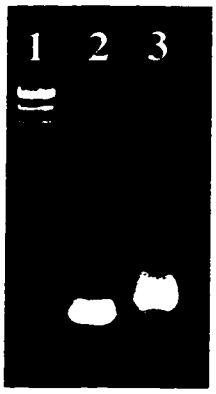
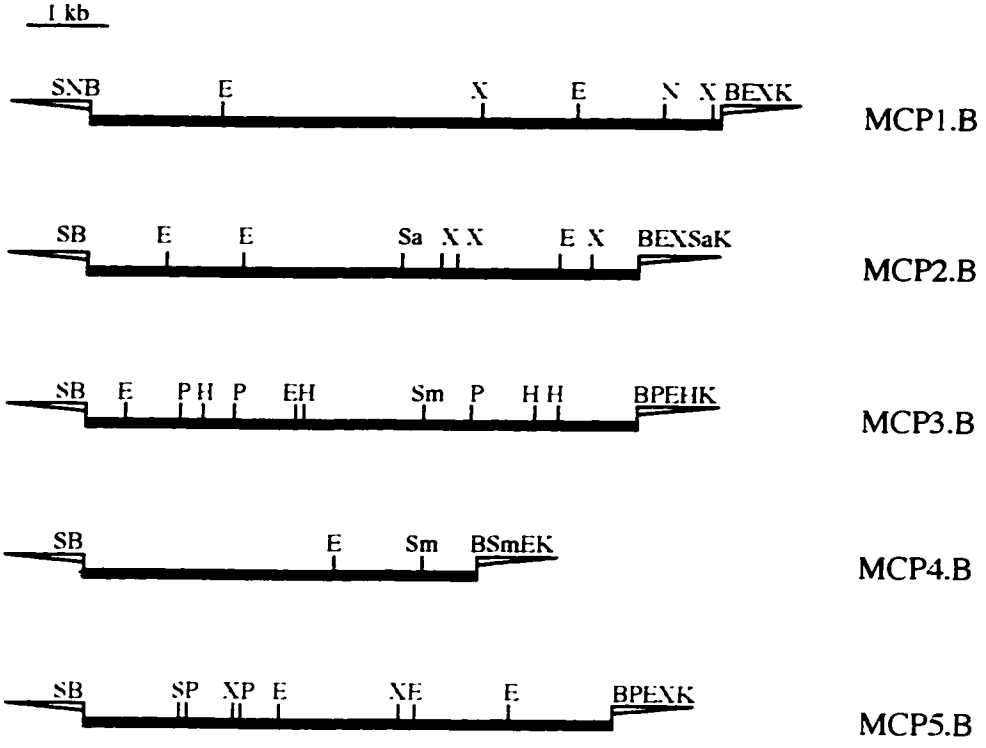


Figure 4-2. A Southern blot of VF39SM genomic DNA digested with *Bam*HI and probed with the *mcp* probe described in figure 4-1. Lane 1 is lambda DNA digested with *Hind* III; lane 2 is VF39SM genomic DNA digested with *Bam*HI. The blot was detected using colour detection as described in the Materials and Methods section.

Figure 4-2 Detection of *mcp*-like genes in VF39SM:

Figure 4-3. The restriction maps presented are of the original *Bam*HI fragments cloned from the size fractionated genomic library of VF39SM. Each fragment was cloned into pBSII SK+. Abbreviations for restriction endonucleases are as follows: B; *Bam*HI, E; *Eco*RI, H; *Hind*III, K; *Kpn*I, N; *Not* I, P; *Pst*I, S; *Sac*I, Sa; *Sal*I, Sm; *Sma*I, and X; *Xho*I.

Figure 4-3 Restriction maps of MCP1.B, MCP2.B, MCP3.B, MCP4.B, and MCP5.B:



4.2 DNA sequencing of the putative *mcp* genes: *mcp-1* to *mcp-5*.

The *Bam* HI fragments originally cloned from the VF39SM partial genomic library were subcloned in pBluescript to isolate smaller DNA fragments which hybridized to the PCR probe. The subclones were then sequenced using T7 or T3 primer. The resultant DNA sequences were subjected to BLASTX (Altschul *et al.*, 1990) homology searches. High homologies were reported to sequences of known MCPs. One of the highest alignment scores occurred with an MCP isolated from *Caulobacter crescentus*, McpA. Upon confirmation that each clone with homology to a *mcp* probe did in fact carry sequences highly related to *mcp* genes, further DNA sequencing was conducted.

The nucleotide sequence of *mcp-1* and its corresponding open reading frame have been deposited into the Genbank database under accession number AF022807. Figure 4-4 provides the DNA sequence of *mcp-1*. The *mcp-1* ORF codes for a protein 716 aa in length with a molecular mass of 76.7 Kda. Based on a Kyte Doolittle plot two transmembrane domains are present in the *mcp-1* ORF (see figure 4-7). The two transmembrane domains span from amino acid residues 18 to 37 and 298 to 317. The area flanked by the transmembrane domains is therefore 261 aa in size and is presumably located in the periplasm.

The *mcp-2* DNA sequence and corresponding open reading frame are listed in figure 4-5 and have been deposited into the Genbank database under accession number AF036168. *mcp-2* has two potential start sites for translation. Each start site is preceded by an equally likely Shine Dalgarno sequence. A Kyte Doolittle plot for both predicted ORFs is shown in figure 4-7. The hydropathy profile of MCP2 using the second start most closely resembles the profiles of MCP1 and MCP3, consequently the second start was chosen as the most probably start site. Based on this start site MCP2 is 626 amino acids in length with a molecular mass of 67.4 Kda. Two transmembrane domains

Figure 4-4. The *mcp-1* gene is 2142 nucleotides in length and has a G/C content of 63.64%. The predicted amino acid sequence of MCP1 is 714 aa in length and is listed in single letter code below the *mcp-1* nucleotide sequence. A putative Shine Dalgarno sequence (AAGAGG) is underlined. The sequence is nearly identical to the RBS consensus sequence (AGGAGGT) and its separation from the ATG start codon (3 nt) falls within the 3-9 nt range predicted for an RBS. The predicted transmembrane domains, and methylation domains of MCP1 are underlined.

Figure 4-4 *mcp-1* gene sequence and predicted protein sequence:

```
                9          18          27          36          45          54
5'  ACG TGA CCA TTC TCC AGA GCG CTG TCG GAC AGG GCG CTG CCG CTG TCG ACG CCG

               63          72          81          90          99         108
    CCG TCA AGT TGA TCC GCA AGG AGA AGG TGC CGC GCG ATT ACA AGG GTC CCT TCG

              117         126         135         144         153         162
    AAT TCG TCA CAC CTG AGA ACA TTG CCA CTT ATC TGC CGA AGA GCC AGT GAG CAT

              171         180         189         198         207         216
    AAG AGG ACT ATG ATG TTG CAT TTC TGG AAT AAA TTC GGC ATT CGC GCA CAG ATC
                --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
                 M   M   L   H   F   W   N   K   F   G   I   R   A   Q   I

              225         234         243         252         261         270
    ACC TCC GGC TTT GTG CCG CTG ATC CTG TTG ATG AGC CTG CTC ACG GTC AGC GCG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     T   S   G   F   V   P   L   I   L   L   M   S   L   L   T   V   S   A

              279         288         297         306         315         324
    ATC TCG GGT ATG AAC GGG CTC GCC TCG ATC TTC TCG TCC TAT CGC GCC ACG GCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   S   G   M   N   G   L   A   S   I   F   S   S   Y   R   A   T   A

              333         342         351         360         369         378
    GGC CAA AGT CTC GCC ATC TCG GAC TAT AGC GAC CAG CTG AAC GAG ATT CAG ATG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   Q   S   L   A   I   S   D   Y   S   D   Q   L   N   E   I   Q   M

              387         396         405         414         423         432
    TCG GCG GAA GCC TTC CGC TCC ACG CCG ACG CAG GCT GTC GTC GAC CGC TTC CGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     S   A   E   A   F   R   S   T   P   T   Q   A   V   V   D   R   F   R

              441         450         459         468         477         486
    GCC GGC GTG AAG GCG TTC GAC GCC GAT GAC CCG CGT TTT GCT GGC AAC AAG GAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   G   V   K   A   F   D   A   D   D   P   R   F   A   G   N   K   D

              495         504         513         522         531         540
    CTG CAG TCC GTT CTT GCG GCG ATC CGC CAG GAC ATC GCC ACT TAT GGC AAG GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   Q   S   V   L   A   A   I   R   Q   D   I   A   T   Y   G   K   A

              549         558         567         576         585         594
    TTC GAA GAA ATA GTT GCC CTG CAG GCC AGG CGT GAT GCC TTG ATC TCC AAG GTC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   E   E   I   V   A   L   Q   A   R   R   D   A   L   I   S   K   V
```

```
          603                 612                 621                 630                 639                 648
ACC GAA TTC GGC CCC TGG ACC AGT ATT GCG CTC AAC GAC GTC GTG CGC AGC GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   E   F   G   P   W   T   S   I   A   L   N   D   V   V   R   S   A

          657                 666                 675                 684                 693                 702
TGG CGC CAG AAC GAT GTG CCG CTG CTG CAG ATG ACG GCG GCG ACA CTG GAG GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 W   R   Q   N   D   V   P   L   L   Q   M   T   A   A   T   L   E   A

          711                 720                 729                 738                 747                 756
TTG AAC CGC AGC CTC TAT TTC TCC GAA CGC TTC GTG CAT TCC GAT GAT TTT GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   N   R   S   L   Y   F   S   E   R   F   V   H   S   D   D   F   A

          765                 774                 783                 792                 801                 810
GCC TAC GAC ACG GCG CAG GCA GCA CTG GCC GAA GCG GTC ACG CTC AAC GAA GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   Y   D   T   A   Q   A   A   L   A   E   A   V   T   L   N   E   A

          819                 828                 837                 846                 855                 864
GCC GCC AAG GCC GCG AAG AAC GAG CTG CAA AAG AAG CGC CTG ATG GGC GCC GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   A   K   A   A   K   N   E   L   Q   K   K   R   L   M   G   A   G

          873                 882                 891                 900                 909                 918
CAG CTG ATG CAG AAC TAC ACC GCC CGT CTC GGC GAC ATG AAG GAC GTG CTG CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   L   M   Q   N   Y   T   A   R   L   G   D   M   K   D   V   L   Q

          927                 936                 945                 954                 963                 972
GCC TCG GGC AAC ATC CGC CAG ACG CAG CTC AGC GTG CTC GCG CCG AAA ATC TCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   S   G   N   I   R   Q   T   Q   L   S   V   L   A   P   K   I   S

          981                 990                 999                1008                1017                1026
GGC GGC TTC AAG GAT CTG CAG GCG ACT GTT ACC GGT GCG CAG AAG ACC CTT GAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   G   F   K   D   L   Q   A   T   V   T   G   A   Q   K   T   L   D

         1035                1044                1053                1062                1071                1080
GGT TCG GTG GAC GCA ACG GTT GCC TCC GCG ACC AGC ACG ACG CTG ATC ATC AGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   S   V   D   A   T   V   A   S   A   T   S   T   T   L   I   I   S

         1089                1098                1107                1116                1125                1134
GGG CTG CTG ATC GTC ATT GGC CTC GTC CTT TCC TAT TTC GTC GGC CGG TTG ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   L   L   I   V   I   G   L   V   L   S   Y   F   V   G   R   L   I
```

```
        1143        1152        1161        1170        1179        1188
TCC TCG GCG GTG CGC AAC ATG GCC CAG TCC ATG GAG CAG CTT GCC CGT GGT GAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   S   A   V   R   N   M   A   Q   S   M   E   Q   L   A   R   G   E

        1197        1206        1215        1224        1233        1242
GAA AGA ATT GTG ATA ACG GGC GTC GAA CAT CGC CAC GAG CTG GGG GCC ATG GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   R   I   V   I   T   G   V   E   H   R   H   E   L   G   A   M   A

        1251        1260        1269        1278        1287        1296
CGT TCG CTG AAG GTT TTC CAG GAA ACG GGG CGC GCC AAG CTG ATC GCG GAA GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   S   L   K   V   F   Q   E   T   G   R   A   K   L   I   A   E   A

        1305        1314        1323        1332        1341        1350
AAT GCC GAA CGC GCC CGC CTG GCG GCC GAA GAA GAG CGG CTC CGC CAG GAG GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   A   E   R   A   R   L   A   A   E   E   E   R   L   R   Q   E   A

        1359        1368        1377        1386        1395        1404
GAG CGG CTC AGC GAC GCG CAG GTG ATG GAG CAT GCC TTC CGC CAG ATC TCA GTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   R   L   S   D   A   Q   V   M   E   H   A   F   R   Q   I   S   V

        1413        1422        1431        1440        1449        1458
GGG CTG GAC GCG CTC TCG AAG GGC GAC CTC ACG GTC CGC GTC GGC GAA GTC GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   L   D   A   L   S   K   G   D   L   T   V   R   V   G   E   V   D

        1467        1476        1485        1494        1503        1512
CAT CGC TAT GTC AGG ATC CGG GAT CAT TTC AAC AAC TCG GTC GCG AGC CTC GAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 H   R   Y   V   R   I   R   D   H   F   N   N   S   V   A   S   L   E

        1521        1530        1539        1548        1557        1566
GAG GCG GTC GAC GCC GTC ATT CGC GCG GTC GGC ACC ATC CGC TCC GGC CTT GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   A   V   D   A   V   I   R   A   V   G   T   I   R   S   G   L   A

        1575        1584        1593        1602        1611        1620
GAA ATC TCC ACC GCC TCC AAC GAT CTC GCC CGC CGC ACC GAG CAG CAG GCA GCT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   I   S   T   A   S   N   D   L   A   R   R   T   E   Q   Q   A   A

        1629        1638        1647        1656        1665        1674
TCG CTG GAG GAG ACC GTC GCG GCG CTG GGT GAA GTG ACC CGC GGC GTC AAT GGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   L   E   E   T   V   A   A   L   G   E   V   T   R   G   V   N   G
```

```
      1683        1692        1701        1710        1719        1728
ACG GCG GAG GGC GCA AGC CGC GCC CAG GGA GTC GTG GCG ACC GCC CGC ACC AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   A   E   G   A   S   R   A   Q   G   V   V   A   T   A   R   T   N

      1737        1746        1755        1764        1773        1782
GCG GAA AAG GGC GGC GAG ATC GTT GCC CGC GCC ATC GAT GCG ATG ACG GAA ATT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   E   K   G   G   E   I   V   A   R   A   I   D   A   M   T   E   I

      1791        1800        1809        1818        1827        1836
CAA AAT TCG TCG TCC AAG ATC GGC AAC ATC ATC AGC GTC ATC GAC GAG ATT GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   N   S   S   S   K   I   G   N   I   I   S   V   I   D   E   I   A

      1845        1854        1863        1872        1881        1890
TTC CAG ACC AAC CTG CTG GCG CTG AAC GCC GGC GTG GAA GCG GCG CGC GCC GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   Q   T   N   L   L   A   L   N   A   G   V   E   A   A   R   A   G

      1899        1908        1917        1926        1935        1944
GAG GCA GGC AAG GGC TTT GCC GTC GTC GCC CAG GAA GTC CGT GAA CTC GCC CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   A   G   K   G   F   A   V   V   A   Q   E   V   R   E   L   A   Q

      1953        1962        1971        1980        1989        1998
CGC TCA GCC AAT GCG GCA AGG GAG ATC AAG CAG CTG ATT TCT ACT TCC TCG GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   S   A   N   A   A   R   E   I   K   Q   L   I   S   T   S   S   A

      2007        2016        2025        2034        2043        2052
CAG GTC AAG ACC GGC GTC CAG CTG GTG GGC GAA TCC GGC CTC TCG CTC GAA CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   V   K   T   G   V   Q   L   V   G   E   S   G   L   S   L   E   Q

      2061        2070        2079        2088        2097        2106
ATT GTC GAG CAG GTC ACC GCC ATG AAT GCG ACC GTG GCC GAG ATC GCC GTT GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   V   E   Q   V   T   A   M   N   A   T   V   A   E   I   A   V   A

      2115        2124        2133        2142        2151        2160
GCC CGC GAG CAG GCG ACA AGC CTG CGC GAG GTC TCG GCT GCC GGC GAC CAG ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   R   E   Q   A   T   S   L   R   E   V   S   A   A   G   D   Q   M

      2169        2178        2187        2196        2205        2214
GAC AAG GTG ACG CAG CAG AAC GCC GCG ATG GTC GAG GAG ACC ACG GCG GCC GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   K   V   T   Q   Q   N   A   A   M   V   E   E   T   T   A   A   A
```

```
       2223        2232        2241        2250        2259        2268
CAG AGC CTG ACA CAT GAA ACC GAA AGC CTT GCC GAA TTG CTG CGG CGG TTC AGA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   S   L   T   H   E   T   E   S   L   A   E   L   L   R   R   F   R

       2277        2286        2295        2304        2313        2322
ACG GGC AGC GGC CGG GTA TCG GAA CAT CGC CAT TAC GCG ATG GCA TCC TGA CGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   G   S   G   R   V   S   E   H   R   H   Y   A   M   A   S   *

       2331        2340        2349        2358        2367        2376
CCT GCC GAT GCA GGC AAG AAA ACG CCG CCG CCT CGA AAG GGC GGT CGG TGA GCG

       2385        2394
GTC ATC GGC CAG CGT CGC CTC 3'
```

Figure 4-5. The *mcp-2* DNA sequence has two potential start codons, the second start site is indicated by bold print. Both start codons are preceded by potential RBS sequences, indicated as underlined sequence. A putative termination sequence is also underlined. Each putative RBS has homology with 5 of the 7 nucleotides of the consensus RBS sequence, AGGAGGT (see below). Based on the second start site *mcp-2* is 2142 nucleotides in length and has a G/C content of 58.1%. The predicted amino acid sequence of MCP2 is 626 aa in length and is listed in single letter code below the *mcp-2* nucleotide sequence. The predicted transmembrane domains, and methylation domains of MCP2 are underlined.

Figure 4-5 *mcp-2* gene sequence and predicted protein sequence:

```
                9          18          27          36          45          54
5'  GTG GCT GCT ATT CTC CCT GAA TTT TCA AGA ACT CAG GGT TCG TAA AGG CGT TTA

               63          72          81          90          99         108
    ATT GAT CGC GGA AGT ATT GGT CAA CTC AGA AGA TCT TCG CAA ATG AGG CGG GAC
                                                           --- --- --- ---
                                                            M   R   R   D

              117         126         135         144         153         162
    CAT TGT TTA GTT AAT ACT AAA TTA TGG AAT CCG GTG CGA GAT CTT ACA TCA GCC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     H   C   L   V   N   T   K   L   W   N   P   V   R   D   L   T   S   A

              171         180         189         198         207         216
    GGC ACG CCG CTC GTC CCA GCT ATC TTC GAG GAG ACC CGC ATG AAA CGC CCG AGT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     G   T   P   L   V   P   A   I   F   E   E   T   R   M   K   R   P   S

              225         234         243         252         261         270
    ATC AAG CAA GCC CTG ATC TTA AAG CTG TCG ATC ATC ACT CTG TTC ATG GTT AGC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   K   Q   A   L   I   L   K   L   S   I   I   T   L   F   M   V   S

              279         288         297         306         315         324
    CTG TCC TAT GTC TCG CTG AGC ACG ATT TCG ACG CTT CGT GCC AAT ACC GAG CAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     L   S   Y   V   S   L   S   T   I   S   T   L   R   A   N   T   E   Q

              333         342         351         360         369         378
    ATC GGG ACC TTC TGG ATG CAG CGG CTT GTG ACG GCT CGC GAG ATC AAG GAT GAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     I   G   T   F   W   M   Q   R   L   V   T   A   R   E   I   K   D   D

              387         396         405         414         423         432
    TTC CTC GAC CTG AAA CTG GTC TAT GCC CAG TAC CTC CTG GAG GAT ACG GCG GAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     F   L   D   L   K   L   V   Y   A   Q   Y   L   L   E   D   T   A   E

              441         450         459         468         477         486
    GAG CGG ATG ATC GGA CAA CAA AAA ATC GAA TCC GCC GGC GCC GCG GTC GAG AAA
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   R   M   I   G   Q   Q   K   I   E   S   A   G   A   A   V   E   K

              495         504         513         522         531         540
    GTC GTC ACC GAA TAC GAA AGG GGT GTC CGC ACC GAG CGT GGA CGC GAA CTG ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   V   T   E   Y   E   R   G   V   R   T   E   R   G   R   E   L   I
```

```
          549        558        567        576        585        594
AAC CAG ATG AAG CCG GAA CTC GCC AAA TAT CGC GCG CTG GCA GAG CAA ATG ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   Q   M   K   P   E   L   A   K   Y   R   A   L   A   E   Q   M   I

          603        612        621        630        639        648
GCG CTT GAA AAT GAC GGG AAG ACG CCT GAA GCA ATC CGT CTT TTC AAG GAA AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   L   E   N   D   G   K   T   P   E   A   I   R   L   F   K   E   N

          657        666        675        684        693        702
ATG GAG CCA CAA GCC GAG CTG GTG AAC AAG GCG GTG GCG GAT CTG GTC ACT TTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 M   E   P   Q   A   E   L   V   N   K   A   V   A   D   L   V   T   F

          711        720        729        738        747        756
ATT CTC AGC CAG GCC GAA TGC TTT GTG GCC GCG AGC GGT GCT TCC GCG CAA TCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   L   S   Q   A   E   C   F   V   A   A   S   G   A   S   A   Q   S

          765        774        783        792        801        810
GCT TTC ATG CTG ACG GCC GCG ATC GCA GCG CTG GCC GTG CTT CTT GCC GTA GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   F   M   L   T   A   A   I   A   A   L   A   V   L   L   A   V   A

          819        828        837        846        855        864
GGA ATC TTA TTT GCG ATA TCG GGC ATC GCC AAC CCA ATC CGA AGC ATC GCC TCA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   I   L   F   A   I   S   G   I   A   N   P   I   R   S   I   A   S

          873        882        891        900        909        918
GCC ATG AGG CGC TTG TCG GAT GGC GAT CTT GAC AGC GAT ATT CCC TAT GCC GGT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   M   R   R   L   S   D   G   D   L   D   S   D   I   P   Y   A   G

          927        936        945        954        963        972
CGC GCC GAC GAA GTC GGC GAA ATG GCC GGC GCG GTT GAA ATC TTT CGT CAG AAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   A   D   E   V   G   E   M   A   G   A   V   E   I   F   R   Q   N

          981        990        999       1008       1017       1026
GCT CTC AAC GTT GTC AGG CTC GAG AAG AAT CCG CTG AAT CCC GCA GCG AGA GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   L   N   V   V   R   L   E   K   N   P   L   N   P   A   A   R   A

        1035       1044       1053       1062       1071       1080
ATG CAG CGC GCG CCG CAC CCC AGC AGC GCG CCG AAC GCG AGG CGG GAA CAA TTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 M   Q   R   A   P   H   P   S   S   A   P   N   A   R   R   E   Q   L
```

```
      1089       1098       1107       1116       1125       1134
CGC TTC GCG ACC ACG ACA TTG GGC GAA GGT CTC CGG CGG CTT GCA TCA GGT GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   F   A   T   T   T   L   G   E   G   L   R   R   L   A   S   G   D

      1143       1152       1161       1170       1179       1188
ATA TCC TTC CAG CTT TCG GAG CAA TTT GCG GCC GAA TAC GAA GCC TTG CGC GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   S   F   Q   L   S   E   Q   F   A   A   E   Y   E   A   L   R   E

      1197       1206       1215       1224       1233       1242
GAC TTC AAC GCT TCG CTC CGG CAA TTG GGC GCG ACG ATC GGC GCA GTG CTC CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   F   N   A   S   L   R   Q   L   G   A   T   I   G   A   V   L   Q

      1251       1260       1269       1278       1287       1296
ACG GTA TAC AGC ATC GAT AAT GGT ACT GGT GAG ATT GCA TCT GCC GCG CAG GAT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   V   Y   S   I   D   N   G   T   G   E   I   A   S   A   A   Q   D

      1305       1314       1323       1332       1341       1350
CTT TCC AAG CGT ACC GAA CAA CAA GCC GCC TCT CTC GAG GAG ACG GCC GCA GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   S   K   R   T   E   Q   Q   A   A   S   L   E   E   T   A   A   A

      1359       1368       1377       1386       1395       1404
CTG GAA GAG ATC ACG TCG AAT GTG ACG ATG GCG ACC AAA CGC ACC GAC GAG GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   E   E   I   T   S   N   V   T   M   A   T   K   R   T   D   E   A

      1413       1422       1431       1440       1449       1458
CGC AAT GTC GCC AAA GAA GCC GAT ATC AGT GCT CAG CGG TCG GCA GCG GTC GTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   N   V   A   K   E   A   D   I   S   A   Q   R   S   A   A   V   V

      1467       1476       1485       1494       1503       1512
TCG CAG GCG GAA GGG GCC ATG CGA CGC ATC GAG GAC AGT TCA CAG CAG ATT TCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   Q   A   E   G   A   M   R   R   I   E   D   S   S   Q   Q   I   S

      1521       1530       1539       1548       1557       1566
AAC ATC ATC GGT GCA ATT GAT GAA ATC GCC TTT CAG ACG AAC CTC CTG GCG CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   I   I   G   A   I   D   E   I   A   F   Q   T   N   L   L   A   L

      1575       1584       1593       1602       1611       1620
AAT GCC GGC GTT GAG GCT GCC CGT GCG GGT GAG GCG GGC AAG GGT TTT GCC GTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   A   G   V   E   A   A   R   A   G   E   A   G   K   G   F   A   V
```

```
      1629        1638        1647        1656        1665        1674
GTC GCC CAG GAA GTC CGT GAG CTC GCC CAG CGA GCC GCT CAA GCG GCC AAG GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   A   Q   E   V   R   E   L   A   Q   R   A   A   Q   A   A   K   E

      1683        1692        1701        1710        1719        1728
ATC AAG GGC TTT ATA CAA AAG TCA TCA GCT GAT GTG GAA AAT GGC GTG AAA CTG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 I   K   G   F   I   Q   K   S   S   A   D   V   E   N   G   V   K   L

      1737        1746        1755        1764        1773        1782
GTT CTT GAA ACC GGA ACC TCG CTC AAG TCG ATC GGT GAG TAC GTT GTC CAT ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   L   E   T   G   T   S   L   K   S   I   G   E   Y   V   V   H   I

      1791        1800        1809        1818        1827        1836
AAC CAA CTC ATG GAT GCG ATT GCC ACA TCG GCG CGT GAG CAG TCG ACT GGA CTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   Q   L   M   D   A   I   A   T   S   A   R   E   Q   S   T   G   L

      1845        1854        1863        1872        1881        1890
GCC GAG ATC AAC ACG GCC GTC AAT CAA ATG GAC CAG GCG ACC CAG CAA AAC GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   E   I   N   T   A   V   N   Q   M   D   Q   A   T   Q   Q   N   A

      1899        1908        1917        1926        1935        1944
GCA ATG GTC GAG CAG TCG ACG GCT GCT GTT GCC TCA TTG TCC TCC GAG GCG GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   M   V   E   Q   S   T   A   A   V   A   S   L   S   S   E   A   G

      1953        1962        1971        1980        1989        1998
CGC CTG CGG GAT CTG GTC AAC CAG TTT CAA TTG GAC GGC GAC AAA AGT GCC GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   L   R   D   L   V   N   Q   F   Q   L   D   G   D   K   S   A   A

      2007        2016        2025        2034        2043        2052
GAC GGG CAG CGC AGC GGG CGG GCC TTT GAA GGC AAT AGG CCA ATC CAC TTG GTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   G   Q   R   S   G   R   A   F   E   G   N   R   P   I   H   L   V

      2061        2070        2079        2088        2097        2106
GCT TCG CGG CGC GTG ACG CAA AGA TGA CGC GCG GCT GAC GTC AGG TCA GAT ACC
--- --- --- --- --- --- --- --- ---
 A   S   R   R   V   T   Q   R   *

      2115        2124        2133        2142        2151        2160
GAA AGA CAT TGA CCT GGC GCC TGG GCG ATC ATC GGC TCG ATC CCT CCA AAG GCA

      2169        2178        2187        2196        2205        2214
GTT GTC CCG CAT CAC NTC GTC GCA AGG CAT CAT GAG CGG CAG CGA TAT TCT CAT

      2223        2232        2241        2250        2259        2268
ATC ATG CCG CAT CGT GTA CGG TCG CGG GAT CTG CGG ACC TAT GGG AAA CTA AGT
```

```
      2277        2286        2295        2304        2313        2322
GCG ATC GAT GAA GAA GAC CTT CAA TCC GGC CTC GGT GCG CCG GCA TTT TGG AAA

      2331        2340        2349        2358        2367        2376
TTA CAA TCA CGG CCT GCT CGT GCC GCC CGG CGC ATC GCT CCT GGT CAC GTC GGG

      2385        2394        2403        2412        2421        2430
CCA GCT TGG TAT CGG TCT CGA CGA TGT TGT TCC GAG CGA CAT CAC CGC GCA GGC

      2439        2448        2457        2466        2475        2484
AGA GCT CTG CTT CGA GGC GAT CAA GGC GAT CCT CGA AGA GGC GGA GAT GAG TTT

      2493        2502        2511        2520        2529        2538
TGC CGA TGT CAT CCG CAT TTC GGG CTT CGT CAC CCG CCG CGA GGA TTT CCC GGC
```

(residues 9-29 and 192-212) are predicted, thereby forming a periplasmic N-terminus that is 163 amino acids in length.

The *mcp-3* gene sequence and predicted open reading frame have been deposited into the Genbank database under accession number U81828 and are listed in figure 4-6. The potential translational start site is not preceded by a particularly strong Shine Dalgarno sequence, however, the gene is expressed in VF39SM as indicated by a promoter-less *lacZ* fusion to *mcp-3* (described in later sections). The predicted protein based on the start codon is 624 amino acids in size and has a molecular mass of 66.3 Kda. A Kyte Doolittle plot predicts two transmembrane regions (see figure 4-7), the first from residues 14 to 32, the second transmembrane domain occurs from residues 179 to 201. Based on these observations, a periplasmic domain of approximately 147 aa should exist in the *mcp-3* ORF.

The high G/C content observed in *mcp-1*, *mcp-2*, and *mcp-3* is consistent with bias to codons with a G+C at the third nucleotide position observed in other rhizobial spp.

The amino acid sequence homology of MCP1, and MCP3 to known MCP proteins is illustrated in figure 4-8. An alignment of these ORFs to the *C. crescentus* McpA sequence (Alley *et al.*, 1992), the *R. leguminosarum* McpA (Brito *et al.*, 1996), and the *E. coli* Tsr sequence (Boyd *et al.*, 1983) indicate that the highest levels of homologies occur within the methylation and signalling domains of the MCPs.

The amino acid sequences predicted from DNA sequences of *mcp-2*, *mcp-4*, and *mcp-5* also show high homology to the C-terminus of known MCP proteins. Figure 4-9 illustrates this point by providing BLASTX alignments of *mcp-2*, *mcp-4*, and *mcp-5* to *C. crescentus* McpA.

Figure 4-6. The *mcp-3* gene is 1872 nucleotides in length and has a G/C content of 61.1%. The predicted amino acid sequence of MCP3 is 624 aa in length and is listed in single letter code below the *mcp-3* nucleotide sequence. A putative Shine Dalgarno sequence (AGGCGC) is underlined. The sequence has weak homology to the RBS consensus sequence (AGGAGGT) but its separation from the ATG start codon (7 nt) does fall within the 3-9 nt range predicted for RBS. The predicted transmembrane domains, and methylation domains of MCP1 are underlined.

Figure 4-6 *mcp-3* gene sequence and predicted protein sequence:

```
                9          18          27          36          45          54
5'  CGG CCG CAA GCG CCG GCA TGC CGA GCC CGA AGA GGC CGT GGC GGA AAA TAG CTG

               63          72          81          90          99         108
    ATC TCC CTG GCG CGA TGG AAG CAC TTT TCC CAC AGC GGT TCG CAA AGC TCG ACG

              117         126         135         144         153         162
    CAG GCG CGC CAG ACT GAA ATG ACC CGC GTG TTC TCC GGC ATA CCA TTC TAC TTT

              171         180         189         198         207         216
    ATG ATT TAT TCA CAT ATC AAA TAC ACA CTG CGT TTG CGC GTT TTG GGT TGC AAT

              225         234         243         252         261         270
    CCG GAC ATG GGC ATG AGG GCT CGT GAT GTG ACC GTC AGC GCG GAT GCC CGT ATT

              279         288         297         306         315         324
    GAG ACG CCG GTC GGG AAC ATG TCC GAC TGA TCC GCA ATT CTA GGC CGC CGC GGA

              333         342         351         360         369         378
    ATG TTC CAA TTT TTG AAA ACG ATG CCT CTG ACA GCC AAG CTG GCG GCG ATT ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     M   F   Q   F   L   K   T   M   P   L   T   A   K   L   A   A   I   I

              387         396         405         414         423         432
    GTT GCC GTC AAT CTC TGC GGC ATT TCC GCT TTC GCC ACC TAT ACC TGG ATG TAC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     V   A   V   N   L   C   G   I   S   A   F   A   T   Y   T   W   M   Y

              441         450         459         468         477         486
    GAA ACC CGG GCA TTG ATC GAT GGC GCC AAG GCG AAC TGG TCC AAG GAT GCG GAG
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     E   T   R   A   L   I   D   G   A   K   A   N   W   S   K   D   A   E

              495         504         513         522         531         540
    CAG TGT GCA TCT CTG GCC GCC GGC GGC GTG AAA TGG GGG AAG GCG AAC GCC GCT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     Q   C   A   S   L   A   A   G   G   V   K   W   G   K   A   N   A   A

              549         558         567         576         585         594
    CGA AAG GCT TAT TCG CTC TAC CGC GAC GAC CCC TCG CTC GAT CTT GTG CAG TTT
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     R   K   A   Y   S   L   Y   R   D   D   P   S   L   D   L   V   Q   F

              603         612         621         630         639         648
    GCC GCA TTC AAC GCC GAA CCT GCC GCC GTC GAT ACA TGG ACA CGC GAC GGC ATC
    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
     A   A   F   N   A   E   P   A   A   V   D   T   W   T   R   D   G   I
```

```
          657         666         675         684         693         702
AGC GGT TTG CCC ACA CCG GGA GAC CTG GCG ACG CGC CTT AGT GCG AAG CCG GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   G   L   P   T   P   G   D   L   A   T   R   L   S   A   K   P   E

          711         720         729         738         747         756
AAA ACG ACG ATT GAT GAC AGT GGA ATA TCC GCT GGC GTG GTC ACG ATC ATC GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   T   T   I   D   D   S   G   I   S   A   G   V   V   T   I   I   A

          765         774         783         792         801         810
CCG CTT CCG CTG GAC AAG TCA GGC AAG GCC GCC GGT TAC ATC GTC ACG AAT TGG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 P   L   P   L   D   K   S   G   K   A   A   G   Y   I   V   T   N   W

          819         828         837         846         855         864
TCT GTC GAA AAA ATC GCT GCC GAA GTC AGG CAG AAG GTT CTC ATT TCG CTG CTC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   V   E   K   I   A   A   E   V   R   Q   K   V   L   I   S   L   L

          873         882         891         900         909         918
ACG CAG TCC GTG ATC ACC GCC TTG GCC GTC ATC GCC TTC CTT CTC GCC ATG CGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 T   Q   S   V   I   T   A   L   A   V   I   A   F   L   L   A   M   R

          927         936         945         954         963         972
AGC CTG GTC GGC CGG CCC ATC AGG GTG ATC AGC GAA CGA ATC AGC GCG TTG CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   L   V   G   R   P   I   R   V   I   S   E   R   I   S   A   L   Q

          981         990         999        1008        1017        1026
AAA GGC GAT CTG GCC TCT CCT GTC ACC TAT AAG GAA AAT GGC GAC GAG ATC GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   G   D   L   A   S   P   V   T   Y   K   E   N   G   D   E   I   G

         1035        1044        1053        1062        1071        1080
TTT CTG GCG CGC GCG TTG GAA GTT TTC CGT CAT GAA GCG ATC GCG AAG GTC GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 F   L   A   R   A   L   E   V   F   R   H   E   A   I   A   K   V   E

         1089        1098        1107        1116        1125        1134
AGA GAG CAG GCC GCT GCC GAG CAG AGC GCT TCG CTC GAC GCC GAA CGG GCG CGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   E   Q   A   A   A   E   Q   S   A   S   L   D   A   E   R   A   R

         1143        1152        1161        1170        1179        1188
AAC GCA TTG TTC ACG GAA GAG GCC AGC AAC ACC CAG CGG CTG GTC ATG ACC GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   A   L   F   T   E   E   A   S   N   T   Q   R   L   V   M   T   A

         1197        1206        1215        1224        1233        1242
CTT GCA AAC TCA CTG GAA AAG CTT GCC GCA GGC GAC TTC TCG ATA CAC CTG GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   A   N   S   L   E   K   L   A   A   G   D   F   S   I   H   L   A
```

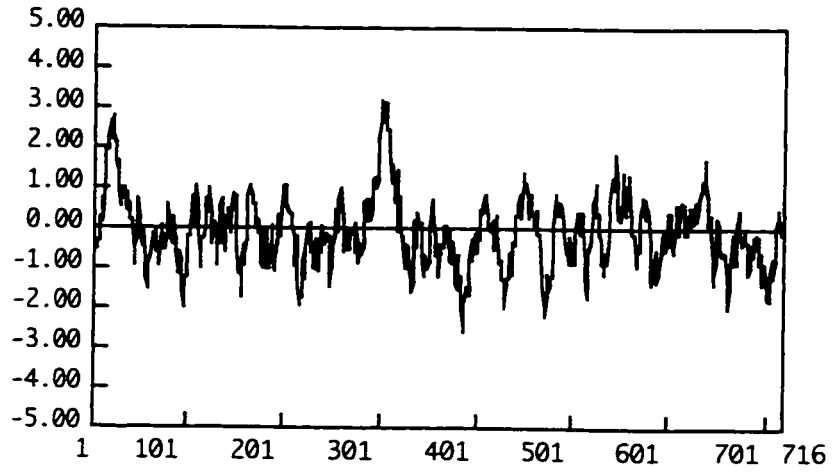
```
      1251        1260        1269        1278        1287        1296
GAT CTC GGT CCT GAA TTC GAT AAA TTG CGG CAG GAT TTC AAC AAC ATG GTC GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   L   G   P   E   F   D   K   L   R   Q   D   F   N   N   M   V   E

      1305        1314        1323        1332        1341        1350
GCG GTC GCG GCT GCG CTG ACA GAG ATC AAG ACC GCC TCG GTC GCG GTC GAA GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   V   A   A   A   L   T   E   I   K   T   A   S   V   A   V   E   G

      1359        1368        1377        1386        1395        1404
GGG TCG AGC GAG CTG GCA TCC TCC GCC GAT CAA CTC GCT AGG CGG ACC GAG CAG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 G   S   S   E   L   A   S   S   A   D   Q   L   A   R   R   T   E   Q

      1413        1422        1431        1440        1449        1458
CAG GCG GCA GCA TTG GAA CAG ACC GCC GCG GCA CTG GAT GAG GTG ACC ACC ACG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 Q   A   A   A   L   E   Q   T   A   A   A   L   D   E   V   T   T   T

      1467        1476        1485        1494        1503        1512
GTC AGA ACA TCG TCG CAG CGA GCC GAA AAT GCC GGC AAG CTG GTC GAG GAA ACC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   R   T   S   S   Q   R   A   E   N   A   G   K   L   V   E   E   T

      1521        1530        1539        1548        1557        1566
AAG CGG AGC GCT CAT GTC TCG GCA ACG GTG GTG CGT GAT GCA ATC GGA GCG ATG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 K   R   S   A   H   V   S   A   T   V   V   R   D   A   I   G   A   M

      1575        1584        1593        1602        1611        1620
GAC CGG ATT CAG ACC TCG TCG AGT CAG ATC GGC CGC ATC ATC GGC GTC ATC GAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 D   R   I   Q   T   S   S   S   Q   I   G   R   I   I   G   V   I   D

      1629        1638        1647        1656        1665        1674
GAA ATC GCC TTC CAG ACG AAC CTG CTG GCG CTG AAT GCC GGC GTC GAG GCG GCG
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 E   I   A   F   Q   T   N   L   L   A   L   N   A   G   V   E   A   A

      1683        1692        1701        1710        1719        1728
CGC GCC GGT GAG GCC GGC AAG GGT TTT GCG GTT GTC GCG CAG GAA GTG CGT GAA
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   A   G   E   A   G   K   G   F   A   V   V   A   Q   E   V   R   E

      1737        1746        1755        1764        1773        1782
CTC GCC CAG CGG TCG GCA AAT GCG GCA AAG GAA ATC AAG AAC CTG ATC AAC GTT
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   A   Q   R   S   A   N   A   A   K   E   I   K   N   L   I   N   V
```

```
      1791        1800        1809        1818        1827        1836
TCC GGC CAG GAA GTT GCC GCG GGC GTC GGG CTG GTG AAC GAA ACC GGC GAC GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 S   G   Q   E   V   A   A   G   V   G   L   V   N   E   T   G   D   A

      1845        1854        1863        1872        1881        1890
TTG CTG AAG ATC GAG GAG CAG ATC AAC CGC ATC AGC GAC AGT ATC GCT TCC ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 L   L   K   I   E   E   Q   I   N   R   I   S   D   S   I   A   S   I

      1899        1908        1917        1926        1935        1944
GTC CAG TCC TAT CGC GAA CAA GCG ACA GGT CTG CAG GAA ATC AAC GGC GCG ATC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 V   Q   S   Y   R   E   Q   A   T   G   L   Q   E   I   N   G   A   I

      1953        1962        1971        1980        1989        1998
AAC CAG ATG GAT CAG GCG ACA CAG CAG AAC GCG GCA ATG GTC GAG GAA ACG AAC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 N   Q   M   D   Q   A   T   Q   Q   N   A   A   M   V   E   E   T   N

      2007        2016        2025        2034        2043        2052
GCG GCC TGC CAG GAA CTG CTG CAG CAG GGA CGC CTT CTG CAG GAC TCG GCC GGC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 A   A   C   Q   E   L   L   Q   Q   G   R   L   L   Q   D   S   A   G

      2061        2070        2079        2088        2097        2106
AGG TTC GTC GTC GGC GCG TCT TCA GCC AGC CAG CCC AGA CCC ATG CAA CCC GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   F   V   V   G   A   S   S   A   S   Q   P   R   P   M   Q   P   A

      2115        2124        2133        2142        2151        2160
CGC CAA TCT CCT CCC GAG CCC AGA GCC TTC GTG CAG CGG CAT GCA GGA AAT GCC
--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 R   Q   S   P   P   E   P   R   A   F   V   Q   R   H   A   G   N   A

      2169        2178        2187        2196        2205        2214
GCC GTC GCC GCT GCT CCC GGT GCC TGG GAG GAG TTC TAA CGT CAT CCT CGT TCG
--- --- --- --- --- --- --- --- --- --- --- --- ---
 A   V   A   A   A   P   G   A   W   E   E   F   *

      2223        2232        2241        2250        2259        2268
ATA AGC CTT ATG CCG CAG CAA CCC ATA ATC AGG AAG GAA ACA CTA TGA AGA AAA

      2277        2286        2295        2304        2313        2322
TCG TGC CCG CAT TTC TTT TGG CCT GCA CCG CAT TTG CCG TGC CCA TGG GCG CAT

      2331        2340        2349        2358
CCA TGG CAC AGG ATG CCA AGC TTG CCC CGA TCT TCG ACT 3'
```

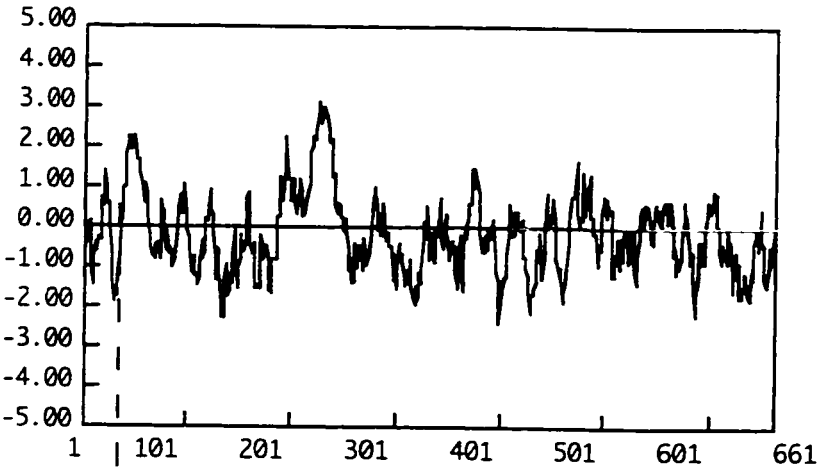
Figure 4-7. Kyte Doolittle hydropathy plots of MCP1, MCP2, and MCP3. The scale on the Y-axis indicates degree of hydrophobicity whereby values above 0.00 increase with increasing degree of hydrophobicity. The X-axis indicates amino acid residue number. The dotted line and * symbol indicate the position of the second start site in *mcp-2*.

To refine the exact locations of the TM regions a TmPred algorithm was used on each DNA sequence.

Figure 4-7 Hydropathy plots of MCP1, MCP2, and MCP3:

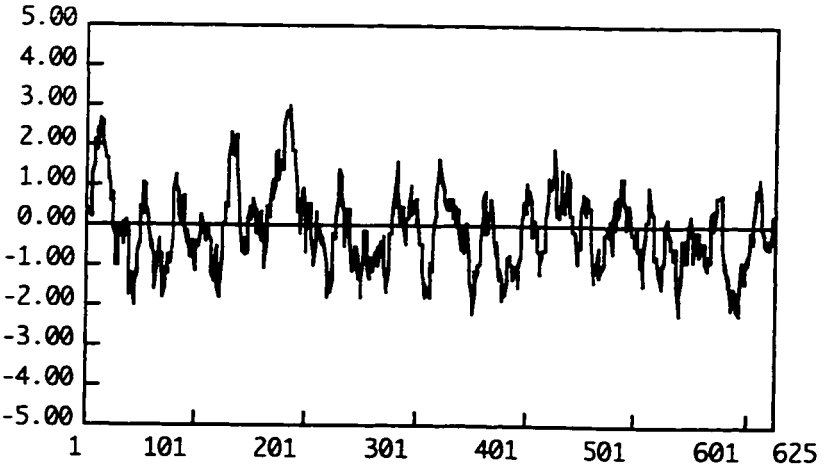


MCP1 Hydrophobicity plot



*

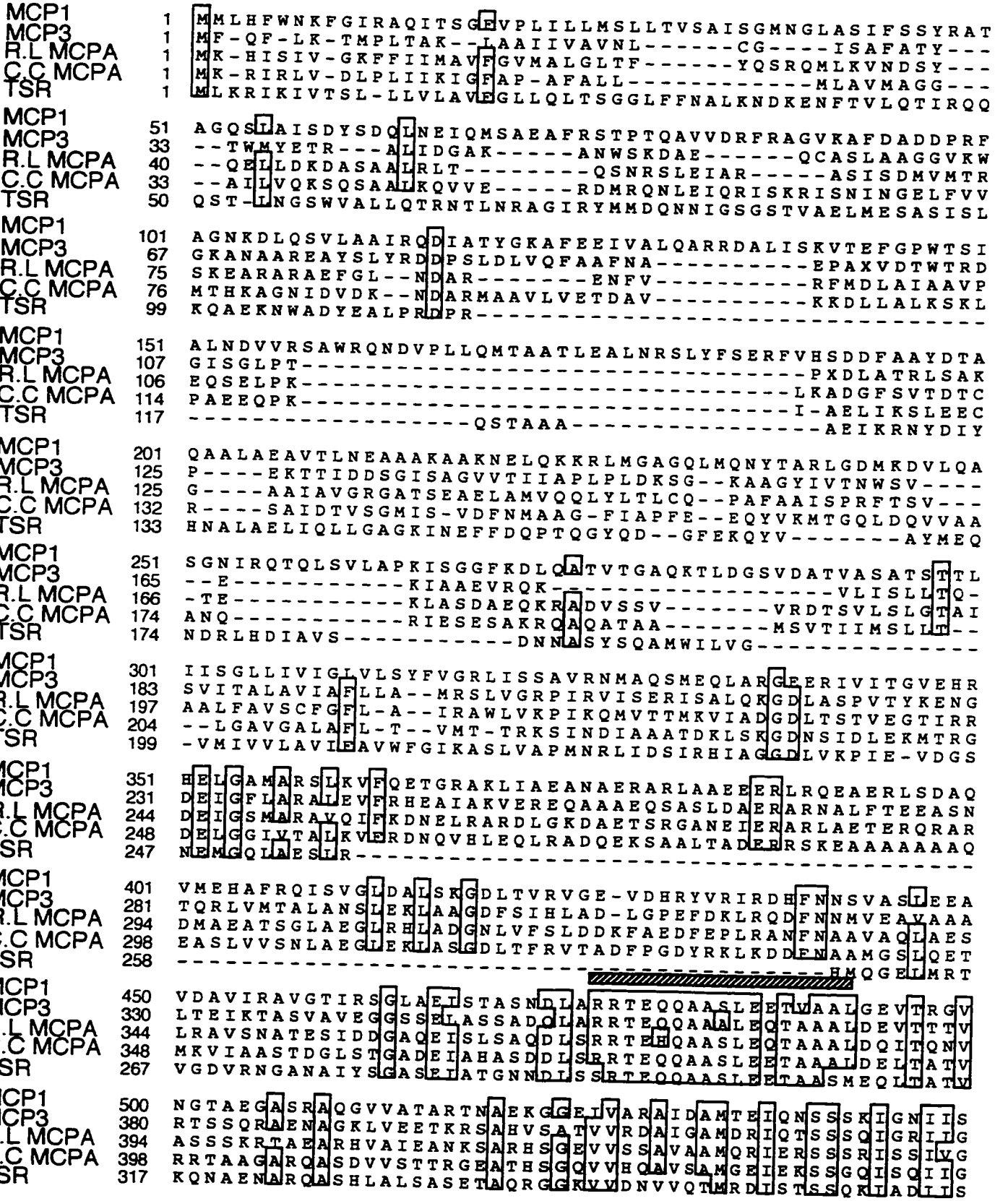
MCP2 Hydrophobicity plot



MCP3 Hydrophobicity plot

Figure 4-8. Amino acid sequence alignment of MCP1 and MCP3 to *R. leguminosarum* McpA, *C. crescentus* McpA, and *E. coli* Tsr. The hatched rectangular boxes indicate the methylation regions of each MCP. The boxed residues indicate identical amino acids amongst the 5 chemotaxis proteins. MCP1 and MCP3 have high homology to the methylation domains and signalling domain of both of the McpA proteins and of Tsr. The sequence alignment was performed using the ClustalV program, and displayed using the SeqVu program (Garvan Institute).

Figure 4-8 ClustalV alignment of MCP1 and MCP3 to other MCP proteins:



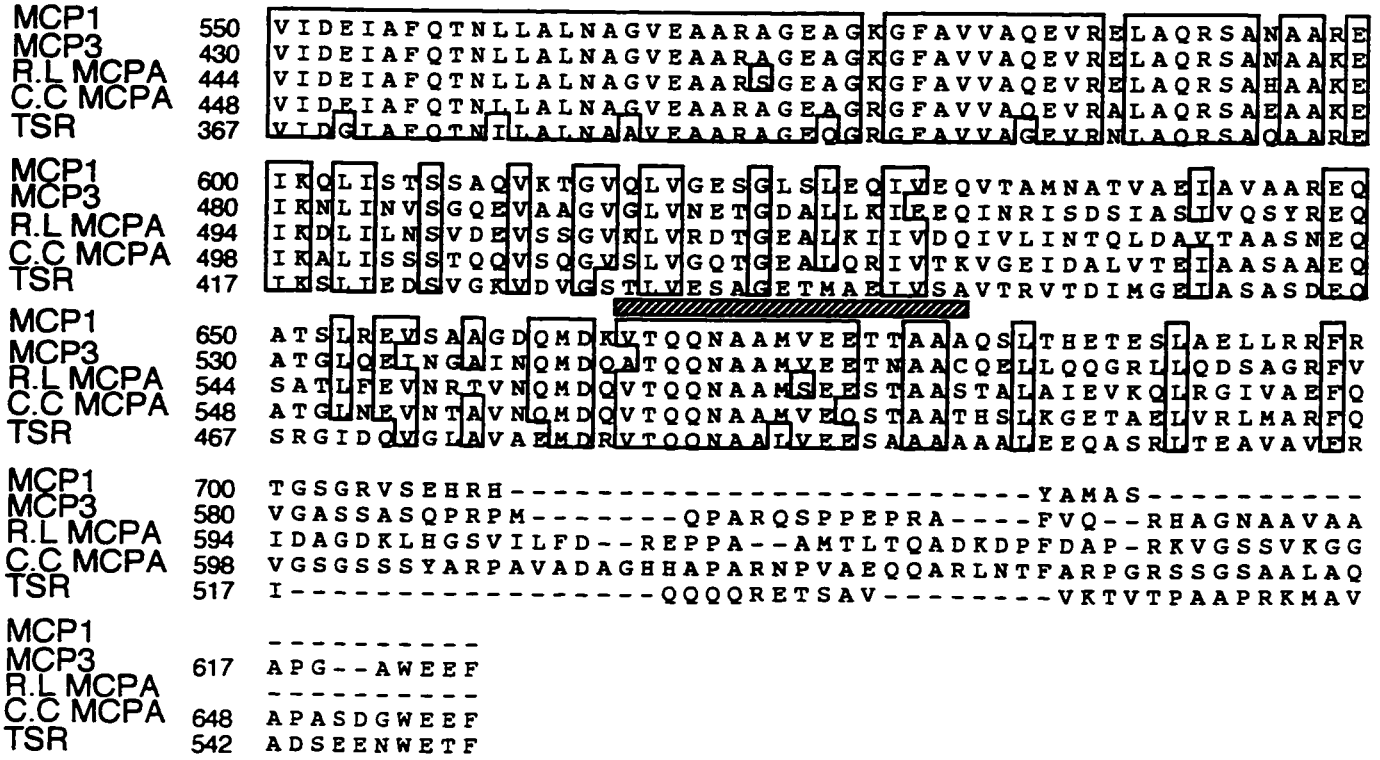

MCP1 550 VIDEIAFQTNLLALNAGVEAARAGEAGKGFAVVAQEVRELAQRSANAARE
MCP3 430 VIDEIAFQTNLLALNAGVEAARAGEAGKGFAVVAQEVRELAQRSANAAKE
R.L MCPA 444 VIDEIAFQTNLLALNAGVEAARSGEAGKGFAVVAQEVRELAQRSAHAAKE
C.C MCPA 448 VIDEIAFQTNLLALNAGVEAARAGEAGRGFAVVAQEVRALAQRSAEAAKE
TSR 367 VIDGIAFQTNILALNAAVEAARAGEQGRGFAVVAGEVRNLAQRSAQAARE
MCP1 600 IKQLISTSSAQVKTGVQLVGESGLSLEQIVEQVTAMNATVAEIAVAAREQ
MCP3 480 IKNLINVSGQEVAAGVGLVNETGDALLKIEEQINRISDSIASIVQSYREQ
R.L MCPA 494 IKDLILNSVDEVSSGVKLVRDTGEALKIIVDQIVLINTQLDAVTAASNEQ
C.C MCPA 498 IKALISSSTQQVSQGVSLVGQTGEALQRIVTKVGEIDALVTEIAASAAEQ
TSR 417 IKSLIEDSVGKVDVGSTLVESAGETMAEIVSAVTRVTDIMGEIASASDEQ
MCP1 650 ATSLREVSAAGDQMDKVTQQNAAMVEETTAAAQSLTHETESLAELLRRFR
MCP3 530 ATGLQEINGAINQMDQATQQNAAMVEETNAACQELLQQGRLLQDSAGRFV
R.L MCPA 544 SATLFEVNRTVNQMDQVTQQNAAMSEESTAASTALAIEVKQLRGIVAEFQ
C.C MCPA 548 ATGLNEVNTAVNQMDQVTQQNAAMVEQSTAATHSLKGETAELVRLMARFQ
TSR 467 SRGIDQVGLAVAEMDRVTQQNAAIVEESAAAAAALEEQASRLTEAVAVFR
MCP1 700 TGSGRVSEHRH----------YAMAS----------
MCP3 580 VGASSASQPRPM--------QPARQSPPEPRA----FVQ--RHAGNAAVAA
R.L MCPA 594 IDAGDKLHGSVILFD--REPPA--AMTLTQADKDPFDAP-RKVGSSVKGG
C.C MCPA 598 VGSGSSSYARPAVADAGHHAPARNPVAEQQARLNTFARPGRSSGSAALAQ
TSR 517 I----------QQQQRETSAV--------VKTVTPAAPRKMAV
MCP1 ----------
MCP3 617 APG--AWEEF
R.L MCPA ----------
C.C MCPA 648 APASDGWEEF
TSR 542 ADSEENWETF

Figure 4-9. The predicted amino acid sequences of MCP2, MCP4, and MCP5 derived from complete or partial DNA sequencing of *mcp-2*, *mcp-4*, and *mcp-5*, were submitted to a BLASTX search (Altschul *et. al.*, 1990). The resultant alignments with *C. crescentus* McpA are shown above. The numbers flanking the McpA sequence represent the amino acid residue numbers of McpA. The methylation regions are represented by bold text. The bold asterisks indicate sites of potential methylation in McpA (Alley *et. al.*, 1992). The MCP2 sequence contains the both the methylation domains and the highly conserved signalling region of McpA, while the sequences from MCP4 and MCP5 contain the first methylation domain.

Figure 4-9 BLASTX alignments of MCP2, MCP4 and MCP5 to *C. crescentus* McpA:

MCP2

```
Identities = 153/383 (39%), Positives = 205/383 (53%)

MCP2:        IRSIASAMRRLSDGDLDSDIPYAGRADEVGEMAGAVEIFRQNALNVVRLEKNPLNPAARA
              I  IA+A  +LS GD   D+    R DE+G +  A+++FR N +++ +L  +    AA
McpA:   222  INDIAAATDKLSKGDNSIDLEKMTRGDELGGIVTALKVFRDNQVHLEQLRADQEKSAALT 281

MCP2:        MQRAPHPSSAPNARREQLRFATTTLGEGLRRLASGDISFQLSEQFAAEYEALREDFNASL
                       +A  A  ++     + L EGL +LASGD++F+++  F  +Y  L++DFNA++
MCPA:   282  ADERRSKEAAAAAAAQEASLVVSNLAEGLEKLASGDLTFRVTADFPGDYRKLKDDFNAAM 341

MCP2:        RQLGATIGAVLQTVYSIDNGTGEIASAAQDLSK**RTEQQAASLEETAAALEEITSNVT**MAT
                L  T+  +  +   +  G  EIA A+ DLS+**RTEQQAASLEETAAAL  E+T+  V**
MCPA:   342  GSLQETMKVIAASTDGLSTGADEIAHASDDLSR**RTEQQAASLEETAAALDELTATVR**RTA 401
                                                        *       *       *

MCP2:        KRTDEARNVAKEADISAQRSAAVVSQAEGAMRRIEDSSQQISNIIGAIDEIAFQTNLLAL
                  +A +V        A  S  VV QA  AM  IE SS QIS IIG IDEIAFQTNLLAL
MCPA:   402  AGARQASDVVSTTRGEATHSGQVVHQAVSAMGEIEKSSGQISQIIGVIDEIAFQTNLLAL 461

MCP2:        NAGVEAARAGEAGKGFAVVAQEVRELAQRAAQAAKEIKGFIQKSSADVENGVKLVLETGT
              NAGVEAARAGEAG GFAVVAQEVR LAQR+A+AAKEIK  I  S+  V  GV LV +TG
MCPA:   462  NAGVEAARAGEAGRGFAVVAQEVRALAQRSAEAAKEIKALISSSTQQVSQGVSLVGQTGE 521

MCP2:        SLKSIGEYVVHINQLMDAIATSAREQSTGLAEINTAVNQMD**QATQQNAAMVEQSTAAVAS**
              +L+ I   V  I+ L+  IA SA EQ+TGL E+NTAVNQMD**Q TQQNAAMVEQSTAA  S**
MCPA:   522  ALQRIVTKVGEIDALVTEIAASAAEQATGLNEVNTAVNQMD**QVTQQNAAMVEQSTAATHS** 581
                                                                        *

MCP2:        **LSSEA**GRLRDLVNQFQLDGDKSA
              **L  E**   L  L+ +FQ+    S+
MCPA:   582  **LKGET**AELVRLMARFQVGSGSSS 604
```

MCP4

Identities = 24/58 (41%), Positives = 37/58 (63%)

ΩSp

```
MCP4:          GIGAGSNEIRSSGDQVSERTEQESVSVEETEAELEEITTTVRDADKRTEEETQVVALT
               G+  G++EI  + D +S RTEQ++ S+EET A L+E+T TVR       + + VV+ T
MCPA:     357  GLSTGADEIAHASDDLSRRTEQQAASLEETAAALDELTATVRRTAAGARQASDVVSTT 414
                                      *       *       *
```

MCP5

Identities = 36/71 (50%), Positives = 49/71 (69%)

```
MCP5:          ADLGQKYAVLRDNFNDALSHLEAAMVKVSAKGTDIGTSKEEIRRASNELSQRTERQAASL
               AD    Y  L+D+FN A+  L+  M  ++A    + T  +EI  AS++LS+RTE+QAASL
MCPA:     324  ADFPGDYRKLKDDFNAAMGSLQETMKVIAASTDGLSTGADEIAHASDDLSRRTEQQAASL 383
                                                                      *

MCP5:          EETSAALDELT
               EET+AALDELT
MCPA:     384  EETAAALDELT 394
                *       *
```

4.3 Evidence for the existence of additional *mcp* genes in *R. leguminosarum* VF39SM.

It is possible that in addition to the 5 *mcp* like genes cloned earlier, there may be more *mcp* genes residing in the VF39SM genome. Consequently studies were undertaken to search for additional *mcps* in VF39SM. Results of these studies suggest the presence of many putative *mcp* genes in VF39SM.

4.3.1 Isolation of putative *mcp* genes from a VF39SM genomic library.

Southern blots of VF39SM genomic DNA probed with the PCR probe derived from PCR amplification using the *mcp-3* primers, C1 and C2, have indicated that more than 5 MCP genes may be present in VF39SM (see figure 4-10). In fact, 14 clones which hybridize strongly to the probe were isolated from a VF39SM complete genomic library (see figure 4-11). All of these except one (cosmid #747) are distinct from clones *mcp-1* to *mcp-5.* Additionally, western blots of VF39SM cell lysates were probed using an antibody which binds to the conserved region of a MCP protein from *H. halobium.* These blots reveal the presence of up to 10 bands which bind to the MCP antibody.

4.3.2 Detection of multiple proteins in VF39SM cell lysates binding to an anti-MCP antibody.

To provide further evidence for the existence of a family of MCPs in VF39SM western blots using an anti-MCP antibody were conducted. Zhang and colleagues (1996) have used the polyclonal antibody HC23 to successfully identify a large family of transducer proteins in *H. salinarium.* Figure 4-12

Figure 4-10. VF39SM genomic DNA digested with *Bam* HI was blotted and probed with the C12 probe. Detection was done using chemiluminescence as described in the Materials and Methods. The size standards are listed on the right in Kb. The grey arrows point to bands which are doublets.

Figure 4-10 Detection of multiple *mcp*-like genes in VF39SM using C12 probe:

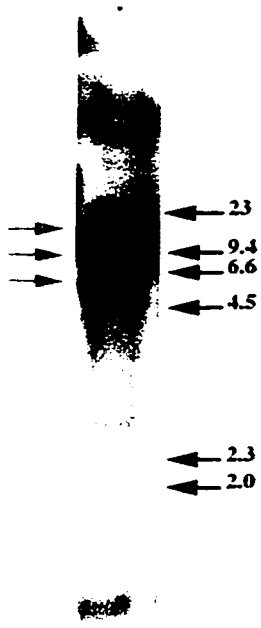
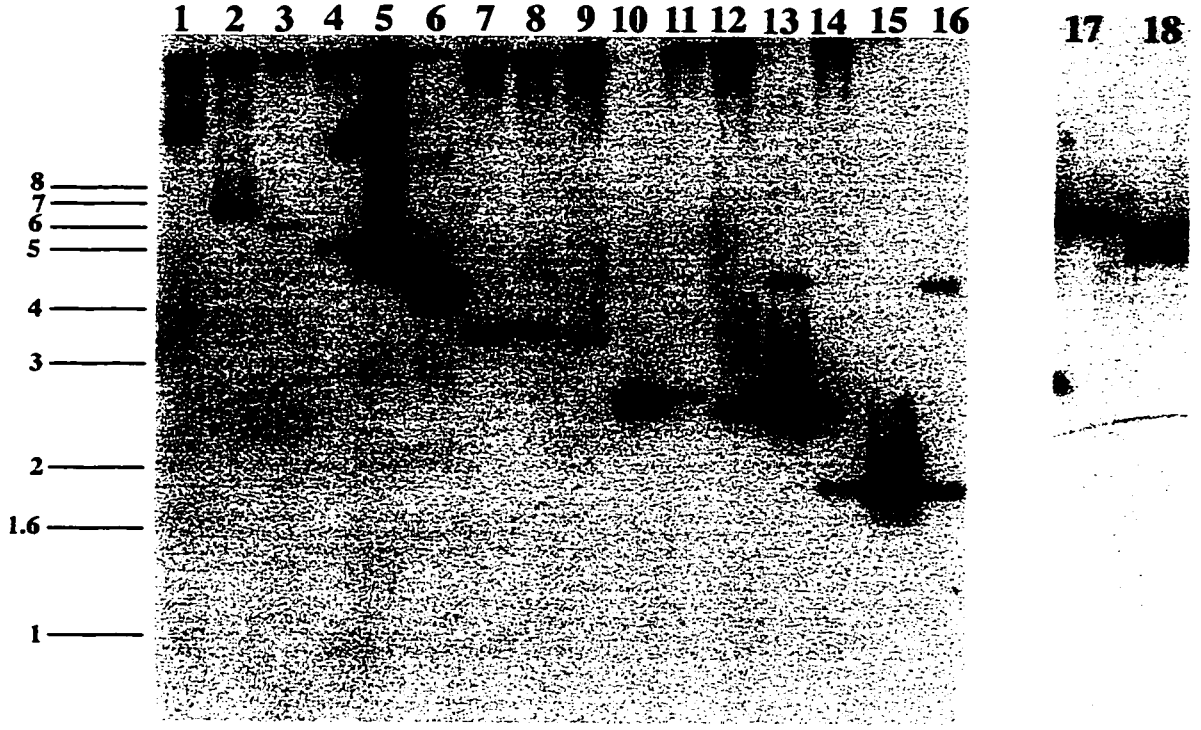


Figure 4-11. The VF39SM genomic library dot blots were probed with the C12 probe as described in the Materials and Methods. DNA from cosmids which hybridized to the probe were digested with restriction enzymes and probed again with the C12 probe. In all cases except in lanes 17 and 18 the DNA was double digested with *Bam* HI and *Eco* RI. The cosmid DNA from lanes 17 and 18 were digested with *Bam* HI. The lane order is as follows: 1, cosmid #525; 2, cosmid #843; 3, cosmid #836; 4 cosmid #938; 5, cosmid #898; 6, MCP2.B; 7, cosmid #75; 8, cosmid #719; 9, cosmid #645; 10, MCP5.B; 11, cosmid #269; 12, cosmid #279; 13, MCP3.B; 14, cosmid #747; 15, MCP1.B; 16, MCP4.B; 17 cosmid #303; 18, cosmid #309.

Figure 4-11 Isolation of VF39SM cosmids with *mcp*-like DNA sequences:



shows the results when cell lysates obtained from strains of VF39SM are used in western blots with the HC23 antibody. The presence of bands in figure 4-12 presents further evidence for the existence of MCPs in VF39SM.

Due to the varying methylation states of MCPs multiple bands could be expected from each MCP. However, the presence of numerous bands lends some support to the conjecture that VF39SM contains a family of MCP-like proteins larger that that seen in the enterics. It is possible that some bands resulted from cross-reactivity with the antibody. Future western blots could include a control strain which lacks MCPs to confirm or deny the possiblity of antibody cross-reactivity. It should be noted that the banding profiles from mutants not expressing genes *mcp-1* through *mcp-4* (labelled VF-MCP1$^-$ to VF-MCP4$^-$ respectatively in the figure legend) appear identical. A possible explanation for this is provided in the discussion.

4.4 Localization of MCPs to *R. leguminosarum* VF39SM plasmids.

Rhizobia characteristically contain large plasmids (Hynes *et al.*, 1989, reviewed by Mercado-Blanco & Toro, 1996). Some are involved in symbiosis while the function of other plasmids has been less well studied VF39SM contains six plasmids, ranging in size from 150 kb to over 600 kb (Hynes *et al.*, 1988). To investigate if any putative *mcp* genes reside on VF39SM plasmids a mixture of DNA probes derived from the six indigenous plasmids of VF39SM were used to probe a blot containing the putative *mcp* clones identified earlier (data not shown). The results of this experiment indicated that *mcp-2* is plasmid localized. Figure 4-13 shows the plasmid localization of *mcp-2*. *mcp-2* was found to reside on pRleVF39f. Additionally, three of the 15 cosmid clones which hybridized to the *mcp-3* probe were of plasmid origin. Cosmid clone #75 was discovered to reside on pRleVF39b (figure 4-14), while cosmid clone #719 resides on the sym plasmid, pRleVF39d (figure 4-15).

Figure 4-12. Cell lysates of VF39SM strains were electrophoresised using SDS-PAGE and western blotted as described in the Materials and methods. Molecular weight standards, in Kda, are provided on the left side of the figure. * denote bands which are likely to be doublets, triplets or greater. Lane order is as follows: 1, VF-MCP1$^-$; 2, VF-MCP2$^-$; 3, VF-MCP3$^-$; 4, VF-MCP4$^-$

Figure 4-12 Identification of putative MCP proteins in VF39SM

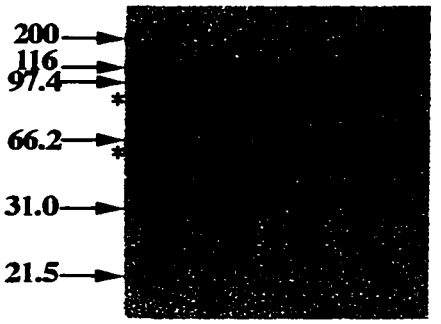


Figure 4-13. A) A typical Eckhardt gel of VF39SM and its plasmid cured derivative strains as created by Hynes & McGregor (1990). The strain in each lane is as follows: 1; VF39SM, 2; LRS39201 (cured of pRleVF39b) 3; LRS39301 (cured of pRleVF39c), 4; LRS39401 (cured of pRleVF39d), 5; LRS39501 (cured of pRleVF39e), and 6; LRS39601 (cured of pRleVF39f).

B) A Southern blot of the Eckhardt gel shown in panel A probed with the 1.5 kb *Eco*RI fragment of MCP2.B. A band is present in all lanes except the lane containing LRS39601.

Figure 4-13 Plasmid localization of *mcp-2*:

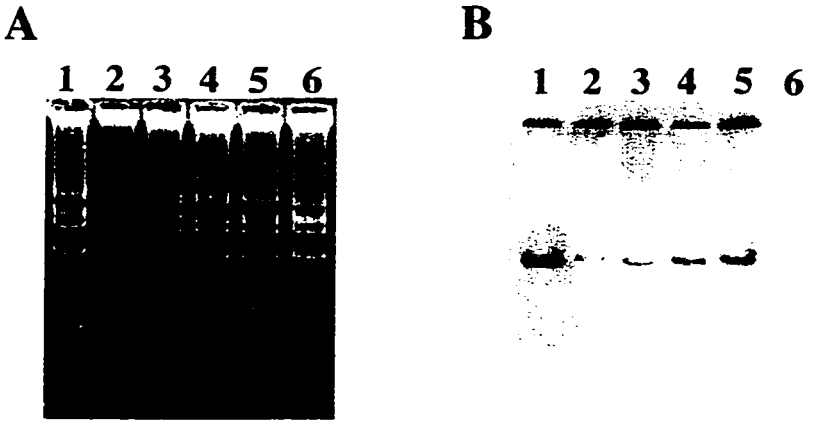


Figure 4-14. A) A typical Eckhardt gel of VF39SM and its plasmid cured derivative strains as created by Hynes & McGregor (1990). The strain order is as follows: 1; VF39SM, 2; LRS39201 (cured of pRleVF39b) 3; LRS39301 (cured of pRleVF39c), 4; LRS39401 (cured of pRleVF39d), 5; LRS39501 (cured of pRleVF39e), and 6; LRS39601 (cured of pRleVF39f).

B) A Southern blot of the Eckhardt gel shown in panel A probed with the 1.5 kb *Eco*RI fragment from cosmid # 75. A band is present in all lanes except the lane containing LRS39201.

Figure 4-14 Plasmid localization of cosmid # 75:

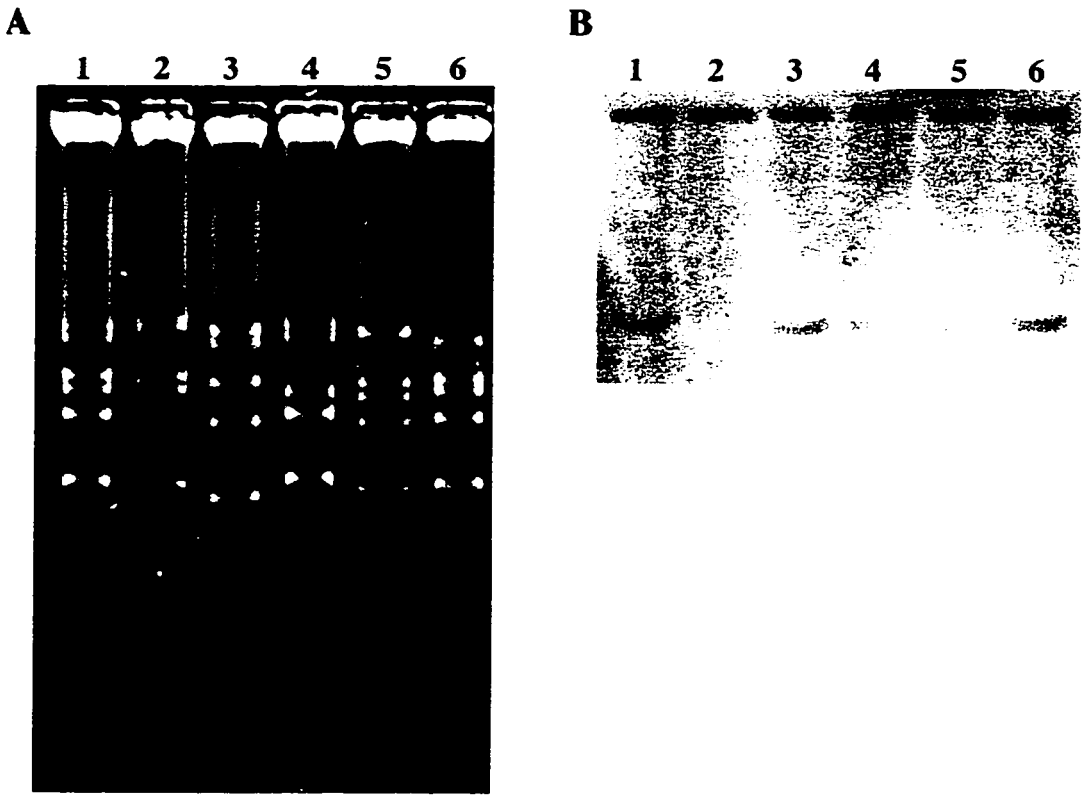
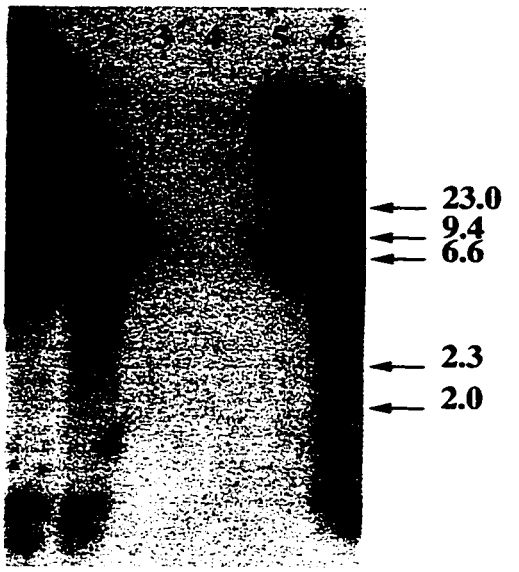


Figure 4-15. Genomic DNA from VF39SM and its plasmid cured derivative strains was digested with *Bam* HI and run on a 0.8% agarose gel. The gel was Southern blotted and probed with a mixture of DIG labelled *Bam* HI fragments isolated from cosmid #719. This figure shows the resultant Southern blot after chemiluminescence detection. Lane 1, VF39SM; lane 2, LRS39301; lane 3, LRS39401; lane 4, LRS39401-c, lane 5 LRS39501; lane 6, LRS39601. The probes were unable to bind DNA from strains cured of pRIeVF39d.

Figure 4-15 Plasmid localization of cosmid # 719:



Localization of the third cosmid clone is complicated by its apparent reiteration. Probes from cosmid #525 hybridize to both pRIeVF39e and pRIeVF39f (figure 4-16).

4.4.1 Identification of catabolic loci on VF39SM plasmids.

Work by Hynes and McGregor (1990), Baldani *et al.* (1992) and Hynes (unpublished) have shown that the genes needed to catabolize a variety of carbon sources are localized on the plasmids of VF39SM. The fact that putative *mcp* genes may also reside on these plasmids raises the interesting question as to whether the catabolic genes and chemotactic transducer for a specific carbon source are linked to the same plasmid. To identify previously unidentified carbon sources whose catabolic genes are plasmid encoded Biolog plates were used. Table 4-1 summarizes the results of this study. LRS39201 was unable to use gluconate, glucuronate, or malonate as sole carbon sources. LRS39401 could not metabolize glycerol or melibiose. LRS39401 was incapable of metabolizing hydroxy-L-proline or alanine. LRS39501 was unable to use histidine, rhamnose, serine or sorbitol as sole carbon sources. LRS39601 did not metabolize citrate, erythritol, ornithine, or proline. This information can be used to select potential carbon sources for swarm plate assays testing the chemotactic abilities of strains carrying mutations in the plasmid encoded *mcp* genes.

4.5 Complementation of a *mcp* deficient *E. coli* strain with the putative *mcp* clones of VF39SM.

The *E. coli* strain RP8611 has deletions in the mcp genes *tar, tap, tsr*, and *trg*. Consequently MCP dependent chemotaxis has been abolished in this strain. As a preliminary investigation to ascertain if the putative *mcp* clones of VF39SM are involved in chemotaxis and then to identify potential ligands these

Figure 4-16. A) An Eckhardt gel with *Agrobacterium* strains carrying each plasmid of VF39. The lanes are as follows: 1; UBAPF2 (plasmid free), 2; UBAPF2::pRleVF39a, 3; UBAPF2::pRleVF39b, 4; UBAPF2::pRleVF39c, 5; UBAPF2::pRleVF39d, 6; UBAPF2::pRleVF39e, 7; UBAPF2::pRleVF39f, and 8; VF39SM.

B) The Southern blot of the Eckhardt gel probed with cosmid 525 fragments. Both lanes 6 and 7 hybridize indicating the probe resides on both the pRleVF39e and pRleVF39f. The resolution of the plasmids for VF39SM was not high enough to distinguish both pRleVF39e and pRleVF39f on a Southern blot. Consequently, there appears to be only one band hybridizing in VF39SM when in fact there are two. This lack of resolution provided the rationale for performing this blot using *Agrobacterium* strains carrying individual plasmids of VF39SM.

Figure 4-16 Plasmid localization of cosmid # 525:

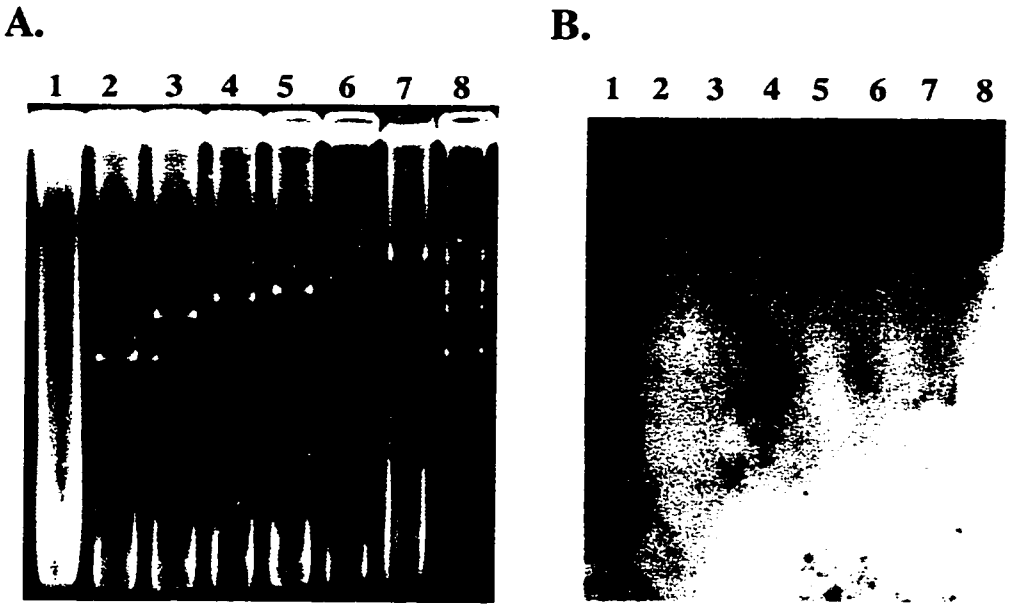


Table 4-1 Carbon compounds whose catabolism is dependent on plasmid encoded loci.

Plasmid	Carbon Source
pRleVF39b	Gluconate
	Glucuronate
	Malonate
pRleVF39c	Glycerol
	Melibiose
pRleVF39d	Alanine
	Hydroxy-L-Proline
pRleVF39e	Histidine
	Rhamnose
	Serine
	Sorbitol
pRleVF39f	Citrate
	Erythritol
	Ornithine
	Proline

Table 4-1. The GN Biolog microtitre plate tests an organism for the metabolism of 95 different carbon sources. VF39SM and its plasmid cured derivative strains were grown in individual GN Biolog Plates over-night. After incubation, the metabolic profiles of the plasmid cured strains were compared to the metabolic profile of VF39SM and differences in growth were recorded. Substances which were metabolized by VF39SM but not by a plasmid cured derivative strain were assumed to have plasmid localized catabolic genes.

clones were mobilized into RP8611. Competent cells of RP8611 were made and transformed with the DNA from cosmid clones listed in figure 4-11, as well as DNA from pRK7813 clones carrying the *Bam* HI fragments of *mcp-1* through *mcp-5*. It was hoped that some of these putative *mcp* clones would complement the mutations in RP8611 and restore chemotaxis to this strain. To test this hypothesis the transformed strains were stab inoculated onto YES swarm plates as described in Materials and methods. The swarm plates were incubated overnight at 28 °C and swarm diameters were compared to RP8611 carrying pRK7813. In all cases no complementation was observed, in fact it appeared that the transformant strains were impaired in their swimming ability.

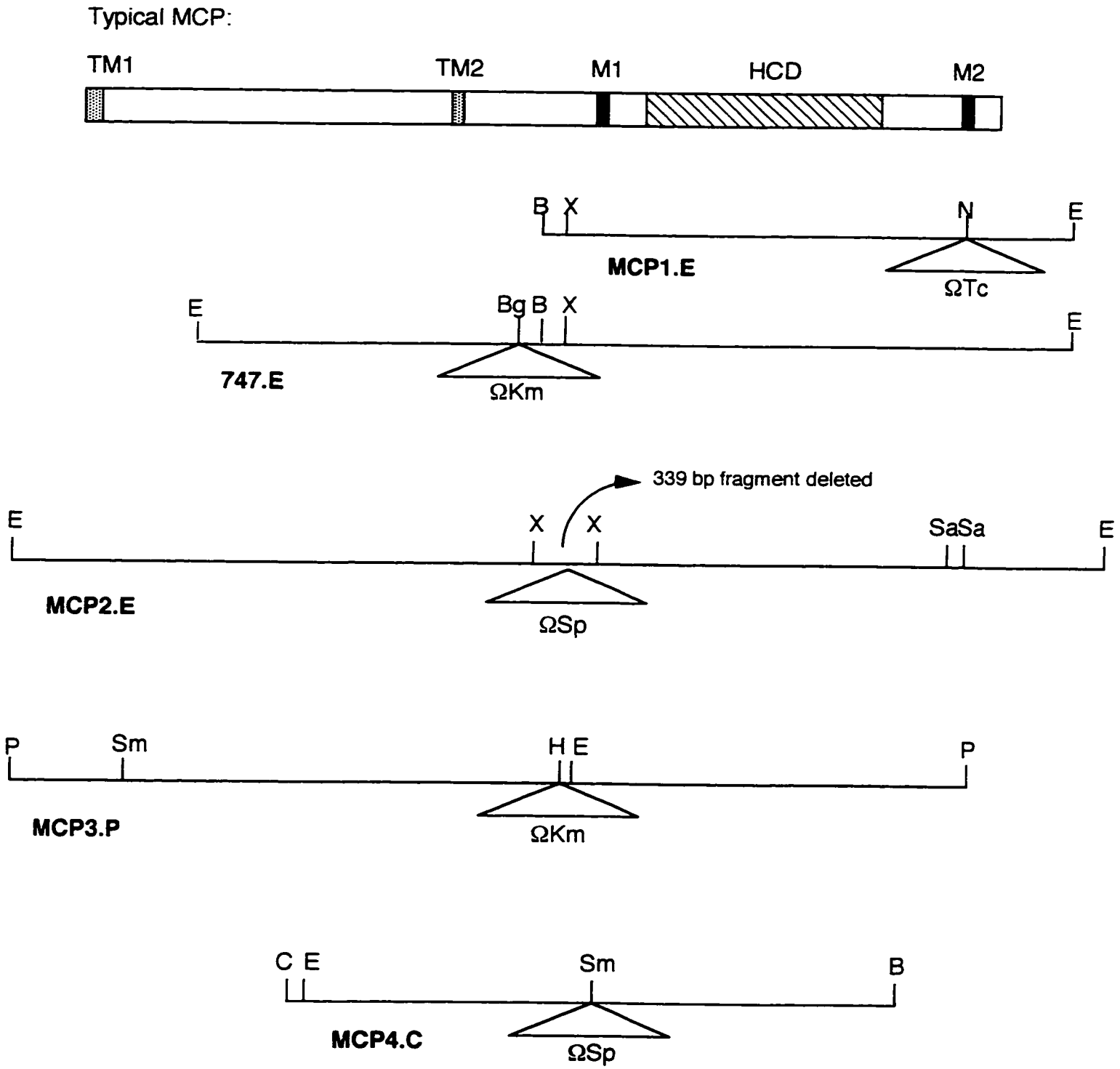
In addition to using YES swarm plates each transformant strain was tested on M9 swarm media supplemented with leucine, histidine and thiamine and containing serine as a source of attractant. This experiment was conducted to ascertain if one of the putative *mcp* clones coded for a MCP which could respond to serine similar to *tsr*. None of the transformant strains exhibited a chemotactic response to the serine swarm plates.

4.6 Insertional mutagenesis of *R. leguminosarum* VF39SM *mcp* genes; *mcp-1* through *mcp-4*.

During this study, *mcp-1*, *mcp-2*, *mcp-3* and *mcp-4* were mutated via insertional mutagenesis. Antibiotic resistance cassettes were used as the selectable markers (Prentki & Krisch, 1984; and Fellay *et al.*, 1987). Restriction endonuclease sites used for insertion of antibiotic resistance cassettes were identified by analyzing the DNA sequence data generated during this study. Figure 4-17 indicates the location of gene disruption in each MCP clone. Insertion of the ΩTc fragment into the *Not*I site of *mcp-1* resulted in gene disruption within the putative second methylation site. Another *mcp-1* mutant was created by inserting a ΩKm fragment into the *Bgl* II site of 747.E, thereby

Figure 4-17. Shown at the top of the figure is a schematic of a typical MCP. TM1 and TM2 represent the transmembrane domains, while M1 and M2 represent the methylation domains. The signalling domain is represented by the hatched box labelled HCD. Below this schematic are the subclones used for VF39SM *mcp* mutagenesis. Based on sequence homology data the subclones have been aligned to the MCP schematic. This allows for visualization of the location of gene disruption in the corresponding protein. The nature of the antibiotic resistance cassette used, and the restriction site used for insertion are indicated on the subclone maps. It should be noted that mutation of *mcp-2* resulted in a 339 bp deletion which removed the coding sequence for the first methylation region. Abbreviations for restriction enzymes are as follows: B, *Bam* HI; Bg, *Bgl* II; C, *Cla* I; E, *Eco* RI; H, *Hind* III; N, *Not* I; P, *Pst* I, Sa, *Sal* I; Sm, *Sma* I; X, *Xho* I.

Figure 4-17 Locations of gene disruption for *mcp-1* through *mcp-4*:



disrupting the gene upstream of the 1st methylation site. Partial digestion of MCP2.E with *Xho* I resulted in a 339 bp excision from *mcp-2* followed by the subsequent insertion of a ΩSp fragment into the *Xho* I site. This mutagenesis resulted in the loss of the first methylation coding sequence and insertional mutagenesis upstream of the signalling domain coding sequence. Insertional mutagenesis of *mcp-3* with ΩKm occurred prior to the first methylation region of *mcp-3* at the *Hind*III site. Lastly, *mcp-4* insertional mutagenesis also was accomplished using a *Sma*I site which is located slightly upstream of the first methylation region. In addition to creating single gene knockouts, a *mcp-2 and mcp-3* double mutant was created. The mutant was created by mating MCP3ΩNm into VF-MCP2$^-$, and then selecting for double recombinants.

4.6.1 Swarm plate analysis of the *mcp* mutants.

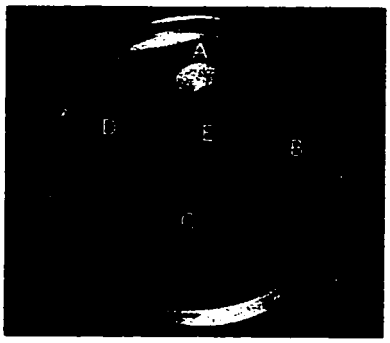
Phenotypic analysis of the mutants was performed using VMM swarm plates. A swarm plate operates on the principle that the bacteria inoculated on the swarm plate are capable of metabolizing the substrate available to them. As the cells metabolize the substrate, the relatively low concentration of substrate in the media allows for a concentration gradient to quickly establish. If the substrate is an attractant, the cells will then move up the gradient due to a chemotactic response. The chemotactic behaviours of the mutants towards various carbon sources were tested. The carbon sources used were: adonitol, arabinose, arginine, aspartate, erythritol, galactose, glutamate, glycerol, histidine, maltose, manitol, melibiose, ornithine, raffinose, ribose, rhamnose, serine, sorbitol, trehalose, trigonelline, and valine. Gluconate and proline were also assayed but were found to be poor chemoattractants. To date phenotypes for the mutants of *mcp-2*, *mcp-3*, and *mcp-4* have not been identified. Conversely mutations in *mcp-1* inhibited chemotaxis to all carbon sources tested. Both *mcp-1* mutants, which were created separately, exhibited this phenotype

Figure 4-18. Swarm plates showing the chemotactic behaviour of the various VF39SM MCP mutants. Strains were labelled as follows: A; VF-MCP1$^-$, B; VF-MCP2$^-$, C; VF-MCP3$^-$, D; VF-MCP4$^-$, and E; VF39SM. Plate **1.)** A yeast extract swarm plate. Plate **2.)** A VMM raffinose swarm plate. Plate **3.)** A VMM adonitol swarm plate. The swarm plates were prepared as described in the Materials and methods.

VF-MCP1$^-$ is impaired in its ability to swarm away from the inoculation point. Cells accumulate in a localized area resulting in a denser circle of growth relative to the wild type and the other mutant strains. This is particularly evident in plate 1. In addition, the swarm diameter of VF-MCP1$^-$ is smaller relative to the other strains. This observation is represented most dramatically in the raffinose swarm plate.

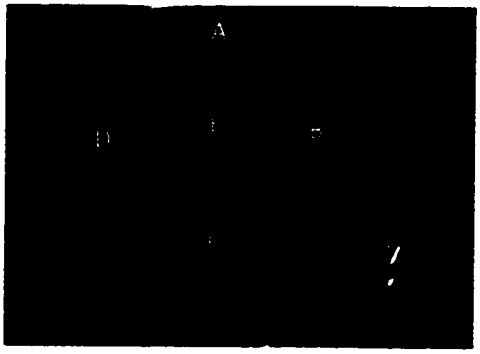
Figure 4-18 Swarm patterns of *mcp* mutants:

1.



2.

3.



indicating that it is a reproducible phenotype. This phenotype does not conform to the *E. coli* model whereby a MCP is responsive to a specific set of ligands.

4.6.2 Nodulation efficacy of the *mcp* mutants.

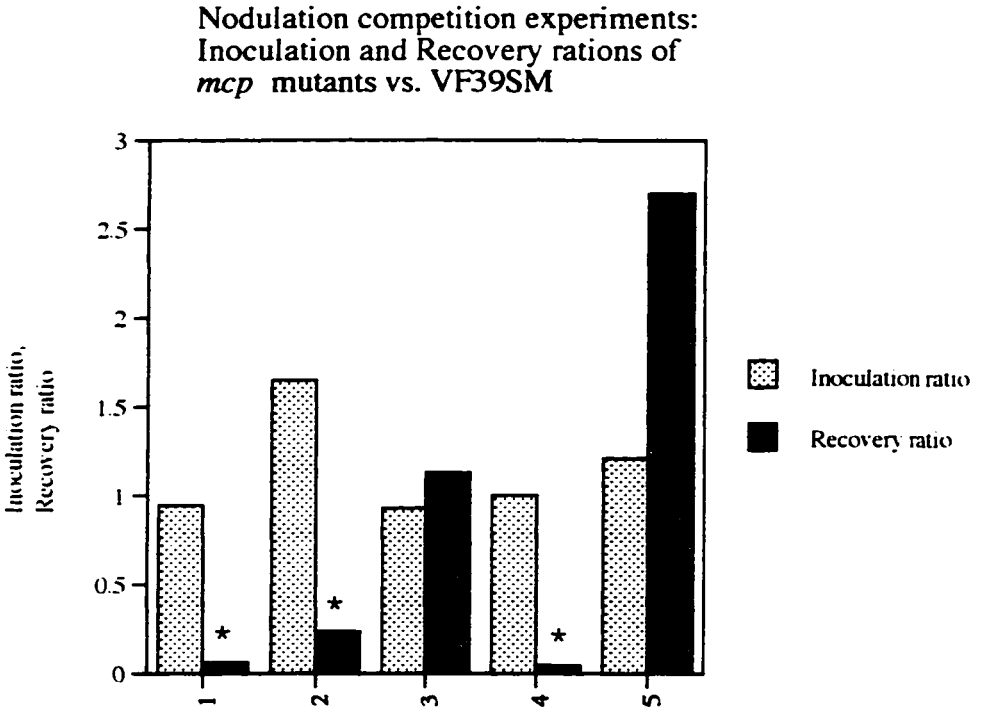
The *mcp* mutants were co-inoculated with wild-type VF39SM onto pea plants. Consequently each strain competed against each other for sites of nodulation. This experiment was conducted to asses the possible role of MCPs during early nodulation. The principle of this type of study is that if a plant is co-inoculated with two strains of rhizobia, formation of more nodules by one strain indicates greater "competitiveness"; this may be due to a variety of factors, including faster growth in the root environment, faster infection rates, or superior ability to migrate towards suitable infection sites (Ames & Bergman, 1981; Gulash *et al.*, 1984; Dharmatilake & Bauer, 1992).

Figure 4-19 indicates the results of the nodule competition experiments. The graph clearly indicates that VF-MCP1$^-$ and VF-MCP2$^-$ were unable to compete with VF39SM in frequency of nodulation. The recovery ratio for VF-MCP3$^-$ indicated that it was able to compete with VF39SM for nodule occupancy. VF-MCP4$^-$ appeared to outcompete VF39SM, it is unlikely that this was due to mutation of *mcp-4*. The inability of the *mcp-2, mcp-3* double mutant to compete with VF39SM for nodule occupancy correlates with the result of the *mcp-2* mutant competition experiment.

The results of the competition experiments using *mcp-1* and *mcp-2* mutants were reproducible based on repeated experimentation.

Figure 4-19. Ratios in the graph are expressed as the mcp mutant/VF39SM. Set 1 represents data from pea plants co-inoculated with VF39SM and VF-MCP1$^-$. Set 2 represents data from pea plants co-inoculated with VF39SM and VF-MCP2$^-$. Set 3 represents data from pea plants co-inoculated with VF39SM and VF-MCP3$^-$. Set 4 represents data from pea plants co-inoculated with VF39SM and VF-MCP2$^-$/3$^-$. Set 5 represents data from pea plants co-inoculated with VF39SM and VF-MCP4$^-$. Where an '*' is indicated the recovery ratio was significantly lower than the initial inoculation ratio. Statistical significance was confirmed using the chi squared test.

Figure 4-19 Nodulation efficacy of *mcp* mutants



4.7 Construction of *lacZ* gene fusions to *mcp-1*, *mcp-2*, and *mcp-3*.

To study the expression and regulation of *mcp* genes in *R. leguminosarum*, a promoterless *lacZ* gene was fused downstream of the *mcp-1*, *mcp-2*, or *mcp-3* promoters. The *lacZ* cassette was from pZ1918 (Schweizer, 1993). This cassette has the promoterless *lacZ* gene flanked by direct repeats of the puc19/18 multiple cloning site, allowing for the insertion of *lacZ* into numerous different restriction sites. The flexibility in restriction site options facilitated the creation of the reporter gene fusions.

4.7.1 Plasmid stability of the fusion constructs.

lacZ fusions with *mcp-3* were carried on the broad host range plasmid pRK7813. A portion of the fusion study involved monitoring fusion expression within the nodule. Results from this study indicated that pRK7813 is not very stable when selection for tetracycline resistance is removed, such as when the fusion strains are nodulating pea plants. Table 4-2 shows the recovery frequency of VF39SM carrying pRK7813::*mcp-3*::*lacZ* after nodulation. The results indicated that on average only 35% of the bacteria recovered from crushed nodules still contained the fusion plasmid. This plasmid instability could cause uncertainty in future *mcp* gene regulation experiments concerned with nodulation. Consequently the *par* locus of RK2, isolated by Weinsteins *et al.* (1992), was integrated into the fusion plasmids to provide greater stability to these plasmids (refer to appendix 4 for a more detailed description of the protocol). The *par* locus was contained within a Tn3 transposon (Tn3-par) and was therefore integrated into the fusion plasmids by transposition. To ensure that Tn3-par did not transpose into the fusion itself it was necessary to select for transposition events where the location of the Tn3-par insertion could be known. Selecting for transposition events where the Tn3 derivative jumped into the Tc^r gene of pRK7813 provided an easy method for identifying transpositions

Table 4-2. pRK7813 stability in VF39SM during nodulation.

Strain	Sm^r isolates	Tc^r isolates	Percent Recovery of Fusion Plasmid
VF39SM pRK7813::*mcp-3* ::*lacZ* R1	21	6	29%
VF39SM pRK7813::*mcp-3* ::*lacZ* R2	21	10	48%
VF39SM pRK7813::*mcp-3* ::*lacZ* W1	21	6	29%

Table 4-2. 21 nodules were picked from plants nodulated by each of the strains listed in the table. The nodules were surface sterilized and crushed as described in the Materials and methods. The macerates were plated on both TY Sm and TY Tc plates and the plates were incubated for 3 days at 28 C. Following incubation plates were scored for growth. pRK7813::*mcp-3*::*lacZ* R1 and R2 are fusion plasmids where the *lacZ* gene was fused to the *mcp-3* promoter in the same orientation, VF39SM pRK7813::*mcp-3*::*lacZ* W1 is a fusion plasmid where the *lacZ* gene was fused to the *mcp-3* promoter in the opposite orientation.

that did not occur within the fusion. Simply, colonies identified containing the Tn3 derivative on the fusion plasmid were screened for Nm^r (encoded by Tn3 derivative) and Tc^s. Table 4-3 shows the recovery frequency of strains carrying the newly formed fusion-par plasmids after nodulation. The *par* locus stabilized the plasmids greatly, in fact 100% of the crushed nodules contained rhizobia carrying the fusion plasmid.

4.7.2 Expression of *mcp* genes during nodulation.

The aim of this study was to determine if *R. leguminosarum* VF39SM *mcp* genes are down-regulated during nodulation. The fusion constructs used in this investigation were MCP-F1/MCP-F2. These constructs were placed in VF39SM and the resultant strains were grown on pea plants and allowed to nodulate. β–galactosidase activity was assayed in bacteroids isolated from crushed nodules. Table 4-4 provides the results of these assays. Clearly, *mcp-3* is down-regulated during nodulation.

After ascertaining that *mcp* genes are down-regulated during nodulation a further investigation was conducted using the strains TP6, TP9 and TP11. These strains contain mutations in regulatory genes that participate in the up-regulation of nitrogen fixation genes during nodulation. The *mcp-3* fusion was introduced into these strains, after-which the strains were grown once again on pea plants and allowed to nodulate. Table 4-4 shows the β–galactosidase activity from bacteroids isolated after nodulation. The down-regulation of *mcp-3* was unaffected in these mutant strains.

Table 4-3 pRK7813 stability in VF39SM during nodulation after insertion of the par locus:

Strain	Sm^r isolates	Nm^r isolates	Percent Recovery of par Plasmid
VF39SM MCP-F1	21	21	100%
TP6 MCP-F1	15	15	100%
TP9 MCP-F1	15	15	100%

* MCP-F1 is the *mcp-3/lacZ* fusion cloned in the pRK7813 *par* plasmid.

Table 4-3. 21 nodules were picked from plants nodulated by each of the strains listed in the table. The nodules were surface sterilized and crushed as described in the Materials and methods. The macerates were plated on both TY Sm and TY Nm plates and the plates were incubated for 3 days at 28 C. Following incubation the plates were scored for growth. In all cases rhizobia carrying the par plasmids could be isolated from each nodule.

Table 4-4 *mcp-3* gene expression during nodulation:

Strain	β-galactosidase activity in Miller Units
VF39SM MCP-F1	91.7 ± 9.11 ($1.67 \times 10^3 \pm 115$)
VF39SM MCP-F2	0 (10.4 ± 1.43)
TP6 MCP-F1	228 ± 44.6
TP6 MCP-F2	0
TP9 MCP-F1	273 ± 39.5
TP9 MCP-F2	0
TP11 MCP-F1	83.2 ± 53.0
TP11 MCP-F2	0
L1	$1.64 \times 10^3 \pm 279$ (55.2 ± 3.04)

* MCP-F1 is the *mcp-3/lacZ* fusion cloned in the pRK7813 *par* plasmid.

* MCP-F2 is the *mcp-3/lacZ* fusion where *lacZ* is fused in the opposite orientation relative to *mcp-3*. Again this fusion was cloned in the pRK7813 *par* plasmid.

Table 4-4. β-galactosidase activity from bacteroids was assayed as described in Materials and methods. Assays were done in triplicate, however the experiment was not repeated. L1 carries a *lacZ* fusion to *fixN* and consequently *lacZ* is expressed during nodulation. The values in the brackets are β -galactosidase activity in Miller Units when the strains were grown in TY broth to stationary phase.

4.7.3 Investigation of succinate as a potential signal controlling the down-regulation of *mcp* gene expression.

Urban and Dazzo (1982) have noted that when *S. meliloti* is grown with succinate as a sole carbon source the cells undergo morphological changes that resemble those seen in bacteroids. When these cells are observed using a microscope they appear very similar to bacteroids observed under a microscope. The same situation is true for VF39SM. When VF39SM is grown with succinate as a carbon source the cells appear similar in morphology to bacteroids (data not shown). Additionally, VF39SM does not swarm normally in succinate swarm plates, and microscopic analysis indicates that the majority of the cells are non-motile. In light of these observations, the possible role of succinate as a regulator of *mcp* gene expression was investigated.

Table 4-5 provides the data concerning *mcp* gene expression in the presence of succinate. Although β-galactosidase activity is lower in cultures grown with succinate it is well above background levels.

4.8 The search for regulatory genes involved in *mcp* gene expression within VF39SM.

Earlier sections described studies concerned with the down-regulation of *mcp* genes in *R. leguminosarum*. To gain a more complete understanding of *mcp* gene expression in *R. leguminosarum* it is necessary to identify genes which may be involved in the regulation of *mcp* genes in *R. leguminosarum* during free living conditions. 400 Tn5 mutants of VF-MCPF4 were screened for loss of β -galactosidase activity on VMM glutamate plates containing X-gal. 2 mutants were identified which no longer expressed β-galactosidase activity based on their white colour on X-gal plates. The plasmid MCP-F1 was mated into both of

Table 4-5 *mcp* gene expression in the presence of succinate:

Strain	Carbon Source	β-galactosidase activity
VF39SM MCP-F1	glutamate	$1.94 \times 10^3 \pm 164$
	glutamate + succinate	$1.06 \times 10^3 \pm 226$
	succinate	$1.13 \times 10^3 \pm 220$
	glucose	$1.59 \times 10^3 \pm 163$
VF39SM MCP-F2	glutamate	5.07 ± 3.42
	glutamate + succinate	7.50 ± 5.20
	succinate	0.00
	glucose	5.58 ± 1.61
VF39SM pFusMCP2	glutamate	683 ± 3.00
	succinate	462 ± 9.00
VF39SM	glutamate	7.54 ± 1.94*
	succinate	0.00

* Activity from indigenous lactose metabolizing enzyme

Table 4-5. Strains were grown in VMM minimal media with sole carbon sources: glutamate, succinate, glucose, or a combination of glutamate and succinate. All carbon sources were used at a final concentration of 5 mM. Cultures were grown to late log phase and then assayed for β-galactosidase activity as described in Materials and methods. Assays were done in triplicate.

these mutants. Once the mutants were carrying MCP-F1 β-galactosidase assays, as described in Materials and methods, were performed on the mutants. In both cases the presence of MCP-F1 complemented each mutant and consequently they both showed high levels of β-galactosidase activity. Therefore both mutations were assumed to be the result of a Tn5 transposition into the *lacZ* gene of *mcp3::lacZ*. Although no regulatory mutants were isolated, the results of this experiment indicate that the methodology described for screening of regulatory mutants is sound and that further isolation and screening of Tn5 mutants can occur.

5.0 Discussion.

The main goal of this research was to characterize chemotaxis in *R. leguminosarum* at the molecular level. At the beginning of this study, little was known, at the molecular level, about the components of the chemotaxis system in *Rhizobium* spp. Consequently, the published literature on other chemotaxis systems was used to design a strategy for studying chemotaxis in *R. leguminosarum*. A key component of a chemotactic signalling pathway are the sensor molecules. Such proteins allow a bacterium to monitor its environment. *Rhizobium* live in an environment diverse in environmental signals. Consequently, it was decided that the focus of this study would be the identification and characterization of chemotransducer molecules in *R. leguminosarum* VF39SM. In particular, since methyl-accepting chemotaxis proteins have been shown to play important roles in the chemotactic behaviour of many genera of bacteria the identification of methyl-accepting chemotaxis proteins was chosen as the focus for this study.

5.1 The identification of *mcp*- like genes in *R. leguminosarum* VF39SM.

The ability to amplify a fragment of VF39SM genomic DNA via PCR using primers derived from the sequence of a *D. vulgaris mcp* gene suggested the presence of *mcp*-like genes in VF39SM. This PCR experiment led to the subsequent cloning of 5 genomic DNA fragments encoding putative *mcp* genes. The assumption that each clone contained *mcp*-like domains was based on DNA hybridization data. The next logical step was to sequence the DNA of these fragments to ascertain if the DNA sequence exhibited extensive homology to known *mcp* genes. DNA sequencing indeed revealed that the cloned fragments of VF39SM DNA exhibited high homology to known *mcp* genes. Referring to ORFs as *mcp* genes on the basis of sequence homology data is not uncommon; *mcpA* of *C. crescentus* (Alley *et al.*, 1992), and *mcpA* of *R. leguminosarum* (Brito *et al.*, 1996). Hence the ORFs cloned and sequenced in this study have been termed *mcp-1* through *mcp-5*.

The complete DNA sequences for the ORFs of *mcp-1*, *mcp-2*, and *mcp-3* have shown that these genes code for putative MCP proteins with all the characteristics common to known MCPs. That is, each contains two predicted transmembrane regions flanking the N-terminal domain which would result in the periplasmic localization of the N-terminal domain. They all show high similarity to both the methylation and signalling domains of known MCPs. The coding regions for the *mcp-4* and *mcp-5* ORFs have not been completely sequenced, however important regions of sequence data have been obtained. The predicted amino acid sequences of these regions indicate that the ORFs of *mcp-4* and *mcp-5* have high similarity to the first methylation region of known MCPs. The fact that all 5 ORFs have high similarity to the methylation region of MCPs is noteworthy. Other sensory transducers such as the *tcp1* and *actB* gene products in *Vibrio cholerae* (Harkey *et al.*, 1994; Everiss *et al.* 1994) have been identified which have high similarity to the signalling domain of

MCPs but do not function as chemoreceptors *per se*. In each case the amino acid similarities of these proteins to the methylation region of Tsr were not very strong. Therefore the fact that the VF39SM ORFs have homology to both the signalling domain and methylation domain strengthens the hypothesis that these ORFs code for MCPs.

The observation that the N-terminal regions of *mcp-1*, *mcp-2*, and *mcp-3* show no homology to MCPs is not surprising. Since the N-terminus determines ligand specificity one would expect high degrees of amino acid sequence variance relating to the variety of ligands. A lack of sequence similarity suggests that MCPs functionally related to *mcp-1*, *mcp-2*, and *mcp-3* have not been identified. For example, if *mcp-1*, *mcp-2*, or *mcp-3* detected serine one might expect notable similarity between its N-terminal domain and the N-terminal domain of Tsr (*E. coli* serine chemoreceptor). In fact, BlastX searches of the N-terminus of *mcp-1*, *mcp-2*, and *mcp-* show little homology to any known proteins. The only exception is that the N-terminus of MCP2 shows very weak homology to a segment of a putative sensor transducer of *Clostridium*, of which the sequence is not yet published (data not shown).

5.1.1 Evidence for the presence of additional *mcp*-like genes in VF39SM.

Probing Southern blots of VF39SM DNA with *mcp*-like probes suggested that VF39SM may contain numerous *mcp*-like genes. In light of these observations, the genomic library of VF39SM was probed with the C1/C2 PCR probe. This experiment revealed that additional clones, distinct from *mcp-1* through *mcp-5*, had homology to the probe derived from the conserved region of *mcp-3*. Therefore VF39SM potentially contains a family of up to 19 *mcp*-like genes. Although the *E. coli* paradigm suggests small numbers of *mcps* are found in bacteria, recent studies in *Desulfovibrio* (Deckers and Voordouw, 1994) and

Halobacterium (Zhang *et al.*, 1996) suggest that large families of *mcp* genes may exist in some bacterial species. Deckers and Voordouw (1994) have proposed that *Desulfovibrio vulgaris* may contain 16 *mcp* genes while Zhang and colleagues (1996) have suggested that 13 *mcp* genes exist in *Halobacterium salinarium*.

What would be the function for the presence of a large family of chemoreceptors in *R. leguminosarum* ? A plausible explanation resides in the fact that rhizobia encounter a diverse range of carbon sources in the rhizosphere and soil and subsequently can grow on a large number of carbon sources (Parke & Ornston, 1984; reviewed by Stowers, 1985). Therefore, it is not surprising to find many reports demonstrating that rhizobia are chemotactic towards numerous carbon sources. (Bowra & Dilworth, 1981; Gaworzewska & Carlile, 1982; Parke *et. al.*, 1985; Armitage *et. al.*, 1988; Dharmatilake & Bauer, 1992; and Robinson and Bauer, 1993). If chemotaxis to these carbon sources is MCP-dependent it seems logical to conclude that multiple MCPs would be necessary to respond to the large repertoire of attractants.

The claim for a large family of *mcp* genes in VF39SM is further supported by the western blot data generated in this study (fig 4-12). The HC23 antibody, raised against a peptide constructed from the conserved signalling domain of MCPs, bound to at least 8 proteins from VF39SM lysates supporting the conjecture that a large family of *mcp* -like genes exists in VF39SM. It is possible that some of the bands appearing in the western blot are due to non-specific cross reactions. However, most of the bands are strong in intensity, making this supposition less likely.

The fact that *mcp* probes hybridize to multiple genes and that antibodies against the HCD recognise multiple proteins of VF39SM does not demonstrate that any these are involved in chemotaxis. As mentioned previously, gene products which have homology to the signalling domain of MCPs but are not

directly involved in chemotaxis do exist. Consequently, further investigation will be required before the role (if any) of these genes in chemotaxis can be commented upon. The work described in this portion of the thesis provides the ground work for continued investigation of these putative *mcp* genes. The clones identified in figure 4-11 and their respective *mcp*-like ORFs are available for characterization via mutagenesis and DNA sequencing.

5.1.1.1 Further interpretation of the western blot shown in figure 4-12.

Identical banding profiles between the cell lysates from each *mcp* mutant is somewhat unsettling. However, it is quite possible that the SDS-PAGE gel used for the western blot was unable to detect any differences in banding patterns. Some of the bands seen in the blot may in fact be multiple bands indicating that banding profiles between the mutants may have been unresolvable with the SDS PAGE conditions used in this study. Western blots with better resolution must be done to clarify this issue. Once this issue is resolved and differential banding patterns can be observed the HC23 antibody can be used for future research with VF39SM putative MCP proteins.

5.2 Significance of plasmid localized *mcp*-like genes in VF39SM.

A large portion of the rhizobial genome is of plasmid origin. In fact, roughly 35-40 % of the VF39SM genome is plasmid derived (Yost and Hynes, unpublished; Hynes and Finan, *in press*). Many catabolic loci have been found to reside on rhizobial plasmids (Tepfer *et al.*, 1988; Gajendiran and Mahavedan, 1990; Charles and Finan 1991; Boivin *et al.*, 1991; Baldani *et al.*, 1992; Saint *et al.*, 1993; Goldman *et al.*, 1994; Soto *et al.*, 1994). If attractants which are catabolized by plasmid encoded enzymes are sensed by MCPs, then it is

tempting to postulate that the genes for these MCPs may be localized to the same plasmid. The first step in addressing this supposition was to ascertain if any of the *mcp*-like genes of VF39SM are plasmid localized. *mcp-2* and three other potential *mcp* genes of VF39SM were determined to be plasmid derived. Other researchers have also identified *mcp*-like genes residing on the plasmids of other *Rhizobium* spp. The *mcpA* gene isolated by Brito and colleagues (1996) is on the sym plasmid of *R. leguminosarum* UPM791. Additionally, DNA sequencing of the nodulation plasmid from *Rhizobium* NGR234 has revealed the presence of two *mcp*-like genes residing on this plasmid (Freiberg *et al.*, 1997). In the closely related bacterial species *Agrobacterium*, a putative *mcp* gene has also been identified which is of plasmid origin. ORF2, described as a *mcpA* homologue has been localized to the Ti plasmid pTi15955 of *Agrobacterium tumefaciens* (Kim *et al.*, 1996). It is interesting that this *mcp* gene maps close to mannopine transport genes (Kim *et al.*, 1996). However, the significance of this is yet to be revealed. To date, the possible link of the plasmid encoded *mcp*-like genes of VF39SM to plasmid encoded catabolic loci has not been determined. However, results from this study may help to elucidate this question. The data generated from Biolog studies and previous studies can be used to aid future investigations focused on determining if plasmid localized MCPs sense attractants whose catabolic genes are also plasmid localized. Table 5-1 summarizes carbon sources which should be considered when searching for ligands of plasmid localized MCPs.

A noteworthy observation resulting from the plasmid localization studies was the reiteration of cosmid 525 on both pRleVF39e and pRleVF39f. The results reported here further support the research of Flores *et al.* (1987) that rhizobia contain extensive genomic reiteration. Additional occurrences of reiterated DNA on the plasmids of VF39SM have been previously documented (Mitsch, 1995; Schlüter *et al.*, 1997). Copies of the operons, *fixGHIS* and *fixNOQP* are found on both pRleVF39c and pRleVF39d. In each case both copies are functional and both must be mutated to abolish nitrogen fixation activity (Mitsch, 1995;

Table 5-1 Summary of carbon sources whose metabolism is plasmid dependent.

Plasmid	Carbon Source	Reference
pRleVF39b	Gluconate	This study
	Glucuronate	This study
	Malonate	This study
pRleVF39c	Glycerol	This study; Orsenik *et al.* (submitted)
	Melibiose	This study
pRleVF39d	Adonitol	Oresnik *et al.* (submitted)
	Alanine	This study
	Hydroxy-L-Proline	This study
	Trigonelline	Hynes, unpublished
pRleVF39e	Histidine	This study
	Rhamnose	This study; Oresnik *et al.* (submitted)
	Serine	This study
	Sorbitol	This study; Oresnik *et al.* (submitted)
pRleVF39f	Arginine	Hynes, unpublished
	Citrate	This study
	Erythritol	This study
	Ornithine	This study; Hynes, unpublished
	Proline	This study

Table 5-1. From this study and the work of others a summary table has been compiled listing the carbon sources that require plasmid encoded loci for their metabolism. The data was gathered by observing differences in the metabolic activities of strains cured of the VF39SM plasmid, and in some instances identifying plasmid localized catabolic genes (Oresnik *et al.* , submitted)

Schlüter *et al.*, 1997). The significance that some putative mcp genes of VF39SM may be reiterated is that such duplication could cause confusion in identifying phenotypes. If both genes are functional yet only one is mutated any phenotypic changes will be masked by the second non mutated copy.

5.3 Phenotypic characterization of the putative *mcp* genes of VF39SM.

DNA sequencing data can implicate gene function through homology comparisons, but sequencing data alone cannot directly prove gene function. Therefore, to connect the putative *mcp* genes to chemotactic behaviours it was necessary to use other experimental methods. The two methods chosen were complementation of the *mcp* deficient *E. coli* strain RP8611, and mutation of selected putative *mcp* genes.

5.3.1 Complementation studies.

RP8611 can be considered chemotactically incompetent, due to the deletions in its *mcp* genes. On YES swarm media or M9 minimal swarm media RP8611 does not swarm away from the point of inoculation; instead it forms a dense area of growth near the inoculation source. Theoretically, it is possible to restore chemotactic ability to this strain through complementation with a functional *mcp* gene. Under this premise, attempts to complement RP8611 with the putative *mcp* genes of VF39SM were conducted. Expression of each putative *mcp* gene was dependent on the native promoter. The low copy number plasmid pRK7813 was used as the vector. A low copy number plasmid was chosen because expression of an *mcp* gene in a high copy number plasmid may produce too many MCP proteins and titrate out signalling components (CheA/CheW) thereby disrupting proper signalling. The cosmid clones,

identified as putative *mcp* genes from figure 4-11, were placed directly into RP8611, as they already resided on pRK7813. However, the *Bam* HI fragments of *mcp-1* through *mcp-5* were carried on pBluescript which is a high copy number plasmid. Therefore these *Bam* HI fragments were first cloned into pRK7813 and then transferred into RP8611. The reconstituted strains of RP8611 were incubated at 28 °C rather than 37 °C since this is the optimum growth temperature for VF39SM; the VF39SM *mcp*-like genes may only be functional at the lower temperature of 28 °C. YES swarm medium was chosen as the initial medium to use in screening for complementation. It is an undefined medium and therefore multiple chemoattractants are present for *mcp* complementation. None of the RP8611 strains carrying VF39SM putative *mcp* clones exhibited restored swarming on YES media. Another complementation attempt was made using M9 minimal swarm media with serine as a chemoattractant. Once again none of the transformants regained swarming ability.

There are numerous reasons as to why the complementation did not work: I) The foreign *mcp* genes may not have been expressed properly in *E. coli*, or *E. coli* CheW/CheA was unable to interact with the foreign MCPs. Many researchers have noted that rhizobial genes are poorly transcribed in *E. coli* when expressed from their natural promoters (Fisher *et al.*, 1987; Bae and Stauffer, 1991; Luka *et al.*, 1996). However, such a complementation strategy did identify a Tsr-like MCP from *R. etli* (Mendoza, personal communication), therefore this explanation is questionable. II) In the YES swarm medium the concentrations of chemoattractants were too low to be detected or the proper chemoattractants were not present in the swarm media. In the serine swarm medium the VF39SM Tsr-like MCP was not present amongst the clones mated into RP8611. III) The putative *mcp* genes do not code for MCPs. This is possible for the cosmid clones but highly unlikely for *mcp-1* through *mcp-5* as DNA sequence data strongly suggest that these clones do code for MCPs.

Since complementation studies were unable to confirm the functional nature of the putative *mcp* genes a different strategy was undertaken.

5.3.2 Insertional mutagenesis of *mcp-1* through *mcp-4*.

5.3.2.1 Swarm plate analysis of VF-MCP1$^-$ through VF-MCP4$^-$.

Phenotypic changes in the swarming behaviour of VF39SM strains with mutations in the *mcp*- like genes would confirm that these genes participate in chemotaxis. *mcp-1* through *mcp-4* were chosen as candidates for mutation as the DNA sequencing data suggested they coded for true MCP proteins and could also be used for planning the insertional mutagenesis strategy.

The results of swarm plate analysis of the mutants suggest that at least *mcp-1* plays a role in chemotaxis. The VF39SM *mcp-1* mutant was blocked chemotactically to various sugars and amino acids. Due to the nature of the insertional mutagenesis used in this study polar mutations may have caused the *mcp-1* mutant phenotype. However, this seems unlikely since transposon mutagenesis directly downstream of *mcp*-1 did not result in an impaired chemotactic phenotype and DNA sequencing downstream did not reveal any homologies to known chemotaxis genes (data not shown). The *mcp-1* mutant phenotype is similar to a result reported by Ward *et. al.* (1995). A *mcpA* mutant of *Rhodobacter sphaeroides* lost chemotaxis to a wide range to carbon sources, rather than a discrete set. In addition, this mutation was only apparent under aerobic conditions; the mutant exhibited normal chemotaxis when grown anaerobically. In both cases the phenotype deviates from the *E. coli* paradigm where a MCP will detect a discrete set of attractants or ligands. Consequently, such phenotypic behaviours suggest that some MCPs in the alpha subgroup of the proteobacteria play considerably different roles in chemotaxis than in the

enterics. An alternate explanation is that MCP1 functions in a similar manner to Tar or Tsr of *E. coli*. *E. coli* chemotaxis requires Tar or Tsr to activate CheA's ability to generate tumbles. Without functional Tsr or Tar *E. coli* is chemotactically incompentent. Therefore, perhaps MCP1 is required for activation of CheA and without it VF39SM chemotactic signalling pathway is impaired.

Swarm plate phenotypes for the other *mcp* mutants have not been observed. It should be noted that VF-MCP2$^-$ was unaffected in its ability to swarm towards arginine, erythritol, glutamate, and ornithine. The catabolic genes for all of these carbon sources are localized to the same plasmid as the *mcp-2* gene. A lack of phenotype for the *mcp-2*, *mcp-3* and *mcp-4* mutants may simply result from not testing the appropriate carbon source. Screening the mutants for loss of chemotaxis to previously untested carbon source attractants may be worth pursuing in the future. However, the great metabolic diversity of *Rhizobium* (Parke & Ornston, 1984; reviewed by Stowers, 1985) and the wide range of chemoattractants available for *Rhizobium* (Bowra & Dilworth, 1981; Gaworzewska & Carlile, 1982; Parke *et. al.*, 1985; Armitage *et. al.*, 1988; Dharmatilake & Bauer, 1992; and Robinson and Bauer, 1993), makes this a formidable task. Additionally, the MCPs potentially coded for by *mcp-2, mcp-3* and/or *mcp-4* may detect root exudate compounds not yet identified which would further complicate phenotypic characterizations.

The extensive reiteration observed in *Rhizobium* spp.(Flores *et. al.*, 1987) may provide an alternative explanation for the lack of swarm plate phenotypes. As alluded to earlier, perhaps some *mcp* genes are reiterated with both copies being functional. Insertional mutagenesis would only remove one copy, the other copy would remain functional and mask a mutant phenotype.

5.3.2.2 The roles of genes *mcp-1* through *mcp-4* in nodulation.

Researchers have suggested that motility and chemotaxis play significant roles in competition between rhizobial strains in the legume rhizosphere (Ames and Bergman, 1981; Caetano-Anolles *et al.*, 1988a; and Bauer and Caetano-Anolles, 1990). The results of the competition experiments in this study support these claims and provide insights into the biological function of *mcp-1* and *mcp-2*.

The results of the competition experiments with VF-MCP1$^-$ indicate that without chemotaxis a *Rhizobium* strain is impaired in its ability to compete for sites of nodulation against a wild-type strain. The competition results with *mcp-2* mutants suggest that subtle changes in chemotaxis can greatly affect the competition ability of a rhizobial strain. *mcp-2* mutants have no apparent chemotactic phenotype in swarm plate assays yet this mutation decreases the strain's ability to compete for nodulation sites. It is possible that *mcp-2* detects a specific plant secreted compound and that chemotaxis to this compound is an important part of the legume-rhizobial signal exchange. Future investigations could include methods to identify chemoattractants that are secreted by the host plant. Using this new repertoire of chemoattractants, swarm plate assays with VF-MCP2$^-$ may reveal the ligands of MCP2.

5.4 Regulation of *mcp* -like genes in *R. leguminosarum* VF39SM during nodulation.

The regulation of gene expression during *Rhizobium*'s transition to the bacteroid state is under investigation in the hope of determining the mechanisms of bacteroid differentiation. In light of the fact that motility genes appear to be down regulated during nodulation (Bergman, personal communication) studying the expression of *mcp* genes in *Rhizobium* during

nodulation may prove to be a useful tool for elucidating the mechanisms of bacteroid differentiation.

Before such tools could be used it was necessary to determine that the *mcp*-like genes of VF39SM are down-regulated during nodulation. *mcp-3* was chosen as the gene to study due to convenience. It was necessary to establish a partition locus in the *mcp-3* fusion construct as selection for the fusion plasmids is not possible once the strains are growing on pea plants; tetracycline could cause adverse effects to the host plant, possibly compromising experimental data. The results of the β–galactosidase assays for bacteroids carrying the *mcp-3* fusions clearly indicate that *mcp-3* is down-regulated during nodulation (table 4-4). Down-regulation is greater than 10 fold when compared to β–galactosidase activity when the strain is grown in the free living state. Similarly, western blots of cell lysates isolated from VF39SM bacteroids exhibit no binding activity to the anti-MCP antibody HC23 (data not shown); further supporting the claim that *mcp* genes are not expressed within bacteroids.

Much of the research regarding gene expression in bacteroids has focused on the nitrogen fixation genes (*nif* genes). Consequently portions of the signalling pathway responsible for the expression of these genes have been elucidated and mutants in the signalling pathway are readily available. *fixK*, *fixL*, and *fnr* are genes involved in the regulation of *nif* genes in *R. leguminosarum* (Patschkowski *et al.*, 1996; Schlüter *et al.*, 1992). Strains carrying mutations in each of these genes were used to determine if *mcp-3* regulation is coupled to nitrogen fixation gene regulation. The results indicate that *mcp-3* regulation can be uncoupled from *nif* gene regulation. These results suggest that *mcp-3* regulation is controlled by a different and potentially unidentified signalling pathway. Other genes of *R. leguminosarum*, such as *ropA* are down-regulated in bacteroids and can be uncoupled from *nif* gene activation (de Maagd *et al.*,

1994). Perhaps *mcp-3* shares a common pathway for down-regulation with such genes.

The environmental conditions encountered within the nodule differ greatly from those encountered in the free living environment. It is possible that one or more of these environmental changes trigger *Rhizobium*'s differentiation to the bacteroid state. Bacteroids are dependent upon dicarboxylic acids such as succinate for their main energy source (McRae *et al.*, 1989). Urban and Dazzo (1982) observed that *R. trifolii,* when grown in succinate, undergoes morphological changes causing it to resemble a bacteroid. This work suggests that succinate may be an environmental cue responsible for triggering some of the changes seen in bacteroids. To investigate if succinate is responsible for the down regulation of *mcp*-like genes in VF39SM, *mcp-2* and *mcp-3 lacZ* fusions were grown in succinate. Strains were grown to late log/stationary phase before assaying for β–galactosidase activity. Chemotaxis genes are expressed presumably at highest levels during late log and stationary phases of growth. The β–galactosidase levels of strains grown in the presence of succinate were consistently lower than those grown in the absence of succinate. However, the β–galactosidase activity levels were still very much higher than the levels of activity observed in bacteroids. This result suggests that succinate does not play a role in the down-regulation of the *mcp*-like genes of VF39SM during nodulation.

Future work using the *lacZ* fusion strains may include sectioning of nodules and subsequent *in situ* staining for *lacZ* activity. Such experiments may indicate during which phase of nodulation genes like *mcp-3* are down-regulated. Additionally, fusions to the other *mcp*-like genes should be assayed for activity in bacteroids to ensure that the down-regulation observed with the *mcp-3* fusion is a general phenomenon and not exclusive to *mcp-3*.

5.5 Identification of genes regulating the expression of *mcp* -like genes in *R. leguminosarum* during free living growth.

In the enterics, *mcp* genes are regulated within a hierarchy that includes the genes involved in flagellar assembly and operation (reviewed by Macnab, 1990) . At the head of the hierarchy is an operon which codes for two sigma factors, FlhD and FlhC. Both are essential for the transcription of the motility genes in the enterics. This operon is subject to catabolite control (Komeda *et al.*, 1975). In *Rhizobium* the regulation of motility genes is relatively uncharacterized. The existence of FlhD and FlhC homologues is unknown. Using a chromosomally integrated *lacZ* fusion to *mcp-3* a strategy was developed to help characterize *mcp* gene expression and potentially motility gene expression in general. Of the Tn5 mutants generated in this study none were found to alter *mcp-3* gene expression. However, only 400 mutants were screened; mutations representing the entire genome of VF39SM have not yet been isolated. Future screenings using the strategy developed in this study may yet reveal genes important for the expression of *mcp*-like genes in *R. leguminosarum*.

6.0 References

Adler, J. (1969). Chemoreceptors in Bacteria. *Science* **166**, 1588-1597.

Adler, J., and Epstein, W. (1974). Phosphotransferase-system enzymes as chemoreceptors for certain sugars in *Escherichia coli* chemotaxis. *Proc. Natl. Acad. Sci. USA* **71**, 2895-2899.

Alley, M. R. K., Maddock, J. R., and Shapiro, L. (1992). Polar localization of a bacterial chemoreceptor. *Genes & Dev.* **6**, 825-836.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Ames, P., and Bergman, K. (1981). Competitive advantage provided by bacterial motility in the formation of nodules by *Rhizobium meliloti J Bacteriol* **148**, 728-729.

Armitage, J. P., Gallagher, A., and Johnston, A. W. B. (1988). Comparison of the chemotactic behaviour of *Rhizobium leguminosrarum* with and without the nodulation plasmid. *Mol. Microbiol.* **2**, 743-748.

Armitage, J. P., Havelka, W., and Sockett, R. E. (1990). Methylation-independent taxis in bacteria. In *The Biology of The Chemotactic Response*, pp. 107-134. Edited by J. P. Armitage and J. M. Lackie. Cambridge, UK: Cambridge University Press.

Armitage, J. P. (1992) Beharioral responses in bacteria. *Annu Rev Physiol* **54**, 683-714.

Ashby, A. M., Watson, M. D., and Shaw, C. H. (1987). A Ti-plasmid determinded function is responsible for chemotaxis of *Agrobacterium tumefaciens* towards the plant wound product acetosyringone. *FEMS Microbiol. Lett.* **41**, 189-192.

Ashby, A. M., Watson, M. D., Loake, G. J., and Shaw, C. H. (1988). Ti plasmid-specified chemotaxis of *Agrobacterium tumefaciens* C58C1 toward *vir* - inducing phenolic compounds and soluble factors from monocotyledonous and dicotyledonous plants. *J. Bacteriol.* **170**, 4181-4187.

Bae, Y. M., and Stauffer, G. V. (1991). Mutations that affect activity of the Rhizobium meliloti *trpE(G)* promoter in *Rhizobium meliloti* and *Escherichia coli*. *J. Bacteriol.* **173**, 5831-5836.

Baldani, J. I., Waver, R. W., Hynes, M. F., and Eardly, B. D. (1992). Utilization of carbon substrates, electrophoretic enzyme patterns, and symbiotic performance of plasmid cured clover *Rhizobium*. *Appl Environ Microbiol* **58**, 2308-2314.

Bauer, W.D. and G. Caetano-Anolles. (1990). Chemotaxis, induced gene expression and competitiveness in the rhizosphere. *Plant and Soil* **129**, 45-52.

Bauer, W.D. (1991). Motility and chemotaxis in the life of rhizobia. In: *Plant Biotechnology and Development.* pp 17-24. Edited by P.M. Gresshoff. Ann Arbor, USA: CRC Press.

Berg, H. C., and Brown, D. A. (1972) Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* **239**: 500-504.

Beringer, J. E. (1974). R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* **84**, 188-198.

Brewin, N. J., Wood, E. A., Larkins, A. P., Galfre, G., and Butcher, G. W. (1986). Analysis of lipopolysaccharides from root nodule bacteroids of *Rhizobium leguminosarum* using monoclonal antibodies. *J. Gen. Microbiol.* **132**, 1959-1966.

Boivin, C., Barran, L. R., Malpica, C. A., and Rosenberg, C. (1991). Genetic analysis of a region of the *Rhizobium meliloti* pSym plasmid specifying catabolism of trigonelline, a secondary metabolite present in legumes. *J Bacteriol* **173**, 2809-2817.

Bollinger, J., Park, C., Harayama, S. and Hazelbauer, G. L. (1984). Structure of the Trg protein: homologies with and differences from other sensory transducers of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **81**, 3287-3291.

Bolton, G. W., Nester, E. W., and Gordon, M. P. (1986). Plant phenolic compounds induce expression of the *Agrobacterium tumefaciens* loci needed for virulence. *Science* **232**, 983-985.

Borkovich, K. A., and Simon, M. I. (1990). The dynamics of protein phosphorylation in bacterial chemotaxis. *Cell* **63**, 1339-1348.

Bowra, B. J., and Dilworth, M. J. (1981). Motility and chemotaxis towards sugars in *Rhizobium leguminosarum*. *J. Gen Microbiol* **126**, 231-235.

Boyd, A., Krikos, A., and Simon, M. (1981). Sensory transducers of *E. coli* are encoded by homologous genes. *Cell* **26**, 333-343.

Boyd, A., Kendall, K., and Simon, M. I. (1983). Structure of the serine chemoreceptor in *Escherichia coli*. *Nature* **301**, 623-626.

Brito, B. Palacios, J-M., Ruiz-Argüeso, T., and Imperial, J. (1996). Identification of a gene for a chemoreceptor of the methyl-accepting type in the symbiotic plasmid of *Rhizobium leguminosarum* bv. viciae UPM791. *Biochimica et Biophysica Acta* **1308**, 7-11.

Burn, J., Rossen, L., and Johnston, A. W. B. (1987). Four classes of mutations in the *nodD* gene of *Rhizobium leguminosarum* biovar. *viciae* that affect its ability to autoregulate and/or activate other *nod* genes in the presence of flavonoid inducers. *Genes & Dev.* **1**, 456-464.

Caetano-Anollés, G., and Bauer, W. D. (1988). Feedback regulation of nodule formation in alfalfa. *Planta* **175**, 546-557.

Caetano-Anollés, G., Wall, L. G., Micheli, A. T. D., Macchi, E. M., Bauer, W. D., and Favelukes, G. (1988a). Role of motility and chemotaxis in efficiency of nodulation by *Rhizobium meliloti*. *Plant Physiol.* **86**, 1228-1235

Caetano-Anollés, G., Crist-Estes, D. K., and Bauer, W. D. (1988b). Chemotaxis of *Rhizobium meliloti* to the plant flavone luteolin requires functional nodulation genes. *J. Bacteriol.* **170**, 3164-3169.

Caetano-Anollés, G., Wrobel-Boerner, E., and Bauer, W. D. (1992). Growth and movement of spot inoculated *Rhizobium meliloti* on the root surface of alfalfa. *Plant. Physiol.* **98**, 1181-1189.

Cangelosi, G. A., Ankenbauer, R. G., and Nester, E. W. (1990). Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. *Proc. Natl. Acad. Sci. USA* **87**, 6708-6712.

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

Clegg, D. O., and Koshland, D. E., Jr. (1984). The role of a signal protein in bacterial sensing: behavioral effects of increased gene expression. *Proc. Natl. Acad. Sci. USA* **81**, 5056-5060.

Conley, M. P., Wolfe, A. J., Blair, D. F., and Berg, H. C. (1989). Both CheA and CheW are required for reconstitution of signalling in bacterial chemotaxis. *J. Bacteriol.* **171**, 5190-5193.

Currier, W. W., and Strobel, G. A. (1977). Chemotaxis of *Rhizobium spp.* to a glycoprotein produced by birdsfoot trefoil roots. *Science* **196**, 434-435.

David, M., Daveran, M. L., Batut, J., Dédieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P., and Kahn, D. (1988). Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. *Cell* **54**, 671-683.

de Maagd, R. R., Rijik, R. D., Mulders, I. H. M., and Lugtenberg, B. J. J. (1989). Immunological characterization of *Rhizobium leguminosarum* outer membrane antigens by use of polyclonal and monoclonal antibodies. *J. Bacteriol.* **171**, 1136-1142.

de Maagd, R. R., Yang, W-C., Goosen-de Roo, L., Mulders, I. H. M., Roest, H. P., Spaink, H. P., Bisseling, T., and Lugtenberg, B. J. J. (1994). Down-regulation of expression of the *Rhizobium leguminosarum* outer membrane protein gene *ropA* occurs abruptly in interzone II-III of pea nodules and can be uncoupled from *nif* gene activation. *Mol. Plant-Microbe Interact.* **7**, 276-281.

Debelle, A., and Sharma, S. B. (1986). Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity. *Nucleic Acids Res.* **14**, 7453-7472.

Deckers, H. M., and Voordouw, G. (1994). Identification of a large family of genes for putative chemoreceptor proteins in an ordered library of the *Desulfovibrio vulgaris* Hildenborough genome. *J Bacteriol* **176**, 351-358.

Deckers, H. M., and Voordouw, G. (1996). The *dcr* gene family of *Desulfovibrio*: implications from the sequence of *dcrH* and phylogenetic comparison with other *mcp* genes. *Antonie van Leeuwenhoek* **70**, 21-29.

Deguchi, Y., Yamato, I., and Anraku, Y. (1989). Molecular cloning of *gltS* and *gltP*, which encode glutamate carriers of *Escherichia coli* B. *J. Bacteriol.* **171**, 1314-1319.

Dénarié, J., Rosenberg, C., Boistard, P., Truchet, G., and Casse-Delbart, F. (1981). Plasmid control of symbiotic properties in *Rhizobium meliloti*. In *Current Perspectives in Nitrogen Fixation. Proceedings of the Fourth International Symposium on Nitrogen Fixation*, pp. 137-141. Edited by A. H. Gibson and W. E. Newton. Amsterdam, Holland. Elsevier/North-Holland Biomedical Press.

Dharmatilake, A. J., and Bauer, W. D. (1992). Chemotaxis of *Rhizobium meliloti* towards nodulation gene-inducing compounds from alfalfa roots. *Appl. Envrion. Microbiol.* **58**, 1153-1158.

Dolla, A., Rongdian, F., Brumlik, M. J., and Voordouw, G. (1992). Nucleotide sequence of *dcr* A, a *Desulfovibrio vulgaris* Hildenborough chemoreceptor gene, and its expression in *Escherichia coli. J Bacteriol* **174**, 1726-1733.

Downie, J. A., Knight, C. D., Johnston, A. W. B., and Rossen, L. (1985). Identification of genes and gene products involved in the nodulation of peas by *Rhizobium leguminosarum. Mol. Gen. Genet.* **198**, 255-262.

Downie, J. A. (1994). Signalling strategies for nodulation of legumes by rhizobia. *Trends in Microbiol.* **2**, 318-324.

Eckhardt, T. (1978). A rapid method for the identification of plasmid deoxyribonucleic acid in bacteria. *Plasmid* **1**, 584-588.

Egelhoff, T. T., Fisher, R. F., Jacobs, T. W., Mulligan, J. T., and Long, S. R. (1985). Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC. DNA* **4**, 241-248.

Everiss, K. D., Hughes, K. J., Kovach, M. E., and Peterson, K. M. (1994). The *Vibrio cholerae acfB* colonization determinant encodes an inner membrane protein that is related to a family of signal-transducing proteins. *Infect. Immun.* **62**, 3289-3298.

Fellay, R., Frey, J., and Krisch, H. (1987). Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vitro insertional mutagenesis of gram-negative bacteria. *Gene* **52**, 147-154.

Fellay, R., Rochepeau, P., Relic, B., and Broughton, W. J. (1995). Signals to and emanating from *Rhizobium* largely control symbiotic specificity. In *Pathogenesis and Host Specificity in Plant Diseases. Histopathological, Biochemical, Genetic and Molecular Bases. Vol I: Prokaryotes*, pp. 199-220. Edited by U. S. Singh, R. P. Singh, and K. Kohmoto. Oxford, UK: Pergamon/Elsevier Science Ltd.

Finan, T.M., Hirsch, A. M., Leigh, J. A., Johansen, E., Kaldau, G. A., Deegan, S., Walker, G. C., and Signer, E. R. (1985). Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. *Cell* **40**, 869-877.

Fisher, R. F., Brierley, H. L., Mulligan, J. T., and Long, S. R. (1987). Transcription of Rhizobium meliloti nodulation genes. Identification of a *nodD* transcription initiation site *in vitro* and *in vivo*. *J. Biol. Chem.* **15**, 6849-6855.

Flores, M., Gonzalez, V., Brom, S., Martinez, E., Pinero, D., Romero, D., Davila, G., Palacios, R. (1987). Reiterated DNA sequences in *Rhizobium* and *Agrobacterium* spp. *J. Bacteriol.* **169**, 5782-5788.

Foster, D. L., Mowbray, S. L., Jap, B. K. and Koshland, E. E. Jr. (1985). Purification and characterization of the aspartate chemoreceptor. *J. Biol. Chem.* **260**, 14850-14858.

Freiberg, C., Fellay, Bairoch, A., Broughton, W. J., Rosenthal, A., and Perret, X. (1997). Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**, 394-401.

Gajendiran, N., and Mahadevan, A. (1990). Plasmid-borne catechol dissimilation in *Rhizobium* sp. *FEMS Microbiol. Ecol.* **73**, 125-130.

Gaworzewska, E. T., and Carlile, M. J. (1982). Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. *J Gen Microbiol* **128**, 1179-1188.

Gegner, J. A., Graham, D. R., Roth, A. F., and Dahlquist, F. W. (1992). Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**, 975-982.

Greck, M., Platzer, J., Sourjik, V., & Schmitt, R. (1995). Analysis of a chemotaxis operon in *Rhizobium meliloti*. *Mol Microbiol* **15**, 989-1000.

Goldmann, A., Boivin, C., Fleury, V., Message, B., Lecoeur, L., Maille, M., and Tepfer, D. (1991) Betaine use by rhizosphere bacteria: genes essential for trigonelline, stachydrine, and carnitine catabolism in *Rhizobium meliloti* are located on pSym in the symbiotic region. *Molec. Plant-Microbe Interac.* **4**, 571-578.

Goy, M. F., Springer, M. S., and Adler, J. (1977). Sensory transduction in *Escherichia coli* : role of a protein methylation reaction in sensory adaptation. *Proc. Natl. Acad. Sci. USA* **74**, 4964-4968.

Gulash, M., Ames, P., Larosiliere, R. C., and Bergman, K. (1984). Rhizobia are attracted to localized sites on legume roots. *Appl Environ Microbiol* **48**, 149-152.

Harkey, G. W., Everiss, K. D., and Peterson, K. M. (1994). The *Vibrio cholerae* toxin-coregulated-pilus gene *tcpI* encodes a homolog of methyl-accepting chemotaxis proteins. *Infect. Immun.* **62**, 2669-2678.

Hawes, M. C., and Smith, L. Y. (1989). Requirement for chemotaxis in pathogenicity of *Agrobacterium tumefaciens* on roots of soil grown pea plants. *J. Bacteriol.* **171**, 5688-5671.

Hayashi, H., Koiwai, O., and Kozuka, M. (1979). Studies on bacterial chemotaxis. Effect of *cheB* and *cheZ* mutations on the methylation of methyl-accepting chemotaxis protein II of *Escherichia coli*. *J. Biochem.* **85**, 1213-1223.

Hazelbauer, G. L., Engstrom, P. and Harayama, S. (1981). Methyl-accepting chemotaxis protein III and transducer gene *trg*. *J. Bacteriol.* **145**, 43-49.

Hazelbauer, G. L., Yaghmai, R., Burrows, G. G., Baumgartner, J. W., Dutton, D. P., and Morgan, D. G. (1990). Transducers: Transmembrane receptor proteins involved in bacterial chemotaxis. In *The Biology of The Chemotactic Response*, pp. 107-134. Edited by J. P. Armitage and J. M. Lackie. Cambridge, UK: Cambridge University Press.

Hess, J. F., Oosawa, K., Matsumura, P., and Simon, M. I. (1987). Protein phosphorylation is involved in bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA* **84**, 7609-7613.

Hess, J. F., Bourret, R. B., and Simon, M. I. (1988a). Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. *Nature*

Hess, J. F., Oosawa, K., Kaplan, N, and Simon, M. I. (1988b). Phophorylation of three proteins in the signalling pathway of bacterial chemotaxis. *Cell* **53**, 79-87.

Holsters, M., Silva, B., Van Vliet, F., Genetello, C., De Block, M., Dhaese, P., Depicker, A., Inze, D., Engler, G., Villarroel, R., Van Montagu, M., and Schell, J. (1980). The functional organization of the nopaline A. tumefaciens plasmid pTiC58. *Plasmid* **3**, 212-230.

Hynes, M. F., Simon, R., and Pühler, A. (1985). The development of plasmid-free strains of *Agrobacterium tumefaciens* by using incompatibility with a *Rhizobium meliloti* plasmid to eliminate pAEC58. *Plasmid* **13**, 99-105.

Hynes, M. F., and McGregor, N. F. (1990). Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. *Mol Microbiol* **4**, **567-574**

Hynes, M. F., and Finan, T. M. (1998). Genetics: general genetic knowledge. In *The Rhizobiaceae*. Edited by H. Spaink, P. J. J. Hooykaas, and A, Kondorosi. NL: Kluwer Academic Publishers. in press

Innes, R. W., Kuempl, P. L., Plazinski, J., Canter-Cremers, H., Rolfe, B. G., and Djordjevic, M. A. (1985). Plant factors induce expression of nodulation and host-range genes in *Rhizobium trifolii*. *Mol. Gen. Genet.* **201**, 426-432.

Jacobs, M. H. J., van de Heide, T., Tolner, B., Driessen, A., J., M., and Konings, W. N. (1995). Expression of the *gltP* gene of *Escherichia coli* in a glutamtae transport-deficient mutant of *Rhodobacter sphaeroides* restores chemotaxis to glutamate. *Mol. Microbiol.* **18**, 641-647.

Johnston, A. W. B., Beynon, J. L., Buchanan-Wollaston, A. B., Setchell, S. M., Hirsch, P. R., and Beringer, J. E. (1978). High frequency transfer of nodulating ability between strains and species of *Rhizobium*. *Nature* **276**, 634-636.

Kannenberg, E. L., and Brewin, N. J. (1989). Expression of a cell surface antigen from *Rhizobium leguminosarum* 3841 is regulated by oxygen and pH. *J. Bacteriol* **171**, 4543-4548.

Kannenberg, E. L., and Brewin, N. J. (1994). Host-plant invasion by *Rhizobium* : the role of cell-surface components. *Trends in Microbiol.* **2**, 277-283.

Kigne, J. W., Diaz, C. L., and dePater, B. S. (1990). Surface interactions between rhizobia and legume root hairs. In *Nitrogen Fixation: Objectives and Acheivements.* pp 187-192

Kim, K. S., Sackett, R. L., and Farrand, S. K. (1996). Gene sequence of ORF2 directly submitted to GenBank. National Center for Biotechnology Information, Bethesda, MD.

Komeda, Y., Suzuki, H., Ishidsu, J-I., and Iino, T. (1975). The role of cAMP in flagellation of *Salmonella typhimurium. Mol. Gen. Genet.* **142**, 289-298.

Kort, E. N., Goy, M. F., Larsen, S. H., and Adler, J. (1975). Methylation of a membrane protein involved in bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA* **72**, 3939-3943.

Kossmann, M., Wolff, C., and Manson, M. D. (1988). Maltose chemoreceptor of *Escherichia coli* : interaction of maltose-binding protein and the Tar signal transducer. *J. Bacteriol.* **170**, 4516-4521.

Krikos, A., Mutoh, N., Boyd, A. and Simon, M. I. (1983). Sensory transducers of *Escherichia coli* are composed of discrete structural and functional domains. *Cell* **33**, 615-622.

Kuo, S. C. and Koshland, E. E. Jr. (1987). Roles of *cheY* and *cheZ* gene products in controlling flagellar rotation in bacterial chemotaxis. *J. Bacteriol.* **169**, 51-59.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Larsen, S. H., Reader, R. W., Kort, E. N., Tso, W. W., and Adler, J. (1974) Change in direction of flagellar rotation is the basisi of the chemotactic response in *Escherichia coli*. *Nature* **249**, 74-77.

Lengeler, J. W., Auburger, A.-M., Mayer, R. and Pecher, A. (1981). The phosphoenolpyruvate-dependent carbohydrate: phosphotransferase system enzyme II as chemoreceptors in chemotaxis of *Escherichia coli* K12. *Mol. Gen. Genet.* **183**, 163-170.

Lengeler, J. W., and Bogler, A. P. (1989). Molecular mechanisms of bacterial chemotaxis towards PTS-carbohydrates. *FEMS Microbiol. Rev.* **63**, 81-92.

Lengeler, J. W., Bockmann, J., Heuel, H., and Titgemeyer, F. (1992). In *Molecular Mechanisms of Transport*, pp. 77-85. Edited by E. Quagliariello and F. Palmieri. New York, USA: Elsevier Science Publishers.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Prome, J. C., and Denarie, J. (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781-784.

Liu, J., and Parkinson, J. S. (1989). Role of CheW protein in coupling membrane receptors to the intracellualr signaling system of bacterial chemotaxis. *Proc. Natl. Acad. Sci USA* **86**, 8703-8707.

Luka, S, Patriarca, E. J, Riccio, A., Iaccarino, M., and Defez, R. (1986). Cloning of the *rpoD* analog from *Rhizobium etli*: *sigA* of *R. etli* is growth phase regulated. *J. Bacteriol.* **178**, 7138-7143.

Lux, R. Jahreis, K., Bettenbrock, K., Parkinson, J. S., and Lengeler, J. W. (1995). Coupling the phosphotransferase system and the methyl-accepting chemotaxis protein-dependent chemotaxis signaling pathways of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **92**, 11583-11587.

Macnab, R. M. (1990). Genetics, structure and assembly of the bacterial flagellum. In *The Biology of The Chemotactic Response*, pp. 77-106. Edited by J. P. Armitage and J. M. Lackie. Cambridge, UK: Cambridge University Press.

McRae, D. G., Miller, R. W., Berndt, W. B., and Joy, K. (1989). Transport of C4-dicarboxylic and amino acids by *Rhizobium meliloti* bacteroids. *Mol. Plant-Microbe Interact.* **2**, 273-278.

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

Miller, J. H. (1972). *Experiments in Molecular Genetics.*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Mitsch, M. J. (1995). *Thesis*, University of Calgary, Calgary, Alberta, Canada.

Morgan, D. G., Baumgartner, J. W., and Hazelbauer, G. L. (1993). Proteins antigenically related to methyl-accepting chemotaxis proteins of *E. coli* detected in a wide range of bacterial species. *J Bacteriol* **175**, 133-140.

Morrison, T. B., and Parkinson, J. S. (1997). A fragment liberated from the *Escherichia coli* CheA kinase that blocks stimulatory, but not inhibitory, chemoreceptor signaling. *J. Bacteriol.* **179**, 5543-5550.

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

Munoz Aguilar, J. M., Ashby, A. M., Richards, A. J. M., Loake, G. J., Watson, M. D., and Shaw, C. H. (1988). Chemotaxis of *Rhizobium leguminosarum biovar phaseoli* towards flavonoid inducers of symbiotic nodulation genes. *J Gen Microbiol* **134**, 2741-2746.

Napoli, C. A., and Hubbell, D. H. (1975). Ultrastructure of *Rhizobium* - induced infection threads in clover root hairs. *Appl. Microbiol.* **30**, 1003-1009.

Nester, E. W., Gordon, M. P., Amasion, R. M., and Yanofsky, M. F. (1984). Grwon gall: a molecular and physiolgical analysis. *Annu. Rev. Plant Physiol.* **35**, 387-413.

Nuti, M. P., Lepidi, A. A., Prakash, R. K., Schilperoort, R. A., and Cannon, F. C. (1979). Evidence for nitrogen fixation (*nif*) genes on indigenous *Rhizobium* plasmids. *Nature* **282**, 533-535.

Oosawa, K., and Imae, Y. (1983). Glycerol and ethylene glycol: members of a new class of repellents of *Escherichia coli* chemotaxis. *J. Bacteriol.* **154**, 104-112.

Oosawa, K., and Imae, Y. (1984). Demethylation of methyl-accepting chemotaxis proteins in *Escherichia coli* induced by the repellents glycerol and ethylene glycol. *J. Bacteriol.* **157**, 576-581.

Palmer, A. C. V., and Shaw, C. H. (1992). The role of VirA and VirG phosphorylation in chemotaxis towards acetosyringone by *Agrobacterium tumefaciens*. *J. Gen. Microbiol.* **138**, 2509-2514.

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

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

Parke, D., Ornston, L. N., and 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*. *J. Bacteriol.* **169**, 5336-5338.

Parkinson, J. S. (1978). Complementation analysis and deletion mapping of *Escherichia coli* mutants defective in chemotaxis. *J. Bacteriol.* **135**, 45-53.

Parkinson, J. S., Parker, S. R., Talbert, P. B., and Houts, S. E. (1983). Interactions between chemotaxis and flagellar gene in *Escherichia coli*. *J. Bacteriol.* **155**, 265-274.

Parkinson, J. S. (1993). Signal transduction schemes of bacteria. *Cell* **73**, 857-871.

Patel, J. J., and Yang, A. F. (1981). Light and electron microscopic studies of nodule structure of alfalfa. *Can. J. Mircrobiol.* **27**, 36-43.

Patschkowski, T., Schlüter, A., and Priefer, U. B. (1996). *Rhizobium leguminosarum* bv. *viciae* contains a second *fnr/fixK* -like gene and an unusual *fixL* homologue. *Mol. Microbiol.* **21**, 267-280.

Pecher, A. , Tenner, I., and Lengeler, J. (1983). The phosphoenolpyruvate-dependent carbohydrate: phosphotransferase system enzymes II, a new class of chemosensors in bacterial chemotaxis. In *Mobility and recognition in cell biology*, pp.517-531. Edited by H. Sund and C. Veeger. Berlin, Germany, Wlter de Gruyter and Co.

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

Postma, P. W., and Lengeler, J. W. (1985). . *Microbiol. Rev.* **49**, 232-269.

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

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

Ravid, S., Matsumura, P., and Eisenbach, M. (1986). Restoration of flagellar clockwise roteation in bacterial envelopes by insertion of the chemotaxis protein CheY. *Proc. Natl. Acad. Sci. USA* **83**, 7157-7161.

Rebbapragada, A., Johnson, M. S., Harding, G. P., Zuccarelli, A. J., Fletcher, H. M., Zhulin, I. B., and Taylor, B. L. (1997). The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox and energy signals for *Escherichia coli* behavior. *Proc. Natl. Acad. Sci. USA* **94**, 10541-10546.

Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., and Rolfe, B. G. (1986). Flavones induce expression of nodulation genes in *Rhizobium*. *Nature* **323**, 632-635.

Reizer, J., Saier, M. H., Deutscher, J., Grenier, F., Thompson, J., and Hengstenberg, W. (1988). .*Crit. Rev. Microbiol.* **15**, 297-338.

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

Robinson, J.B. and Bauer, W. D.(1993). Chemotaxis of Rhizobium meliloti towards C4-dicarboxylic acids and its relation to carboxylate transport functions. *J. Bacteriol.* **175**, 2284-2291.

Rossen, L., Johnston, A. W. B., and Downie, J. A. (1984). DNA sequence of the *R. leguminosarum* nodulation genes *nodA,B* and *C* required for root hair curling. *Nucleic Acids Res.* **12**, 9497-9508.

Rossen, L., Shearman, C. A., Johnston, A. W. B., and Downie, J. A. (1985). The *nodD* gene of *R. leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nodABC* genes. *EMBO J.* **4**, 3369-3374.

Rossenberg, C., Boistard, P., Dénarié, J., and Casse-Delbart, F. (1981). Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol. Gen. Genet.* **184**, 326-333.

Roth, L. E., and Stacey, G. (1989). Bacterium release into host cells of nitrogen-fixing soybean nodules: the symbiosome membrane comes from three sources. *Eur. J. Cell Biol.* **49**, 13-23.

Rowsell, E. H., Smith, J. M, Wolfe, A., and Taylor, B. L. (1995). CheA, CheW, and CheY are required for chemotaxis to oxygen and sugars of the phosphotransferase system in *Escherichia coli*. *J. Bacteriol.* **177**, 6011-6014.

Russo, A. F. and Koshland, D. E. Jr. (1983). Separation of signal transduction and adaptation functions of the aspartate receptor in bacterial sensing. *Science* **220**, 1016-1020.

Saint, C. P., Wexler, M., Murphy, P. J., Tempe, J., Tate, M. E., and Murphy, P. J. (1993). Characterization of genes for synthesis and catabolism of a newrhizopine induced in nodules by Rhizobium meliloti Rm220-3: extension of the rhizopine concept. *J. Bacteriol.* **175**, 5205-5215.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory

Sanders, D. A., Gillece-Castro, B. L., Stock, A. M., Burlingame, A. L., and Koshland, D. E., Jr. (1989). Identification of the site of phosphorylation of the chemotaxis response regulator CheY. *J. Biol. Chem.* **264**, 21770-21778.

Schlüter, A., Patschkowski, T., Unden, G., and Priefer, U. B. (1992). The *Rhizobium leguminosarum* FnrN protein is functionally similar to *Escherichia coli* Fnr and promotes heterologous oxygen-dependent activation of transcription. *Mol. Microbiol.* **6**, 3395-3404.

Schlüter, A., Patschkowki, T., Quandt, J., Selinger, L. B., Weidner, S., Krämer, M., Zhou, L., Hynes, M. F., and Priefer, U. B. (1997). Functional and regulatory analysis of the two copies of the *fixNOQP* operon of *Rhizobium leguminosarum* strain VF39. *Mol. Plant Microbe Interact.* **10**, 605-616.

Schofield, P. R., and Watson, J. M. (1986). DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. *Nucleic Acids Res.* **14**, 2891-2903.

Schweizer, H. P. (1993). Two plasmids, X1918 and Z1918, for easy recovery of the xylE and lacZ reporter genes. *Gene* **134**, 89-91.

Segall, J. Ishihara, A. and Berg, H. C. (1985). chemotactic signalling in filamentous cells of *Escherichia coli*. *J. Bacteriol.* **161**, 51-59.

Shaw, C. H. (1991). Swimming against the tide: chemotaxis in *Agrobacterium*. *BioEssays* **13**, 25-29.

Shearman, C. A., Rossen, L., Johnston, A. W. B., and Downie, J. A. (1986). The *Rhizobium leguminosarum* nodulation gene *nodF* encodes a protein similar to acyl-carrier protein and is regulated by *nodD* plus a factor in pea root exudate. *EMBO J.* **5**, 647-652.

Silverman, M., and Simon, M. (1974a). Flagellar rotation and the mechanism of bacterial motility. *Nature* **249**, 73-74.

Silverman, M., and Simon, M. (1974b). Characterization of *Escherichia coli* flagellar mutants that are insensitive to catabolite repression. *J. Bacteriol.* **120**, 1196-1203.

Silverman, M., and Simon, M. (1977). Chemotaxis in *Escherichia coli*: methylation of *che* gene products. *Proc. Natl. Acad. Sci. USA* **74**, 3317-3321.

Sindhu, S. S., Brewin, N. J., and Kannenberg, E. L. (1990). Immunochemical analysis of lipopolysaccharides from free-living and endosymbiotic forms of *Rhizobium leguminosarum*. *J. Bacteriol.* **172**, 1804-1813.

Slocum, M. K., and Parkinson, J. S. (1983). Genetics of methyl-accepting chemotaxis proteins in *Escherichia coli*: organization of the *tar* region. *J. Bacteriol.* **155**, 565-577.

Sockett, R. E., Armitage, J. P., and Evans, M. C. W. (1987). Methylation-independent and methylation-dependent chemotaxis in *Rhodobacter sphaeroides* and *Rhodospirillum rubrum*. *J. Bacteriol.* **169**, 5808-5814.

Soto, M. J, Zorzano, A., Garcia-Rodriguez, F. M., Mercado-Blanco, J., Lopez-Lara, I. M., Olivares, J., and Toro, N. (1994). Identification of a novel Rhizobium meliloti nodulation efficiency *nfe* gene homolog of Agrobacterium ornithine cyclodeaminase. *Mol. Plant Microbe Interact.* **7**, 703-707.

Sourjik, V., and Schmitt, R. (1996). Different roles of CheY1 and CheY2 in the chemotaxis of *Rhizobium* meliloti. *Mol. Microbiol.* **22**, 427-436.

Spaink, H. P., Sheeley, D. M., van Brussel, A. A., Glushka, J., York, W. S., Tak, T., Geiger, O., Kennedy, E. P., Reinhold, V. N., and Lugtenberg, B. J. (1991). A novel highly unsarturated fatty acid moiety of lipo-oligosaccharide signals determiens host specificity of *Rhizobium*. *Nature* **354**, 125-130.

Springer, W. R., and Koshland, D. E. Jr. (1977). Identification of a protein methyltransferase as the *cheR* gene product. *Proc. Natl. Acad. Sci. USA* **74**, 533-537.

Springer, M. S., Goy, M. F., and Adler, J. (1979). Protein methylation in behavioural control mechanisms and in signal transduction. *Nature* **280**, 279-284.

Stachel, S. E., Messens, E., wan Montagu, M. and Zambryski, P. (1985). Identification of the signal molecules produced by wounded plant cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. *Nature* **318**, 624-629.

Stachel, S. E., and Zambryski, P. (1986). *virA* and *virG* control the plant induced activation of the T-DNA transfer process of *A. tumefaciens*. *Cell* **46**, 325-333.

Stock, J. B., and Koshland, D. E. Jr. (1978). A protein methylesterase involved in bacterial sensing. *Proc. Natl. Acad. Sci. USA* **75**, 3659-3663.

Stock, A. M., Wylie, D. C., Mottonen, J. M., Lupas, A. N., Ninfa, A. J., Schutt, C. E., and Stock, J. B. (1988) Phosphoproteins involved in bacterial signal transduction. *Cold Spring Harb. Symp. Quant. Biol.* **53**, 49-57.

Stokkermans, T. J., and Peters, N. K. (1994). *Bradyrhizobium elkanii* lipo-oligosaccharide signals induce complete nodule structures on Glycine soja Siebold et Zucc. *Planta* **193**, 413-420.

Stowers, M. D. (1985). Carbon metabolism in *Rhixobium* species. *Ann Rev Microbiol* **39**, 89-108.

Temper, D., Goldman, A., Pamboukdjian, N., Maille, M., Lepingle, A., Chevalier, D., Denaire, J., and Rosenberg, C. (1988). A plasmid of *Rhizobium meliloti* 41 encodes catabolism of two compounds from root exudate of *Calystegium sepium*. *J Bacteriol* **170**: 1153-1161.

Torok, I., Kondorosi, E., Stepkowski, T., Posfai, J., and Kondorosi, A. (1984). Nucleotide sequence of *Rhizobium meliloti* nodulation genes. *Nucleic Acids Res.* **12**, 9509-9524.

Tu, J. C. (1975). Structural similarity of the membrane envelopes of rhizobial bacteroids and the host plasma membrane as revealed by freeze-fracturing. *J. Bacteriol.* **122**, 691-694.

Tully, R. E., and Terry, M. E. (1985). Decreased exopolysaccharide synthesis by anaerovic and symbiotic cells of *Bradyrhizobium japonicum*. *Plant Physiol.* **79**, 445-450.

Turgeon, B. G. and Bauer, W. D. (1982). Early events in the infection of soybean by *Rhizobium japonicum* II. Time course and cytology of the initial infection process. *Can. J. Bot.* **60**, 152-161.

Turgeon, B. G. and Bauer, W. D. (1985). Ultrastructure of infection thread development during the infection of soybean by *Rhizobium japonicum*. *Planta* **163**, 328-349

Turner, A. K., De La Cruz, F., and Grinsted, J. (1990). Temperature sensitivity of transposition of class II transposons. *J. Gen. Microbiol.* **136**, 65-67.

Urban, J. E., and Dazzo, F. B. (1982). Succinate-induced morphology of *Rhizobium trifolii* 0403 resembles that of bacteroids in clover nocules. *Appl. Envrion. Microbiol.* **44**, 219-226.

Van Brussel, a. A. N., Planque, K., and Quispel, A. (1977). The wall of *Rhizobium leguminosarum* in bacteroid and free-living forms. *J. Gen. Microbiol.* **101**, 51-56.

Van Larebeke, N., Engler, G., Holsters, M., Van den Elsacker, S., Zaenen, I., Schilperoort, R. A. , and Schell, J. (1974). Large plasmid in Agrobacterium tumefaciens essential for crown gall-inducing ability. *Nature* **252**, 169-170.

Vande Broek, A. and Vanderleyden, J. (1995). The role of bacterial motility, chemotaxis, and attachment in bacteria-plant interactions. *Mol Plant-Microbe Interact* **8**, 800-810.

VandenBosch, K. A., Brewin, N. J., and Kannenberg, E. L. (1989). Developmental regulation of a *Rhizobium* cell surface antigen during the growth of pea root nodules. *J. Bacteriol.* **171**, 4537-4542.

Vesper, S. J. and Bauer, W. D. (1985). Studies on the attachment of rhizobia to soybean roots. *Symbiosis* **1**, 139-162.

Vincent, J. M. (1970). A manual for the practical study of root-nodule bacteria. IBF handbook No. 15 Blackwell Scientific Publications, Oxford.

Wang, Y0P., Birkenhead, K., Boesten, B., Manian, S., and O'Gara, F. (1989). Genetic analysis and regulation of the *Rhizobium meliloti* genes controlling C_4-dicarboxylic acid transport. *Gene* **85**, 135-144.

Ward, M. J., Bell, A. W., Hamblin, P. A., Packer, H. L., and Armitage, J. P. (1995a). Identification of a chemotaxis operon with two *cheY* genes in *Rhodobacter sphaeroides*. *Mol. Microbiol.* **17**, 357-366.

Ward, M. J., Harrison, D. M., Ebner, M. J., and Armitage, J. P. (1995b). Identification of a methyl-accepting chemotaxis protein in *Rhodobacter sphaeroides*. *Mol Microbiol* **18**, 115-121.

Weinstein, M., Roberts, R. C., and Helindki, D. R. (1992). A region of the broad-host range plasmid RK2 causes stable *in planta* inheritance of plasmids in *Rhizobium meliloti* cells isolated from alfalfa root nodules. *J. Bacteriol.* **174**, 7486-7489.

Yamamoto, K., Macnab, R., and Imae, Y. (1990). Repellent response functions of the Trg and Tap chemoreceptors of *Escherichia coli*. *J. Bacteriol.* **172**, 383-388.

Zhang, W., Brooun, A., McCandless, J., Banda, P., and Alam, M. (1996). Signal transduction in the Archaeon *Halobacterium salinarium* is processed through three subfamilies of 13 soluble and membrane-bound transducer proteins. *Pro. Natl Acad Sci USA* **93**, 4649-4654.

Zhulin, I. B., Rowsell, E. H, Johnson, M. S., and Taylor, B. L. (1997). Glycerol elicits energy taxis of *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **179**, 3196-3201.

Appendices

1.0 Bacteriological media and other commonly used solutions.

LB

Tryptone	10.0g
Yeast extract	5.0 g
NaCl	10.0g
H_2O	1.0 L

(12.5g agar for solid media)

TY

Tryptone	5.0g
Yeast extract	3.0g
$CaCl_2$	0.5g
H_2O	1.0 L

PH

Peptone	4.0g
Yeast extract	0.5g
Tryptone	0.5g
$CaCl_2$	0.2g
$MgSO_4$	0.2g
H_2O	1.0 L

VMM

Solution A

K_2HPO_4	1.0g
KH_2PO_4	1.0g
H_20	1.0 L

Solution B

$FeCl_3$	0.1g
$MgSO_4$	2.5g
$CaCl_2$	1.0g
H_20	1.0 L

Solution C

Biotin	0.01g
Thiamine	0.01g
Ca Pantathenate	0.01g
H_20	1.0 L

890 mL of Solution A is combined with 100 mL of Solution B, 10 mL of Solution C and appropriate carbon and nitrogen sources if desired.

M9 Minimal Media

Tris borate buffer

Tris base	216.0g
Boric acid	110.0g
EDTA	14.88g
H_20	20.0 L

Plant medium

Solution A

$CaCl_2$	294.0g
H_20	1.0 L

Solution B

KH_2PO_4	136.0g
H_20	1.0 L

Solution C

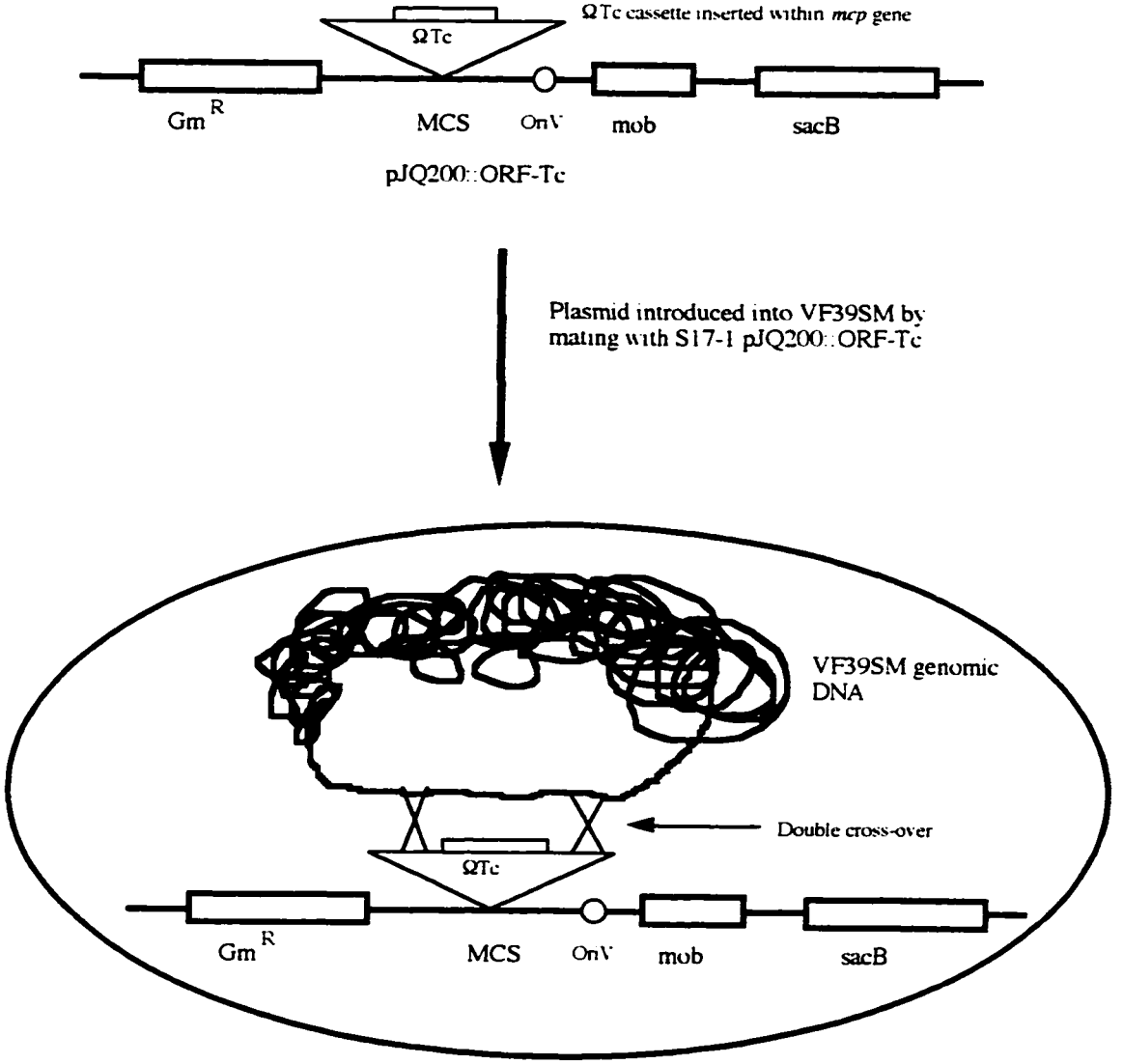
Fe-citrate	6.7g
H_20	1.0 L

Solution D

$MgSO_4$	123.0g
K_2SO_4	87.0g
$MnSO_4$	0.338g
H_2BO_4	0.247g
$ZnSO_4$	0.288g
$CuSO_4$	0.1g
$CoSO_4$	0.056g
Na_2MoO_4	0.048g
H_20	1.0 L

1.0 mL of each solution is added to 2.0 L of dH_20.

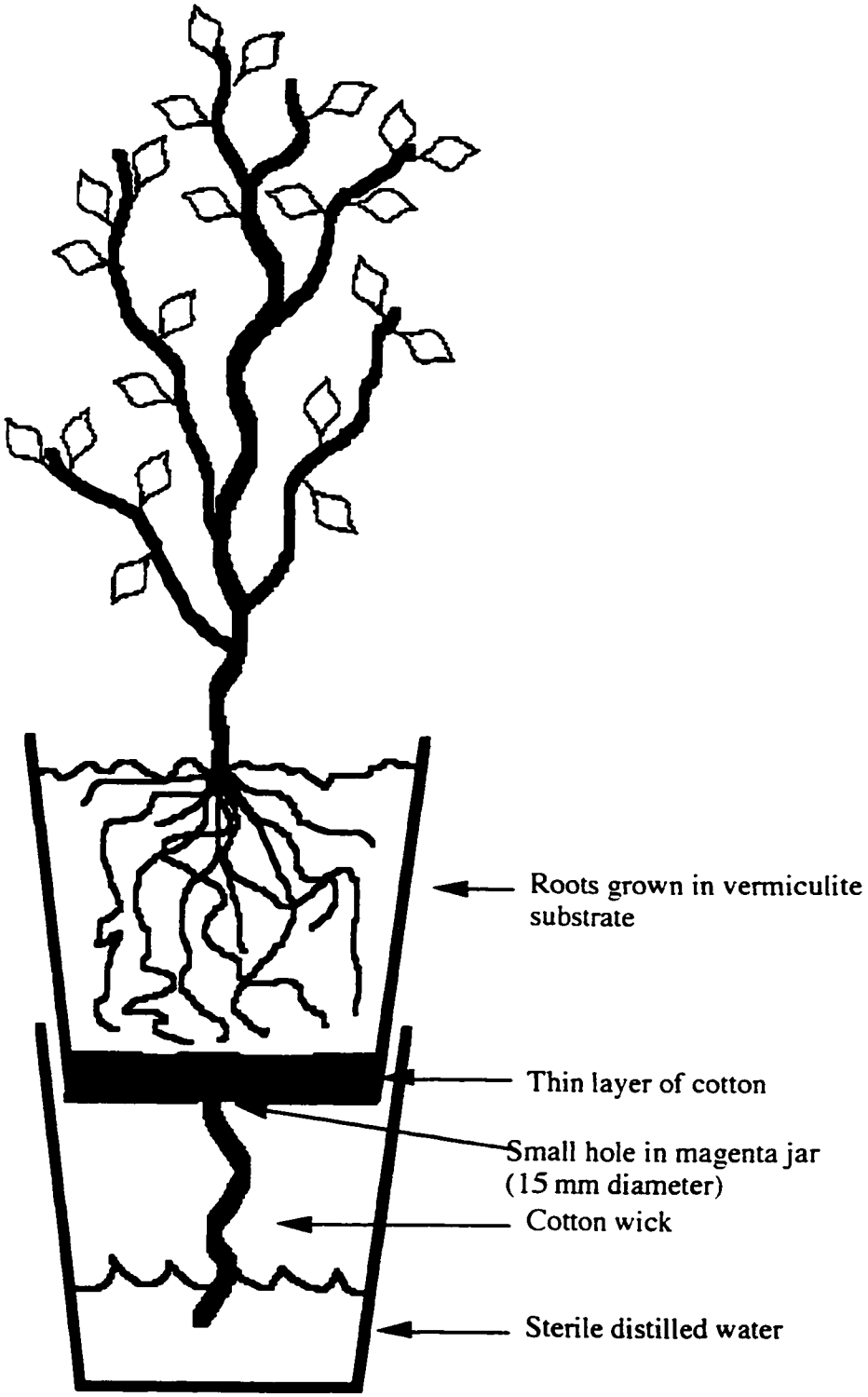
2.0 Strategy used to introduce mutated *mcp* genes into VF39SM.



Selection for VF39SM recipients with gene replacment accomplished by plating on TY or VMM plates containing **Sm, Tc,** and **5% sucrose**.

Explanations for use of selection agents:
I) Sm selects against the S17-1 donor strain; VF39SM is Sm resistant.
II) Tc selects for recombinational events since the pJQ200::ORF-Tc plasmid is incapable of replicating in VF39SM due to the OriV.
III) Sucrose selects for double recombinational events, since expression of the sacB gene in VF39SM is lethal when sucrose is present.

3.0 Schematic of apparatus used to cultivate Trapper pea plants.



4.0 Introduction of a par stability locus into broad host range plasmids.

To create the MCP-F1 and MCP-F2 plasmids the generalized protocol supplied by A. Kent (University of Wisconsin, Madison, Wisconsin) was used. The protocol, as it pertained to this study, is described below:

1. Competent cells of HB101, pSShe were prepared and transformed with pRK7813::mcp-3::lacZ R.O. and pRK7813::mcp-3::lacZ W.O. Selection was for Cm, Sm, and Tc. This resulted in HB101 strains carrying both the pSShe plasmid and either of the two mcp-3 fusion plasmids. pSShe must be present as a helper plasmid for Tn3 transposition.

2. Competent cells of the strains created in step 1 were prepared and transformed with pTn3PAR DNA. Selection was with Km, Cm, and Sm. The plates were incubated at 28 C. At 28 C Tn3 transposition is enhanced, temperatures above 28 C result in fewer transpositions (Turner *et al.*, 1990).

3. Colonies arising from the transformation in step 2 were pooled together by placing 1 ml of LB media onto the agar plate and scrapping the cells from the colonies using a spreader. The 1 ml was collected and the pool of colonies was mated with C2110. C2110 is *polA* and consequently pSShe and pTn3PAR are unable to replicate in this strain. Selection was for Nal (selects against donor) and Km (selects for fusion plasmids carrying the Tn3PAR).

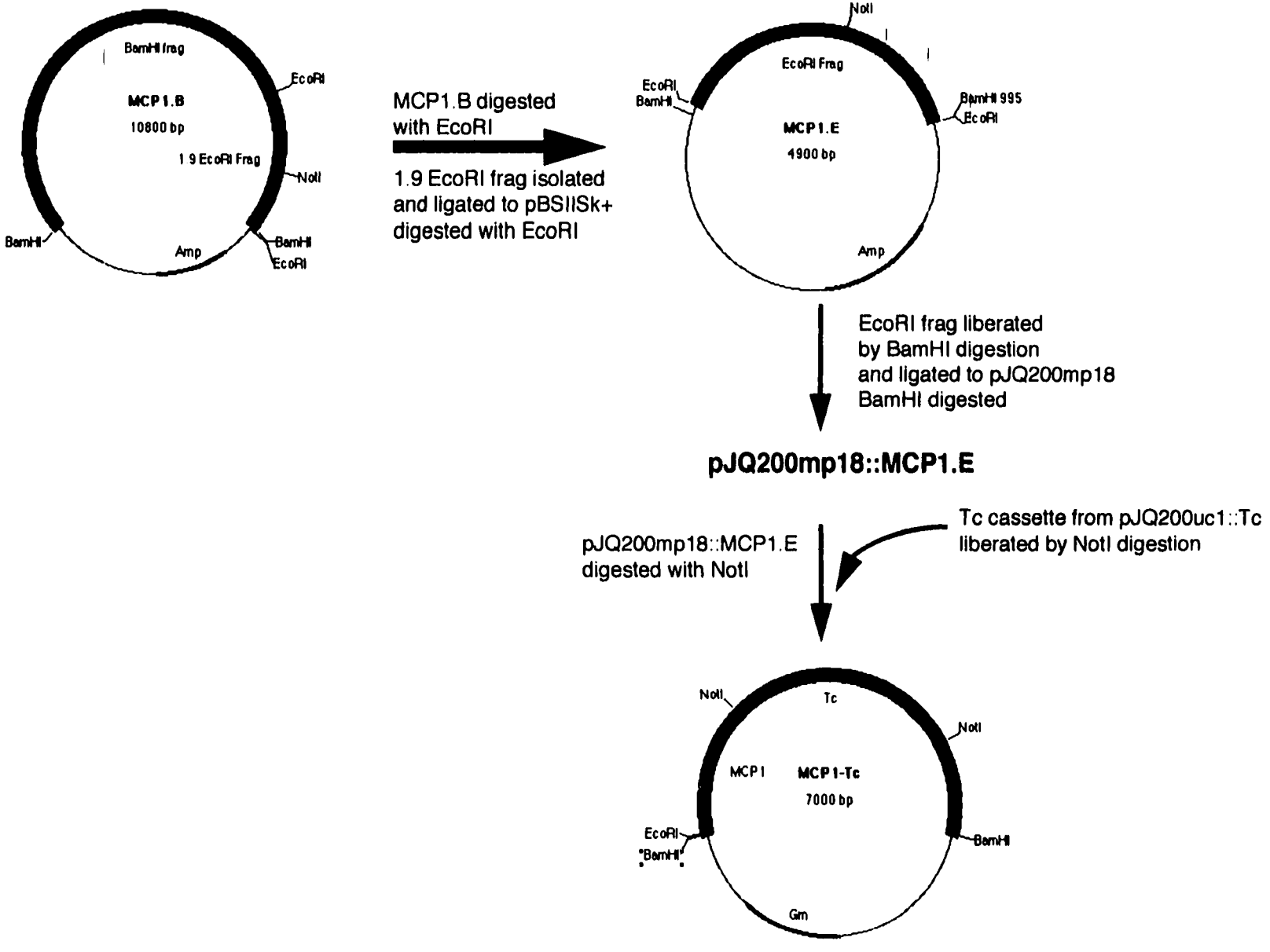
4. C2110 colonies arising after the mating were screened for Tc sensitivity. Colonies which were Tcs had the Tn3PAR insert within the Tc gene of pRK7813. This selection ensured that the Tn3PAR did not transpose into the fusion itself. Since the helper plasmid pSShe was eliminated in the mating the

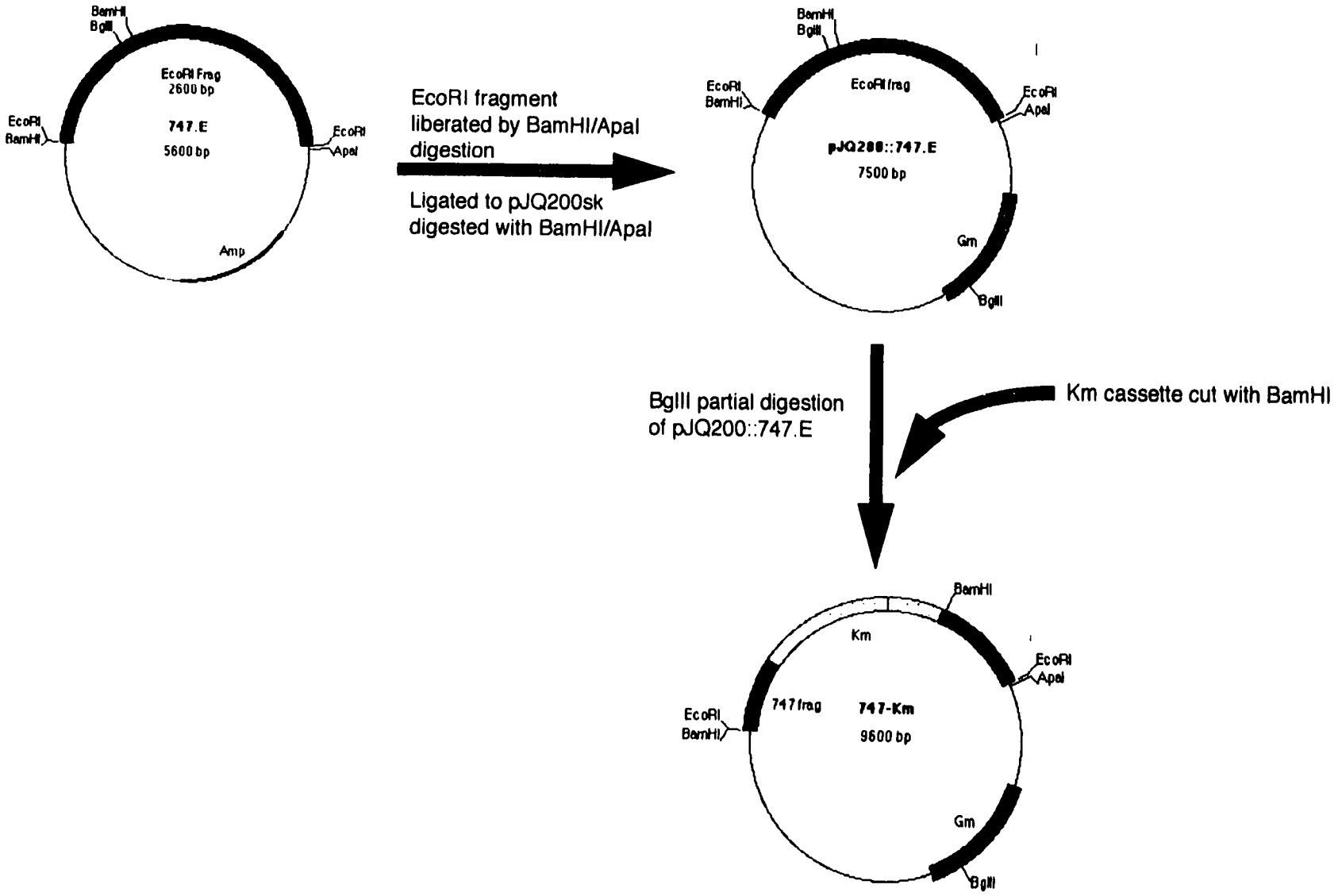
Tn3PAR is unable to transpose further and remains stable on the fusion plasmid.

5. C2110 strain carrying the newly formed MCP-F1 or MCP-F2 was mated with VF39SM. Selection was for Sm (selects against donor) and Nm (selects for fusion plasmid).

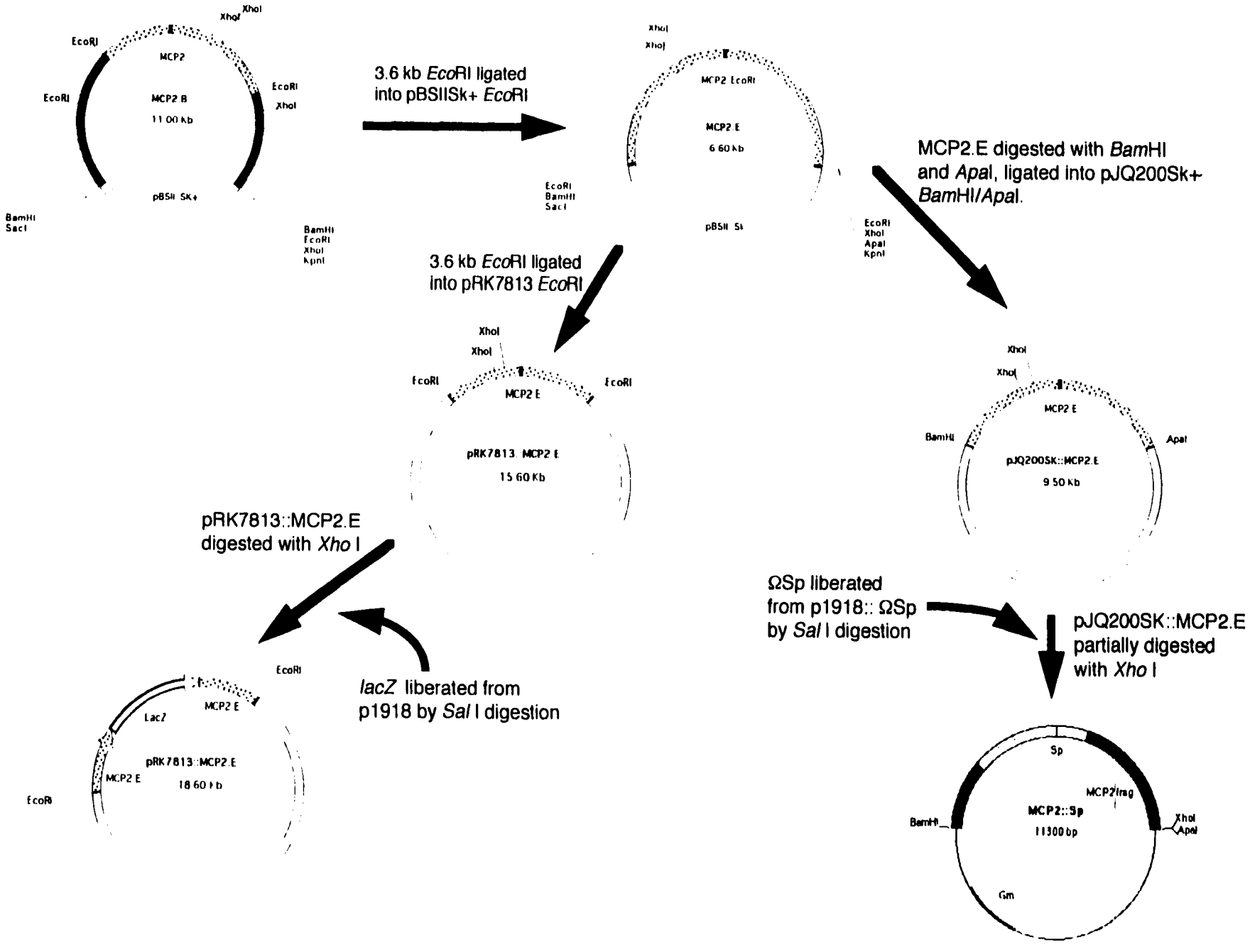
5.0 Pertinent plasmid maps and construction details:

5.1 Plasmids constructed during the study of *mcp-1*

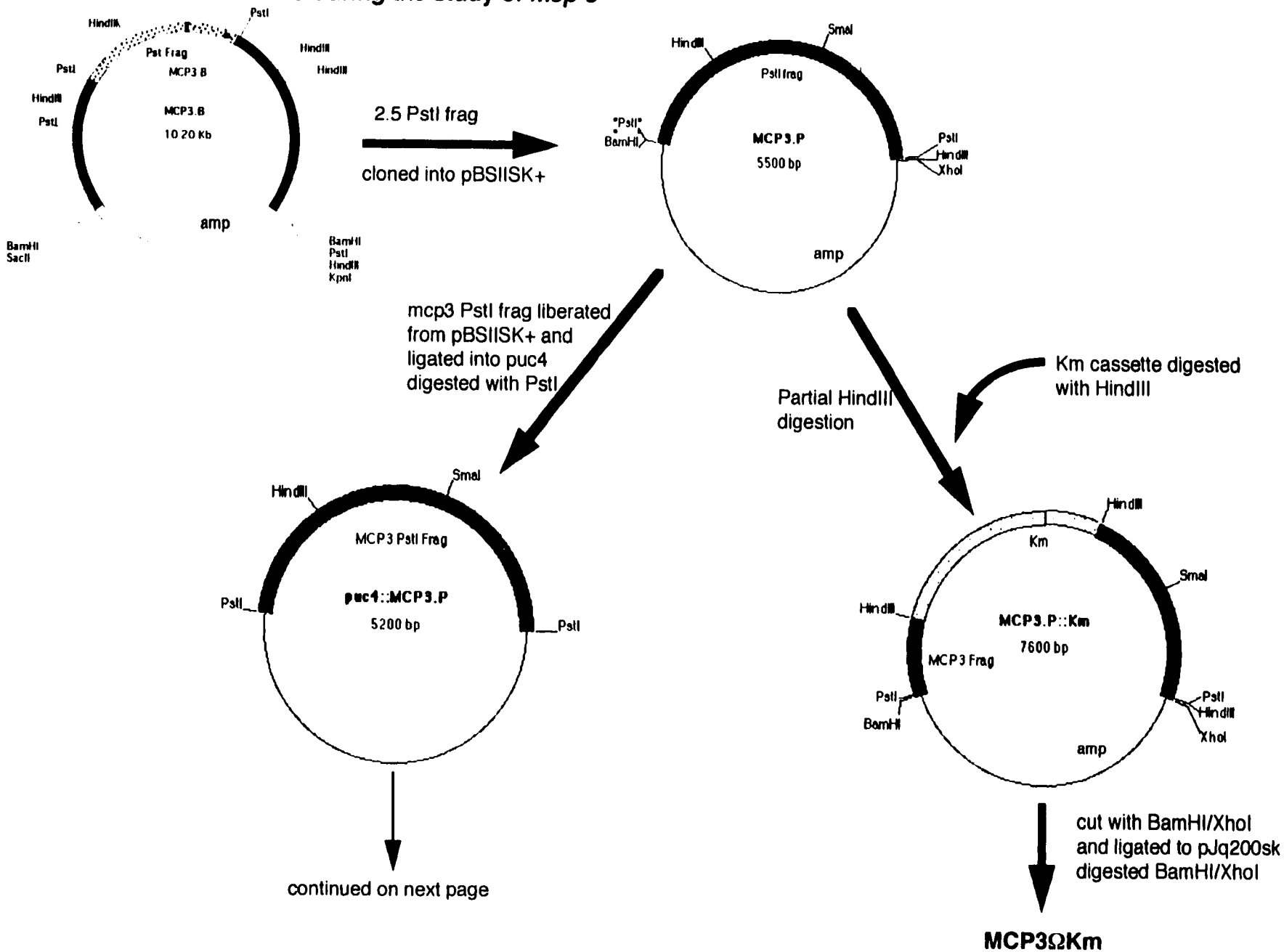


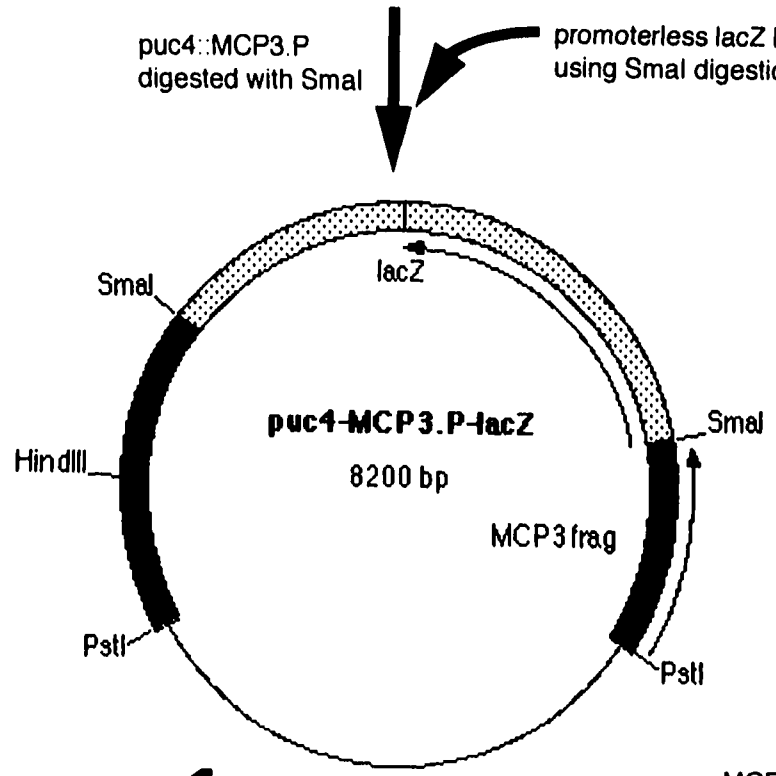

BamHI
BglII
EcoRI Frag
2600 bp
747.E
5600 bp
EcoRI
BamHI
EcoRI
ApaI
Amp
EcoRI fragment
liberated by BamHI/ApaI
digestion
Ligated to pJQ200sk
digested with BamHI/ApaI
BamHI
BglII
EcoRI
BamHI
EcoRI frag
EcoRI
ApaI
pJQ200::747.E
7500 bp
Gm
BglII
BglII partial digestion
of pJQ200::747.E
Km cassette cut with BamHI
BamHI
Km
EcoRI
ApaI
747 frag
747-Km
9600 bp
EcoRI
BamHI
Gm
BglII

5.2 Plasmids constructed during the study of *mcp-2*



5.3 Plasmids constructed during the study of *mcp-3*





* lacZ gene was also inserted in the oppositie orientation to create a negative control.

MCP3-lacZ cassette liberated by PstI digestion and ligated to pRK7813, PstI digested

MCP3-lacZ cassette liberated by PstI digestion and ligated to pJQ200mp18, PstI digested

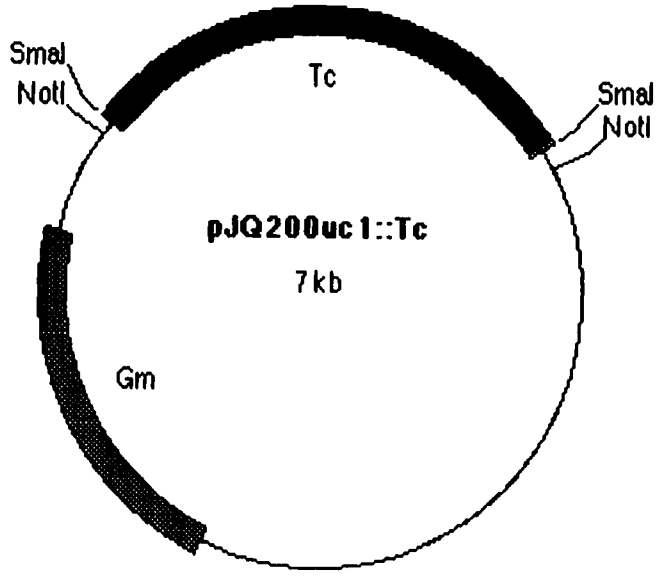
pRK7813::MCP3-lacZ

-The par locus was subsequently introduced into this plasmid

pJQ200::MCP3-lacZ

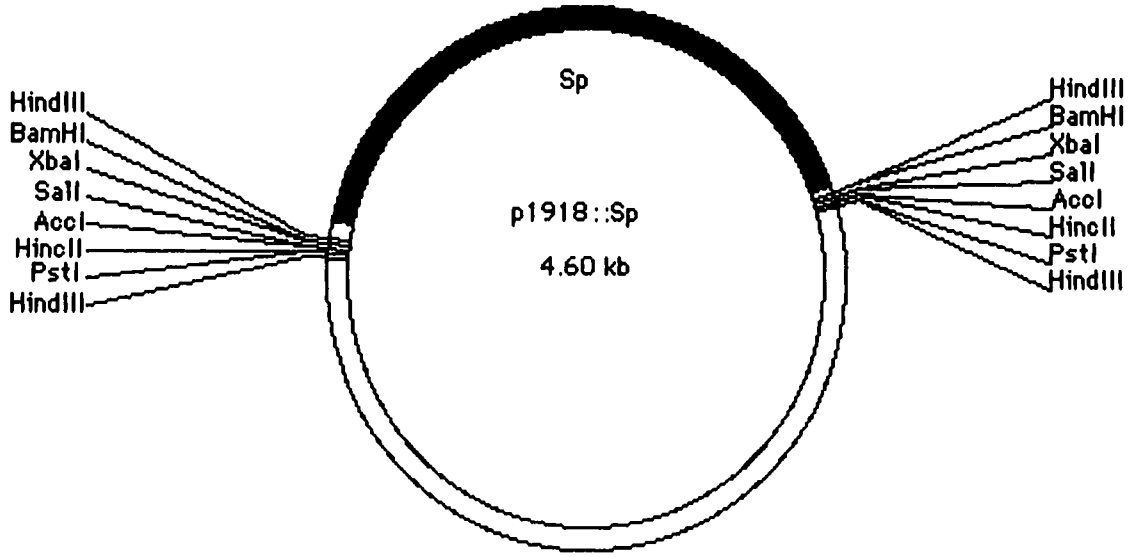
-Used to integrate a mcp3 reporter gene fusion into the VF39SM genome

5.4 Plasmid map of pJQ200uc1::Tc



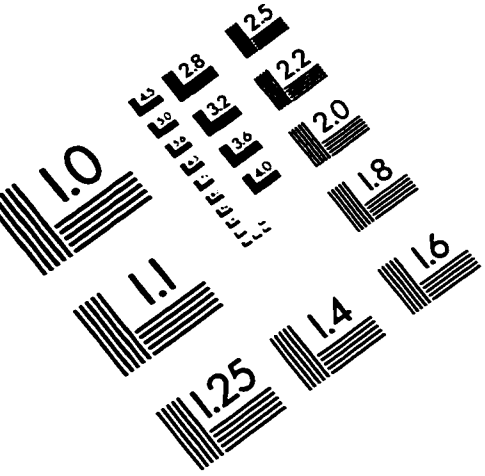
-Constructed by digesting ΩTc with *Sma* I and ligating to pJQ200uc1 digested with *Sma* I.

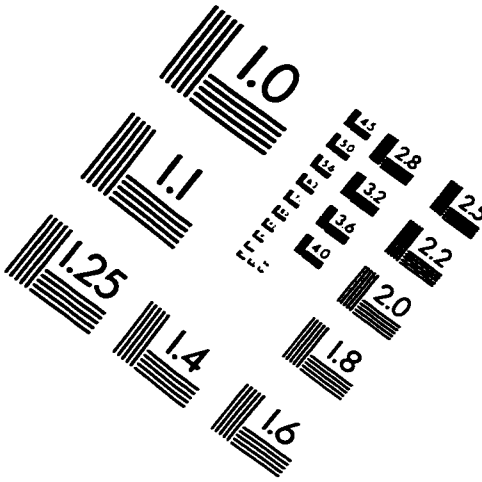
5.5 Plasmid map of p1918::Sp

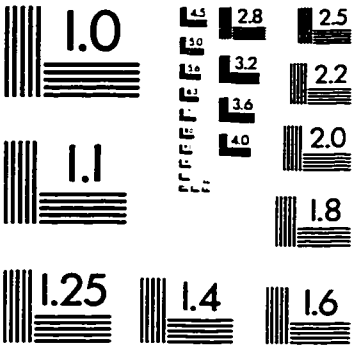


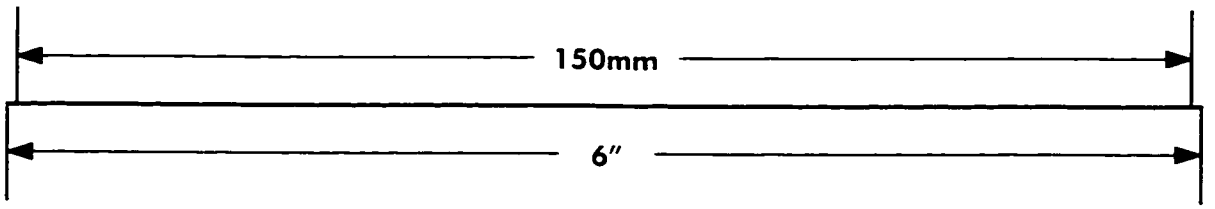
-Constructed by digesting ΩSp with *Bam* HI and ligating to p1918 digested with *Bam* HI.

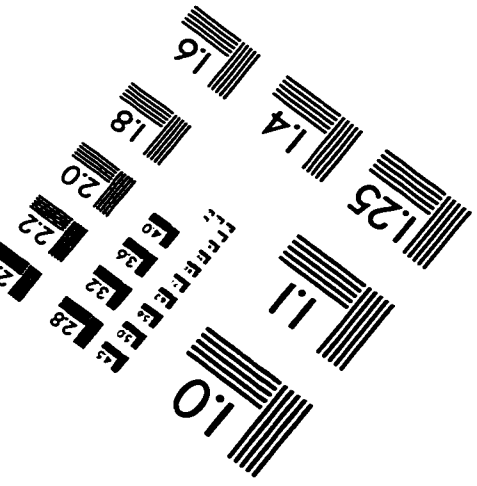
IMAGE EVALUATION
TEST TARGET (QA-3)

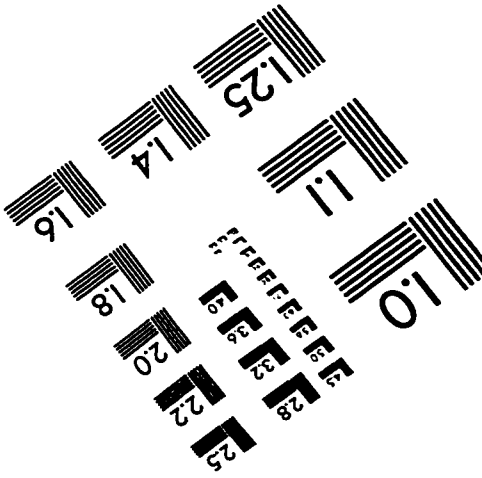












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989