Identification and analysis of a flagellar gene operon of Agrobacterium tumefaciens

Kelly, Victoria Elizabeth

Identification and analysis of a flagellar gene operon of *Agrobacterium tumefaciens*

by

Victoria Elizabeth Kelly.

A thesis submitted to the Department of Biological Sciences

University of Durham

In accordance with the requirements for the degree of

Master of Science

December 1998

For My Family.

Declaration

I declare that the work contained within this thesis submitted by me for the degree of Master of Science is my own original work, except where otherwise stated, and has not been submitted previously for a degree at this or any other University.

Identification and analysis of a flagellar gene operon of *Agrobacterium tumefaciens*.

Victoria Elizabeth Kelly

MSc 1998

Abstract

The cosmid, pDUB1900, from a representative genomic library of the *Agrobacterium tumefaciens* C58C1 chromosome has previously been partly mapped and found to contain genes concerned with flagellar structure, assembly and function. In this study a region of 4470 nucleotides from the largest *Hin*dIII fragment of pDUB1900 was sequenced completely in both directions. This sequenced region unified two previously identified operons which were transcribed in oppposite directions, *motA* leftward and *flgB* rightward. This sequence was compared against databases to identify homologous regions. The comparison results and other computer analyses revealed two flagellar gene homologues, *flgF* and *fliI*, and two open reading frames, *orfX* and *orfZ* with no significant sequence identity to any previously identified genes. Computer analysis also indicated that all four ORFs may be transcribed together as a putative operon, independent of the two flanking operons.

Previously identified *A. tumefaciens* flagellar genes were found to be homologous to flagellar genes in *Rhizobium meliloti* and have the same gene order, therefore, the four ORFs identified here were also compared to *R. meliloti* DNA. The ORFs were isolated, radiolabelled and used as probes against Southern blots containing *R. meliloti* DNA from the pMB7 plasmid and pRZ cosmids, known to contain motility genes. All four of the *A. tumefaciens* ORFs hybridised to homologous DNA in *R. meliloti* and were found in the same order in both species. Recently released DNA sequence from *R. meliloti*, homologous to this region, confirmed these findings. This suggests that there is a high degree of conservation between the two species at the molecular level with regards to motility and chemotaxis.

Mutants were created in the last ORF (*orfZ*) of the putative operon to determine the effects on motility. The *orfZ* mutants produced were all motile suggesting that the gene product of *orfZ* is not essential for flagella function. Alternatively, enough of *orfZ* may have remained before the insertion site to ensure it was still functional and thus, the putative operon transcribed as normal.

Acknowledgements

This thesis would not have been possible without the support of the following people. I thank Dr. Charlie Shaw for acting as my supervisor during my MSc. Additionally, I would like to thank the Department of Biological Sciences (University of Durham) for providing half of the necessary funding.

Special thanks must go to the following people; I am grateful to Mrs. Christine Richardson for her help and advice with the electron microscopy, to Miss. Emma Wright for her help and support especially during the final stages of writing my thesis, to Mr. Paul Sidney for his photographic expertise and also to Miss. Gillian Storey for all her help, particularly in ensuring the DNA Sequencing Service has continued to run smoothly during these final months. Lastly, thanks must go to Dr. Bill Deakin for providing an enormous amount of help and advice throughout.

Finally, I must thank my family and friends for all of their support. Thanks to my husband, Paul, and to Scooby for providing some necessary light relief and also to Paul for his 'computer delivery service'. Lastly, I would like to thank my Mam and Dad not only for their financial help but more importantly for their unwavering support and encouragement, it has always been greatly appreciated.

Abbreviations

A_{260}	=	absorbance at 260nm
A_{280}	=	absorbance at 280nm
Amp	=	ampicillin
ATP	=	adenosine triphosphate
bp	=	base pairs
BSA	=	bovine serum albumin
CCW	=	counterclockwise
CW	=	clockwise
dNTP	=	deoxyribonucleoside triphosphate
EDTA	=	ethylenediaminetetraacetic acid
Gm	=	gentamyicn
kb	=	kilobase pairs
kDa	=	kilodaltons
M_r	=	molecular weight
PEG	=	polyethylene glycol
Rhizobium meliloti	=	*Sinorhizobium meliloti*
Rif	=	rifampicin
RNAase	=	ribonuclease
SDS	=	sodium dodecyl sulphate
Sm	=	streptomycin
Tc	=	tetracycline
Tris	=	tris(hydroxymethyl)aminomethane
X-gal	=	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
UV	=	ultraviolet
5'	=	5' terminal phosphate of DNA molecule
3'	=	3' terminal hydroxyl of DNA molecule

Contents

Figures

Chapter 1

Introduction

1.1 *Agrobacterium tumefaciens* - a brief introduction

Agrobacterium tumefaciens is a Gram-negative, rod shaped bacterium of approximately 1.5-3.0 μm in length and 0.6-1.0 μm in width. It is a member of the family *Rhizobiaceae*, which is a heterogeneous group of Gram-negative, aerobic, non-spore forming rods. The genus *Agrobacterium* is generally said to consist of four species, namely *A. tumefaciens*, *A. rhizogenes*, *A. rubi* and *A. radiobacter*. The first three species cause Crown Gall Tumour, Hairy Root and Cane Gall Tumour, respectively, with *A. radiobacter* being avirulent [Nester and Kosuge, 1981]. *A. tumefaciens* is a soil bacterium that is able to genetically transform susceptible plant cells, leading to the neoplastic (cancerlike) disease entitled Crown Gall. This agronomically important disease affects most dicotyledonous plants and a few monocotyledonous, and thus, *A. tumefaciens*-induced infection has been extensively investigated by molecular techniques.

A. tumefaciens is a rhizosphere dwelling bacterium. The rhizosphere is generally defined as the volume of soil that is adjacent to and influenced by the plant root. Hence it is the physical location in soil where plants and microorganisms interact. The rhizosphere hosts higher populations and a greater diversity of microorganisms than the surrounding soil [New and Kerr, 1972] thus the competition for nutrients is great. *A tumefaciens* has evolved to create itself a competitive advantage over other soil dwelling microorganisms and therefore, carve an influential niche for itself within the rhizosphere.

A. tumefaciens is a motile bacterium with 2 polar flagella and 2-4 lateral flagella. Motile microorganisms display behavioural responses which operate principally to guide them away from locations in the environment that cannot support optimal energy generation and growth and towards those that can do so. Microbial movement occurs in response to such stimuli as oxygen, light or chemical compounds (aerotaxis, phototaxis and chemotaxis, respectively) and depends upon sensory transduction and energization of the flagellar motors. Bacteria respond to stimuli by altering the balance of runs and tumbles (see **1.2**). The principal factor regulating chemotaxis is the concentration gradient of the chemoeffector; a cell which senses an increase in attractant, or a decrease in repellent, represses tumbling and continues to swim in the favourable direction, while an increase in repellent enhances the probability of tumbling. *A. tumefaciens* has developed a highly sensitive chemotaxis system [Ashby, 1988; Ashby, *et al.*, 1987; Loake, *et al.*, 1988] enabling it to respond to changes in the environment and move towards plant wound sites.

After attaching itself to the plant, *A. tumefaciens* is able to transfer DNA from its Tumour Inducing (Ti) plasmid to the plant chromosome [Ashby, *et al.*, 1988]. The

resulting expression of this T-DNA in the plant cell, leads to the production of crown gall cells. These cells in turn produce unusual metabolites, termed opines, which can be used by *A. tumefaciens* as its sole carbon, nitrogen and energy sources [Tempe and Goldman, 1982].

Thus, for *A. tumefaciens* to operate successfully in the rhizosphere it has evolved a system which can recognize compounds exuded from plant wound sites, respond to them, attach itself to the plant, transfer part of its Ti-plasmid to the plant chromosome thereby ensuring the synthesis and release of opines and hence the continued existence of the bacterium. All of these factors, and how they operate and are controlled, will be discussed in greater detail in relation to both *A. tumefaciens* and other bacteria.

1.2 An example of a model chemotactic system

For chemotaxis to occur the microorganism must be motile. The motile behaviour of the bacterium *Escherichia coli* has been one of the most extensively studied and here provides an example of a model system. *E. coli* cells are able to propel themselves by the rotation of helical flagella [Berg and Anderson, 1973; Shaw, 1990]. Each cell contains on average 8 peritrichous flagella, but in practice can be anywhere between 0 and 15. Each flagellum consists of a basal body which is embedded in the surface of the cell, a hook which acts as a universal joint between the filament and the cell [Macnab, 1987a] and a helical filament that can be between 0-20 μm in length [Macnab, 1987b]. The motor in the cell envelope uses the transmembrane proton potential to provide energy for rotation of the filament [Manson, *et al.*, 1977]. Swimming is the result of a counterclockwise rotation of the flagella (CCW). When reversed, that is the flagellar motor rotates clockwise (CW), a 'tumble' occurs. In an isotropic environment a random motility pattern consisting of straight runs for about one second followed by tumbles can be observed.

A CCW rotation of the flagellar motor produces a left-handed structure which allows the flagella to form a bundle that propels the bacterium forward in a run. When any of the flagella switch to a CW rotation a conformational change occurs. A right-handed helical structure is produced. The transition from a left-handed helical structure to a right-handed causes the bundle to break and the bacterium to go into a tumble. After reversion back to a CCW rotation, the bacterium then resumes swimming in a different, random direction [Berg and Brown, 1972]. Full conversion to right-handed filaments can result in slower and less smooth runs, but this is a less stable configuration and does not last for long [Eisenbach, 1990; Macnab and Ornston, 1977]. When the flagellar motor pauses the flagellum becomes stationary. This may be due to an incomplete switching between CCW

and CW rotations and hence, preventing the less stable configuration being produced. *E. coli* modulates the direction of rotation of the flagellar motor in response to environmental cues. Tumbling is suppressed when the cell is moving in a favourable direction.

As mentioned previously, the principal factor regulating chemotaxis is the concentration gradient of the chemoeffector. There are various chemoattractants for *E. coli* which the cell responds to. They include certain amino acids, sugars, carboxylic acids and oligopeptides, fumarate and nitrate [Adler, 1975; Taylor and Johnson, 1993]. Repellents include hydrophobic amino acids, indole, ethanol, glycol and inorganic ions such as Co^{2+}, Ni^{2+} and S^{2-} [Macnab, 1987b]. Changes in pH and temperature of the environment may also elicit a chemotactic response in *E. coli* [Macnab, 1987b]. Additionally, *E. coli* responds to the level of oxygen in the environment. This aerotactic response ensures that the bacterium moves to an environment where the oxygen concentration is at an optimum level for growth [Taylor and Zhulin, 1998].

E. coli senses temporal gradients of chemical stimuli by continually monitoring the concentration of chemoeffectors. If the present conditions are more favourable than the immediate past conditions, CW rotation will be suppressed producing a longer run. When the bacterium becomes used to the environment it returns to a random pattern of runs and tumbles, sensing new chemoeffectors. By this process, the bacteria accumulate in a favourable environment [Shaw, 1990].

1.3 Methylation-dependent chemotaxis

A chemotactic response occurs when a receptor on the surface of the cell recognises a chemoeffector, generating a rapid excitation phase, an adaptive phase and finally a return to a random walk. In methylation-dependent chemotaxis the receptors are involved in the initiation of the excitation phase via a cytoplasmic signalling pathway and their methylation is responsible for adaptation.

1.4 Methyl-accepting chemotaxis proteins (MCPs)

E. coli is known to possess four MCPs which recognise the indicated attractants; Tsr, serine; Tar, aspartate and maltose; Trg, ribose, galactose and glucose; and Tap, dipeptides [Manson, *et al.*, 1998]. Salmonella lacks Tap but has Tcp, a citrate sensor [Manson, *et al.*, 1998; Yamamoto and Imae, 1993] The MCPs also produce responses to temperature and pH, and act as receptors for various repellents.

The four proteins are all integral membrane proteins of approximately 60 kDa in size. Their sequences are very similar and hence they are thought to all have the same structural domains [Bollinger, *et al.*, 1984; Boyd, *et al.*, 1981]. The amino-terminal is the receptor region, with two hydrophobic transmembrane sequences surrounding a hydrophilic ligand-binding domain which extends into the periplasm [Hazelbauer, *et al.*, 1990; Manson, *et al.*, 1998]. A large cytoplasmic domain is responsible for intracellular signalling and adaptation. The MCP's exist as homodimers with or without ligands [Manson, *et al.*, 1998].

Serine, aspartate and citrate bind directly to the MCP's, whereas maltose, ribose, galactose, glucose and dipeptides bind to their specific binding proteins which then dock with the appropriate membrane MCP [Manson, *et al.*, 1998]. The MCP's are very sensitive, a change in occupancy of only four out of the several hundred that are present per cell eliciting a response [Block, *et al.*, 1982].

Communication between the MCPs and the flagellar switch utilises four proteins. CheA- a histidine protein kinase, CheY- the response regulator, CheW- the receptor-coupling factor and CheZ- an enhancer of CheY-PO_4 dephosphorylation [Manson, *et al.*, 1998]. CheA functions as a dimer, with two CheW proteins binding per CheA dimer, the resulting complex associating with a dimeric receptor. In this complex, CheA autophosphorylation is highly stimulated, which results in an increased phosphotransfer from CheA-PO_4 to CheY. CheY-PO_4 binds to the FliM protein in the flagellar motor-switch complex resulting in a CW rotation [Manson, *et al.*, 1998; Welch, *et al.*, 1993]. CheZ prevents the accumulation of CheY-PO_4 by accelerating the decay of the unstable aspartyl-phosphate residue. CheY-PO_4, under steady-state conditions, is maintained at a level that produces the random walk [Manson, *et al.*, 1998]. In aerotaxis the sensor, Aer, is the oxygen transducer and also signals *via* CheA and CheY.

The binding of an attractant to an MCP initiates a conformational change that results in the suppression of CheA activity. The level of CheY-PO_4 falls and cells tumble less frequently. This response ensures that cells increase the length of their runs as they enter areas of higher attractant concentration.

Two further proteins, the methyltransferase CheR and the methylesterase CheB are required for adaptive methylation. CheR methylates the MCPs at specific residues. CheB is a target for phosphotransfer from CheA and CheB-PO_4 removes methyl groups from the MCPs. When an attractant binds to an MCP, CheA activity falls and ultimately so do CheB-PO_4 levels. During steady state, methyl addition by CheR balances methyl removal by CheB-PO_4 to achieve an intermediate level of receptor methylation which maintains

run-tumble behaviour [Manson, *et al.*, 1998]. Increased methylation restores the ability of the receptor to stimulate CheA [Manson, *et al.*, 1998], a conformational change produces the CW rotation and the bacterium tumbles.

Calcium ions (Ca^{2+}) have also been indicated to play a role in *E. coli* chemotaxis [Tisa and Adler, 1992; Tisa, *et al.*, 1993]. Experimental evidence has suggested that increased intracellular Ca^{2+} levels resulted in cells tumbling and a decrease resulted in smooth swimming. It is thought possible that Ca^{2+} maintains the phosphorylated state of CheY, prolonging the tumble signal as there is evidence that Ca^{2+} inhibits the dephosphorylation of CheY-PO_4 [Lukat, *et al.*, 1990].

1.5 Methylation-independent chemotaxis

The response to sugars transported by the phosphoenolpyruvate-phosphotransferase system (PTS) shows a special type of chemotaxis. Transport of the sugar into the cell is required to produce a behavioural response [Taylor and Zhulin, 1998]. The methyl-accepting chemotaxis proteins are not required and therefore, the process is methylation-independent [Taylor and Lengeler, 1990]. Once in the cell, the sugar is then phosphorylated. The phosphate group is transferred *via* a family of proteins (EII) which interact with CheA, thus eliciting a chemotactic response.

1.6 Flagella structure and gene nomenclature in the model systems

The number of flagella on a cell varies from species to species, however, their operation is much the same, the flagellum rotates pushing the organism forward. The enteric bacteria *E. coli* and *S. typhimurium* have been most closely studied and the structure and function of their flagella have been found to be practically identical [Macnab, 1992].

Morphologically, the flagellum is divided into three parts namely the filament, the hook and the basal structure [Aizawa, 1996]. In *E. coli* and *S. typhimurium* almost 50 gene products are called upon to build the three parts of the flagellum from the 'inside out' [Harshey and Taguchi, 1996]. The basal body is constructed first. It is a multiprotein assembly, anchored in the outer and inner membranes, consisting of four rings (M, S, P and L) and an axial rod (R) [Sosinsky, *et al.*, 1992]. The hook and helical filament are external to the cell. They are constructed after the basal body. The filament is the biggest

part of the flagellum and is rotated to act as a propeller. The hook works as a universal joint joining the filament and basal body [Aizawa, 1996].

The nomenclature used for the genes associated with this flagellum assembly and function is determined by the phenotype of mutations in those genes and chromosomal position. Phenotype has been categorised into three groups:- Fla- (nonflagellate), Mot- (nonmotile) and Che- (nonchemotactic). The Fla- group is the largest as nonflagellate can mean anything from having everything except the filament to having no flagellar structure at all. The genes in this Fla- group have been subdivided into three classes depending on their position on the chromosome, *flg* (region I), *flh* (region II) and *fli* (region IIIa or IIIb). Unlike *E. coli,* however, *S. typhimurium* has a fourth class *flj*. The Mot- class is limited to only two genes, *motA* and *motB*. Here a complete flagellum is present but no rotation occurs. The phenotype concerned with defects in receptors, the sensory transduction chain and flagellar switch proteins is designated Che-. The gene symbol *che* however, is only given to genes whose products are a part of the intracellular signalling pathway as the receptor genes have their own symbols and the switch proteins' genes have a Fla- null phenotype [Macnab, 1992].

1.7 Flagellar Structure

As mentioned previously, the flagellum consists of three main parts; the filament, hook and a basal body. A diagram showing the general structure of the flagellum of *E. coli* and *S. typhimurium* is shown in figure 1.7.1. The genes which are responsible for encoding the structural components are also included in the diagram.

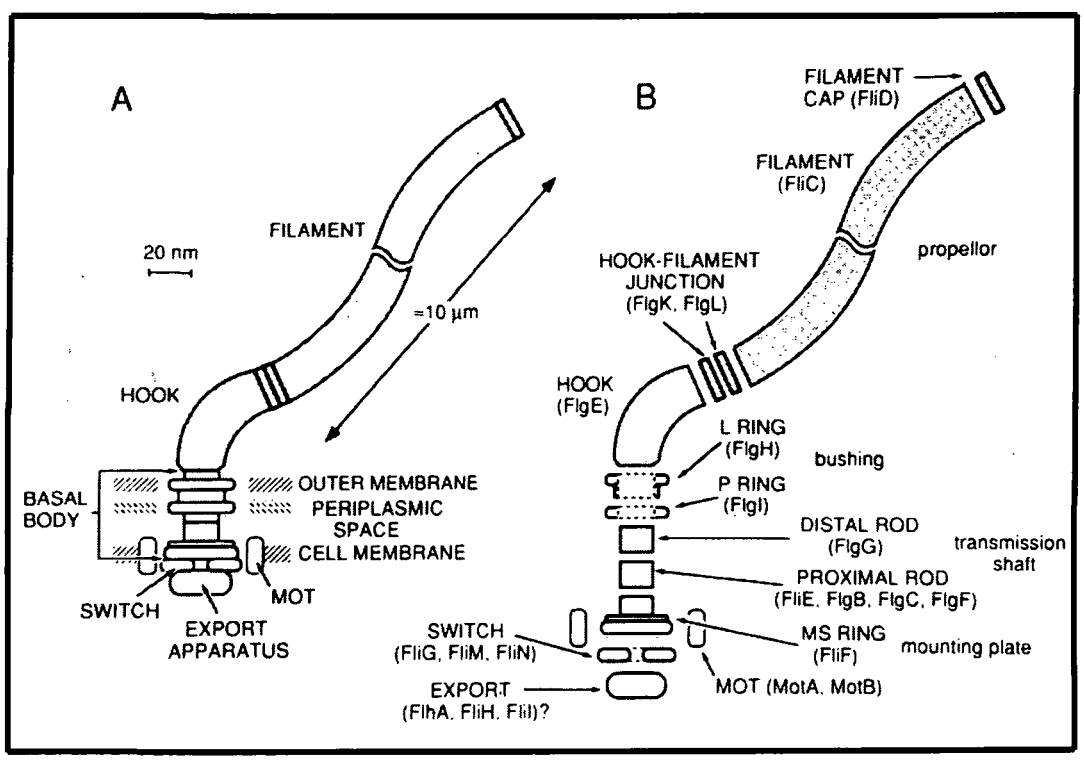
The filament is composed of a single protein, flagellin encoded by the *fliC* gene [Macnab, 1987]. Flagellin has to be translocated from its site of synthesis in the cytoplasm across the cell membrane. The flagellin proteins are exported through the central channel in the proximal part of the filament for assembly at the distal tip [Macnab,R.M., 1990]. The protein FliD, plays an essential role in filament growth. The filament has a cap structure at its tip to prevent its loss during assembly, encoded for by *fliD*. This cap structure plays a major part in the polymerisation of flagellin preventing it from diffusing away so the filament can be completed [Ikeda, *et al.*, 1996]. Two other proteins FlgK and FlgL, present at the proximal end of the filament, are thought to be needed to join the filament to the hook [Homma, *et al.*, 1986]. These two minor proteins are also known as HAP1 and HAP3 (hook-associated proteins) [Aizawa, 1996]. In addition, *S. typhimurium* has a further protein FljB, an alternative, serotypically distinct flagellin.

Figure 1.7.1 The flagellum structure of *E. coli* and *S. typhimurium*

A shows the general structure of the flagellum and location of the components in relation to the cell. The export apparatus is assumed to be at the base of the flagellar although it is not shown on the diagram.

B shows a more detailed version labelling the substructures and the encoding proteins. The order of the proximal rod proteins, FlgB,C,F and FliE is unknown.

The figure was reproduced with permission from R. M. Macnab [1992].



The hook, which precedes the filament, contains only one protein, FlgE. It is structurally similar to the filament with a central channel along the hook axis to allow for the export of hook and filament subunits [Morgan, *et al.*, 1993]. In addition, the hook and filament have the same subunit packing and a comparison of the amino acid sequences show that both possess heptad repeats of hydrophobic amino acid residues [Morgan, *et al.*, 1993]. The hook in contrast to the filament seems to have a fairly well defined length [Kawagishi, *et al.*, 1996; Macnab, 1987a] probably to allow for a sufficient bend angle and provide stability to the filament.

The basal body is embedded in the cell surface and is made up of a number of proteins. It appears to be of cylindrical symmetry with a rod joined to the hook and two pairs of rings [Macnab, 1992]. The L ring is embedded in the outer membrane and is composed of FlgH with the P ring thought to be in the peptidoglycan layer composed of FlgI [Francis, *et al.*, 1994]. These two rings form a large pore through which the rod passes [Jones,*et al.*, 1989]. The MS ring in the inner membrane is made from a single protein, FliF and is thought to act as a mounting bracket for the motor [Francis, *et al.*, 1994]. An additional basal body protein, FliE is present although its location is unknown [Sosinsky, *et al.*, 1992]. It is thought to link the MS ring to the proximal rod proteins FlgB, FlgC and FlgF [Muller, *et al.*, 1992]. The protein FlgG forms the distal portion of the rod.

The MotA and MotB proteins are arranged circumferentially around the MS ring pair [Macnab, 1987]. They are integral to the cell membrane and are believed to be necessary for motor rotation. MotA forms a proton channel believed to be involved in the transmembrane proton conduction which drives the flagellum [Harshey and Toguchi, 1996]. MotB forms a complex with MotA connecting it to the cell wall [Stolz and Berg, 1991].

Rotational direction of the flagellum is determined by the switch complex. The switch contains subunits of three proteins FliG, FliM and FliN which form a bell-like structure (C ring) at the flagellar base [Francis, 1992; Khan, *et al.*, 1992]. FliM is believed to be involved mainly with the switching process and probably interacts with FliG during this action. FliN is also postulated to be involved in the rotation of the flagella [Irikura, *et al.*, 1993].

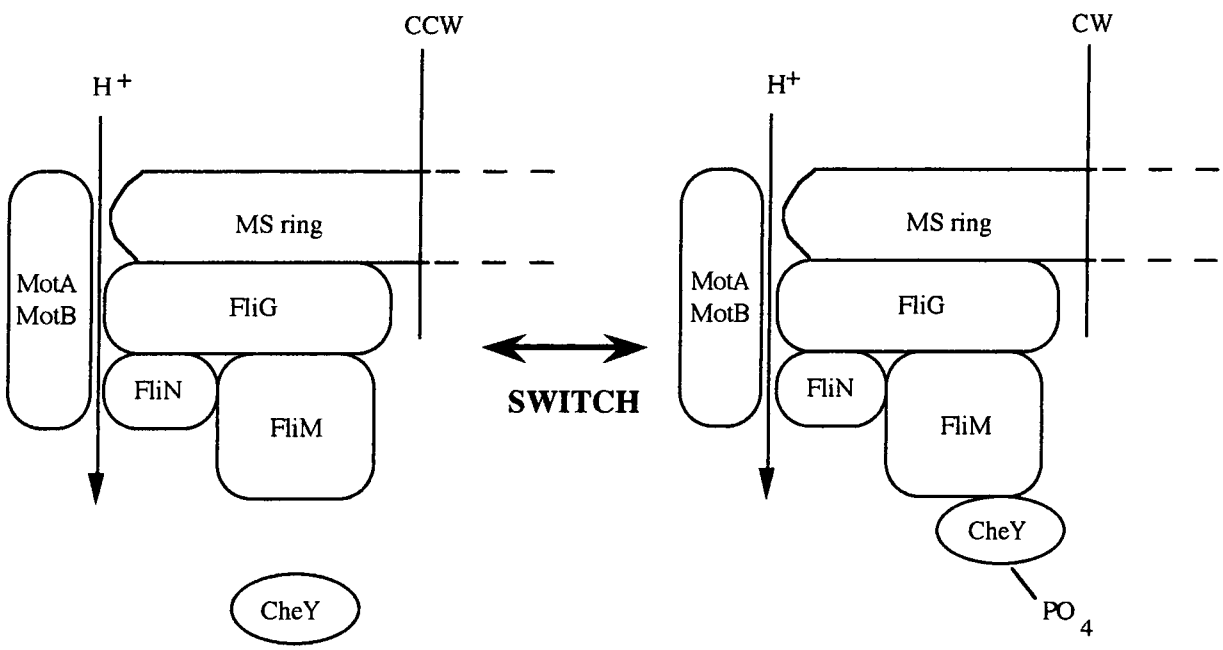
Mutation studies of the three switch proteins have allowed for a fairly comprehensive hypothesis to be put forward with regards to the processes of assembly, rotation and switching. A diagram showing these mechanisms was drawn by Irikura *et al.*, [1993] and can be seen in Figure 1.7.2.

Figure 1.7.2 The switch-motor complex

A CheY is not interacting with the switch, therefore, a CCW rotational state is adopted.

B Phosphorylation of CheY causes it to bind to FliM, a CW state is adopted resulting in FliG bringing about CW rotation.

The figure was reproduced with permission from Irikura *et al.* [1993].



In *E. coli* and *S. typhimurium* , flagella and related chemotaxis genes are found at a few locations on their chromosomes where they form clusters [Macnab, 1990]. These clusters are known as flagellar regions and there are four major ones, I, II, IIIa and IIIb. Within each region there are several operons containing up to nine genes which are expressed as a regulatory hierarchy or regulon [Aizawa, 1996; Macnab, 1992]. The expression of the lower classes of operons is governed by the expression of the higher classes of operon. Figure 1.7.3 shows a diagram of the four main flagellar regions and the classes within them. The operons are divided into four classes, class 1, 2, 3a and 3b. Class 1 contains only one operon and is the master operon governing the expression of the other operons by activating those in class 2 [Aizawa, 1996]. The class 2 operons contain genes for structural units that are required in the early to middle stages of flagellar synthesis. *fliA*, a class 2 gene, codes for a flagellum-specific sigma factor thought to solely control the class 3b operons and part control, along with the master operon, the class 3a operons [Aizawa, 1996; Macnab, 1992].

Activation of the master operon is under the control of c-AMP via the regulatory protein CAP [Macnab, 1990; Silverman and Simon, 1974]. The products of the class 1 operon are the proteins FlhC and FlhD and act as positive regulators of the class 2 operons by an, as yet, unknown mechanism [Macnab, 1992].

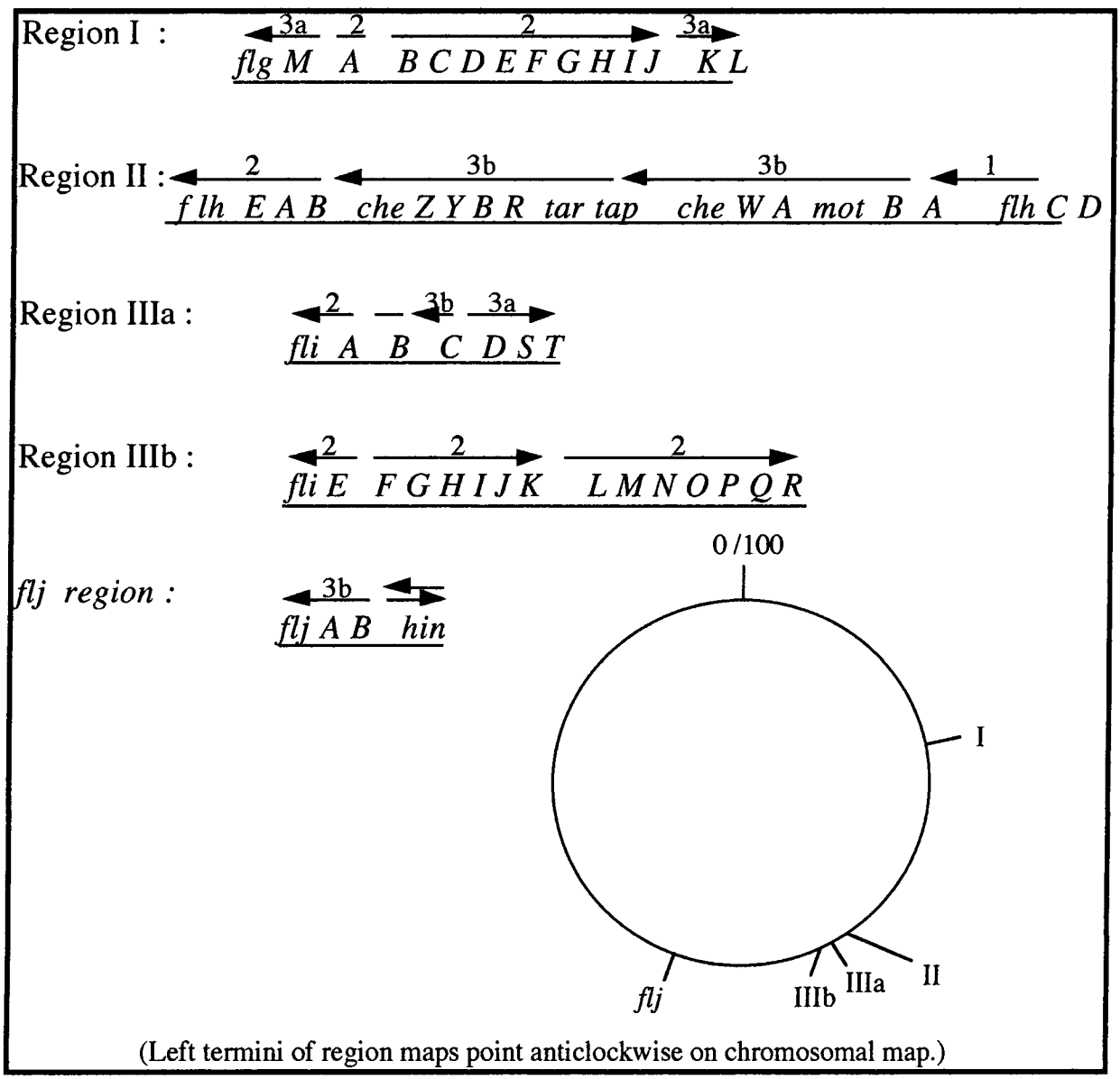
FliA can produce expression of class 3a and 3b operons, although class 3a operons can also be expressed in the absence of FliA. This expression occurs as long as the master operon is expressed suggesting the mechanism is probably the same.

There are also negative regulatory systems which exert an effect on the operons. The protein FlgM is known to block the FliA mediated expression of class 3a and 3b operons. Formation of a complex of FlgM with FliA results in a weakened affinity of FliA for the core polymerase[Gillen and Hughes, 1991; Macnab, 1992]. The activity of FlgM appears to be controlled by flagellar assembly, that is, FlgM is functional during flagellar assembly until the hook has been completed but then a counteracting regulation occurs and the class 3a and 3b operons are expressed. Immature flagellar structure is used to free FliA from the FliA-FlgM complex. FlgM is exported through the central channel out into the external medium leaving FliA free to activate the class 3a and 3b operons [Hughes, *et al.*, 1993; Koncz, *et al.*, 1989] and complete flagellum assembly. It is believed that as the filament lengthens, FlgM ceases to be exported out of the cell and hence the FlgM anti-sigma factors bind once more to FliA inactivating the class 3a and 3b operons [Hughes, *et al.*, 1993; Koncz, *et al.*, 1989] halting flagellar assembly.

Figure 1.7.3 The flagellar and chemotactic genes of *E. coli* and *S. typhimurium*

The four main chemotactic and flagellar gene regions (I, II, IIIa and IIIb) can be seen on the diagram. The *S. typhimurium* specific *flj* region is shown. In *S. typhimurium* the *tap* MCP-receptor gene is not located at this position. The transcriptional directions of the operons are marked above the genes, where they are known. Numbers 1, 2, 3a and 3b show the regulatory classes of the operons.

The figure was reproduced with permission from R. M. Macnab [1992] and W. J. Deakin [1994].



(Left termini of region maps point anticlockwise on chromosomal map.)

A second example of a negative regulatory system is that of *fliDST*. Mutations in the *fliDST* operon have been shown to produce elevated expression of class 3a and 3b operons [Kutsukake, *et al.*, 1990]. This would suggest that the operon has a negative regulatory activity in wild-type cells. This activity, known as *rflA* activity has an unknown mechanism [Kutsukake, *et al.*, 1990; Macnab, 1990] although *fliS* and *fliT* are thought to be involved [Kawagishi, *et al.*, 1992].

1.8 Biogenesis of the flagellum in *E. coli* and *S. typhimurium*

Presumably such a complex organelle as the flagellum would require numerous sophisticated assembly mechanisms. Many of the mechanisms involved in the actual assembling of the flagellum have yet to be identified although much of the order of the assembly process has now been catalogued [Macnab, 1987a; Macnab, 1992]. Biogenesis is on the whole linear from proximal to distal with the filament last to be assembled.

Figure 1.8.1 [Kubori, *et al.*, 1992; Macnab, 1992] shows the morphological pathway of flagellar assembly. Assembly is presumed to begin with the membrane protein, FliF inserting into the cell membrane and self-assembling into the MS ring complex [Aizawa, 1996; Macnab, 1992]. This is followed by the assembly of the switch proteins, FliG, FliM and FliN at the rim of the M ring extending towards the cytoplasm [Aizawa, 1996]. A FliF-FliG fusion protein appears to be assembled at the same time inferring sequential assembly of the switch complex in wild-type cells [Macnab, 1992]. Defects in the switch genes cause the failure of subsequent construction [Aizawa, 1996].

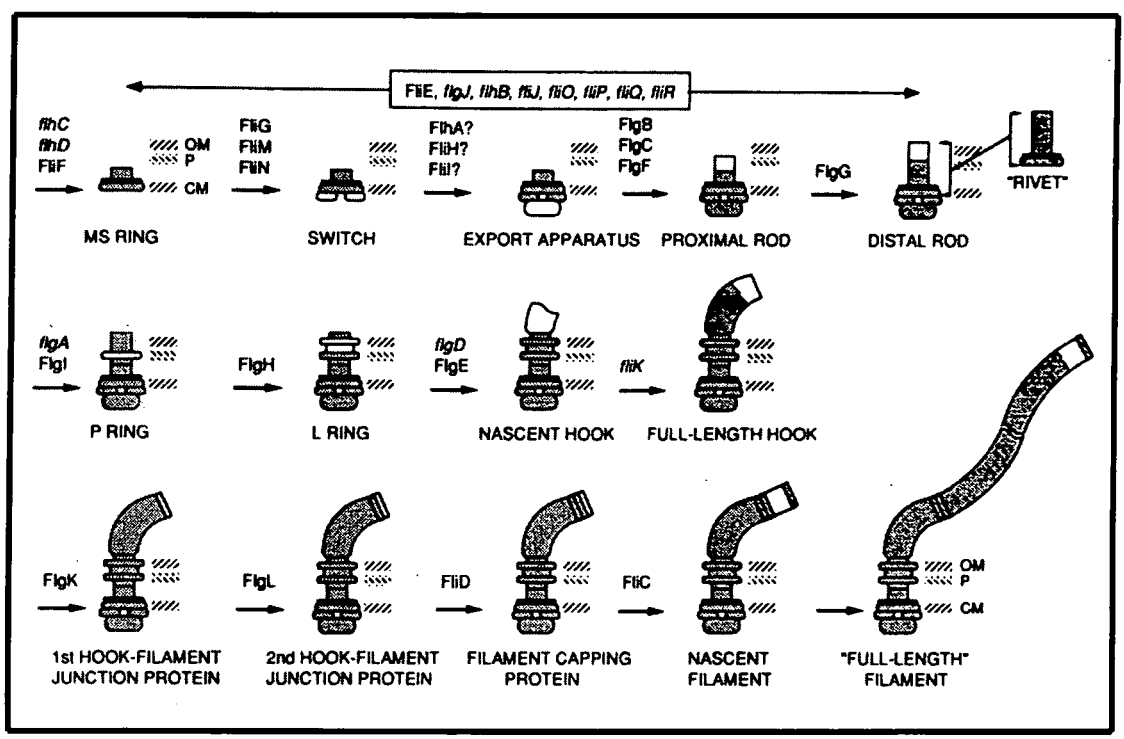
Much of the flagellum lies outside the cytoplasmic membrane and therefore, many flagellar proteins must be transported across the membrane. A channel in the filament, hook and rod is proposed to be responsible for this export of proteins. Rod formation is prevented in *flhA* , *fliH* and *fliI* mutants [Kubori, *et al.*, 1992] suggesting that these gene products are involved in flagellum-specific export. This apparatus must be completed before rod formation can begin, the flagellum-specific export pathway is discussed more fully later.

Four proteins comprise the rod (FlgB, FlgC, FlgG and FlgF) with FlgF being the major component although all four must be present at the same time for rod formation to occur [Aizawa, 1996]. It is believed that FlgG is assembled after the other rod proteins and is found at the distal end [Jones and Macnab, 1990; Kubori, *et al.*, 1992]. FliE must also be present for the rod to be assembled [Macnab, 1992]. It is a component of the basal

Figure 1.8.1 The biogenesis of the flagellum of *E. coli* and *S. typhimurium*

The various stages of assembly can be seen along with the genes and proteins required for each stage. Roman letters denote the proteins which are known to assemble into a structure, the genes with unknown product function are written in italics. *flgD* should not be in italics as the function of FlgD is now known. Each additional feature is shown in white with the preceding structure shaded. The cell membrane (CM), periplasmic space (P) and outer membrane (OM) can be seen. Each stage is discussed more fully in the text.

The figure was reproduced with the permission of R. M. Macnab [1992].



body and is hypothesised to act as an adaptor between the MS ring and the rod [Muller, *et al.*, 1992].

When the rod becomes mature, hook growth begins at the tip. This is halted when the hook reaches the outer membrane [Aizawa, 1996]. Assembly of the outer P and L rings is now necessary to form a hole through the outer membrane. The P and L rings are believed to be exported by the Sec-dependent pathway [Pugsley, 1993] before assembling near the centre of the rod [Aizawa, 1996; Macnab, 1992]. The mechanism of this action is not yet known.

For hook synthesis to begin *in vivo*, the FlgD protein must be present. FlgD is added to the rod structure and remains at the distal tip during hook synthesis [Ohnishi, *et al.*, 1994]. FlgD acts as a scaffolding protein allowing for hook protein subunits (FlgE) to polymerise into hooks under the FlgD cap [Aizawa, 1996; Ohnishi, *et al.*, 1994]. Upon completion of the hook FlgD is lost and replaced by FlgK [Ohnishi, *et al.*, 1994].

FlgK and the other hook-associated protein FlgL (HAP1 and HAP3) separate the hook from the filament. Once these are asssembled and the filament cap protein FliD is present, filament synthesis and assembly can begin [Aizawa, 1996; Ohnishi, *et al.*, 1994]. The flagellin subunits are exported through the central channel to the distal end where filament formation begins. This assembly requires the protein FliD (HAP2) to act as a scaffolding protein and allow for flagellin subunits to be added beneath a FliD cap [Aizawa, 1996]. Filament length is not strictly regulated like hook length, but the rate of elongation slows as the filament lengthens. This is presumably due to frictional force as the flagellin subunits pass through the central channel [Macnab, 1992].

The MotA and MotB proteins are not needed for flagellar assembly and the point at which they assemble around the basal body is not known. However, as their main function is with the flagellar switch it is assumed that they are assembled early in the pathway [Macnab, 1992].

1.9 The flagellum-specific export pathway

The flagellum-specific export apparatus is proposed to be located at the base of the flagellum, as the physical pathway is through the nascent structure itself [Macnab, 1990; Vogler, *et al.*, 1991]. Experiments with temperature sensitive mutants revealed four flagellar genes which failed to produce reassembled flagella when mutated [Dreyfus, *et al.*, 1993; Macnab, 1990]. The protein products of these genes are needed before rod

synthesis after the M ring has been assembled and previously three of the genes had unknown functions (*flhA*, *fliH* and *fliI*). The fourth protein, FliN is a part of the switch-motor complex [Irikura, *et al.*, 1993; Macnab, 1990] and is thought to interact indirectly with the flagellum-specific export pathway [Vogler, *et al.*, 1991].

fliI encodes a protein that has been shown to have sequence similarity to the catalytic β subunit of the F_OF_1 ATPases and hence may be involved in supplying the necessary energy for the export pathway [Dreyfus, *et al.*, 1993]. In addition, FliI is estimated to have a high copy number, approximately 1500 per cell, and is suggested to interact dynamically with the flagella [Dreyfus, *et al.*, 1993].

FliI and FlhA have been found to have sequence similarity to some virulence proteins involved in the protein-export pathways of mammalian and plant pathogens [Dreyfus, *et al.*, 1993]. FliI has about 40% sequence similarity to Spa47 of *Shigella flexneri* [Dreyfus, *et al.*, 1993; Venkatesan, *et al.*, 1992] and 50% similarity to HrpB6 of *Xanthomonas campestris* [Dreyfus, *et al.*, 1993]. FlhA has been shown to be very similar to LcrD of *Yersinia pestis*, InvA of *S. typhimurium*, VirH of *S. flexneri* and HrpC2 of *X. campestris*, all thought to be connected in the translocation of proteins across cell membranes [Dreyfus, *et al.*, 1993; Galan, *et al.*, 1992]. The function of FliH has yet to be determined.

All of these proteins, along with a further flagellar protein FliP, are believed to belong to a superfamily of proteins involved in the control, export and assembly of proteins. It is believed that further components of these systems remain to be found [Dreyfus, *et al.*, 1993].

1.10 Motility and chemotaxis systems in other bacteria

The systems discussed up to now have been those most studied. However, the systems are not uniform throughout the microorganism world. Flagellar synthesis and rotation in *Bacillus subtilis* appears to be similar to the model systems with considerable homologies between the genes involved in the processes [Aizawa, *et al.*, 1990; Carpenter, *et al.*, 1992]. The organisation of the flagellar genes is very different to that of the Gram negative bacteria [Bischoff and Ordal, 1992]. Although these genes are found in five regions on the *B. subtilis* chromosome [Ordal, *et al.*, 1993], the majority of them are clustered in one big operon, the major *che/fla* operon, spanning over 26 kb of DNA [Carpenter, 1992]. This region contains genes encoding proteins involved in flagellar structure and assembly, intracellular chemotaxis, some with homologies to *E. coli* genes

of unknown function and some that appear unique to *B. subtilis* [Ordal, *et al.*, 1993]. Analysis of the MCP-dependent intracellular signalling system in *B. subtilis* uncovers the most striking differences. In a reverse operation to that which occurs in *E. coli* and *S. typhimurium*, it is the attractant rather than the repellent which induces CheA to phosphorylate CheY, producing CCW rotation and smooth swimming [Harshey and Toguchi, 1996; Ordal, *et al.*, 1993]. The addition or removal of high concentrations of attractants or repellents can produce methanol. Methyl groups are transferred from receptors to another protein which is then thought to be demethylated at the flagellum resulting in the methanol production. A CheZ homologue has not been found in *B. subtilis* and the CheY-PO_4 dephosphorylation mechanism has not been elucidated [Bischoff and Ordal, 1992; Ordal, *et al.*, 1993]. FliY, FliM and FliG seem to comprise the switch complex in *B. subtilis*, but no homologue of *fliN* has been found. Although FliY in *B. subtilis* is homologous to FliM and FliN of *E. coli* and *S. typhimurium* , it is a much bigger protein [Bischoff and Ordal, 1992]. Despite these differences, biochemical and genetic evidence point to parallels in the flagellar apparatus and control mechanisms between the bacteria.

Vibrio parahaemolyticus is a marine bacterium and human pathogen which produces two distinct types of flagella, each adapted for a different type of motility. In a liquid medium, a single polar flagellum propels the bacterium forward in a swimming motion. Under viscous conditions or on surfaces, synthesis of a second flagellar system is induced. The swimmer cell differentiates into a swarmer cell with many lateral flagella [Atsumi, *et al.*, 1992; McCarter and Wright, 1993]. The two systems are distinct from each other with no shared structural components. The polar flagellum is composed of two different types of flagellin subunits with a sheath formed from the outer membrane [McCarter and Wright, 1993]. The lateral flagella are similar to those of *E. coli* and *S. typhimurium*, they are arranged peritrichously and are made of only one flagellin subunit LafA [McCarter and Wright, 1993]. The energy source for the polar flagellar motor is the sodium-motive force, whereas the lateral flagella are driven by the proton-motive force [Atsumi, *et al.*, 1992]. Although the two flagellar systems remain distinct there is believed to be a common chemosensory apparatus [Sar, *et al.*, 1990].

Rhodobacter sphaeroides has a single subpolar flagellum which rotates unidirectionally in a clockwise rotation. Periodically the flagellar rotation pauses, cell body reorientation occurs by Brownian motion and the direction of swimming changes [Packer and Armitage, 1993]. *R. sphaeroides* responds chemotactically to a wide range of attractants such as amino acids, sugars and organic acids, resulting in an accumulation in regions of increased nutrient concentration [Armitage,J.P., *et al.*, 1990; Packer, *et al.*, 1996]. *R. sphaeroides* also shows a second response, chemokinesis, to an increase in the

concentration of the limiting metabolite. This produces an increase in the speed of flagellar rotation [Armitage,J.P., 1993]. Chemokinesis produces a dispersal of the bacterial population [Armitage,J.P., *et al.*, 1990] and is totally independent to the chemotaxic response [Packer and Armitage, 1994]. Photosynthetic bacteria like *R. sphaeroides* are also phototactic unlike *E. coli*.

Caulobacter crescentus has a single polar flagellum made up of three different types of flagellin subunits [Driks, *et al.*, 1989]. *C. crescentus* has been found to have a very similar hierarchial control of flagellar biogenesis to *E. coli* and *S. typhimurium* [Stephens and Shapiro, 1993]. No class I homologues have been identified and hence the class 2 genes, whose transcription is coupled to chromosomal replication, are the first flagellar genes expressed [Stephens and Shapiro, 1993]. Classes 3 and 4 contain most of the structural components of the flagellar and chemotaxis systems, which are transcribed with specific sigma factors and *trans*-acting factors [Stephens and Shapiro, 1993].

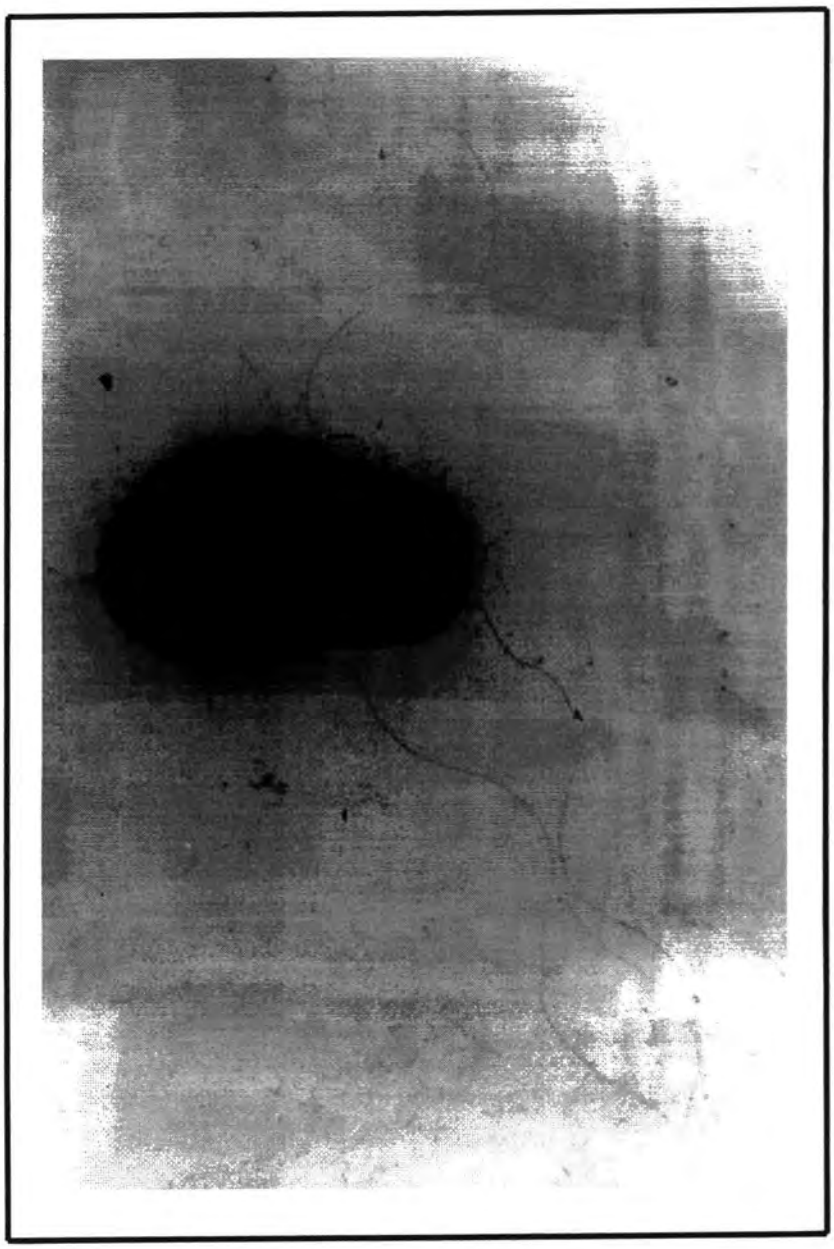
Flagellar motility is known to be an essential requirement for *Helicobacter pylori* to colonise the human gastric mucosa. The genetic organisation of the flagellar apparatus is different from that of other bacteria in numerous respects. There are two flagellin genes which code for distinctly different flagellin proteins, they are preceded by different promoters and are not linked on the chromosome [Suerbaum, 1995]. Very little is known about flagellar gene expression regulation, however, evidence suggests that the regulatory network of flagellar biosynthesis is distinctly different to that of the model systems [Suerbaum, 1995]. The most significant contrast is that mutant strains of the hook (*flgE*) gene still express the FlaA and FlaB flagellin proteins although the reason why has not yet been established [Suerbaum, 1995].

1.11 Flagella pattern and environmental effects in *Agrobacterium tumefaciens*

A. tumefaciens has an unusual flagellation pattern [Shaw, 1990]. The standard pattern is of two polar or sub-polar flagella at one pole only and 2-4 lateral flagella. The filaments all show a typical sinusoidal curvature when viewed with an electron microscope. In contrast to *E. coli* the flagella appear to rotate unidirectionally in a clockwise direction [Loake, 1989; Loake, *et al.*, 1988]. Runs of *A. tumefaciens* are long with relatively infrequent tumbles. The strain C58C1 has been observed to be an active swimmer and thus derivatives of it have been used for this study. Figure 1.11.1 shows an electron micrograph of a wild-type *A. tumefaciens* cell.

Figure 1.11.1 ***Agrobacterium tumefaciens***

An electron micrograph of a wild-type *A. tumefaciens* cell. The two polar flagella and two lateral flagella can be seen. The photograph was taken by Dr. C.H. Shaw.



A. tumefaciens has been shown to respond to sugars abundant in plant extracts [Loake, *et al.*, 1988], sucrose producing the most acute response. Oligosaccharides in general produce greater responses than monosaccharides probably due to the breakdown of oligosaccharides in the periplasm and the synergistic effects of the component monosaccharides [Loake, *et al.*, 1988]. *Agrobacteria* have been shown to increase in number around roots [Schroth, *et al.*, 1971] probably as a result of the release of fixed carbon molecules by plant cells into the rhizosphere [Hawes, 1990]. Amino acids have been shown to be poor chemoattractants, with only arginine, valine and glutamic acid producing a chemotactic response [Loake, *et al.*, 1988]. Finally, research indicates that plant wound exudate-specific phenolic compounds evoke chemotactic responses in *A. tumefaciens*. Two classes of chemoattractants have been identified [Ashby, 1988; Ashby, *et al.*, 1988]. The strongest response was produced by acetosyringone, sinapinic acid and syringic acid, a group which require the presence of a Ti-plasmid [Ashby, 1988; Ashby, *et al.*, 1988]. The second group did not require a Ti-plasmid but needed to be present in much higher concentrations for a response to occur [Ashby, 1988; Ashby, *et al.*, 1988]. This suggests the involvement of two signalling pathways, the first which requires the Ti-plasmid and the second which is chromosomally encoded.

1.12 Environmental effects on other rhizosphere bacteria

Colonization of plant roots by bacteria is not well understood. Laboratory experiments have shown that many plant pathogenic and soil borne bacteria are susceptible to chemoattractants although this data has not been supported by *in situ* studies [Hawes, *et al.*, 1988]. Research has indicated that *Rhizobium meliloti* uses chemotaxis and motility to spread in soil [Soby and Bergman, 1983] and *Pseudomonas fluorescens* uses flagella action for potato colonization [deWeger, *et al.*, 1987]. Rhizobia respond to chemoattractants within plant root exudates and as a result become localised at certain areas of roots [Currier and Strobel, 1976; Gaworzewska and Carlile, 1982; Gulash, *et al.*, 1984]. Chemoattractants, including amino acids, sugars, flavones such as luteolin and naringen [Aguilar, *et al.*, 1988; Caetano-Anolles, *et al.*, 1988] are secreted by cells in the zone of elongation. Rhizobia migrate to this region and it is here that the root is most susceptible to infection [Bhuvaneswari, *et al.*, 1981].

Agrobacterium-plant cell interaction

1.13 Attachment

Once *A. tumefaciens* arrives at the wound site on the plant, it must attach itself to the plant to allow transformation to occur. Attachment is a two step process, firstly the bacterium binds loosely to the surface of the host cell in a reversible process mediated by rhicadhesin, a bacterial protein [Sheng and Citovsky, 1996; Smit, *et al.*, 1989]. Secondly, the bound bacterium synthesises cellulose filaments which anchor it securely to the plant in a tight irreversible binding [Matthysse, 1983; Sheng and Citovsky, 1996]. Other agrobacteria can become trapped at this junction forming an aggregate which is likely to increase the chance of tumour production occurring [Matthysse, 1983]. Each plant cell may bind only a finite number of bacteria and, hence, specific binding sites are thought to be involved [Lippincott and Lippincott, 1969]. The identity of these receptors remains to be found, although experimental evidence suggests that a vitronectin-like molecule found on the cell surface of many plants may play an important role [Sheng and Citovsky, 1996]. The genes involved in attachment are located in the *Agrobacterium* chromosome and include *chvA, chvB, pscA* (or *exoC*) and *att* [Hooykaas and Beijersbergen, 1994]. In addition to their attachment roles, the *chv* (chromosomal virulence) genes are necessary for *vir* gene induction. Genes involved include *chvD, chvE, Ivr*-221, -223, -225 and *miaA* [Gray, *et al.*, 1992; Huang, *et al.*, 1990].

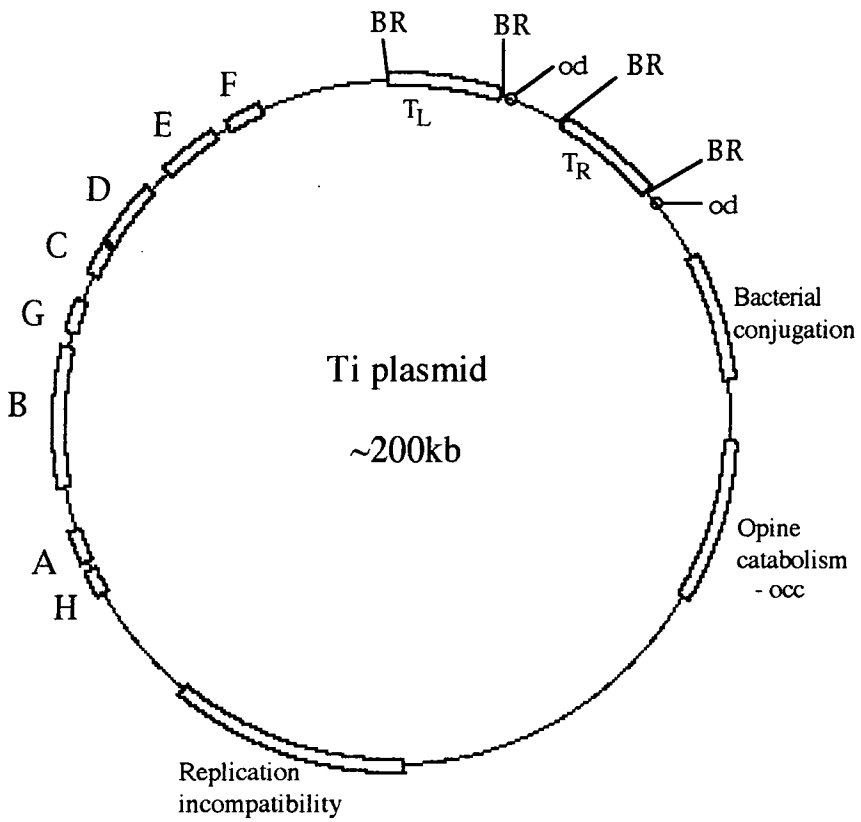
1.14 Ti- plasmid

Three genetic components of *Agrobacterium* are necessary for plant cell transformation. They are the T-DNA, the virulence region and the previously mentioned chromosomal virulence genes. The T-DNA is a discrete segment of DNA located on the 200 kb Ti plasmid. As mentioned previously crown gall tumours are characterised by the production of opines, tumour specific compounds formed by condensation of an amino acid and a keto acid or a sugar [Hooykaas and Beijersbergen, 1994]. The synthesis of the opines is catalysed by opine synthases which are encoded by the T-DNA. Agrobacteria are classified into strains based on the kind of opines produced in the tumours, namely, ocotopine, nopaline, succinamopine and leucinopine [Hooykaas and Beijersbergen, 1994]. The four groups of Ti plasmids range in size from 190-240 kb and are present in the cell at a low copy number [Melchers and Hooykaas, 1987]. The Ti plasmid consists of five distinct regions:- the T-DNA, the *vir* region, the opine metabolism region, a bacterial conjugation region and a region for replication and incompatibility [Binns and Thomashow, 1988]. Figure 1.14.1 shows a genetic map of an octopine- type Ti plasmid. The T-DNA can be found in different locations, for example in the nopaline Ti plasmid the T-DNA is a continuous stretch of approximately 22 kb whereas the octopine-specific T-DNA is made up of three independently transported T-DNAs [Sheng and Citovsky, 1996].

Figure 1.14.1 An octopine-type Ti plasmid

The diagram shows the relative locations of the T-DNA, the *vir* genes and the border repeats. Additionally, the approximate positions of the loci responsible for plasmid replication and incompatibility, bacterial conjugation and opine catabolism can be seen.

The diagram is not drawn to scale.



od = overdrive sequence

BR = border repeats

The T-DNA regions are flanked by border repeats. The borders are conserved 25 bp sequences that delimit the transferred segments [Sheng and Citovsky, 1996]. Genetic studies have shown that the right border repeat is essential for T-DNA transfer, but the left border is not [Sheng and Citovsky, 1996].

1.15 The virulence region

The virulence region is approximately 35 kb in length with over twenty proteins encoded for, most of which are required for T-DNA transfer [Binns and Thomashow, 1988]. VirA and VirG act as a signal transduction system. They sense signal molecules released by wounded plant cells and activate the expression of other *vir* genes resulting in the initiation of T-DNA transport. *virA, virB, virD* and *virG* must be present for tumour formation to occur, with *virC, virE, virF* and *virH* only needing to be present for tumour formation on certain plant hosts [Hooykaas and Beijersbergen, 1994]. *virH* may be involved in the detoxification of the harmful phenolics secreted by some wounded plant cells [Sheng and Citovsky, 1996].

For transfer to occur, the T-DNA is nicked at the border sequences. VirD1 and VirD2 are believed to act together as an endonuclease that carries out site and strand specific nicks in the border repeats [Sheng and Citovsky, 1996]. This action generates a linear single-stranded copy of the T-DNA region, termed the T-strand [Sheng and Citovsky, 1996]. VirC1 can enhance the formation of the T-strand from octopine Ti plasmids [Willmitzer, *et al.*, 1983].

It is believed that the T-strand is transferred into the plant cell as a protein-nucleic acid complex called the T-complex [Sheng and Citovsky, 1996]. The T-complex has at least three components, the T-strand DNA molecule containing the genetic information and the VirD2 and VirE2 proteins which protect the T-strand. VirD2 is covalently attached to the T-strand at its 5' end. It may act as a pilot protein within the plant cell, as it has two nuclear targeting signals [Herrera-Estrella, *et al.*, 1990]. VirE2 is believed to coat the single stranded T-DNA and protect it from external nucleolytic activity [Sheng and Citovsky, 1996].

A direct passageway between *Agrobacterium* and the plant is necessary for the T-complex to be transferred. It is believed that the *virB* operon, which contains 11 open reading frames, forms a channel to allow for T-complex transfer, termed the VirB channel [Sheng and Citovsky, 1996]. Two of the VirB proteins, VirB4 and VirB11 have homology

to ATPases and are thought to provide energy for the transport through the channel [Sheng and Citovsky, 1996].

1.16 *vir* gene induction

As mentioned previously VirA and VirG act as a two-component regulatory system which recognises signal molecules. It is still unclear if the phenolic signals (from plant wound sites) are directly sensed by VirA or if another receptor protein acts as an intermediate. Some experiments have suggested that the plant phenolics interact with the chromosomally encoded proteins p10 and p21 which then bind VirA. However, recent data indicates that VirA recognises the signals directly, because of the inability to isolate non-VirA mutants that are unable to bind phenolic compounds [Sheng and Citovsky, 1996]. Sugar enhancers of *vir* gene induction have been demonstrated to interact with ChvE, a chromosomally encoded glucose/galactose binding protein, which then interacts with VirA [Sheng and Citovsky, 1996].

After interaction with a signal compound, VirA autophosphorylates at its His-474 residue. The phospho-histidine high energy phosphate bond is then transferred from VirA to the Asp-52 or Asp-8 residue in VirG [Hooykaas and Beijersbergen, 1994; Sheng and Citovsky, 1996]. In order to activate expression of the other *vir* genes, VirG binds to the *vir* box. The *vir* box is a conserved 12 bp sequence present in the 5'-non-coding regions of all *vir* operons (TNCAATTGAAAPy). After VirG has bound to a *vir* box, phosphorylation may allow interactions with other proteins such as RNA polymerase to bring about transcription [Sheng and Citovsky, 1996].

1.17 T-DNA integration

After the T-complex has been imported into the plant cell, it needs to be integrated into the plant cell chromosome [Sheng and Citovsky, 1996]. At this time little is known about the mechanisms involved in this process although the T-strand associated proteins VirD2 and VirE2 are thought to be involved [Sheng and Citovsky, 1996].

1.18 Chemotactic and motility genes in *A. tumefaciens*

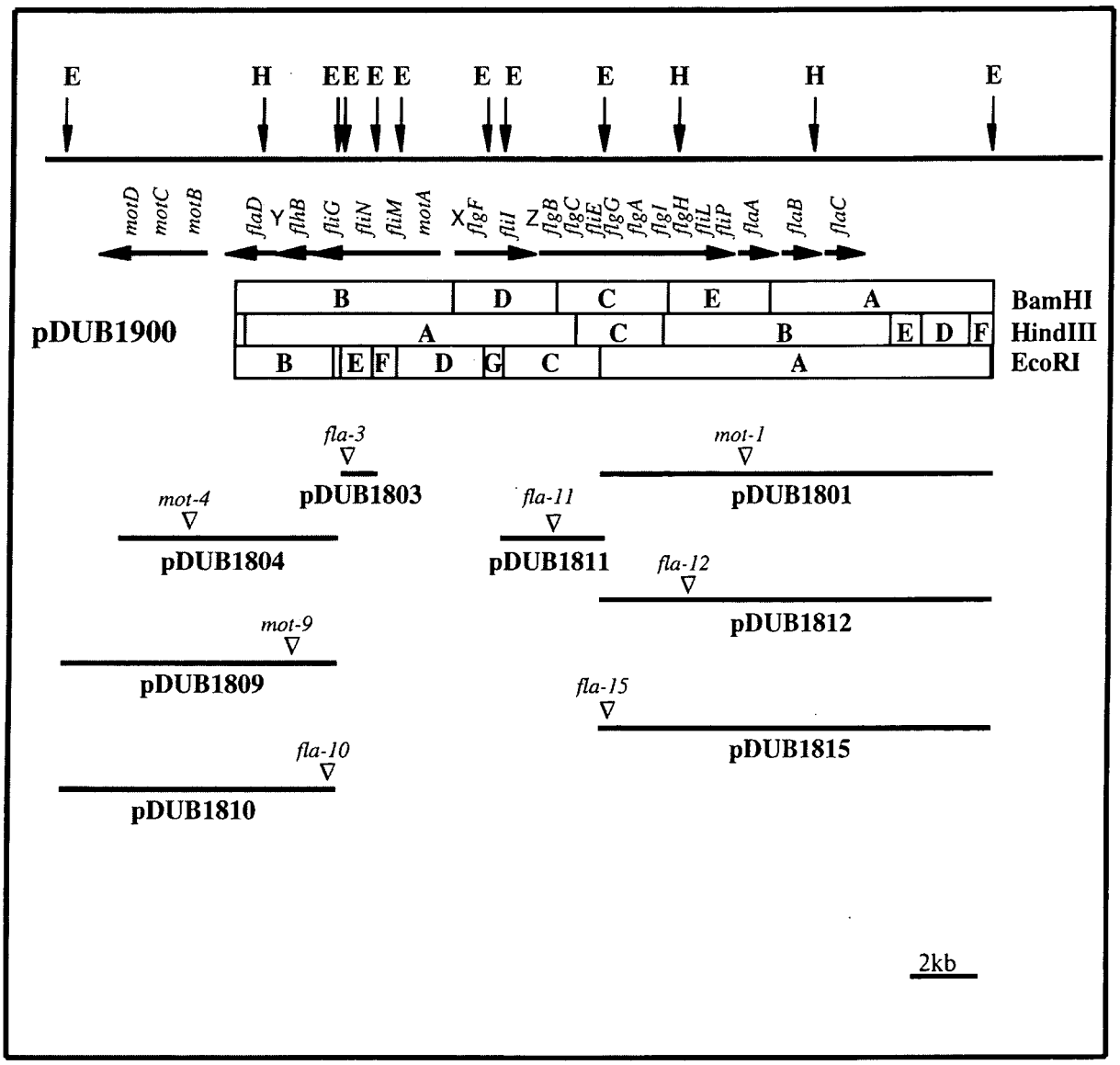
Numerous independent behavioural mutants have been created using the transposon Tn5 [Loake, 1989; Shaw, *et al.*, 1991]. Chemotactic mutants were identified and used to study the genes involved in chemotaxis and motility in *A. tumefaciens*.

A study by Shaw *et al.*, [1991] isolated twenty chemotactic and motility mutants. These were analysed by light and electron microscopy and their swarming ability determined by inoculation onto swarm agar plates. Of the twenty studied, seven were said to be *mot* mutants (non-motile but possess flagella), a further seven were *fla* mutants (non-motile with no visible flagella) and a *che* mutant was isolated. This *che* mutant (*che-2*) proved to be motile, but was found to tumble excessively. One mutant had a phenotype of tandemly paired cells attached extracellularly at the poles. This mutant swarmed more slowly than the wild-type and was designated *tpc*-17 (*t*andem *p*aired *c*ells). The final four mutants were indistinguishable from the wild-type in their chemotactic behaviour to L-broth in assays. They did however, produce intermediate sized swarms and thus were designated *ssw* mutants (*s*mall *sw*arms).

A Ti-plasmid was conjugated into each of the mutants. They were all found to be virulent when inoculated directly onto wounded sunflower seedlings. Thus chemotaxis and motility are not essential for virulence. The method of deWeger *et al.* [1987] was used to determine that *mot-1* seemed incapable of colonising newly grown shoots which seems to support the assumption that motility and chemotaxis are needed for rhizosphere colonisation [Hawes and Smith, 1989; Shaw, *et al.*, 1991].

A cosmid library of C58C1 DNA in pLAFR3 was constructed [Shaw, *et al.*, 1991]. The library was screened with radioactively labelled probes from the DNA sequences flanking Tn5 from some of the chemotactic and motility mutants created [Shaw, 1993; Shaw, *et al.*, 1991]. Seven cosmids were isolated. (pDUB1900 - 1906) pDUB1900 was found to either hybridise to or complement nine of the mutants namely, *mot-1, fla-3, mot-4, mot-9, fla-10, fla-11*, *mot-12*, *mot-14* and *fla-15*. Originally, these were all thought to have separate Tn5 insertion sites giving different phenotypes, but a study by W. J. Deakin [1994] has shown that *mot-12* and *mot-14* have the same Tn5 insertion site and the phenotype was wrong consequently the *mot-14* nomenclature was dropped.

Restriction mapping of pDUB1900 has been carried out using the enzymes *Hin*dIII, *Bam*HI, *Eco*RI and *Hpa*I. Figure 1.18.1 shows the restriction map produced.


E
H
E E E E
E E
E
H
H
E
motD motC motB
flaD Y flhB fliG fliN fliM motA X flgF fliI Z flgB flgC fliE flgG flgA flg1 flgH fliL fliP flaA flaB flaC
pDUB1900
B D C E A
A C B E D F
B E F D G C A
BamHI
HindIII
EcoRI
fla-3
pDUB1803
mot-1
pDUB1801
mot-4
pDUB1804
fla-11
pDUB1811
fla-12
pDUB1812
mot-9
pDUB1809
fla-15
pDUB1815
fla-10
pDUB1810
2kb

1.19 Aims

The aims of this study were to sequence a part of the pDUB1900 cosmid clone (labelled in figure 1.18.1), upstream of the putative flagellar gene operon identified by Deakin [1994]. This sequence data was to be analysed for the presence of any open reading frames, and the databases searched with these to try and identify any homologies to known flagellar genes. The flagellar operon described by Deakin [1994] (containing *flgI, flgH, fliP* and ORFs A, B and C) was mapped in *R. meliloti* and found to have the same gene order, therefore, any identified homologues or unknown open reading frames present in the *A. tumefaciens* sequence were to be compared to *R. meliloti* DNA to determine if they were also conserved between the two species, and if so, in the same order. Finally, a number of other flagellar operons in *A. tumefaciens* contain open reading frames with no homology to other flagellar genes. Therefore, any unidentified open reading frames located in this study were to be analysed to try and determine their function by making mutants using gene replacement mutagenesis methods.

Chapter 2

Materials and Methods

2.1 Materials

All inorganic chemicals were purchased from BDH Chemicals Ltd., Poole, Dorset, U.K. and were of AnalaR quality, unless specified otherwise.

All organic chemicals and enzymes were purchased from Sigma Chemicals Plc., Poole, Dorset, U.K. unless specified otherwise.

Agar bacteriological (no.1) and yeast extract were from Oxoid Ltd., Basingstoke, Hants., U.K..

Agarose was from BRL, Gaithersburg, U.S.A..

Sephadex G-50 was from Pharmacia Fine Chemicals, Uppsala, Sweden.

Filter paper (3MM) and laboratory sealing film were from Whatman International Ltd., Maidstone, U.K..

Fuji RX-100 X-ray film was from Fuji Photo Film Co., Ltd., Japan.

Lab M Nutrient Broth (no.2), Lab M Nutrient Agar, nylon hybridisation transfer membranes and radiochemicals were from Amersham Ltd., Bury, U.K..

Minisart filters were from Sartorius GmbH, Postfach 3243, D-3400 Göttingen, Germany.

Nitrocellulose discs (25 mm, 0.22 μm pore size) for tri-parental mating were from Millipore (UK) Ltd., Watford, U.K..

Polaroid film was from Polaroid (UK) Ltd., St. Albans, Hertfordshire, U.K..

Restriction endonucleases, corresponding buffers, T4 DNA ligase, Klenow enzyme, X-gal and wild type λ DNA were from NBL, Cramlington, Northumberland, U.K., Boehringer Mannheim (UK) Ltd, Lewes, U.K., or New England Biolabs, CP Labs Ltd., Bishop's Stortford, Hertfordshire, U.K..

Sodium chloride was from Riedel-de Haen, Seelze, Germany.

Trypticase Peptone was from BBL, Cockeysville, U.S.A..

Vacuum grease was from Dow Corning S. A., Seneffe, Belgium.

2.2 Bacterial strains and plasmids

***Agrobacterium tumefaciens* strain**

C58C1	rifampicin resistant ; lab stock

***Escherichia coli* strain**

DH5α	*supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 hi-1 relA1* ; lab stock

Plasmids

pUC19 and pUC18	general *E. coli* vectors, ampicillin resistant
pBluescript SK+ and KS+	phagemids derived from pUC19, ampicillin resistant
pRK2013	helper plasmid for tri-parental matings in *A. tumefaciens*, has the ColE replicon containing the transfer function of RK2, kanamycin resistant
pDUB1900	a C58C1 cosmid library clone containing a *Bam* HI chromosomal fragment in the vector pLAFR-3 (carries cloned behavioural genes), tetracycline resistant
pDUB2033	Neomycin cassette
pRZ1, pRZ2 and pRZ4	pLAFR-1 broad host range vector containing cloned behavioural genes of *R. meliloti* , tetracycline resistant
pJQuc1	basic vector used for the gene replacement mutagenesis of *A. tumefaciens* possesses *sacB* and gentamycin resistance

Methods

2.3 Standard biochemical and molecular biological techniques

All standard techniques are common practise in the department of Biological Sciences, University of Durham or are based on those in Molecular Cloning: A Laboratory Manual by Sambrook *et al.*, [1989] unless a reference is specifically given.

2.4 Glassware, plasticware, reagents and general apparatus

All apparatus coming directly or indirectly into contact with enzymes, nucleic acids, bacteria and sensitive or sterile reagents were sterilised first by autoclaving at 121 °C for 20 minutes. All solutions for DNA manipulations, with the exception of gels, were autoclaved where possible or made up using sterile water and then put through a sterilising filter. All water used was distilled or deionised.

2.5 Procedures employed for the growth of bacteria

2.5.1 Media and conditions

The following types of media were used throughout this work;

Lab M nutrient broth no. 2 (LM broth)	25 g made up to 1 litre with distilled water gave final concentrations of 10 $g.l^{-1}$ balanced peptone no. 1, 10 $g.l^{-1}$ beef extract, 5 $g.l^{-1}$ NaCl, pH 7.5 +/- 0.2
LB broth	10 $g.l^{-1}$ trypticase peptone, 5 $g.l^{-1}$ yeast extract, 5 $g.l^{-1}$ NaCl
Chemotaxis media	0.1 mM EDTA pH 7.0, 10 mM phosphate buffer pH 7.0 (Phosphate buffer is 1 M KH_2PO_4 adjusted to pH 7.0 with 1 M KOH)

Lab M nutrient agar (LM agar)	28 g made up to 1 litre with distilled water gave final concentrations of 5 $g.l^{-1}$ peptone, 3 $g.l^{-1}$ beef extract, 8 $g.l^{-1}$ NaCl, 12 $g.l^{-1}$ agar no.2, pH 7.3 +/- 0.2.

Except for when Lab M nutrient agar (LM agar) was used, agar was added at a concentration of 1 % for solid plates or 0.16 % for swarm plates.

Antibiotics were added to cooled media after autoclaving to produce the following concentrations.

For *E. coli*; ampicillin 50 $\mu g.ml^{-1}$, gentamycin 15 $\mu g.ml^{-1}$, kanamycin 50 $\mu g.ml^{-1}$, neomycin 50 $\mu g.ml^{-1}$ and tetracycline 15 $\mu g.ml^{-1}$. When selecting for the inactivation of the β-galactosidase gene by insertion of DNA fragments into the multiple cloning sites of pUC19 plasmids, 40 µl of 20 $mg.ml^{-1}$ X-gal (in DMF) were spread over the surface of agar plates.

For *A. tumefaciens*; gentamycin 100 $\mu g.ml^{-1}$, kanamycin 25 $\mu g.ml^{-1}$, neomycin 100 $\mu g.ml^{-1}$, rifampicin 100 $\mu g.ml^{-1}$ and tetracycline 10 $\mu g.ml^{-1}$.

For *R. meliloti*; as *A. tumefaciens* and streptomycin 100 $\mu g.ml^{-1}$.

Aseptic technique was always employed when handling bacterial cultures. Liquid cultures were inoculated with either a flamed loop or a sterile cocktail stick. Solutions and bacterial cultures were spread onto agar plates using a glass spreader which had first been sterilised with 70 % ethanol. Liquid bacterial cultures were incubated on an orbital shaker at 200 rpm, at the following temperatures-

E. coli	37 °C
Agrobacterium and Rhizobium strains	28 °C

2.5.2 Measurement of bacterial growth

The optical density of bacteria was measured at 600 nm on a Beckman DU7500 spectrophotometer.

2.5.3 Storage of bacterial strains

In the short term (approximately 2 months) bacterial colonies were stored on agar plates kept inverted at 4 °C, protected from light and sealed with laboratory sealing film (Whatman). Permanent storage was carried out at -80 °C. A single colony was inoculated into LM broth with the appropriate antibiotic selection and it was incubated overnight at the required temperature. 500 μl of the overnight culture was added to 500 μl of 80 % glycerol and it was immediately transferred on ice to the -80 °C freezer for storage.

2.6 Isolation of DNA

2.6.1 Quick plasmid minipreps

This quick method was used to prepare plasmid DNA when a large number of minipreps needed to be carried out. The DNA produced was generally of a low yield and quality, but after digestion with restriction enzymes and, hence, the correct plasmid containing strain identified, alkaline lysis minipreps or specialised kits were employed to produce plasmid DNA of a better yield and quality.

1.5 ml of an overnight culture was spun in a microfuge for 1 minute to harvest the cells. The pellet was resuspended in 100 μl of ice-cold miniprep solution 1 (10 mM EDTA pH 8.0, 1 % glucose, 25 mM Tris-HCl pH 8.0). This was left at room temperature for 5-10 minutes before the addition of 200 μl of miniprep solution 2 (0.2 M NaOH, 1 % SDS). The tube was incubated on ice for 5-10 minutes. Following this, 150 μl of miniprep solution 3 (11.5 ml of glacial acetic acid, 28.5 ml of distilled water, 60 ml of 5 M potassium acetate) was added and the tube incubated on ice for a further 5-10 minutes. The tube was then microfuged for 10 minutes and the supernatant transferred to a fresh tube. This was repeated to ensure no precipitate was carried over. To the final supernatant, 2 volumes of 100 % ice-cold ethanol were added, and the tube placed at -20 °C for 5 minutes to precipitate the DNA. The DNA was collected by centrifugation for 5 minutes. The resulting pellet was washed in 1 ml of 70 % ice-cold ethanol, air-dried and resuspended in 50 μl of TE buffer (1 mM EDTA pH 8.0, 10 mM Tris.HCl pH 8.0) with RNAase A added to a final concentration of 20 $\mu g.ml^{-1}$ (see 2.7.1).

2.6.2 Alkaline lysis with polyethylene glycol precipitation plasmid minpreps

This method is that as primarily described by Sambrook *et al.*, [1989], and produces relatively pure plasmid DNA. It is essentially the same as the quick plasmid miniprep up until after the addition of miniprep solution 3. Having been spun for 10 minutes to remove bacterial debris, the clear supernatant was transferred to a fresh tube and RNAase A was added to a final concentration of 20 $\mu g.ml^{-1}$. This was incubated at 37 °C for 20 minutes before being extracted with an equal volume of TE saturated phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1). It was spun for 2 minutes in a microfuge and the upper aqueous phase was transferred to a fresh tube. 2 volumes of 100 % ethanol were added to precipitate the DNA, and the tube was incubated at -20 °C for 2 hours. The DNA was collected by centrifugation at 12,000 xg for 10 minutes, washed twice in 70 % ethanol and dried briefly. The pellet was taken up in 16 μl of distilled water and 8 μl of 2 M NaCl and 20 μl of 13 % polyethylene glycol were added. This was mixed well and incubated on ice for 20-30 minutes. The DNA was collected again by centrifugation at 12,000 xg for 10 minutes. The final pellet was washed 3 times with 200 μl of ice-cold 70 % ethanol, dried briefly and resuspended in 20 μl of distilled water.

2.6.3 Large scale plasmid preparation

Two 50 ml overnight cultures were spun at 7000 xg for 7 minutes in a Beckman Centrifuge using a JA20 rotor to pellet the cells. Each pellet was then resuspended in 3.335 ml of ice-cold solution I (1 % glucose, 10 mM EDTA pH 8.0, 25 mM Tris pH 8.0) and 20 μl of 10 mg/ml RNAase A was added. 3.335 ml of solution II was added (0.2 M NaOH, 1 % SDS) and they were mixed well. A further 3.335 ml of ice-cold solution III (6 ml of 5 M KAc, 1.15 ml of glacial acetic acid, 2.85 ml of distilled water) was added and they were mixed again. The tubes were spun at 15000 rpm using an F0650 rotor, at 4 °C for 5 minutes. The supernatants were transferred to new tubes and 20 ml of phenol:chloroform (24:1) was added to each. After a good mix, the tubes were spun at 3500 rpm for 3 minutes. The aqueous layers were transferred to fresh tubes and 20 ml of ethanol was added to them. The tubes were left on ice for 15 minutes before being spun at 15000 rpm for 15 minutes at 4 °C to pellet the DNA. The pellets were washed with 5 ml of 70 % ethanol, spun for 3 minutes at 15000 xg and allowed to air dry. The two final DNA pellets were each resuspended in 500 μl of either TE buffer or distilled water to give a total volume of 1 ml.

2.6.4 Preparation of bacterial chromosomal DNA

This method utilises a modified version of that of Dhaese *et al.*, [1979].

A 5 ml LM broth culture was grown to stationary phase. A 400 μl aliquot of this was transferred to an eppendorf tube and the cells harvested by centrifugation for 1 minute in a microfuge. The supernatant was removed and the pellet left for 30 minutes at -20 °C. The pellet was then resuspended in 200 μl of TE buffer, 8 μl of 10 mg.ml^{-1} aqueous stock of lysozyme [Sambrook, *et al.*, 1989] was added and the tube was incubated at 37 °C for 30 minutes. 40 μl of 4 M sodium perchlorate, 24 μl of 10 % SDS and 8 μl of Proteinase K (20 mg.ml^{-1} aqueous stock) were added to lyse the cells. The tube was incubated at 45 °C for 2 hours before the DNA was precipitated by the addition of 2 volumes of 100 % ethanol. The DNA was pelleted by centrifugation for 5 minutes in a microfuge, it was washed in 70 % ethanol, dried and resuspended in 500 μl of TE buffer. The DNA solution was extracted twice with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform:isoamyl alcohol (24:1). The final aqueous layer was precipitated with ethanol as previously described and the pellet resuspended in 50 μl of TE buffer with RNAase A added to a final concentration of 20 μg.ml^{-1}.

2.7 DNA manipulation techniques

2.7.1 RNAase treatment of DNA solutions

RNA was removed from a DNA solution to prevent contamination, by treatment with RNAase A. A stock solution of RNAase was prepared (Pancreatic RNAase A was dissolved in 15 mM NaCl, 10 mM Tris.HCl pH 7.5 at a concentration of 10 mg.ml^{-1}. The solution was boiled for 15 minutes and allowed to cool slowly to room temperature to inactivate any DNAases present. Aliquots were stored at -20 °C). This was added to the DNA solution to a final concentration of 20 μg.ml^{-1}. Incubation was carried out at 37 °C for the required length of time. Larger amounts of plasmid DNA or chromosomal DNA were treated with 50 μg.ml^{-1} RNAase A for 1 hour at 37 °C. The enzyme was removed by phenol:chloroform extraction before the DNA was precipitated. The majority of the RNAase treatments were carried out during restriction enzyme digestions.

2.7.2 Phenol-chloroform extraction of DNA solutions

The following method was taken from Deakin [1994].

Proteins were removed from DNA solutions by extracting twice with ~ 1 volume of phenol:chloroform:isoamyl alcohol saturated with TE buffer. (The phenol:chloroform:isoamyl alcohol, 25:24:1, was equilibrated 3 times with TE buffer and stored under TE in a light-proof bottle at 4 °C.) Followed by one extraction with chloroform:isoamyl alcohol (24:1) to remove any traces of phenol from the DNA solution. The layers were mixed by brief vortexing and separated by centrifugation for 2 minutes in a microfuge. The DNA remains in the top aqueous phase in each case.

2.7.3 Removal of proteins from DNA solutions using silica fines

Preparation of silica fines

The silica fines used for this protocol were a gift from Professor Nigel Robinson and were prepared as follows. 250 ml of silica 325 mesh powder was resuspended in distilled water to a final volume of 500 ml. This was stirred for 1 hour and then left for another hour to settle. The suspension was centrifuged in a Beckman J2-HS centrifuge using a JA-14 rotor at 5000 xg. The resulting pellet was resuspended in 150 ml of distilled water and 150 ml of nitric acid. The silica fines were heated to 98 °C and left to cool to room temperature, before being washed repeatedly with sterile distilled water until a pH greater than 5.5 was reached. The silica fines were stored as a 50 % slurry in distilled water at 4 °C.

Removal of proteins

The DNA solution was made up to a minimum volume of 200 μl with sterile distilled water. Two volumes of sodium iodide solution were added to it in an eppendorf tube (90.8 g NaI and 1.5 g Na_2CO_3 dissolved in distilled water, filter sterilised and saturated with 0.5 g Na_2SO_3 before being stored in a light-proof bottle at 4 °C). 5 μl of silica fines were added to the DNA solution, it was vortexed briefly and then incubated for 10 minutes at room temperature with constant agitation. The tube was spun for 15 seconds in a microfuge, the supernatant removed and the silica fine pellet washed with 70 % ethanol. Having been dried, the silica fine pellet was resuspended in 50 μl of TE buffer and incubated at 37 °C, with occasional shaking, for 10 minutes. The tube was again microfuged for 15 seconds and then the DNA containing supernatant was removed to a fresh tube.

2.7.4 Alcohol precipitation of DNA

DNA solutions were made to 0.3 M sodium acetate by the addition of $^{1}/_{10}$th volume of 3 M sodium acetate (pH 4.8), then 2 volumes of ethanol were added and mixed by vortexing. Chromosomal DNA solutions were left at room temperature for 15 minutes and plasmid DNA solutions were placed at -80 °C for 30-60 minutes. The precipitated DNA was pelleted by centrifugation for 10 minutes in a microfuge. The pellet was washed in 70 % ethanol and vacuum dried before being resuspended in sterile distilled water or TE buffer.

2.8 Gel electrophoresis of nucleic acids

2.8.1 Agarose gel electrophoresis

DNA fragments of varying size were electrophoresed on large 180 x150 mm maxigels (volume 200 ml) or 77 x55 mm minigels (volume 50 ml) ranging in concentration from 0.7-2 % (w/v),[See figure 2.8.1.1]. Horizontal slab gels submerged in buffer were used throughout. The required amounts of agarose and 1x TAE buffer (50 x stock:- 242 g Tris, 100 ml EDTA pH 8.0, 0.57.1 ml glacial acetic acid per litre) were mixed and boiled to dissolve the agarose. The solution was cooled to approximately 60 °C and 10 mg.ml^{-1} ethidium bromide was added to a final concentration of 0.2 μg.ml^{-1}. After the agarose had been poured into a gel mould with the required well comb, and the gel had set, the gel was put into a tank and covered with 1x TAE buffer containing 0.2 μg.ml^{-1} ethidium bromide in order to visualise the DNA bands. Samples were mixed with 0.2 volumes of 6x gel-loading buffer (6x gel-loading buffer contains 0.25 % xylene cyanol FF, 0.25 % bromophenol blue and 40 % sucrose in distilled water. The buffer was filter sterilised and stored at 4 °C.) prior to loading. The gels were electrophoresed at 5-10 V.cm^{-1} for the required amount of time.

Small DNA fragments upto approximately 300 bp, were analysed on MetaPhor Agarose gels (FMC BioProducts), which were prepared and used according to the supplier's protocol.

Figure 2.8.1.1 Approximate size range of double stranded DNA fragments analysed on an agarose gel of given concentration

Concentration of agarose gel (%)	Size range of ds DNA analysed (kb)
0.7	10 - 0.8
1.0	7 - 0.5
1.2	6 - 0.4
1.5	3 - 0.2
2.0	2 - 0.1

2.8.2 Analysis of band patterns on gels to determine fragment sizes and concentration

The DNA within the agarose gels was visualised under long wave ultraviolet light on a transilluminator (UVP Inc.), and photographed with a Polaroid RP4 Land camera (using a red filter) onto Polaroid 667 film. Restriction digests of lambda DNA were run on the gels alongside the sample DNA to act as size markers.

***Pst*I digested lambda DNA produces DNA fragments of the following sizes (in kb)**

14.05, 11.49, 5.07, 4.75, 4.51, 2.84, (2.56, 2.46, 2.44), 2.14, 1.99, 1.70, 1.16, 1.09, 0.81, 0.52, 0.47, 0.45, 0.34...

***Hin*dIII digested lambda DNA produces DNA fragments of the following sizes (in kb)**

23.13, 9.42, 6.56, 4.36, 2.32, 2.03, 0.56, 0.13

The fragments enclosed in brackets run together on agarose gels. To determine DNA concentration a known standard was also run out on the gel. 1 μl of pGem 3Z (Perkin Elmer) gives a band of 200 ng which can be compared to the sample DNA.

2.9 Isolation of DNA fragments from agarose gels

2.9.1 Silica fines

The silica fines used in this method were as described for removal of proteins from DNA solutions. The DNA band was excised from the agarose gel using a sterile scalpel blade and placed in an eppendorf tube. 800 μl of NaI solution was added to it and the tube was placed at 70 °C for 5 minutes or until all the agarose gel had melted. It was left for 5 minutes to cool to room temperature before the addition of 5 μl of silica fines. The silica fines extraction was as described previously.

2.10 Standard enzymatic methods

2.10.1 Restriction endonuclease digestions

Digestions were carried out according to the enzyme manufacturer's instructions. Plasmid DNA was usually digested in a total volume of 10-30 μl. 0.1 volumes of the supplied 10x concentrated enzyme buffer were added along with 5 units of restriction endonuclease and enough sterile distilled water to make up the volume. The digestion was incubated at the required temperature for 1-2 hours depending on the reaction volume. For digestions using more than one restriction enzyme with differing buffers, the reaction was buffered with one-phor-all buffer PLUS (Pharmacia). Chromosomal DNA was digested in bigger volumes, 100-200 μl, using appropriate amounts of buffer and sterile distilled water and 10 units of restriction enzyme for every microgram of DNA. A layer of mineral oil was added to the top to prevent evaporation and hence maintain buffering conditions. The large volume digestions were incubated overnight at the desired temperature.

For analysis of the digestions by gel electrophoresis, 0.2 volumes of 6x gel-loading buffer were added. (see 2.8.1).

2.10.2 Filling in 3'-recessed termini

After isolation from an agarose gel, the DNA was resuspended in 10 -15 μl of sterile distilled water. A 1 μl solution containing all 4 dNTPs at 1 mM was added to it along with 2 μl of 10x Klenow buffer (0.5 M Tris.HCl pH 7.6 , 0.1 M MgCl2), 1 μl of Klenow fragment and enough sterile distilled water to make the total volume upto 20 μl. The tube was incubated for 30 minutes at room temperature and then the Klenow fragment was either inactivated by heating to 70 °C for 5 minutes or removed by the silica fines method (see 2.7.3).

2.10.3 Ligation

DNA fragments with compatible cohesive or blunt termini were ligated using T4 DNA ligase. A maximum of 300 ng of DNA were added to the vector at a ratio of 3:1 (insert:vector). 0.1 volumes of 10x ligase buffer (0.66 M Tris.HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 10 mM ATP) and for cohesive termini 1 unit of DNA ligase were added. Incubation was carried out overnight at 4 °C. 3 units of ligase were added for blunt-ended termini and the reaction incubated overnight at 15 °C. The ligation reaction was used immediately after incubation to transform competent *E. coli* cells.

2.11 Transformation of strains of *E.coli* with plasmids

E. coli cells were made competent for transformation with plasmids essentially using a modified procedure of that described by Hanahan, [1983].

2.11.1 Preparation of competent cells

Cultures of the relevant strain were grown overnight in 5 mls of LM broth, at 37°C, from single colonies, and were used to inoculate, at a ratio of 1:100, fresh aliquots of LM broth. These were then grown on a fast shaker for increased aeration at 37 °C until an OD_{600} of 0.3-0.35 was reached. After being chilled on ice, the cells were harvested at 4000 xg for 7 minutes at 4 °C. The supernatant was carefully removed and the cells resuspended in $^2/_5$ the original culture volume of solution A (30 mM potassium acetate, 100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride and 15 % glycerol. The solution was adjusted with 0.2 M acetic acid to pH 5.8 and filter

sterilised).Great care was taken to maintain everything at 4 °C. After 5 minutes the cells were isolated as before, resuspended in 1/25 of the original culture volume of solution B (10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$ and 15 % glycerol. The solution was adjusted with KOH to pH 6.5 and filter sterilised.) and stored on ice for 15 minutes. The competent cells were transferred in 200 μl aliquots into prechilled eppendorfs, frozen in liquid nitrogen and stored at -80 °C.

2.11.2 Transformation procedure

Transformations were affected by adding the DNA, up to 250 ng/200 μl of competent cell suspension. The mixture was incubated at 4 °C for 15-45 minutes and heat shocked at 42 °C for 90 seconds, followed by the addition at room temperature of 800 μl of pre-warmed (to 37 °C) LM broth. Incubation ,with occasional shaking, took place at 37 °C for 1 hour. Samples of this culture, usually 1/10th and 9/10ths of the tube, were then spread on pre-warmed nutrient agar plates containing the relevant antibiotics.

2.12 DNA sequencing procedures

DNA sequencing was carried out on an Applied Biosystems 373 Stretch DNA Sequencer by both dye primer and dye terminator chemistries according to the manufacturer's protocols on double stranded plasmid DNA.

2.12.1 Preparation of double stranded DNA templates for sequencing

To produce high quality DNA sequence data it was imperative that the DNA template to be sequenced was as pure as possible. Qiagen QIAprep spin miniprep kits and Promega Wizard Miniprep DNA Purification Systems were used according to the manufactures' protocols to produce good quality DNA for sequencing. 1 μl of the DNA template was run out on an agarose gel alongside 1 μl of pGem 3Zf standard (200ng.μl^1,Perkin Elmer) to determine it's concentration. 0.5 μg of DNA in total was required for dye terminator reactions and 1.0 μg of DNA in total was required for dye primer chemistry.

2.12.2 DNA sequence handling and computer programs

DNA sequence data was processed on Apple Macintosh computers as was all other computing. The following software was used -

Microsoft word	word processing	
DNA Strider TM 1.2	sequence analysis	Christian Marck

Also used was the gcg package of programs. Sections of the DNA were examined using the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package mounted on the SEQNET VAX 3600 at SERC, Daresbury U.K., and then compared against the Gen Bank and EMBL databases using the FASTA program. Open reading frames in the sequence were identified using DNA Strider and translated into their protein sequence before being compared against the OWL database using the SEQNET program SWEEP.

2.13 Polymerase chain reaction methods

2.13.1 Standard PCR

The polymerase chain reaction, [Saiki, *et al.*, 1988] with a thermostable polymerase enzyme, was performed using the following protocol.

For a 25 μl reaction the following were combined;

2.5 μl of 10 x buffer (supplied with the enzyme)
2 μl of 2.5 mM each dNTP
40 pmol. of forwards oligo
40 pmol. of reverse oligo
2.5 units of Taq DNA polymerase
$MgCl_2$ conc. was varied between 1 mM and 5 mM
50-100 ng of substrate DNA
distilled water to 25 μl

This was mixed well, spun briefly and approximately 50 μl of mineral oil added to the top. The reaction was incubated on a thermal cycler (Perkin Elmer)

The standard conditions were :-

94 °C	5 minutes	
94 °C	1 minute	
50 °C	1 minute	X 25 cycles
72 °C	45 seconds	
72 °C	5 minutes	

The annealing temperature was varied depending on the Tm value of the primers according to Sambrook's equation

Tm=2(A+T) + 4(G+C0

2.14 DNA hybridisation procedures

2.14.1 The transfer of nucleic acids from gel to filter membrane

The method used was primarily that of Southern [1975] and the DNA was transferred to Hybond-N (Amersham) nylon membranes, according to the manufacturer's instructions. The DNA in an agarose gel was photographed alongside a ruler for band size estimation later. The DNA in the agarose gel was made single stranded by gently shaking the gel in denaturation buffer (1.5 M NaCl, 0.5 M NaOH) for 30 minutes. The gel was then rinsed twice with distilled water and neutralised by gentle shaking in neutralising buffer (1.5 M NaCl, 0.5 M Tris.HCl pH 7.2, 0.001 M EDTA) for 30 minutes before being given 2 final rinses in distilled water. For the transfer of large DNA fragments, (greater than 10 kb) the gel was subjected to an initial soak in 0.25 M HCl for 15 minutes, to partially depurinate the DNA before being rinsed twice with distilled water and denatured.

Gels were blotted in the conventional manner. For single sided (one-way) blots a 10 x SSC reservoir was set up (20 xSSC, 3.0 M NaCl, 0.3 M Na.citrate pH 7.0). A glass plate was placed over this reservoir to provide a platform for the blot to be assembled. A long piece of Whatman 3 mm paper was presoaked in 10 xSSC and then placed on the platform with each end dipping into the reservoir buffer. The gel was placed, wells uppermost, on top of the paper and a piece of Hybond-N nylon membrane, cut to the

same size as the gel and presoaked in 10 xSSC, was placed on top of it. At no time were the filters handled except with rubber gloves or blunt-ended forceps and only at the edges. Great care was taken to exclude any air bubbles to avoid uneven transfer of DNA before 3 sheets of Whatman 3 mm paper, presoaked in 10 xSSC and cut to the same size as the gel, were placed on top. Finally, 2 layers of disposable nappies were placed on top, covered with a glass plate and a 1 kg weight placed on the very top. Care was taken to avoid "short circuit" liquid paths around the gel. Double-sided (two-way) blots were essentially set up the same except that the blots were created by "sandwiching" the agarose gel between two piles on each side of the Hybond-N nylon filter, 3 pieces of Whatman 3 mm paper, 2 layers of disposable nappies and 2 glass plates. A 1 kg weight was placed on the top glass plate. Each type of blot was left for at least 16 hours to allow for DNA transfer. The apparatus was then dismantled, the position of the wells marked on the nylon membrane and the DNA fixed onto the membrane by exposure to UV light for 2 minutes wrapped in clingfilm on a transilluminator.

2.14.2 Labelling of DNA fragments

Radio-labelling

DNA fragments were labelled with [α-^{32}P] dCTP by the random primer labelling method using an Amersham Multiprime kit. 30-50 ng of DNA in a total volume of 28 μl were boiled for 5 minutes and cooled on ice for 2 minutes. To the DNA, 10 μl of labelling buffer, 5 μl of random hexanucleotide primers, 5 μl of [α-^{32}P] dCTP (equivalent to 50 μCi) and 2 μl of Klenow enzyme were added. The reaction was left to proceed either at room temperature overnight or for 2-3 hours at 37 $^{\circ}$C. A purification step was then carried out using Sephadex G-50 (Pharmacia). The labelled DNA was boiled for 5 minutes and cooled to 65 $^{\circ}$C before being used.

2.14.3 Sephadex G-50 column chromatography

Column chromatography with Sephadex G-50 fractionates DNA roughly by size such that the smaller the molecules the slower they pass through the column. Normally the columns were run in sterile 10 ml glass pipettes with a plug of siliconized glass wool at the bottom using approximately 6 ml of Sephadex. The columns were prepared by slurrying Sephadex in excess elution buffer (150 mM sodium chloride, 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.1 % SDS), and heating this mix to 65 $^{\circ}$C for 2 hours. The column was settled, washed with 1 ml of elution buffer and run using elution buffer. The

DNA was loaded on by allowing the column to run nearly dry then adding the DNA with 100 μl of elution buffer. Successive 100 μl fractions were collected by eluting with the elution buffer. The progress of the DNA through the column could be followed by a Geiger counter for ^{32}P labelled DNA and fractions of approximately 1 ml were pooled. Labelled DNA was detected by scintillation counting and is eluted in the first peak. The second peak contains unreacted nucleotides *etc.*

Fluorescein-labelling

The *Gene Images* Random Prime Labelling Module from Amersham Life Science was used. The kit was designed to label nucleic acid pobes by random prime labelling. Nonamers of random sequence are used to prime DNA synthesis on a denatured DNA template in a reaction catalyzed by the (exonuclease-free) Klenow fragment of *E.coli* DNA polymerase I. Fluorescein-11-dUTP (Fl-dUTP) partially replaces dTTP in the reaction so that a fluorescein-labelled probe is generated (kit protocol). The DNA to be labelled was diluted to a concentration of between 2-25 ng/μl. It was denatured by heating for 5 minutes in a boiling water bath, then held on ice for 2 minutes. In an ice bath 50 ng of the DNA was added to the following reagents; 10 μl of Nucleotide mix, 5 μl of Primer, 1 μl of Klenow enzyme solution and water to a final reaction volume of 50 μl. The reaction mix was incubated at 37 $^{\circ}$C for 1 hour. Prior to use the probe was denatured by boiling for 5 minutes and snap cooled on ice.

2.14.4 Hybridisation of labelled nucleic acid probes to filter bound nucleic acids

These procedures were used to detect nucleic acids bound to Southern blots. Hybridisation reactions were carried out using Techne Hybridisation tubes in a Techne Hybridiser HB-1 oven.

Hybridisation of radiolabelled probes

The nylon filter was placed inside the hybridisation tube and 200 μl per cm^2 of filter of pre-hybridisation solution (5 xSSC, 5 x Denhardt's solution, 0.5 % SDS and 100 μg.ml^{-1} of denatured salmon sperm DNA) were added. All air bubbles were carefully removed before the tube was incubated with continuous rotation at 65 $^{\circ}$C in the oven for 2 hours. The denatured labelled probe was added to the tube and the incubation continued for at least another 12 hours at 65 $^{\circ}$C.

Hybridisation of fluorescein-labelled probes

The same method as for radiolabelling was employed except that the hybridisation buffer used was 5 xSSC, 0.1 % SDS, 5 % dextran sulphate and a 20-fold dilution of liquid block (supplied with kit). 0.125 ml/cm^2 of filter of buffer was used in the hybridisation.

2.14.5 Washing of probed Southern blots

Radiolabelled blots

During the washing process care was taken to ensure that the filters never became dry and all washes were carried out within the hybridisation tube. For high stringency washes, the filter was washed twice in 2 xSSC, 0.1 % SDS for 10 minutes at room temperature, then once in 0.1 xSSC, 0.1 % SDS for 15 minutes at 65 °C. Low stringency washes were carried out using 2 xSSC alone at 42 °C. After each wash, the filter was checked with a Geiger counter and the washing stopped when sufficient non-specific radiolabelled probe had been removed. Finally, the filters were wrapped in clingfilm.

Fluorescein-labelled blots

All stringency washes were carried out at 65 °C. The filter was firstly washed in 2-5 ml per cm^2 of membrane, pre-heated 1 xSSC, 0.1 % SDS for 15 minutes in the hybridisation tube and secondly, in 0.5 xSSC, 0.1 % SDS for a further 15 minutes. After washing, the blot was not allowed to dry out and detection was carried out immediately.

2.14.6 Autoradiography

Detection of radiolabelled blots

The filter was taped to some Whatman 3 mm paper and placed in a purpose-made cassette. Radioactive bands were detected on the filter by exposing Fuji RX-100 x-ray film to the filter. The x-ray film was sensitised by exposure to a low intensity flash of light and laid on the filter. Exposure was carried out at -80 °C for varying lengths of time

(generally overnight). After exposure, the position of the wells were marked on the film and then it was developed with Ilford Phenisol developer for upto 5 minutes and fixed with Kodak Unifix fixer for 2 minutes. The size of any hybridising fragments were calculated by referring back to the original photograph of the gel and a ruler.

Detection of fluorescein-labelled blots

The *Gene Images* CDP-*Ster* detection module (Amersham Life Science) was used to detect the fluorescein-labelled probes in Southern blots. Following hybridisation, the detection module allows detection of hybrids (after a blocking step) by incubation with an anti-fluorescein alkaline phosphatase (AP) conjugate.After washing off the excess conjugate, probe-bound AP is used to catalyze lightproduction by enzymic decomposition of a stabilized dioxetane substrate (kit protocol). The detection process was carried out as descibed in the manufacturer's instructions. Exposure of film to the blot varied between 1 hour and 3 days at room temperature. The film was developed as previously described for radiolabelled blots, and the size of any hybridising fragments estimated using the same method.

2.14.7 Stripping of probes from nylon filters

Radiolabelled probes

Radioactive probes needed to be stripped from nylon filters to allow for reprobing of the filters. The filter was washed with 0.4 M NaOH for 30 minutes at 45 °C, followed by a wash with 0.1 xSSC, 0.1 % SDS and 0.2 M Tris-HCl pH 7.5 for 30 minutes at 45 °C. It was important that the filter had never been allowed to completely dry out. The filter was checked that all the probe had been removed by reexposing x-ray film.

Fluoresceinlabelled probes

The filter was not allowed to dry during or after hybridisation and washing as with radiolabelled blots. The membrane was rinsed in 5 xSSC for 1-2 minutes, it was then added to a boiling solution of 0.1 % SDS, using approximately 5 ml of SDS solution per cm^2 of membrane. It was placed on a bench-top shaker for 10 minutes, before being

repeated twice with freshly boiling SDS each time. The filter was then ready for prehybridising, probing and detecting as before.

2.15 Mutagenesis

To prevent transcription of a functional wild-type gene product, a specific gene was replaced with an insert containing the DNA sequence of the gene and an added selectable marker. The vector used for this study was pJQuc1. It incorporates *sacB* from *Bacillus subtilis* , which is inducible by sucrose and is lethal when expressed in Gram-negative bacteria and also an antibiotic-resistant marker, namely, gentamicin resistance. In addition, the vector carries the *mob* region from the plasmid RP4 allowing for *A. tumefaciens* conjugation.[Quandt and Hynes, 1993].

The *A. tumefaciens* genes were subcloned into pJQuc1 and were selected by inactivation of β-galactosidase. A neomycin resistance cassette from pDUB2033 was introduced into the *A. tumefaciens* genes to disrupt them. Positive transformants were selected by resistance to both gentamicin and neomycin, and were then conjugated into *A.tumefaciens.*

The vector is only able to replicate in enterobacteria and hence functions as a "suicide" vector in *A. tumefaciens*. The vector is maintained in *A. tumefaciens* by homologous recombination. To achieve gene replacement, rare double recombination events leaving only the mutated gene have to be identified. Both the integrated and replacement recombinants carry resistance to neomycin and , thus, cannot be selected for directly. The pJQuc1 vector carries a conditional lethal gene which discriminates between the integration of the vector and the replacement recombinants. The lethal gene is the sucrose-inducible *sacB* of *B. subtilis* and its expression has been shown to be lethal in a wide range of Gram-negative bacteria in the presence of 5 % sucrose [Gay, 1985]. Thus *A. tumefaciens* cells which had been replaced with a mutated gene, were selected for on LM-agar plates containing 5 % sucrose and neomycin. Rifampicin was also added to the agar to ensure *Agrobacterium* colonies only were selected.

2.16 Conjugation

A culture of *A. tumefaciens* C58C1 was grown overnight to reach approximate mid-log phase. This culture would act as the recipient strain during tri-parental matings [Ditta, *et al.*, 1980]. The donor plasmid culture and an *E. coli* pRK2013 helper plasmid culture were also grown to approximate mid-log phase. A 0.22 μm nitrocellulose filter

was placed on the centre of an LM agar plate warmed to 28 °C. 100 μl of all three cultures were mixed and pipetted onto the filter. The upright plate was incubated overnight at 28 °C. Following incubation, the filter was placed into a universal containing 10 mls sterile 10 mM $MgSO_4$. Three different dilutions of cell suspension were plated out on LM agar/sucrose plates containing neomycin and rifampicin antibiotic selection. The plates were incubated at 28 °C for 2-3 days.

2.17 Confirmation of gene replacement

2.17.1 The 3-keto-lactose assay for Agrobacterium

The putative *Agrobacterium* colonies were streaked onto plates containing 2 % glucose, 2 % $CaCO_3$, 1 % yeast extract and 1 % agar. They were incubated at 28 °C overnight and then a loopful of the bacteria was steaked onto plates containing 1 % lactose, 0.1 % yeast and 1 % agar. They were incubated at 28 °C for 2 days. After growth of the bacteria, the plates were flooded with Benedict's reagent (173 g sodium citrate, 100g sodium carbonate and 17.3 g copper sulphate per litre) and incubated at room temperature for approximately 2 hours. A yellow zone around colonies was indicative of lactose being converted into 3-keto-lactose and, hence, presumably *A. tumefaciens*. [Bernaerts and De Ley, 1963].

2.17.2 Swarm plates

To detemine if a gene affecting motility in *A. tumefaciens* had been replaced, a loopful of bacteria was inoculated into the centre of a swarm plate containing 25 ml of LM broth and 0.16 % agar. The swarm plates were incubated upright for 2-3 days at 28°C. The size of the swarm of the putative gene replacement bacteria was compared with that of the wildtype.

2.17.3 Southern blotting

Finally, to confirm that gene replacement had definately taken place, the chromosomal DNA was isolated and digested with restriction enzymes. An enzyme was chosen which would produce different size bands in the wild-type genomic DNA. Southern blots were then carried out using radiolabelled probes to detect size differences between fragments and hence confirm gene replacement (see 2.14.1).

2.18 Microscopy

2.18.1 Light microscopy

A loopful of bacteria was resuspended in chemotaxis media (0.1 mM EDTA pH 7.0 with 1 M KOH) on a glass microscope slide and observed using a Nikon Optiphot microscope under phase contrast optics.

2.18.2 Electron microscopy

A loopful of bacteria was resuspended in chemotaxis media and a 50 μl aliquot was transferred to a Formvar-coated grid. The grids were left for 30 minutes and then the excess media was drained off using blotting paper, the grids never being allowed to dry completely. A 50 μl drop of 1 % uranyl acetate was added to the grid, left for 30 seconds and then drained off using blotting paper as before. Finally, the grids were air-dried. A Philips EM400 was used to view and photograph the grids.

Chapter 3

DNA sequencing of the cosmid clone pDUB1900.

3.1 DNA Sequenced

Approximately 4.5 kb of DNA from the largest *Hin*dIII fragment of pDUB1900 was sequenced. This sequence incorporated the whole of the *Bam*HI fragment D and part of the *Eco*RI fragment D and is shown in figure 3.1.1.

The DNA fragments to be sequenced were usually subcloned into pUC19 or pBluescript KS+/SK+ vectors and sequenced using the universal M13 forward and reverse primers. Occasionally, custom-synthesised oligonucleotides were employed to finish the sequence of certain areas.

Primer α

5' CGA ATC GGC AAT CGG ATC G 3'

Primer β

5' GCC AGC GTT GAC GGT CGC 3'

Primer γ

5' GCC AAG AAG GCG TGG ACG 3'

Primer δ

5' GCC TTT GAG AGC ATT TCC TCG 3'

The positions of these four primers are marked on figure 3.1.2.

Figure 3.1.2 shows the individual sequencing reactions used to sequence the region. The entire region was sequenced fully in both directions with overlapping DNA fragments. The generated DNA sequences were analysed and aligned using DNA Strider and DNAid (see methods). Certain predicted restriction enzyme sites were confirmed by digesting the DNA with enzymes and comparing band sizes produced with those expected. Figure 3.1.3 gives the entire sequence of 4470 nucleotides in the sense strand.

Figure 3.1.1 The region of pDUB1900 that was sequenced

The boxes show the *Bam*HI, *Hin*dIII and *Eco*RI sites within a section of the chromosomal insert of pDUB1900. The top row of boxes show the *Bam*HI sites, the middle row the *Hin*dIII sites and the bottom row the *Eco*RI sites. The labelled line at the bottom shows that all of *Bam*HI fragment D and part of *Eco*RI fragment D was sequenced. The labelled arrows along the top show the position and order of previously identified genes.

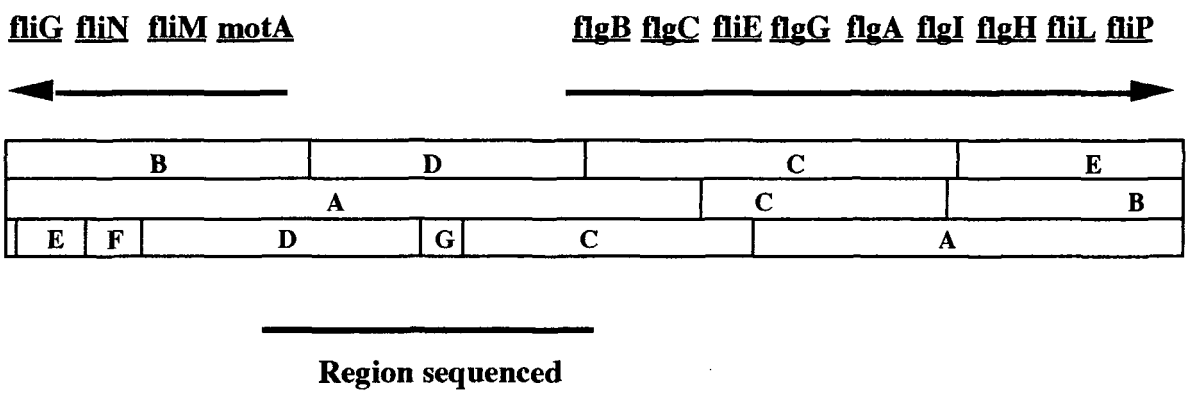


Figure 3.1.2 Individual sequencing reactions.

The box shows the region of pDUB1900 that was sequenced in this study. The *Bam*HI (B), *Eco*RI (E), *Pst*I (P) and *Stu*I (S) sites are marked on it. The individual sequencing reactions are shown by arrows below the box. The custom-synthesised oligonucleotides are also shown, they are labelled as described in the text.

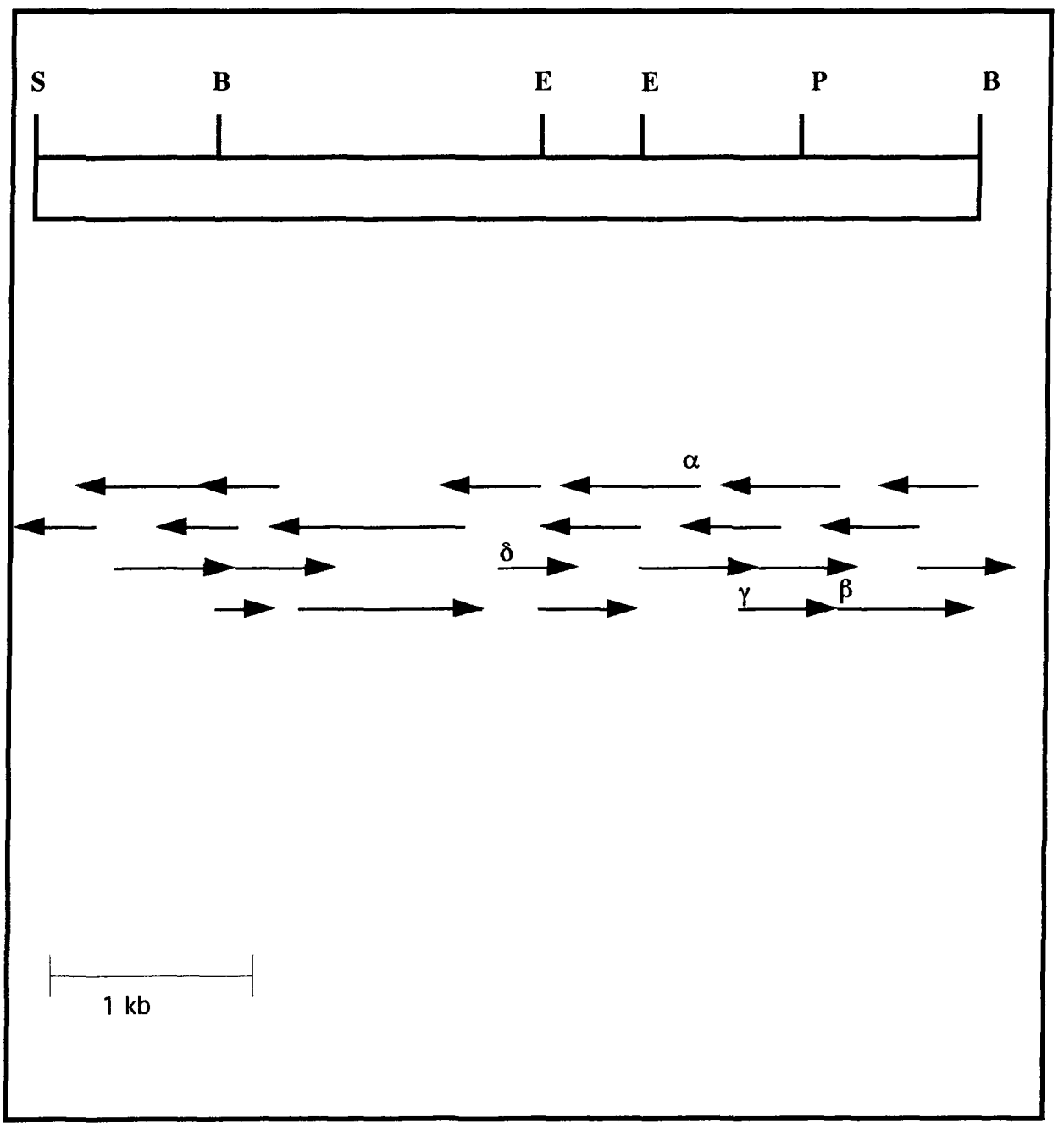


Figure 3.1.3 The entire sense strand of DNA sequenced in pDUB1900

This is the entire sense strand of DNA sequenced in this study. The positions of the genes are given later in figures 3.2.2, 3.3.2, 3.4.2 and 3.5.2.

The GenBank accession number of this sequence is U95165 dated 1st October 1997.

```
          10         20         30         40         50         60         70         80

   1 AGGCCTCGCC GAGCGCCTTG CCCGAATCCT TCACGACCTT CATCGGGTTC GCCATGATGA AGCCGCCGAG ACCGGCGCCT 80
  81 CCGATGATCA TCAATTCGAA CGGCTGAACC AGCACGTTCA AATGACCGCC CATCGCCATA TAGCCGCCGA TGATGCAGCC 160
 161 GAAGGTGATT ATAAGTCCAA TTACAATATT CATTGATCGA CCACGCCCGA ACAGTGTTCC ATGGCGCCAT GTCTAGAAAA 240
 241 TGTGGCTTGC GTGAACCTGT CCGCAATGCC TCCTGCCCCG CCCTCACCTC AGCCAGTTTC GCACAAGCAA GCCGCGACAT 320
 321 TATGCCGACA AGGGCGGACG TGCGCCTTGA AATCGGGATC AAACGAGAGC GGCGGTTCAG TGACTTCCAC CTATACCAGC 400
 401 TACAGACTGA TCAGTCAGGA TATCGGCAAA TCGCTCGAGC GGGTCTCGAA GCAGCCAGAC GTGGCACGCG AGACCGAATA 480
 481 TTACCGTGAA AAGATCGGCA GCGTGAAATC CATCGACGAT TTCATGGCCG ATACACGTCT TTACAATTAC GCGCTGAAAG 560
 561 CCCACGGCCT GGAAGACATG GCCTATGCAA AAGCCTTCAT CCGCAAGGTG TTGACGGAAG GGGCGAGCGA TAAGAACGCC 640
 641 TTCGCCAACA AGCTCTCCGA CAATCGTTAT GCCGAGCTTG CAAAGTCGCT GGATTTCGCG GGCCTGGGTG CGGCGGCAAC 720
 721 CGCGACCGAG GCGGCCAAAT CAGGTGTCAT CGGCAACTAC GCCCGCCAGA CGCTGGAACA GGAAGCGGGC GACGACAATA 800
 801 ACGGCGTCCG GCTCGCCCTT TATTTCGAGC GCAAGGCACC GACAATCAAA TCCGGCCTCG ATTTTCTGGC CGACGACGCG 880
 881 CTTGCGCAGG TCTTCCGCAC AACGTTCAAC CTGCCTGACG CATTTGCTGC TGCTGATGTC GACAAGCAGG CCGCCCTCAT 960
 961 CGAAAAGAGC ATCAATATCA AGGATCTGCA GGATCCTGAA AAGGTCGGCA AGTTGCTCGA GCGTTTCACC ATCATGTGGG 1040
1041 AAATGCAAAA CCCCTCGACG ACCTACGATC CCCTGGCCGT TTTCGGCTCC TCCAGCGGCT ACGGCATTTC CCCCGACCTG 1120
1121 CTGATTTCCA TCAACTCCCT GAAACTCGGA GGCAAATGAT ATGCAATCCG GACTATATGT CGCCCTGTCC TCGCAGATCG 1200
1201 CGCTGGAGCG CCGTCTCACC ACCATTTCTG ACAATATGGC GAATGTGAAC ACGGTCGGCT TCCGCGGCTC CGAGGTAAAA 1280
1281 TTCGACGAAA TGGTCGCCAA GAACCACAAT GACATGAATG CCAAGGTGGC TTTCGTGTCG CAGGGCAACG ACTATCTCTC 1360
1361 CACACGCCAG GGCGCTTTCG AGCAGACCGG CAATTCCTTC GATTTCGCCA TCAAGGGCGA CGCCTGGTTT GCGCTGGATA 1440
1441 CGCCGGATGG CCAGATCCTG ACGCGCGACG GCCGTTTCAC CATGCGCCCG GACGGCGCAT TGATTTCGTC CAGCGGTTAT 1520
1521 CCCGTCCTCG ACGCGGGCGG CGGCCCGATC AGGCTCAATC CGAATGGCGG CCCGATCACC GTCGGCCTCG ACGGCGCCAT 1600
1601 CAGGCAGAAC GACACCATCG CCGCCTCGCT CGGCATCTTC CAGGCGGATT TTTCAAAGGG CTTCCTGCGT CACCCCAATA 1680
1681 GCGGCGTGAA GCCCGTGGCC CAGCCCGTTC CGGTCGTCAA CAATCATGAG GTCGGCGTCG TTCAGGGTTA TCTCGAGCAG 1760
1761 TCGAACGTCA ACGGCATTTC CCAGATGACG CAGCTCATCC AGGTCAACCG CGCCTTTGAG AGCATTTCCT CGATGATGCG 1840
1841 CGATACCGAA TCGACCTTCG GCGAGGGCAT CAAGACGCTC GGCGGCGCGC GTTGAGGAAC TGACGGATGA CAATGCCGGA 1920
1921 ATCCATGCTC TCGGAATCCA AGCTCTCGGG CTGGGCCATT TCGCCGAAGC TCGCCCAGCT CGCCAGCCTT GCCGGGCACT 2000
2001 ATGCCGATCC GGAGTTCTCG GTGGCGCATG GCGGCCATGT CCGCACCATC GCCGCCGGGC ACTATACGGT GTCCGGCCTG 2080
2081 TCGCGGCATG TGCGGCTCGG CGAATTCGTT GCCCATCGCA GCGCAACCGG CATTCATCTC GGCGAAGTCG TGCGCGTGGA 2160
2161 ACCGGATATC TGCTATGTCT GCCCTATCGA GCCCGGCGAG CCGATCGGCA TCCACGACAC CGTCATCCGC AAGGGTGCCT 2240
2241 TCCGCGTCTC GCCCGACGAG AGCTGGTGCG GGCGCACCAT CAATGCGCTT GGCGAGCCGA TCGACGGCCA GGGGCCGCTT 2320
2321 GCTTCCGGCA TCGTGCGCCG TTCGATTTCC AACAATGCGC CGCCCTCGAT GACCCGTAAA AGGGTGGAGA CACCGTTCAA 2400
2401 GACGGGCGTG CGGGCTATCG ATATTTTCTC ACCGCTCTGC CTTGGGCAGC GTCTCGGTAT TTTCGCCGGT TCGGGCGTGG 2480
2481 GCAAATCCAC GCTGCTGTCG ATGCTCGCCA AAGCCGATGC CTTCGACAAG GTGGTGATTG CGCTTGTCGG CGAACGTGGC 2560
2561 CGCGAAGTGC GCGAATTCAT CGAAGACACG ATGGGCGATA ATATGAGCAA GTCCGTCGCC GTCGTCGCGA CGAGCGACGA 2640
2641 GAGCCCGATG CTGCGCAAGA TGGCCCCGCT TTCCGCCGTC ACCATTGCCG AGCATTTCCG CGATCAGGGC GACAATGTCC 2720
2721 TTCTCATCAT CGACAGCGTG ACGCGATTTG CCCATGCGAT CCGCGAAGTG GCGGTCGCTT CGGGCGAACC GCCCGTCGCA 2800
2801 CGCGGTTATC CCGCCTCGGT CTTCACCGAA CTGCCGCGGC TGCTGGAACG GGCAGGACCG GGCGCGGAAG GCACCGGCAC 2880
2881 CATCACCGCC ATCGTCTCCA TCCTCGTGGA TGGTGACAAT CACAACGATC CGATTGCCGA TTCGACCCGT GGTATTCTCG 2960
2961 ACGGCCATAT CGTGCTGGAT CGCAGCCTTG CCGAAGAGGG CCGCTATCCG CCGATCAATC CGCTCGCCTC GATCTCGCGT 3040
3041 CTTGCCAAGA AGGCGTGGAC GCCGGATCAG GAAAAGCTGG TCTCGCGCCT GAAGGCGCTG GTGCATCGTT TCGAGGAAAC 3120
3121 GCGGGATCTT CGCCTCATCG GCGGTTACCG GCCCGGAACC GACCCCGATC TCGACATGGC GGTGAAGCAG GTTCCGATCA 3200
3201 TCTATGAAAC GCTGAAACAG CTACCGGACG AACCGGCGGC ACAGGATGCC TATGCCGATC TGGCGACCGC ACTACGCGGT 3280
3281 GGAGCACAGA ATGGCCAGCC CCAGGTCAAT CCGAGAATGA GAGGCTGACC CCGTGAACAA CAAGCGGAAC GACAATAGCC 3360
3361 ACGATGAAGA CGTGCTGTCC AAGCCGATCT TCTTTACGCC CGACCGCATT CTCGCATGGA CGGGCATCGC GCTTGCCGCT 3440
3441 GCAGCAGCCT TTTTCCCCTG GTACGTCTTC TTCAACGAGG AGAAATTCGG CATGAAGGTG GCGGGCGGCG ATCGCACCCG 3520
3521 CGATCTGCCT CATACCGGAC CGCGGGAAGT CTTCAGTGTC TCGCCCATGG CCATGACCAA CCGTAACAAG GAAGACACCC 3600
3601 CGCCAGCGGC GGAGCTGCCC GACATGCTGA CGACGGCGAC CGTCAACGCT GGCGGCAAGG AGAGACAGAA CAATGTGGCA 3680
3681 GCCGGGATCG AAGACCAGCC CTTCCCCGGC CAGACATCGT TCCGTCTGCT ACATGTCTCC AATGGCCGCG CCCTGATCGA 3760
3761 AGACAGTTCG GGCATGTATA TGGTGCGGGT CGGTTCCACA TTGCCTGACA ATGCCTTGCT GACGAAGATC GAGCAGCGCA 3840
3841 ATGGCGAATG GCTGATCGAG ACATCCTCCG GAAAAACCTA TCATCCGGAA AAATGAGAGA ATTTCTCCCC GCTAAAAAGG 3920
3921 CGCGCCACAC ATCCGGCGCG CTTTTTTGTT TGTGAGGGAC ACGCAAGCAA TGCGTCTGTT TCCATTGTTC CACCCCTTCA 4000
4001 TCCTTCTCCG CAAGTCTGAC GCAAGATTAA CCCCGTAGGT TCGGGCCATA CGAGATGGAG AAGTCTATGC AACCGATTCA 4080
4081 ACTGTTCGAA CTGGCATCCC GCCAAGCGGA ATGGCTGAGC GTGCGGCAGG AAGTCGTGGC GACCAACATC GCCAACGCCA 4160
4161 ACACACCCAA GTTCCACGCC AAAGATGTCA GTCCCTTCGA AGCGGTCATG CAGGCCACCA GCCAGCAGGT CGGCATGGCC 4240
4241 AGGACAAACC CTGTCCACCT CGGTGAAAGC ACACTCAGCG AGAATATCGC GGTGCGGGAC AACCCGGTCA ACAACGAGAT 4320
4321 CGGCATGCAG GAATCGGGCA ATTCGGTCGG GCTTGCGGAA GAAATGACCA AGACCGGCGA GATCAAGCGG CAGTACGACC 4400
4401 TGAACGCCAG TCTCGTCAAA TCCTTCCATC GCATGATGCT GATGACGGTT AAGAGGTAAA CCATGGATCC           4470
          10         20         30         40         50         60         70         80
```

The DNA sequence produced was compared against nucleic acid and protein databases to identify homologous regions. As described in the methods section, the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package was used to manipulate sections of the DNA sequence and compare them against the GenBank and EMBL databases using the FASTA program. In addition, DNA Strider was used to identify open reading frames (ORFs) in the DNA sequence. These ORFs were translated into their protein sequence and were compared against the OWL database using the SEQNET programme SWEEP.

The entire sequenced region was analysed using the UWGCG program TESTCODE for potential protein coding sequences. The results of this testcode are shown in figure 3.1.4.

As a result of using TESTCODE and from the database searches, four open reading frames were identified, flanked by partial sequence from *motA* upstream [Deakin, *et al.*, 1997b] and *flgB* downstream [Deakin, *et al.*, 1997a]. For the purpose of this study the four open reading frames only, will be analysed. The positions and sizes of these open reading frames are given in figure 3.1.5. Two of the open reading frames identified were found to contain sequence identical to previously identified genes/proteins. For this study, these two open reading frames have been denoted *fliI* and *flgF* and will be discussed in further detail later. The two remaining open reading frames were found to contain no homologies to any current genes/proteins identified in the databases. For the purpose of this study they have been denoted *orfX* and *orfZ* and will also be discussed in greater detail.

Figure 3.1.6 shows the position and order of this putative operon in relation to those already identified.

Further analysis of all open reading frames will be discussed later in this chapter and also in chapters 4 and 5. In addition, reasons will be given to support the view that these ORFs are not thought to be transcribed as part of the putative operon slightly upstream and downstream of them, but infact form a smaller operon of their own. Finally, the possible presence and locality of these ORFs in *R. meliloti* will be discussed.

Figure 3.1.4 A graphical representation showing the results of the TESTCODE algorithm on the sequenced DNA of pDUB1900.

The y-axis of the graph plots the TESTCODE algorithm scores. The x-axis of the graph gives the nucleotide scale. Scores greater than 9.5 shown above the top line on the graph, indicate regions of the DNA that TESTCODE has a 95 % confidence are coding regions. Scores below 7.4 shown below the bottom line of the graph are regions of the DNA that TESTCODE is 95 % confident of being non-coding. The 'window of vulnerability' lies between these two values in which the algorithm is unable to make any significant predictions.

The position of the four open reading frames are labelled below the graph along with *motA* and *flgB*.

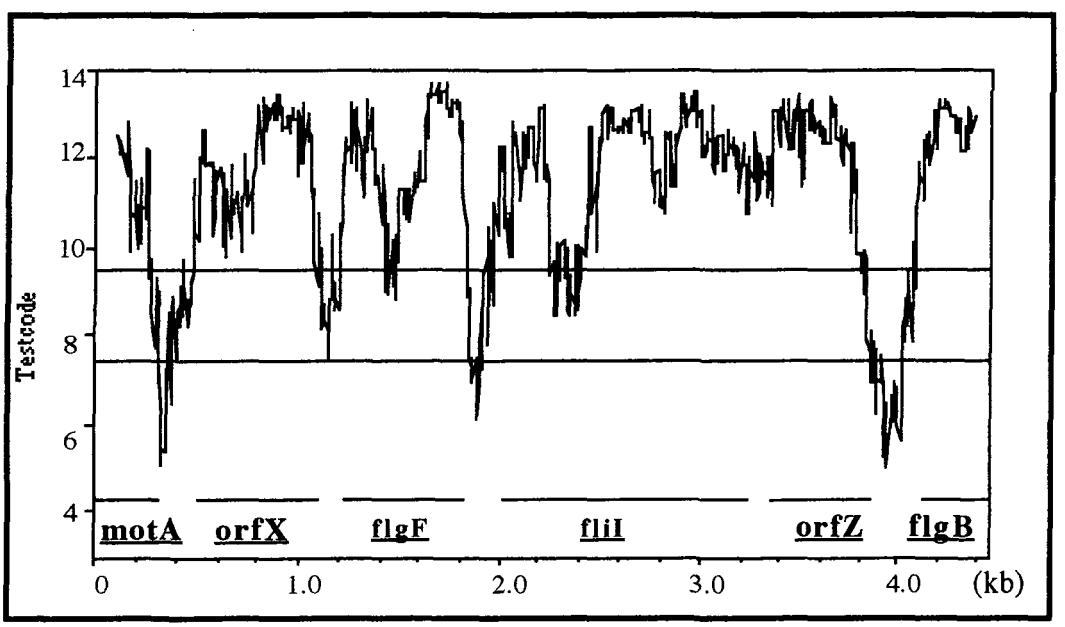


Figure 3.1.5 A diagram showing the open reading frames identified within the sequenced region using DNA Strider.

The DNA Strider program analysed all six phases of the DNA sequence producing the diagram of the open reading frames. The six phases are numbered at the side of the diagram from 3 to -3. Possible start codons are represented by the smaller bars and stop codons by the full bars. The numbers across the top and bottom of the diagram show the nucleotide scale.

The open reading frames with homology to previously identified genes are labelled *flgF* and *fliI* . The two unidentified open reading frames (*orfX* and *orfZ*) are also shown.

All of the open reading frames are believed to have an ATG start codon with the exception of *orfZ* which is believed to start with GTG, reasons for this assumption are discussed in the text.

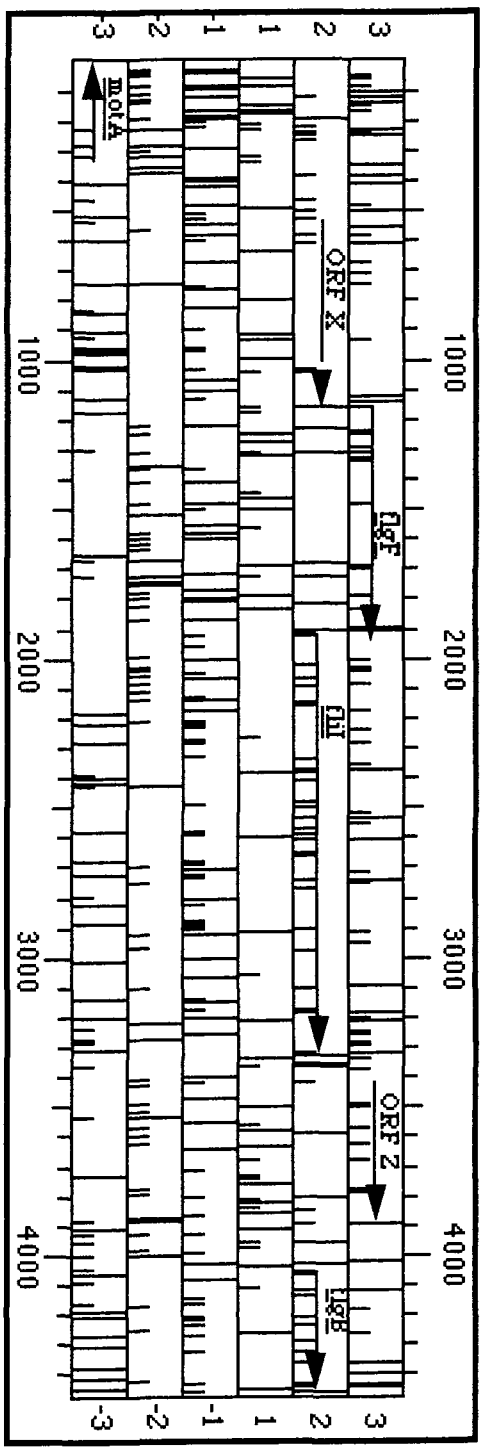
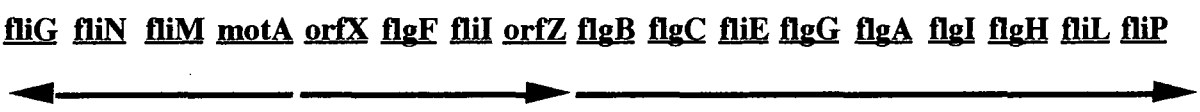

motA
ORF X
flgF
flil
ORF Z
flgB
1000
2000
3000
4000
-3
-2
-1
1
2
3

Figure 3.1.6 Position and order of the putative operon

The diagram shows the position and order of the four open reading frames, *orfX*, *flgF*, *fliI* and *orfZ* in relation to the flanking operons. The known gene order of the flanking operons are labelled above the arrows.



Further analysis of the sequenced flagellar homologues and unidentified open reading frames.

The flagellar homologues, *flgF* and *fliI*, and the unidentified open reading frames, *orfX* and *orfZ*, within the putative operon identified earlier (3.1) are here studied in more depth.

3.2 The *flgF* homologue

This is a 735 base pair open reading frame starting at nucleotide position 1161 and ending at position 1895. The gene product, when translated, is 244 amino acids in length with a predicted molecular weight of 26,119 Da. When the OWL database was searched with this gene product it was found to have sequence similarity to the FlgF protein from a number of bacteria; 56.769 % identity in a 229 amino acid overlap to the FlgF protein of *R. meliloti* and 34.632 % identity in a 231 amino acid overlap to the FlgF protein of *C. crescentus*. In addition, it also showed a degree of similarity to the FlgG protein in several bacteria, for example, 28.405 % identity in a 257 amino acid overlap to the FlgG protein of *E. coli*. FlgF and FlgG are both flagellar basal-body rod proteins and are structurally related. A multiple alignment of the homologues was carried out using the SEQNET program CLUSTALV [Higgins and Sharp, 1988], the results of which are shown in figure 3.2.1. As can be seen the sequence similarity is throughout their entire lengths.

Figure 3.2.2 shows the DNA sequence of the *flgF* homologue along with the predicted gene product.

FlgF is a proximal rod protein belonging to the family of axial proteins which are located extracellularly or within the cell envelope. In *B. subtilis* and *S. typhimurium* the FlgF protein has a consensus sequence ANNLAN. Whether this sequence is a signal involved in the export of the proteins or, rather, represents a structural constraint of a component of the axial filament remains to be demonstrated. The conservation of such a sequence in gram-positive and -negative bacteria that have diverse cell envelope organisation may argue in favour of the structural constraint [Albertini, *et al.*, 1991]. Many other bacteria have been shown to have this consensus sequence or a similar one, the *A. tumefaciens* FlgF homologue has the sequence SDNMAN at amino acid positions 23-28, this shows a high degree of conservation to the consensus sequence with three identities and three conservative substitutions.

Figure 3.2.1 Multiple alignment of the *A. tumefaciens* FlgF homologue to those of *R. meliloti* and *C. crescentus* and to the FlgG protein of *E. coli*.

```
CF              MDNALYVGLSRQMTVRRELDIVANNIANANTTGFKVEDLMVRTEQAKPAKTLDGSSPVKF
EG              MISSLWIAKTGLDAQQTNMDVIANNLANVSTNGFKRQRAVFEDLLYQTIRQPGAQSSEQT
RF              MQTGLYVALSSQMALEKRLNTLADNIANSNTVGFRATEVKFNQVLG---DTKPTK--VSY
FLGF            MQSGLYVALSSQIALERRLTTISDNMANVNTVGFRGSEVKFDEMVAKNHNDMNAK--VAF
                *  .* .. .   .    . ...*.** .* **.

CF              VMDTGVRRNF----------TQGPMTKTGGDYDLAINGMGFFKVQANGGER-YTRDGRF
EG              TLPSGLQIGTGVRPVATERLHSQGNLSQTNNSKDVAIKGQGFFQVMLPDGSSAYTRDGSF
RF              VSEGEEFLST----------KTGAL-RTGSALDFAIKGDAWFSIDTPGGPA-LTRDGRF
FLGF            VSQGNDYLST----------RQGAFEQTGNSFDFAIKGDAWFALDTPDGQI-LTRDGRF
                                    *   .*    * **.* . * .   .*    **** *

CF              TTNPEGILVTQAGAPVLDDGGGQITIDPRLGPVTVGKDGIVS------QGAIRVGRIGLV
EG              QVDQNGQLVTAGGFQV----QPAITIPANALSITIGRDGVVSVTQQGQAAPVQVGQLNLT
RF              TLTETGELVTIKGYPVLDAGGPPIQLNGGAGEIAVGADGAIH------QNGVQTCLLGLY
FLGF            TMRPDGALISSSGYPVLDAGGGPIRLNPNGGPITVGLDGAIR------QNDTIAASLGIF
                     * *..  *  *       * .      ...* ** .              . .

CF              RPDDLSTFAKDGDNLYRNT----TNTAPQPVT-DAQIH-QGMLEASNVQPVIEITKLIEI
EG              TFMNDTGLESIGENLYTETQSSGAPNESTPGLNGAGLLYQGYVETSNVNVAEELVNMIQV
RF              EADFSKGFMRYDNSSVMPA----AQPNRR-DRFDVGVM-QGFLEESNVNGIQEMSQLIMI
FLGF            QADFSKGFLRHPNSGVKPV----AQPVPVVNNHEVGVV-QGYLEQSNVNGISQMTQLIQV
                       .    ..         .            . ** .* ***.   .. ..* .

CF              QRAYESVAKMMDNTAELSRTPSSVWARSTRER
EG              QRAYEINSKAVSTTDQMLQKLTQL--------
RF              TRAFDNVTALMRATVRDRSIRPSRRSAAG-P-
FLGF            NRAFESISSMMRDT--ESTFGEGIKTLGG-AR
                 **..  .  .  *
```

CF= *C. crescentus* EG= *E. coli* RF= *R. meliloti* FLGF= *A. tumefaciens*

The astericks indicate identities, and the conservative substitutions are denoted by dots.

Figure 3.2.2 The DNA sequence of the *flgF* homologue and the predicted gene product.

The *flgF* coding region begins at nucleotide position 1161 and finishes at position 1895. The start and stop codons are shown in bold. The predicted protein sequence is shown beneath the DNA sequence.

The stop codon of upstream *orfX* is overlined and labelled X (nucleotide positions 1157-1159). The start codon of downstream *fliI* is overlined and labelled I (nucleotide positions 1907-1909).

A possible ribosome binding site GGA is shown in bold 10 base pairs upstream of the start codon.

```
881 cttgcgcaggtcttccgcacaacgttcaacctgcctgacgcatttgctgctgctgatgtcgacaagcaggccgccctcat 960

961 cgaaaagagcatcaatatcaaggatctgcaggatcctgaaaaggtcggcaagttgctcgagcgtttcaccatcatgtggg 1040

1041 aaatgcaaaaccccctcgacgacctacgatcccctggccgttttcggctcctccagcggctacggcatttcccccgacctg 1120
                                    rbs      X
1121 ctgatttccatcaactccctgaaactcggaggcaaatgat ATG CAA TCC GGA CTA TAT GTC GCC CTG TCC 1190
   1                                          M  Q  S  G  L  Y  V  A  L  S  10

1191 TCG CAG ATC GCG CTG GAG CGC CGT CTC ACC ACC ATT TCT GAC AAT ATG GCG AAT GTG AAC 1250
  11 S  Q  I  A  L  E  R  R  L  T  T  I  S  D  N  M  A  N  V  N  30

1251 ACG GTC GGC TTC CGC GGC TCC GAG GTA AAA TTC GAC GAA ATG GTC GCC AAG AAC CAC AAT 1310
  31 T  V  G  F  R  G  S  E  V  K  F  D  E  M  V  A  K  N  H  N  50

1311 GAC ATG AAT GCC AAG GTG GCT TTC GTG TCG CAG GGC AAC GAC TAT CTC TCC ACA CGC CAG 1370
  51 D  M  N  A  K  V  A  F  V  S  Q  G  N  D  Y  L  S  T  R  Q  70

1371 GGC GCT TTC GAG CAG ACC GGC AAT TCC TTC GAT TTC GCC ATC AAG GGC GAC GCC TGG TTT 1430
  71 G  A  F  E  Q  T  G  N  S  F  D  F  A  I  K  G  D  A  W  F  90

1431 GCG CTG GAT ACG CCG GAT GGC CAG ATC CTG ACG CGC GAC GGC CGT TTC ACC ATG CGC CCG 1490
  91 A  L  D  T  P  D  G  Q  I  L  T  R  D  G  R  F  T  M  R  P  110

1491 GAC GGC GCA TTG ATT TCG TCC AGC GGT TAT CCC GTC CTC GAC GCG GGC GGC GGC CCG ATC 1550
 111 D  G  A  L  I  S  S  S  G  Y  P  V  L  D  A  G  G  G  P  I  130

1551 AGG CTC AAT CCG AAT GGC GGC CCG ATC ACC GTC GGC CTC GAC GGC GCC ATC AGG CAG AAC 1610
 131 R  L  N  P  N  G  G  P  I  T  V  G  L  D  G  A  I  R  Q  N  150

1611 GAC ACC ATC GCC GCC TCG CTC GGC ATC TTC CAG GCG GAT TTT TCA AAG GGC TTC CTG CGT 1670
 151 D  T  I  A  A  S  L  G  I  F  Q  A  D  F  S  K  G  F  L  R  170

1671 CAC CCC AAT AGC GGC GTG AAG CCC GTG GCC CAG CCC GTT CCG GTC GTC AAC AAT CAT GAG 1730
 171 H  P  N  S  G  V  K  P  V  A  Q  P  V  P  V  V  N  N  H  E  190

1731 GTC GGC GTC GTT CAG GGT TAT CTC GAG CAG TCG AAC GTC AAC GGC ATT TCC CAG ATG ACG 1790
 191 V  G  V  V  Q  G  Y  L  E  Q  S  N  V  N  G  I  S  Q  M  T  210

1791 CAG CTC ATC CAG GTC AAC CGC GCC TTT GAG AGC ATT TCC TCG ATG ATG CGC GAT ACC GAA 1850
 211 Q  L  I  Q  V  N  R  A  F  E  S  I  S  S  M  M  R  D  T  E  230
                                                              I

1851 TCG ACC TTC GGC GAG GGC ATC AAG ACG CTC GGC GGC GCG CGT TGA ggaactgacggatgacaatg 1915
 231 S  T  F  G  E  G  I  K  T  L  G  G  A  R  *                                  245

1916 ccggaatccatgctctcggaatccaagctctcgggctgggccatttcgccgaagctcgcccagctcgccagccttgccgg 1995

1996 gcactatgccgatccggagttctcggtggcgcatggcggccatgtccgcaccatcgccgccgggcactatacggtgtccg 2075

2076 gcctgtcgcggcatgtgcggctcggcgaattcgttgcccatcgcagcgcaaccggcattcatctcggcgaagtcgtgcgc 2155

2156 gtggaaccggatatctgctatgtctgccctatcgagcccggcgagccgatcggcatccacgacaccgtcatccgcaaggg 2235
```

3.3 The *fliI* homologue

The *fliI* coding region begins at nucleotide position 1907 and finishes at position 3328. The gene product encoded by this open reading frame has 473 amino acids and a predicted M_r of 50,775 Da. It was found to have sequence similarity to the FliI protein of several other bacteria; 81.96 % identity over a 449 amino acid overlap to FliI of *R. meliloti*, 48.404 % identity over a 376 amino acid overlap to FliI of *C. crescentus* and 47.2 % identity over a 318 amino acid overlap to FliI of *S. typhimurium*. In addition, there was significant similarity to the HrpB6 protein from *Xanthomonas campestris*, an ATP synthase known to be 'related' to FliI.

The extent of the similarity between the gene products is shown in a multiple alignment, produced using the SEQNET program CLUSTALV [Higgins and Sharp, 1988] in figure 3.3.1. The alignment shows the similarity between the proteins to be throughout their entire lengths with the greatest divergence between them being in the initial 100 amino acids.

Figure 3.3.2 shows the DNA sequence of the *fliI* homologue and the predicted gene product.

As mentioned in Chapter One, flagellin sub-units are transported through the central channel of the growing flagellar structure. In *E. coli* and *S. typhimurium* the product of the *fliI* gene is required for this to occur. The exact function of FliI is not known, however, its resemblance to a component of ATP-driven proton translocases suggests that it is an ATPase that may energise the export of flagellar sub-units across the cytoplasmic membrane. The *A. tumefaciens* FliI homologue contains the very highly conserved A and B motifs, described by Walker, *et al.* [1982], which in $F_1\beta$ are believed to make major contributions to the nucleotide-binding site of the β-subunit [Dreyfus, *et al.*, 1993]. It seems likely that in *A. tumefaciens*, as in other bacteria, FliI binds and hydrolyses ATP during export of the flagellar sub-units.

Figure 3.3.1 Multiple alignment of the *A. tumefaciens* FliI homologue to those of *R. meliloti*, *C. crescentus* and *S. typhimurium* and to HrpB6 from *X. campestris*.

```
CI              MRS----LIAAVERIDP----LTI---YGRVAAVNGLLIEVRGGLTRLAVGARVEIERFG
RI              FLIIMTMPESMLS--ESKLSGWAISPKLAQLASLAGHYADPEFSVAHGGHVRTIAAGHYT
STI             MTTRLTRWLTALDNFEAKMALLPAVRRYGRLTRATGLVLEATG--LQLPLGATCIIERQD
XI              MLAETPLLETTLER---ELATLAVGRRYGKVVEVVGTMLKVAG--VQVSLGEVCELRQRD
FLII            ----MTMPESMLS--ESKLSGWAISPKLAQLASLAGHYADPEFSVAHGGHVRTIAAGHYT
                         . .                ...    *          .          .

CI              Q---KPLPAEVVGFRETRALL-MPFGPVEGVGPGAEI--RIVPEGAVV------RPTK--
RI              VSGLSR-HVRLGEFVAHRSATGIHLGEVVRVEPDICYVCPIEPGEPIGIHDTVIRKGAFR
STI             GPETKEVESEVVGFNGQRLFL-MPLEEVEGILPGARVYARNGHGDGLQ------SGKQLP
XI              GTLLQR--AEVVGFSRDLALL-APFGELIGL-------SRETRVIGLG------RPLAVP
FLII            VSGLSR-HVRLGEFVAHRSATGIHLGEVVRVEPDICYVCPIEPGEPIGIHDTVIRKGAFR
                         . *          . . .                .

CI              -----AWLGRIINAFGEPIDGLGPLPQGEVPYPLKTAPPPAHARGRVGERLDLGVRSMNV
RI              VSPDESWCGRTINALGEPIDGQGPLASGIVRRSISNNAPPSMTRKRVETPFKTGVRAIDI
STI             LGP--ALLGRVLDGGGKPLDGLPA-PDTLETGALITPPFNPLQRTPIEHVLDTGVRAINA
XI              VGP--ALLGRVLDGLGEPSDGQGA-IACDTWVPIQAQAPDPMRRRLIEHPMPTGVRIVDG
FLII            VSPDESWCGRTINALGEPIDGQGPLASGIVRRSISNNAPPSMTRKRVETPFKTGVRAIDI
                     .  ** ... * * **  .         ..   .  .  *  .      *** ..

CI              FTTTCRGQRLGIFAGSGVGKSVLLSMLAKEATCDAVVVGLIGERGREVREFVEETLGEEG
RI              FSPLCLGQRLGIFAGSGVGKSTLLSMLAKADAFDKVVIALVGERGREVREFIEDTMG-DN
STI             LLTVGRGQRMGLFAGSGVGKSVLLGMMARYTRADVIVVGLIGERGREVKDFIENILGPDG
XI              LMTLGEGQRMGIFAAAGVGKSTLMGMFARGTQCDVNVIVLIGERGREVREFIELILGADG
FLII            FSPLCLGQRLGIFAGSGVGKSTLLSMLAKADAFDKVVIALVGERGREVREFIEDTMG-DN
                .     ***.*.**..***** *..* *.    *  *. *.*******..*.*  .* .

CI              LRRAVVVVATSDEPALTRRQAAYMTLAISEFMRDQDQEVLCLMDSVTRFAMAQREIGLAA
RI              MSKSVAVVATSDESPMLRKMAPLSAVTIAEHFRDQGDNVLLIIDSVTRFAHAIREVAVAS
STI             RARSVVIAAPADVSPLLRMQGAAYATRIAEDFRDRGQHVLLIMDSLTRYAMAQREIALAI
XI              LARSVVVCATSDRSSIERAKAAYVGTAIAEYFRDRGLRVLLMMDSLTRFARAQREIGLAA
FLII            MSKSVAVVATSDESPMLRKMAPLSAVTIAEHFRDQGDNVLLIIDSVTRFAHAIREVAVAS
                  ..* . * .* ... *  ..     *.*  **..  ** ..**.**.* * **...*

CI              GEPPTTKGYTPTVFTELPKLLERAGPGPIRPDGTTAAPITALFTVLVDGDDHNEPIADAT
RI              GEPPVARGYPASVFTELPRLLERAGPGA-----EGTGTITAIVSILVDGDNHNDPIADST
STI             GEPPATKGYPPSVFAKLPALVERAGNG-IHGGGS----ITAFYTVLTEGDDQQDPIADSA
XI              GEPPTRRGFPPSVFAELPRLLERAGMG---ESGS----ITAFYTVLAEDDTGSDPIAEEV
FLII            GEPPVARGYPASVFTELPRLLERAGPGA-----EGTGTITAIVSILVDGDNHNDPIADST
                **** .*. ..**. ** *.**** *           ***. ..* ..*   .***.

CI              RGILDGHIVMERAIAERGRFPAINVLKSISRTMPGCQHPHERDIVKGARQVMSAYSNMEE
RI              RGILDGHIVLDRSLAEEGRYPPINPLASISRLAKKAWTPDQEKLVSRLKALVHRFEETRD
STI             RAILDGHIVLSRRLAEAGHYPAIDIEASISRAMTALITEQHYARVRLFKQLLSSFQRNRD
XI              RGILDGHLILSREIAAKNQYPAIDVLASLSRVMSQIVPYDHSQAAGRLRRLLAKYNEVET
FLII            RGILDGHIVLDRSLAEEGRYPPINPLASISRLAKKAWTPDQEKLVSRLKALVHRFEETRD
                *.*****... * .*   ..*.*.   *.**        ..       . ..  .

CI              LIRIGAYRAGADPVVDRAIRLNPAIEAFLSQDKEEATSLDDSFGMLGQILQSEY------
RI              LRLIGGYRPGTDPDLDMAVKQVPIIYETLKQLPDEPAA-QDAYADLATALRGGAQNGQPQ
STI             LVSVGAYAKGSDPMLDKAITLWPQLEAFLQQGIFERADWEDSLQALDLIFP-TV------
XI              LVQVGEYRQGSDAVADEAIDRIDAIRDFLSQPTDQLSAYENTLELLTSVTD-DA------
FLII            LRLIGGYRPGTDPDLDMAVKQVPIIYETLKQLPDEPAA-QDAYADLATALRGGAQNGQPQ
                *  .* *  *.*.  * *.     .   * *   . .  ...   *

CI              -------
RI              VNPRMRG
STI             -------
XI              -------
FLII            VNPRMRG
```

CI=*C. crescentus* RI= *R. meliloti* STI= *S. typhimurium* XI=*X.campestris*

FLII=*A. tumefaciens* * = identities dots = conservative substitutions

Figure 3.3.2 The DNA sequence of the *fliI* homologue and the predicted gene product.

The *fliI* coding region begins at nucleotide position 1907 and finishes at position 3328. The start and stop codons are shown in bold. The predicted protein sequence is shown beneath the DNA sequence.

The stop codon of upstream *flgF* is overlined and labelled F at nucleotide positions 1893-1895.

The start codon of downstream *orfZ* is overlined and labelled Z at nucleotide positions 3372-3374. 8 base pairs upstream of the start codon a possible ribosome binding site AGGA is shown in bold.

The sequences underlined with a dotted line at nucleotide postions 2447-2493 and 2684-2734 are the nucleotide binding motifs, Walker boxes, that are very highly conserved [Goodfellow *et al.*, 1996; Walker *et al.*, 1982].

```
1601 caggcagaacgacaccatcgccgcctcgctcggcatcttccaggcggatttttcaaagggcttcctgcgtcaccccaata 1680

1681 gcggcgtgaagcccgtggcccagcccgttccggtcgtcaacaatcatgaggtcggcgtcgttcagggttatctcgagcag 1760

1761 tcgaacgtcaacggcatttcccagatgacgcagctcatccaggtcaaccgcgcctttgagagcatttcctcgatgatgcg 1840

                                                                F
1841 cgataccgaatcgaccttcggcgagggcatcaagacgctcggcggcgcgcgttgaggaactgacgg ATG ACA ATG   1915
   1                                                          rbs      M   T   M     3

1916 CCG GAA TCC ATG CTC TCG GAA TCC AAG CTC TCG GGC TGG GCC ATT TCG CCG AAG CTC GCC  1975
   4 P   E   S   M   L   S   E   S   K   L   S   G   W   A   I   S   P   K   L   A    23

1976 CAG CTC GCC AGC CTT GCC GGG CAC TAT GCC GAT CCG GAG TTC TCG GTG GCG CAT GGC GGC  2035
  24 Q   L   A   S   L   A   G   H   Y   A   D   P   E   F   S   V   A   H   G   G    43

2036 CAT GTC CGC ACC ATC GCC GCC GGG CAC TAT ACG GTG TCC GGC CTG TCG CGG CAT GTG CGG  2095
  44 H   V   R   T   I   A   A   G   H   Y   T   V   S   G   L   S   R   H   V   R    63

2096 CTC GGC GAA TTC GTT GCC CAT CGC AGC GCA ACC GGC ATT CAT CTC GGC GAA GTC GTG CGC  2155
  64 L   G   E   F   V   A   H   R   S   A   T   G   I   H   L   G   E   V   V   R    83

2156 GTG GAA CCG GAT ATC TGC TAT GTC TGC CCT ATC GAG CCC GGC GAG CCG ATC GGC ATC CAC  2215
  84 V   E   P   D   I   C   Y   V   C   P   I   E   P   G   E   P   I   G   I   H   103

2216 GAC ACC GTC ATC CGC AAG GGT GCC TTC CGC GTC TCG CCC GAC GAG AGC TGG TGC GGG CGC  2275
 104 D   T   V   I   R   K   G   A   F   R   V   S   P   D   E   S   W   C   G   R   123

2276 ACC ATC AAT GCG CTT GGC GAG CCG ATC GAC GGC CAG GGG CCG CTT GCT TCC GGC ATC GTG  2335
 124 T   I   N   A   L   G   E   P   I   D   G   Q   G   P   L   A   S   G   I   V   143

2336 CGC CGT TCG ATT TCC AAC AAT GCG CCG CCC TCG ATG ACC CGT AAA AGG GTG GAG ACA CCG  2395
 144 R   R   S   I   S   N   N   A   P   P   S   M   T   R   K   R   V   E   T   P   163

2396 TTC AAG ACG GGC GTG CGG GCT ATC GAT ATT TTC TCA CCG CTC TGC CTT GGG CAG CGT CTC  2455
 164 F   K   T   G   V   R   A   I   D   I   F   S   P   L   C   L   G   Q   R   L   183
                                                                        ---------

2456 GGT ATT TTC GCC GGT TCG GGC GTG GGC AAA TCC ACG CTG CTG TCG ATG CTC GCC AAA GCC  2515
 184 G   I   F   A   G   S   G   V   G   K   S   T   L   L   S   M   L   A   K   A   203
     ---------------------------------------------------

2516 GAT GCC TTC GAC AAG GTG GTG ATT GCG CTT GTC GGC GAA CGT GGC CGC GAA GTG CGC GAA  2575
 204 D   A   F   D   K   V   V   I   A   L   V   G   E   R   G   R   E   V   R   E   223

2576 TTC ATC GAA GAC ACG ATG GGC GAT AAT ATG AGC AAG TCC GTC GCC GTC GTC GCG ACG AGC  2635
 224 F   I   E   D   T   M   G   D   N   M   S   K   S   V   A   V   V   A   T   S   243

2636 GAC GAG AGC CCG ATG CTG CGC AAG ATG GCC CCG CTT TCC GCC GTC ACC ATT GCC GAG CAT  2695
 244 D   E   S   P   M   L   R   K   M   A   P   L   S   A   V   T   I   A   E   H   263
                                                                    ---------------

2696 TTC CGC GAT CAG GGC GAC AAT GTC CTT CTC ATC ATC GAC AGC GTG ACG CGA TTT GCC CAT  2755
 264 F   R   D   Q   G   D   N   V   L   L   I   I   D   S   V   T   R   F   A   H   283
     ---------------------------------------------------

2756 GCG ATC CGC GAA GTG GCG GTC GCT TCG GGC GAA CCG CCC GTC GCA CGC GGT TAT CCC GCC  2815
 284 A   I   R   E   V   A   V   A   S   G   E   P   P   V   A   R   G   Y   P   A   303

2816 TCG GTC TTC ACC GAA CTG CCG CGG CTG CTG GAA CGG GCA GGA CCG GGC GCG GAA GGC ACC  2875
 304 S   V   F   T   E   L   P   R   L   L   E   R   A   G   P   G   A   E   G   T   323

2876 GGC ACC ATC ACC GCC ATC GTC TCC ATC CTC GTG GAT GGT GAC AAT CAC AAC GAT CCG ATT  2935
 324 G   T   I   T   A   I   V   S   I   L   V   D   G   D   N   H   N   D   P   I   343

2936 GCC GAT TCG ACC CGT GGT ATT CTC GAC GGC CAT ATC GTG CTG GAT CGC AGC CTT GCC GAA  2995
 344 A   D   S   T   R   G   I   L   D   G   H   I   V   L   D   R   S   L   A   E   363

2996 GAG GGC CGC TAT CCG CCG ATC AAT CCG CTC GCC TCG ATC TCG CGT CTT GCC AAG AAG GCG  3055
 364 E   G   R   Y   P   P   I   N   P   L   A   S   I   S   R   L   A   K   K   A   383

3056 TGG ACG CCG GAT CAG GAA AAG CTG GTC TCG CGC CTG AAG GCG CTG GTG CAT CGT TTC GAG  3115
 384 W   T   P   D   Q   E   K   L   V   S   R   L   K   A   L   V   H   R   F   E   403
```

```
3116 GAA ACG CGG GAT CTT CGC CTC ATC GGC GGT TAC CGG CCC GGA ACC GAC CCC GAT CTC GAC  3175
 404 E   T   R   D   L   R   L   I   G   G   Y   R   P   G   T   D   P   D   L   D    423

3176 ATG GCG GTG AAG CAG GTT CCG ATC ATC TAT GAA ACG CTG AAA CAG CTA CCG GAC GAA CCG  3235
 424 M   A   V   K   Q   V   P   I   I   Y   E   T   L   K   Q   L   P   D   E   P    443

3236 GCG GCA CAG GAT GCC TAT GCC GAT CTG GCG ACC GCA CTA CGC GGT GGA GCA CAG AAT GGC  3295
 444 A   A   Q   D   A   Y   A   D   L   A   T   A   L   R   G   G   A   Q   N   G    463

3296 CAG CCC CAG GTC AAT CCG AGA ATG AGA GGC TGA ccccgtgaacaacaagcggaacgacaatagccacga 3364
 464 Q   P   Q   V   N   P   R   M   R   G   *                                       474
          Z
3365 tgaagacgtgctgtccaagccgatcttctttacgcccgaccgcattctcgcatggacgggcatcgcgcttgccgctgcag 3444

3445 cagccttttttcccctggtacgtcttcttcaacgaggagaaattcggcatgaaggtggcgggcggcgatcgcacccgcgat 3524

3525 ctgcctcataccggaccgcgggaagtcttcagtgtctcgcccatggccatgaccaaccgtaacaaggaagacaccccgcc 3604
```

3.4 *orfX*

This is a possible 636 base pair open reading frame, beginning at nucleotide position 524 and finishing at position 1159, upstream of the *flgF* homologue. The gene product of *orfX* is 212 amino acids long with a predicted M_r of 22,828 Da. When the OWL database was searched with this gene product it produced no significant sequence identity to any previously identified proteins. It did however, show significant similarity to an unidentified open reading frame in *R. meliloti*, entitled Orf20. It has 64.29 % identity over 210 amino acids. The extent of this similarity is shown in a gap alignment in figure 3.4.1. The alignment shows the similarity between the ORFs to be throughout their entire lengths.As *orfX* has no known identified homology it is impossible to state definitely where it begins, however reasons will be discussed later to support the opinion that it begins at nucleotide position 524.

Figure 3.4.2 shows the DNA sequence of *orfX* and the putative protein it encodes.

The TESTCODE program predicted at the 95 % confidence level that *orfX* is a coding region. The guanine and cytosine content of *orfX* was 57.2 % and this is significantly higher than that of non-coding regions.

Figure 3.4.1 A gap alignment between OrfX of *A. tumefaciens* and Orf20 of *R. meliloti.*

The amino acid sequence on the top line is that of OrfX from *A. tumefaciens* and the amino acid sequence of Orf20 from *R. meliloti* is shown below. Identities are indicated by double dots, and conservative substitutions by single dots.

```
                                          10        20        30
                                  MADTRLYNYALKAHGLEDMAYAKAFIRKVLT
                                   X:.:::::::.:::::::::::::::.::::.
ERVSAQPDVARETEYYLSKIGSIKTLDDFFADSRLYNYAMKAHGLEDMAYAKAFMRKVLA
   80         90        100       110       120       130

       40        50        60        70        80        90
EGASDKNAFANKLSDNRYAELAKSLDFAGLGAAATATEAAKSGVIGNYARQTLEQEAGDD
:: .. .:::::::.:.::  :..:::.::  : :::: : :..::  .:.:::::::.::..
EGIDSDDAFANKLADGRYKALVESLNFARHGEAATAFERAQKGVAEKYTRQTLEQKAGEE
  140       150       160       170       180       190

      100       110       120       130       140       150
NNGVRLALYFERKAPTIKSGLDFLADDALAQVFRTTFNLPDAFAAADVDKQAALIEKSIN
:.:::::::::.: :::: .: ...:::.::..: ::...::  ::::::::.::   : .:.
NTGVRLALYFQRMAPTIANGYEIIADEALSHVVRTVLQLPAEFAAADVDRQAEAYEAAID
  200       210       220       230       240       250

      160       170       180       190       200       210
IKDLQDPEKVGKLLERFTIMWEMQNPSTTYDPLAVFGSSSGYGISPDLLISINSLKLGGK
.::.:::.:....:.::: .::..: :  ::::::::.:::::::::::..::.::::X.
FKDFQDPQKIAEFLDRFTALWEIDNSSDGYDPLAVFGASSGYGISPDLLLTINNLKLGGR
  260       270       280       290       300       310
```

Figure 3.4.2 The DNA sequence of *orfX* and the putative protein it encodes.

orfX is postulated to begin at nucleotide position 524 and finish at position 1159. The start and stop codons are shown in bold. The putative protein sequence is shown beneath the DNA sequence.

The start codon of downstream *flgF* is overlined and labelled F at nucleotide positions 1161-1163.

```
   1 aggcctcgccgagcgccttgcccgaatccttcacgaccttcatcgggttcgccatgatgaagccgccgagaccggcgcct  80

  81 ccgatgatcatcaattcgaacggctgaaccagcacgttcaaatgaccgcccatcgccatatagccgccgatgatgcagcc 160

 161 gaaggtgattataagtccaattacaatattcattgatcgaccacgcccgaacagtgttccatggcgccatgtctagaaaa 240

 241 tgtggcttgcgtgaacctgtccgcaatgcctcctgccccgccctcacctcagccagtttcgcacaagcaagccgcgacat 320

 321 tatgccgacaagggcggacgtgcgccttgaaatcgggatcaaacgagagcggcggttcagtgacttccacctataccagc 400

 401 tacagactgatcagtcaggatatcggcaaatcgctcgagcgggtctcgaagcagccagacgtggcacgcgagaccgaata 480

 481 ttaccgtgaaaagatcggcagcgtgaaatccatcgacgatttc ATG GCC GAT ACA CGT CTT TAC AAT TAC  550
   1                                              M   A   D   T   R   L   Y   N   Y    9

 551 GCG CTG AAA GCC CAC GGC CTG GAA GAC ATG GCC TAT GCA AAA GCC TTC ATC CGC AAG GTG  610
  10 A   L   K   A   H   G   L   E   D   M   A   Y   A   K   A   F   I   R   K   V   29

 611 TTG ACG GAA GGG GCG AGC GAT AAG AAC GCC TTC GCC AAC AAG CTC TCC GAC AAT CGT TAT  670
  30 L   T   E   G   A   S   D   K   N   A   F   A   N   K   L   S   D   N   R   Y   49

 671 GCC GAG CTT GCA AAG TCG CTG GAT TTC GCG GGC CTG GGT GCG GCG GCA ACC GCG ACC GAG  730
  50 A   E   L   A   K   S   L   D   F   A   G   L   G   A   A   A   T   A   T   E   69

 731 GCG GCC AAA TCA GGT GTC ATC GGC AAC TAC GCC CGC CAG ACG CTG GAA CAG GAA GCG GGC  790
  70 A   A   K   S   G   V   I   G   N   Y   A   R   Q   T   L   E   Q   E   A   G   89

 791 GAC GAC AAT AAC GGC GTC CGG CTC GCC CTT TAT TTC GAG CGC AAG GCA CCG ACA ATC AAA  850
  90 D   D   N   N   G   V   R   L   A   L   Y   F   E   R   K   A   P   T   I   K  109

 851 TCC GGC CTC GAT TTT CTG GCC GAC GAC GCG CTT GCG CAG GTC TTC CGC ACA ACG TTC AAC  910
 110 S   G   L   D   F   L   A   D   D   A   L   A   Q   V   F   R   T   T   F   N  129

 911 CTG CCT GAC GCA TTT GCT GCT GCT GAT GTC GAC AAG CAG GCC GCC CTC ATC GAA AAG AGC  970
 130 L   P   D   A   F   A   A   A   D   V   D   K   Q   A   A   L   I   E   K   S  149

 971 ATC AAT ATC AAG GAT CTG CAG GAT CCT GAA AAG GTC GGC AAG TTG CTC GAG CGT TTC ACC 1030
 150 I   N   I   K   D   L   Q   D   P   E   K   V   G   K   L   L   E   R   F   T  169

1031 ATC ATG TGG GAA ATG CAA AAC CCC TCG ACG ACC TAC GAT CCC CTG GCC GTT TTC GGC TCC 1090
 170 I   M   W   E   M   Q   N   P   S   T   T   Y   D   P   L   A   V   F   G   S  189

1091 TCC AGC GGC TAC GGC ATT TCC CCC GAC CTG CTG ATT TCC ATC AAC TCC CTG AAA CTC GGA 1150
 190 S   S   G   Y   G   I   S   P   D   L   L   I   S   I   N   S   L   K   L   G  209
                 F
1151 GGC AAA TGA tatgcaatccggactatatgtcgccctgtcctcgcagatcgcgctggagcgccgtctcaccaccattt 1227
 210 G   K   *                                                                      212

1228 ctgacaatatggcgaatgtgaacacggtcggcttccgcggctccgaggtaaaattcgacgaaatggtcgccaagaaccac 1307

1308 aatgacatgaatgccaaggtggctttcgtgtcgcagggcaacgactatctctccacacgccagggcgctttcgagcagac 1387
```

3.5 *orfZ*

This is a 524 base pair open reading frame, beginning at nucleotide position 3372 and finishing at position 3896, between the *fliI* homologue upstream and *flgB* downstream. Unlike the three other open reading frames, *orfZ* is believed to start with a GTG codon, reasons will be discussed later to support this assumption. The gene product of *orfZ* is 175 amino acids in length with a predicted M_r of 19,131 Da and as with *orfX* has no significant sequence identity to any other identified proteins in the OWL database. However, it does show significant similarity to an unidentified open reading frame in *R. meliloti*, entitled Orf23. It has 49.375 % identity over 160 amino acids, shown in a gap alignment in figure 3.5.1. The alignment shows the similarity between the ORFs to be throughout their entire lengths.

orfZ is predicted at the 95 % confidence level to be a coding region by the UWGCG program TESTCODE. The guanine and cytosine content of *orfZ* was 59.8 %, which is high, relative to predicted non-coding regions, again suggesting that *orfZ* like *orfX* is a coding region. TESTCODE predicted a non-coding region between *orfZ* and *flgB*, therefore, the DNA sequence after *orfZ* was searched for a possible transcription termination signal. A hairpin-loop was found at nucleotide positions 3914-3926 and 3935-3947 which may serve this purpose.

Figure 3.5.2 shows the DNA sequence of *orfZ* and the putative protein it encodes.

Figure 3.5.1 A gap alignment between OrfZ of *A. tumefaciens* and Orf23 of *R. meliloti.*

The amino acid sequence on the top line is that of OrfZ from *A. tumefaciens* and the amino acid sequence of Orf23 from *R. meliloti* is shown below. Identities are indicated by double dots and conservative substitutions by single dots.

```
                        10        20        30        40
                VLSKPIFFTPDRILAWTGIALAAAAAFFPWYVFFNEEKFGMKVAGG
                          :.::.  :.:.::.:.:.:::::.. :::.:
MKLTDFDADEIVRQRRRETRMPLIDKILGAIGLAMAACATFLPWYVYLHPEKFSMPSLWQ
        10        20        30        40        50        60

  50        60        70        80        90        100
DRTRDLPHTGPREVFSVSPMAMTNRNKEDTPPAAELPDMLTTATVNAGGKERQNNVAAGI
  ::::::.   :...:::::.:::. ..:    .:   : X::::: .  ..  ...  :
GTTRDLPERPDRRILSVSPLAMTDMDEE----TAAAVDRLTTATVPGLDQDPAEDTDMGA
        70        80             90        100       110

 110       120       130       140       150       160
EDQPFPGQTSFRLLHVSNGRALIEDSSGMYMVRVGSTLPDNALLTKIEQRNGEWLIETSS
.:::::::. ::.:.::.::::::::.:::::.::.::::. :. .:.:.:.:.. :X.
HDQPFPGKESFKLMHVANGRALIEDASGMYIVRIGSVLPDNSRLATFEERDGRWVMITSK
 120       130       140       150       160       170

 170
GKTYHPEKZ
:. ..
GEIFEAK
 180
```

Figure 3.5.2 The DNA sequence of *orfZ* and the putative protein it encodes.

orfZ begins at nucleotide position 3372 and finishes at position 3896. The start and stop codons are shown in bold. The putative protein sequence is shown beneath the DNA sequence.

The stop codon of upstream *fliI* is overlined and labelled I at nucleotide positions 3326-3328.

A possible hairpin loop is underlined and labelled H at nucleotide positions 3914-3926 and 3935-3947.

```
2881 catcaccgccatcgtctccatcctcgtggatggtgacaatcacaacgatccgattgccgattcgacccgtggtattctcg 2960

2961 acggccatatcgtgctggatcgcagccttgccgaagagggccgctatccgccgatcaatccgctcgcctcgatctcgcgt 3040

3041 cttgccaagaaggcgtggacgccggatcaggaaaagctggtctcgcgcctgaaggcgctggtgcatcgtttcgaggaaac 3120

3121 gcgggatcttcgcctcatcggcggttaccggcccggaaccgaccccgatctcgacatggcggtgaagcaggttccgatca 3200

3201 tctatgaaacgctgaaacagctaccggacgaaccggcggcacaggatgcctatgccgatctggcgaccgcactacgcggt 3280
                                                    I
3281 ggagcacagaatggccagccccaggtcaatccgagaatgagaggctgaccccgtgaacaacaagcggaacgacaatagcc3360

3361 acgatgaagacGTG CTG TCC AAG CCG ATC TTC TTT ACG CCC GAC CGC ATT CTC GCA TGG ACG  3422
   1           V   L   S   K   P   I   F   F   T   P   D   R   I   L   A   W   T    17

3423 GGC ATC GCG CTT GCC GCT GCA GCA GCC TTT TTC CCC TGG TAC GTC TTC TTC AAC GAG GAG  3482
  18 G   I   A   L   A   A   A   A   A   F   F   P   W   Y   V   F   F   N   E   E    37

3483 AAA TTC GGC ATG AAG GTG GCG GGC GGC GAT CGC ACC CGC GAT CTG CCT CAT ACC GGA CCG  3542
  38 K   F   G   M   K   V   A   G   G   D   R   T   R   D   L   P   H   T   G   P    57

3543 CGG GAA GTC TTC AGT GTC TCG CCC ATG GCC ATG ACC AAC CGT AAC AAG GAA GAC ACC CCG  3602
  58 R   E   V   F   S   V   S   P   M   A   M   T   N   R   N   K   E   D   T   P    77

3603 CCA GCG GCG GAG CTG CCC GAC ATG CTG ACG ACG GCG ACC GTC AAC GCT GGC GGC AAG GAG  3662
  78 P   A   A   E   L   P   D   M   L   T   T   A   T   V   N   A   G   G   K   E    97

3663 AGA CAG AAC AAT GTG GCA GCC GGG ATC GAA GAC CAG CCC TTC CCC GGC CAG ACA TCG TTC  3722
  98 R   Q   N   N   V   A   A   G   I   E   D   Q   P   F   P   G   Q   T   S   F   117

3723 CGT CTG CTA CAT GTC TCC AAT GGC CGC GCC CTG ATC GAA GAC AGT TCG GGC ATG TAT ATG  3782
 118 R   L   L   H   V   S   N   G   R   A   L   I   E   D   S   S   G   M   Y   M   137

3783 GTG CGG GTC GGT TCC ACA TTG CCT GAC AAT GCC TTG CTG ACG AAG ATC GAG CAG CGC AAT  3842
 138 V   R   V   G   S   T   L   P   D   N   A   L   L   T   K   I   E   Q   R   N   157

3843 GGC GAA TGG CTG ATC GAG ACA TCC TCC GGA AAA ACC TAT CAT CCG GAA AAA TGA gagaattt3904
 158 G   E   W   L   I   E   T   S   S   G   K   T   Y   H   P   E   K   *            175

3905 ctccccgctaaaaaggcgcgccacacatccggcgcgctttttttgtttgtgagggacacgcaagcaatgcgtctgtttcca 3984
          H                   H
3985 ttgttccacccttcatccttctccgcaagtctgacgcaagattaaccccgtaggttcgggccatacgagatggagaagt 4064

4065 ctatgcaaccgattcaactgttcgaactggcatcccgccaagcggaatggctgagcgtgcggcaggaagtcgtggcgacc 4144

4145 aacatcgccaacgccaacacacccaagttccacgccaaagatgtcagtcccttcgaagcggtcatgcaggccaccagcca 4224

4225 gcaggtcggcatggccaggacaaaccctgtccacctcggtgaaagcacactcagcgagaatatcgcggtgcgggacaacc 4304

4305 cggtcaacaacgagatcggcatgcaggaatcgggcaattcggtcgggcttgcggaagaaatgaccaagaccggcgagatc 4384

4385 aagcggcagtacgacctgaacgccagtctcgtcaaatccttccatcgcatgatgctgatgacggttaagaggtaaaccat 4464

4465 ggatcc                                                                        4470
```

Chapter 4

A comparative study of the flagellar gene homologues and open reading frames of *Agrobacterium tumefaciens* and homologous DNA sequences in *Rhizobium meliloti.*

As mentioned previously, *A. tumefaciens* shows a great deal of similarity to the bacterium *R. meliloti*. This similarity stretches as far as the genes responsible for cell motility. Numerous studies to date have shown that DNA sequences flanking motility mutants in *A. tumefaciens*, when used as radiolabelled probes for Southern blots, are homologous to DNA sequences in *R. meliloti* genomic DNA and pRZ cosmids [Brown, 1992; Deakin, 1994; Shaw, et al., 1991]. When the databases were initially searched with the four open reading frames, the *R. meliloti* sequence homologous to this region was not known. Therefore, it was decided to make Southern blots of *R. meliloti* DNA and probe them with the four identified open reading frames from Chapter 3. One set of blots contains restriction enzyme digests of three pRZ cosmids, namely pRZ1, 2 and 4. The blot used for this study was kindly loaned by Dr. W. Deakin and is taken from his work on the Molecular Characterisation Of Flagellar Genes From *A. tumefaciens* [1994]. The second set of blots were carried out using DNA kindly given by Dr. K. Bergman, and are of the plasmid pMB7.

4.1 The pRZ cosmids

The cosmids pRZ1, pRZ2 and pRZ4 were isolated, digested with *Eco*RI and *Bam*HI and electrophoresed on a large 0.7 % agarose gel. A photograph of the gel was taken and then the DNA transferrred to a nylon filter by Southern blotting [Deakin, 1994]. The photograph of the gel is shown in Figure 4.1.1. The three pRZ cosmids overlap each other, covering a sequence region of approximately 35 kb. Within this region is a stretch of DNA of roughly 25 kb in length which complements numerous chemotactic and motility mutants in *R. meliloti*. Restriction digests using the enzymes *Bam*HI and *Eco*RI have been carried out to map the sequence and the locations of the complement mutations have been identified [Ziegler, et al., 1986]. Figure 4.1.2 shows the original map of this region.

The blot corresponding to the photograph shown in figure 4.1.1 was probed four times with fluorescein labelled DNA fragments. The size and location of the sections of DNA used to make the probes can be seen in figure 4.1.3. As no suitable restriction enzyme sites were present, two primers were designed and a PCR reaction was carried out to isolate a section of *orfZ* DNA to use as a probe. The two primers are shown below and their locations are marked on figure 4.1.3.

Primer Probe Z1 **GCGATCTGCCTCATACCG**

Primer Probe Z2 **CGATCAGCCATTCGCCATTGC**

Figure 4.1.1 Photograph taken with permission from Deakin [1994] showing an agarose gel containing the digests of the pRZ cosmids and chromosomal DNA from *R. meliloti* prior to Southern blotting.

LANE 1	*Hin*dIII cut λ DNA	
LANE 2	pRZ1	*Bam*HI
LANE 3	pRZ1	*Eco*RI
LANE 4	pRZ2	*Bam*HI
LANE 5	pRZ2	*Eco*RI
LANE 6	pRZ4	*Bam*HI
LANE 7	pRZ4	*Eco*RI
LANE 8	*R. meliloti* genomic DNA cut with *Eco*RI	
LANE 9	*Pst*I cut λ DNA	

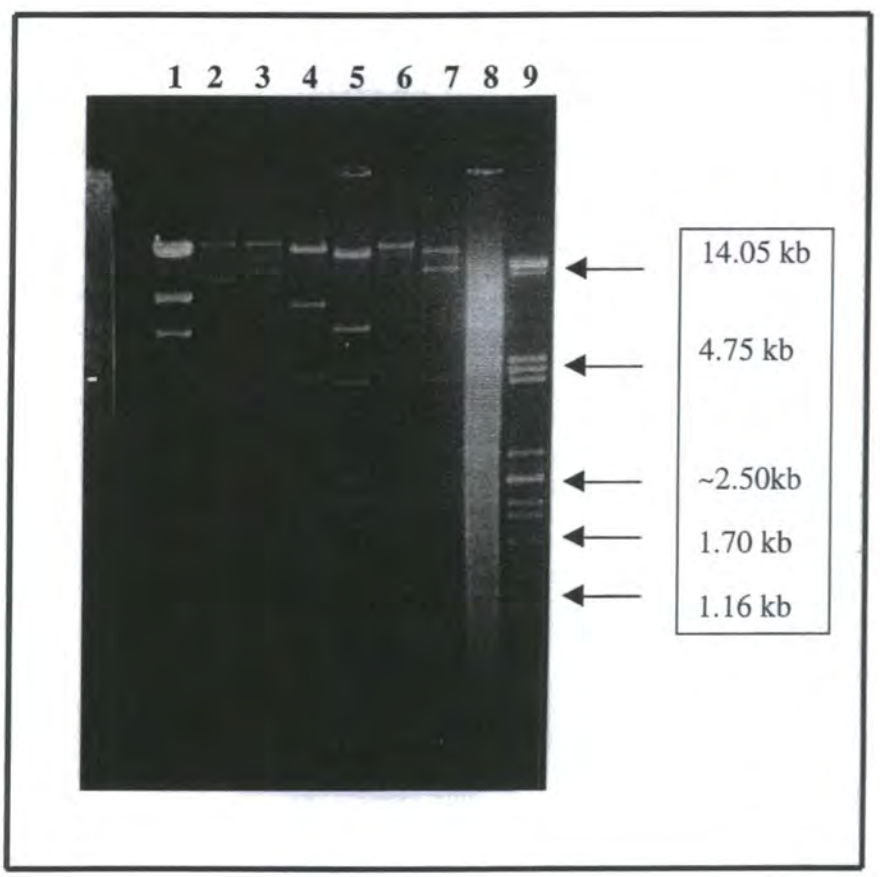


Figure 4.1.2 A map showing the location of the three pRZ cosmids and their complementing behavioural mutants within *R. meliloti*

The diagram was redrawn with the permission of Deakin [1994] and Ziegler *et al.* [1986]. At the top of the diagram several of the *R. meliloti* behavioural mutants can be seen. Beneath them are the pRZ1, pRZ2 and pRZ4 cosmids. The thick line shows an approximate restriction map of the region carried out with *Bam*HI (B) and *Eco*RI(R). The *Eco*RI fragments labelled RA, RB and RC, and the *Bam*HI fragment labelled BA are discussed in Deakin [1994]. The two arrows labelled A and B are the flagellin genes *flaA* and *flaB* also in Deakin [1994]. The three cosmids were created by cloning *Eco*RI partially digested fragments of chromosomal DNA from *R. meliloti* into pLAFR1.

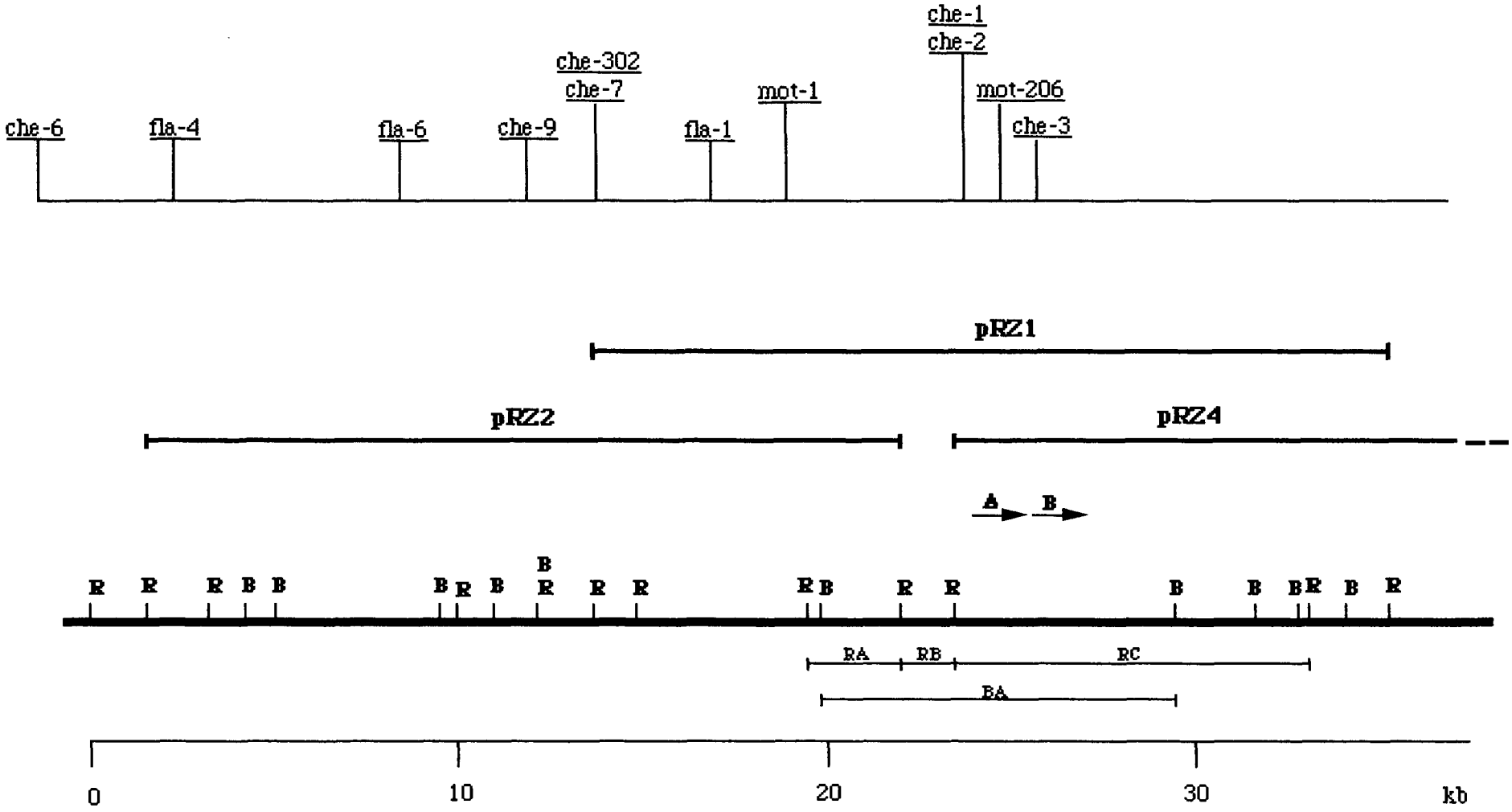
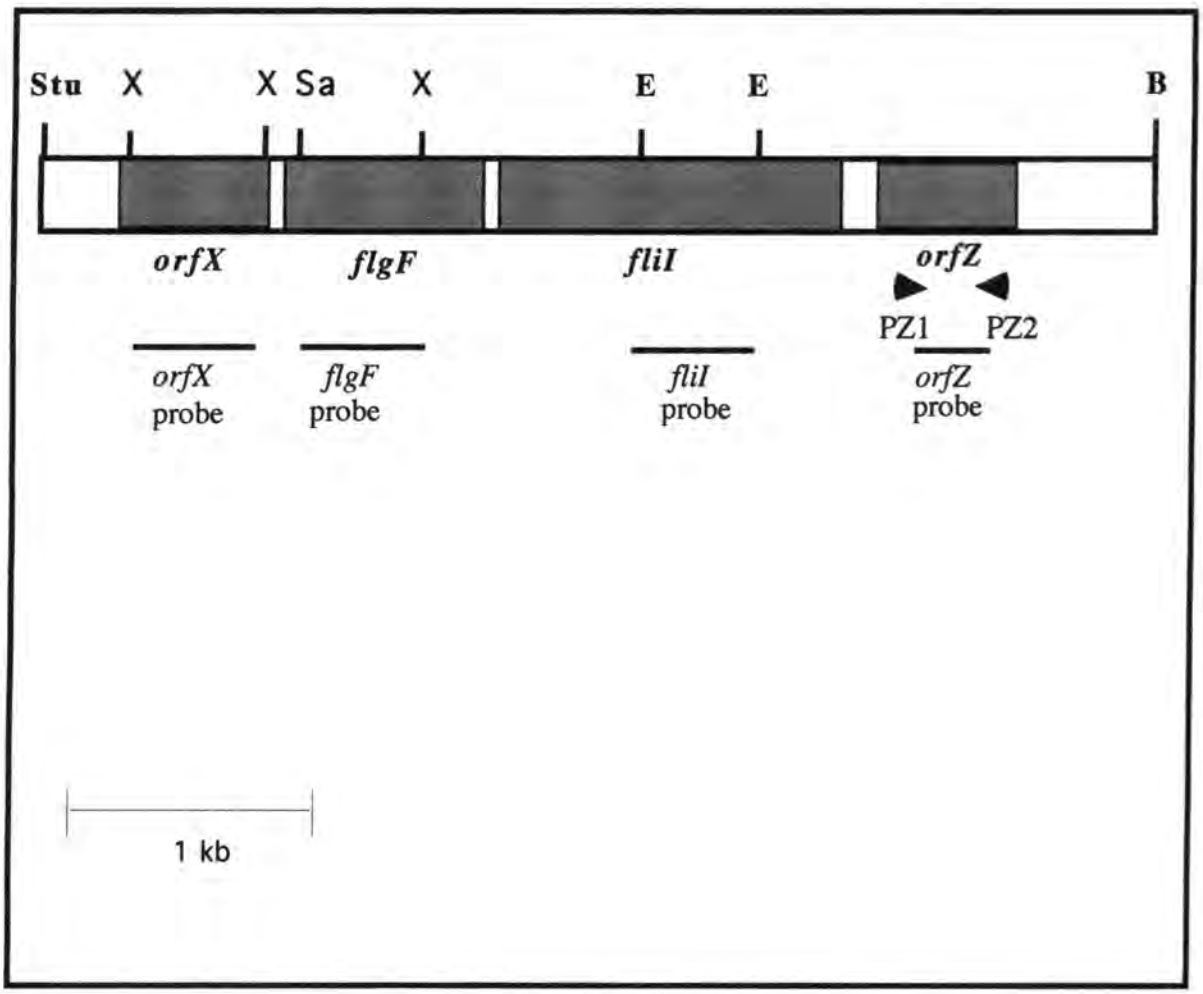

che-1
che-2
che-302
che-7
mot-1
mot-206
che-6
fla-4
fla-6
che-9
fla-1
che-3
pRZ1
pRZ2
pRZ4
A
B
R
B
RA
RB
RC
BA
0
10
20
30
kb

Figure 4.1.3 Size and location of labelled DNA

The four shaded boxes show the full length genes. The lines at the bottom of the diagram show the locations and sizes of the DNA used to make the four probes. The restriction enzyme sites used to cut out the fragments from pDUB1900 are shown. X= *Xho*I, Sa= *Sac*II, E= *Eco*RI. The two primer sites, PZ1 and PZ2, used to make the *orfZ* probe are also labelled.



All four of the fluorescein labelled *A. tumefaciens* open reading frames hybridised to two of the three pRZ cosmids. No bands were produced in the lanes containing digests of pRZ4, indicating that the four open reading frames show homologies to sequences of DNA upstream of *flaA* and *flaB* [Deakin, 1994] shown on figure 4.1.2.

The fluorescein labelled fragment of *orfX* hybridised to bands in both pRZ1 and pRZ2. The *Bam*HI digest of pRZ1 that the *orfX* probe hybridised to was > 25 kb. This was presumably because the DNA, in addition to the Rhizobium DNA, contained the pLAFRI vector (~23 kb). In each of the *Eco*RI digests a band of ~ 4.4 kb was hybridised to. The only band of this size common to both pRZ1 and pRZ2 has been labelled on figure 4.1.4 as **A**. The *Bam*HI fragment of pRZ2 that the *orfX* probe hybridised to was ~8.4 kb. On the map, the *Bam*HI fragment that this most closely matched was 8.5 kb in size, this has been labelled on figure 4.1.4 as **D** and is assumed to be the same *Bam*HI fragment as was partly hybridised to in the pRZ1 *Bam*HI digest. In the *Eco*RI digest of the genomic DNA, the *orfX* probe hybridised to a faint band of ~ 4.4 kb, as with the other *Eco*RI digests. The results of this blot can be seen in figure 4.1.5.

The majority of the bands present on the *flgF* blot (figure 4.1.5) are identical to those on the *fliI* blot (figure 4.1.5) indicating that these two open reading frames are in close proximity to one another. Both the *flgF* and *fliI* probes hybridised to a *Bam*HI fragment in the pRZ1 digestion which was > 25 kb. This is assumed to be the same band as with *orfX* and contains mainly vector DNA. The probes also hybridised to an 8.4 kb band in the *Bam*HI digest of pRZ2. This again is believed to be the same fragment as with *orfX* and is labelled **D** on figure 4.1.4. Both *fliI* and *flgF* probes hybridised to a band of ~ 1.2 kb in the *Eco*RI digests of pRZ1 and pRZ2. There is only one *Eco*RI fragment of this size that is present in both pRZ1 and pRZ2, it is labelled **B** on figure 4.1.4. This would agree with the results of the *Bam*HI digestion of pRZ1 which suggested that *flgF* and *fliI* are located at one end of pRZ1. The same size *Eco*RI fragments were hybridised to faintly in the genomic DNA digestions. Additionally, the *fliI* probe hybridised fairly strongly to a band of approximately 2.2 kb in the *Eco*RI digest of pRZ2. This may indicate that *fliI* slightly spans the *Eco*RI site and crosses into the region labelled **C** on figure 4.1.4.

The labelled *orfZ* probe hybridised to an 8.4 kb band in the *Bam*HI digest of pRZ1 indicating that *orfZ* must be located on the same *Bam*HI fragment as the three other open reading frames. The *orfZ* probe hybridised to a band of ~ 2.2 kb in the *Eco*RI digest of pRZ2. This is presumably the same band as was faintly hybridised to by *fliI* and is labelled **C** on figure 4.1.4. Therefore, it would seem likely that *orfZ* is directly to the left of *fliI* in pRZ2. The labelled *orfZ* probe did not hybridise to a band in the *Bam*HI digest of pRZ1 as would agree with the original map shown in figure 4.1.2, it did however, hybridise to a band

Figure 4.1.4 **The sites of hybridisation in the pRZ cosmids**

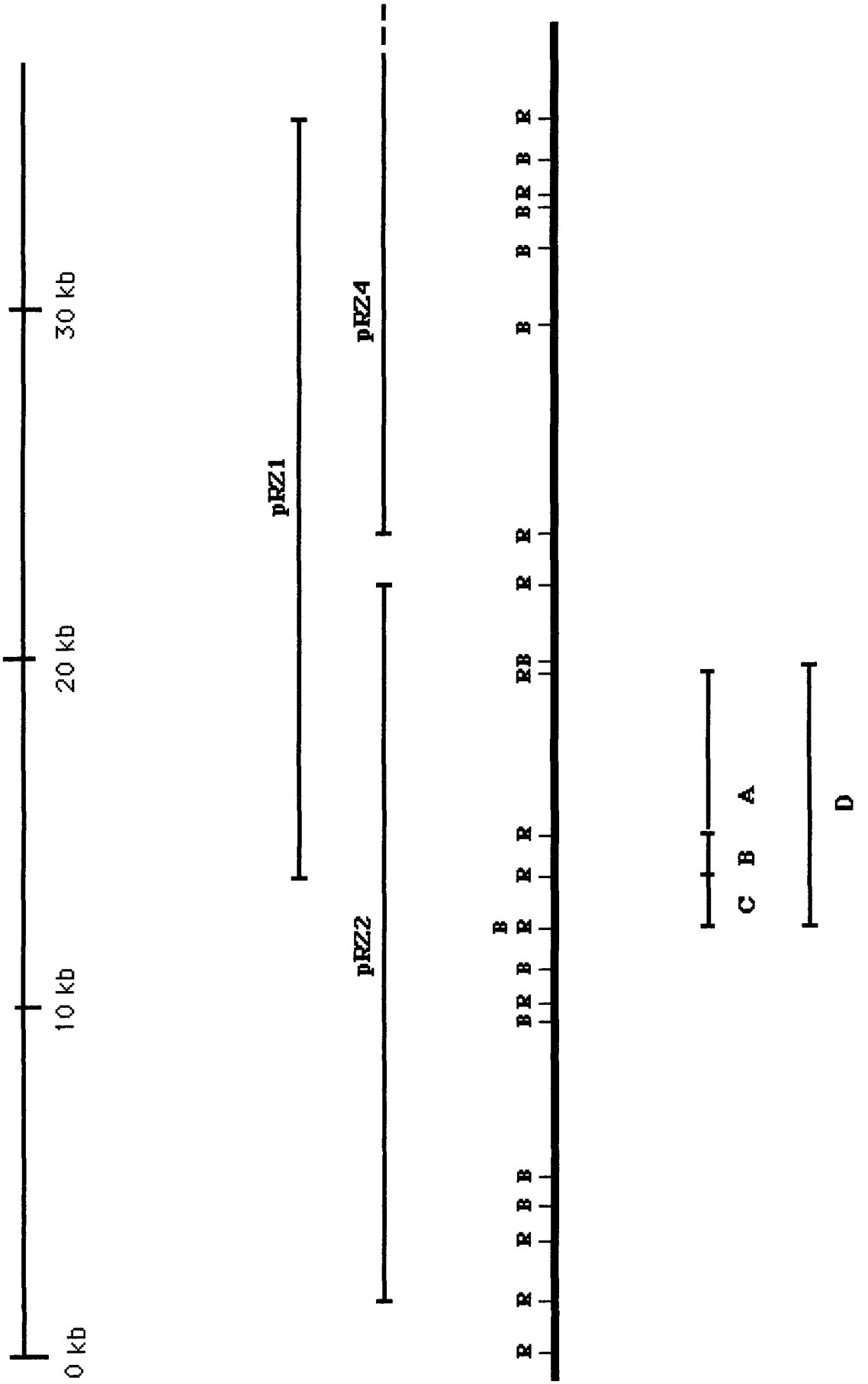


Figure 4.1.5 pRZ cosmid DNA and *R. meliloti* genomic DNA Southern blots probed with *orfX*, *flgF*, *fliI* and *orfZ* fluorescein-labelled fragments.

Blot A ***orfX* probe on agarose gel (figure 4.1.1)**

LANE 2	pRZ1	*Bam*HI
LANE 3	pRZ1	*Eco*RI
LANE 4	pRZ2	*BamHI*
LANE 5	pRZ2	*EcoRI*
LANE 6	pRZ4	*Bam*HI
LANE 7	pRZ4	*Eco*RI
LANE 8	*R. meliloti* genomic DNA cut with *Eco*RI	

Blot B ***flgF* probe on agarose gel (figure 4.1.1)**

LANE 2	pRZ1	*Bam*HI
LANE 3	pRZ1	*Eco*RI
LANE 4	pRZ2	*BamHI*
LANE 5	pRZ2	*EcoRI*
LANE 6	pRZ4	*Bam*HI
LANE 7	pRZ4	*Eco*RI
LANE 8	*R. meliloti* genomic DNA cut with *Eco*RI	

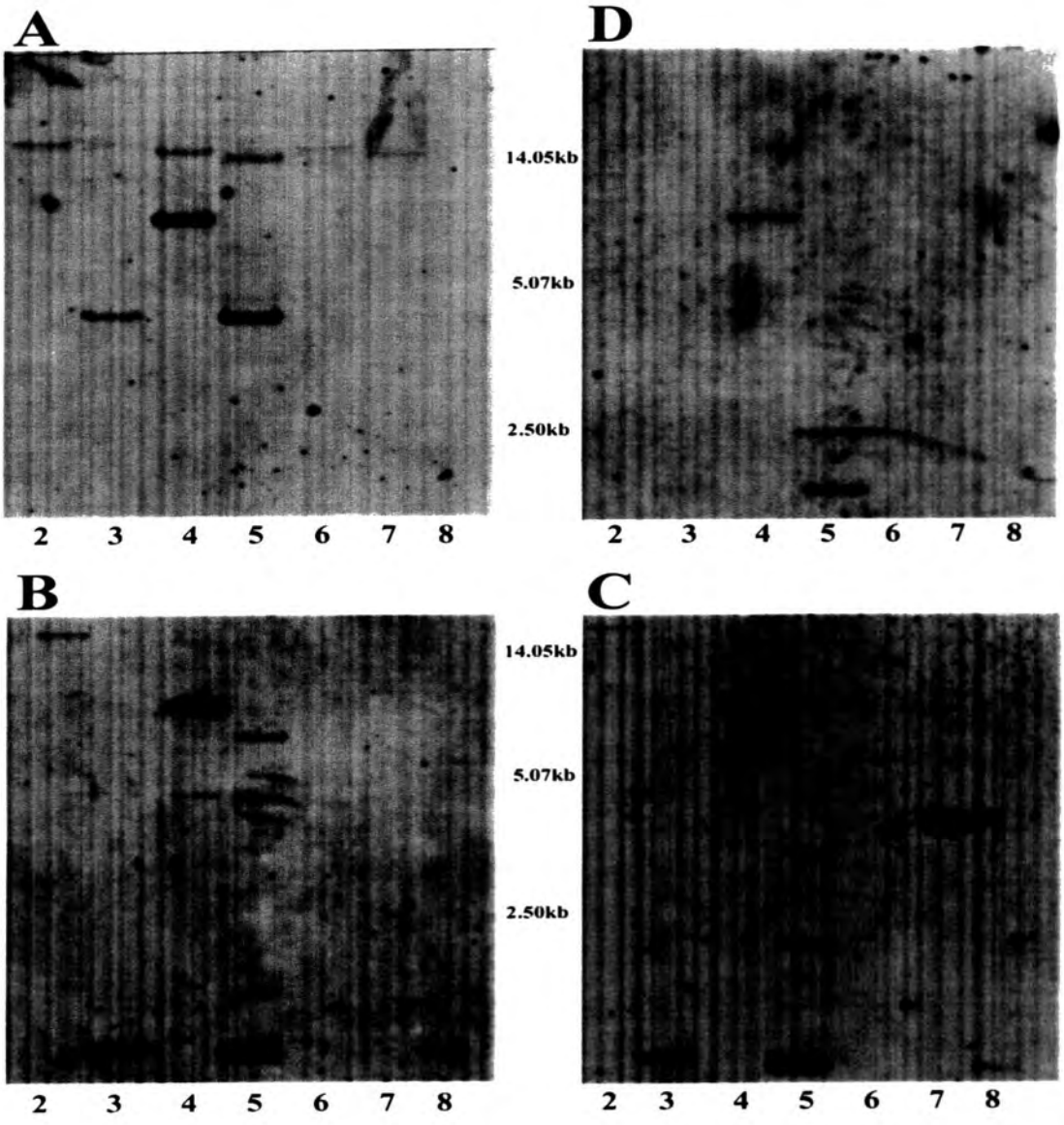
Blot C ***fliI* probe on agarose gel (figure 4.1.1)**

LANE 2	pRZ1	*Bam*HI
LANE 3	pRZ1	*Eco*RI
LANE 4	pRZ2	*BamHI*
LANE 5	pRZ2	*EcoRI*
LANE 6	pRZ4	*Bam*HI
LANE 7	pRZ4	*Eco*RI
LANE 8	*R. meliloti* genomic DNA cut with *Eco*RI	

Blot D ***orfZ* probe on agarose gel (figure 4.1.1)**

LANE 2	pRZ1	*Bam*HI
LANE 3	pRZ1	*Eco*RI
LANE 4	pRZ2	*BamHI*
LANE 5	pRZ2	*EcoRI*
LANE 6	pRZ4	*Bam*HI
LANE 7	pRZ4	*Eco*RI
LANE 8	*R. meliloti* genomic DNA cut with *Eco*RI	

The films were exposed to the blots for 48 hrs.

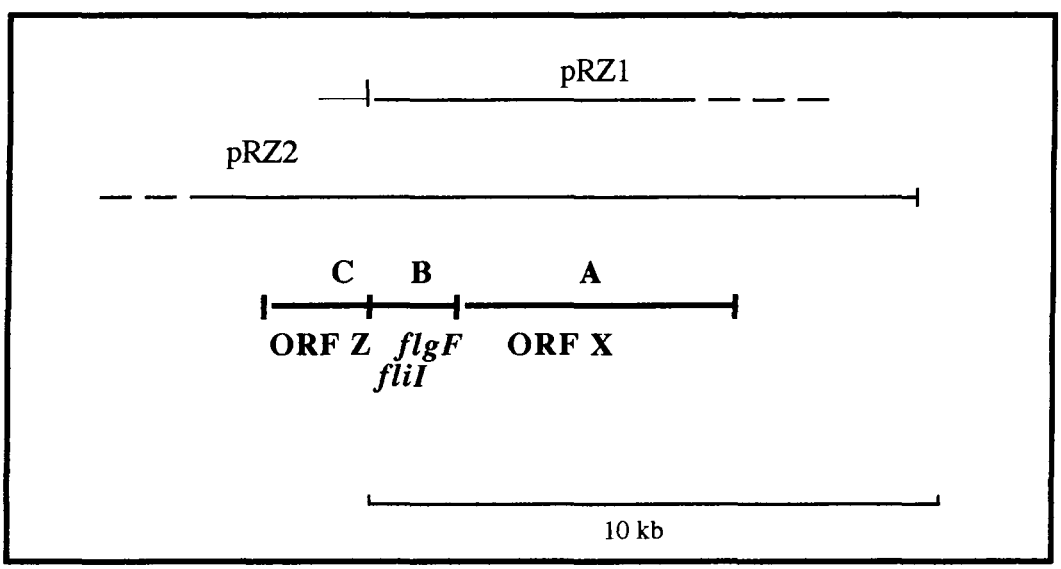

A
D
14.05kb
5.07kb
2.50kb
2 3 4 5 6 7 8
B
C
14.05kb
5.07kb
2.50kb
2 3 4 5 6 7 8

of 2.2 kb in the *Eco*RI digest of pRZ1 which appears to dispute this. It is impossible to say at this stage which of these results is correct without further investigation. All the bands on this blot were weak, with the genomic DNA homology not showing up at all, this may suggest that the pRZ1/*Bam*HI homology was also too weak to register, this band being very weak on the three other blots. If this thory is correct, then pRZ1 must extend slightly further upstream than was shown in figure 4.1.2. The *orfZ* blot is shown in figure 4.1.5.

Faint bands can be seen in a number of lanes on the *orfX* blot at greater than 25 kb. These are thought to be as a result of genomic contamination of the cosmids due to an inefficient clean-up method.

These results show that all four of the *A. tumefaciens* flagellar genes and unidentified open reading frames were found to be homologous to regions of DNA sequence within *R. meliloti*. More importantly the gene order appears to be the same in *R. meliloti* as in *A. tumefaciens* , although from the original map in figure 4.1.2, transcription appears to be in the opposite direction relative to *flaA*. Figure 4.1.6 shows the postulated order of the four open reading frames in *R. meliloti*.

Figure 4.1.6 The postulated order of the four open reading frames in *R. meliloti*



It was decided to try and confirm these results by probing Southern blots of a different plasmid, pMB7, which also covered this region and had a more up-to-date map.

4.2 The pMB7 plasmid

The pMB7 plasmid was isolated, digested with *Eco*RI and *Bam*HI and electrophoresed on a large 0.7 % agarose gel. A photograph of the gel was taken and can be seen in figure 4.2.1. The DNA was transferred from the gel to a nylon filter by Southern blotting. The pMB7 plasmid covers a region of approximately 8.4 kb. This region is entirely contained within pRZ2 and most of pRZ1. A map of the pMB7 plasmid with various restriction sites identified is shown, with the permission of Dr. K. Bergman in figure 4.2.2. As can be seen from the map, there have been some amendments since the original map shown in figure 4.1.2 was constructed. Notably there are two extra *Eco*RI sites. These have been marked on figure 4.2.2 with a *. The blot corresponding to the photograph shown in figure 4.2.1 was split into 4 identical strips, with each separate blot now containing pMB7 digested with *Bam*HI and *Eco*RI. The blots were probed with four radiolabelled DNA fragments. The size and location of the sections of DNA used to make the probes were the same as for the pRZ cosmids blots and can be seen in figure 4.1.3.

All four of the radiolabelled *A. tumefaciens* open reading frames hybridised to an 8.4 kb band in the *Bam*HI digest of the pMB7 plasmid. This band was presumably the pMB7 insert that had been cut from the pUC19 vector in the *Bam*HI digestion. The band is labelled **D** in figure 4.2.2. This indicated that the four open reading frames showed homologies to sequences of DNA located on the pMB7 plasmid and in the same region as was found with the pRZ blots.

In the *Eco*RI digestion of pMB7, the radiolabelled *orfX* probe hybridised to a fragment of ~ 4.4 kb. There was a band on the pMB7 map of a similar size and hence, thought to be the same fragment. This band has been labelled **A** on figure 4.2.2. In both the *Bam*HI and *Eco*RI digests, *orfX* hybridised to two much fainter bands of 2.65 kb and 3.1 kb respectively. Due to the size of these bands they were thought to be as a result of contamination of the probe with pUC19 vector DNA. These bands were also hybridised to in each of the other three pMB7 blots. The results of the *orfX* blot are shown in figure 4.2.3.

As with the pRZ blots, the *flgF* and *fliI* probes showed practically identical homologies with pMB7. In the *Eco*RI digests, both probes hybridised to a fragment of approximately 1.2 kb. On the pMB7 map (figure 4.2.2) the nearest fragment to this size is to the left of the *orfX* hybridisation site and is labelled **B**. The results of the *flgF* and *fliI* blots can be seen in figure 4.2.3.

Figure 4.2.1 Photograph showing an agarose gel containing the digests of the pMB7 plasmid DNA from *R. meliloti* prior to Southern blotting.

LANE 1	pMB7	*Bam*HI
LANE 2	*Pst*I cut λ DNA	
LANE 3	pMB7	*Eco*RI
LANE 4		
LANE 5	pMB7	*Bam*HI
LANE 6	*Pst*I cut λ DNA	
LANE 7	pMB7	*Eco*RI
LANE 8		
LANE 9	pMB7	*Bam*HI
LANE 10	*Pst*I cut λ DNA	
LANE 11	pMB7	*Eco*RI
LANE 12	pMB7	*Bam*HI
LANE 13	*Pst*I cut λ DNA	
LANE 14	pMB7	*Eco*RI

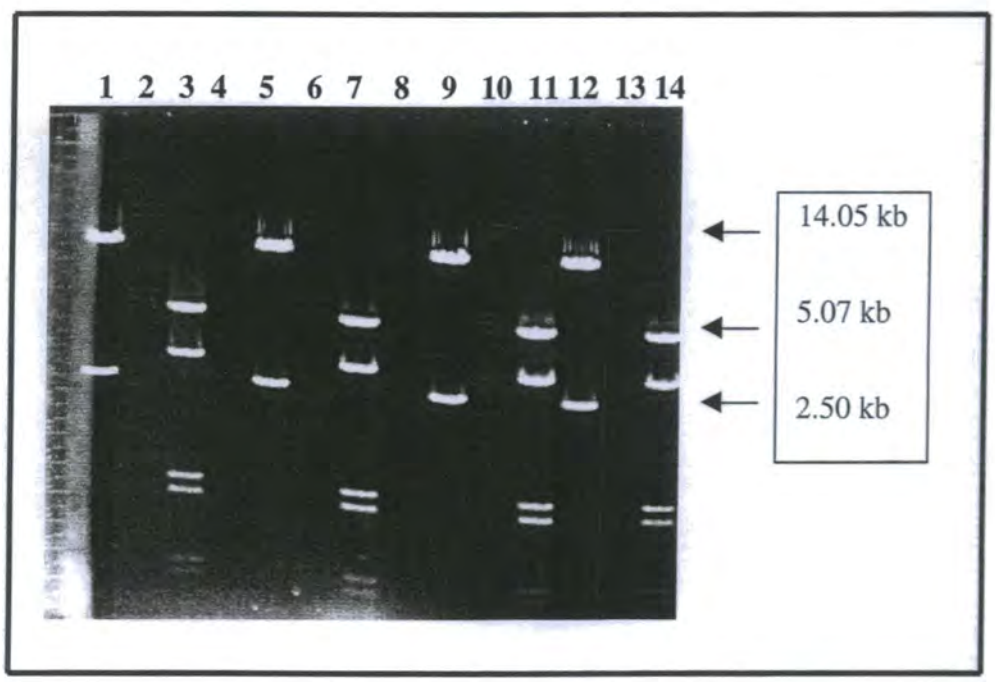


Figure 4.2.2 A map of the pMB7 plasmid with *Eco*RI and *Bam*HI restriction sites marked

The diagram was redrawn with the permission of Dr. K. Bergman. At the top of the diagram the corresponding positions of the pRZ1 and pRZ2 plasmids can be seen. The thick line shows an approximate restriction map of the region carried out with *Eco*RI (E) and *Bam*HI (B). The labelled fragments are discussed in the text. The * denotes the two extra *Eco*RI sites mentioned in the text.

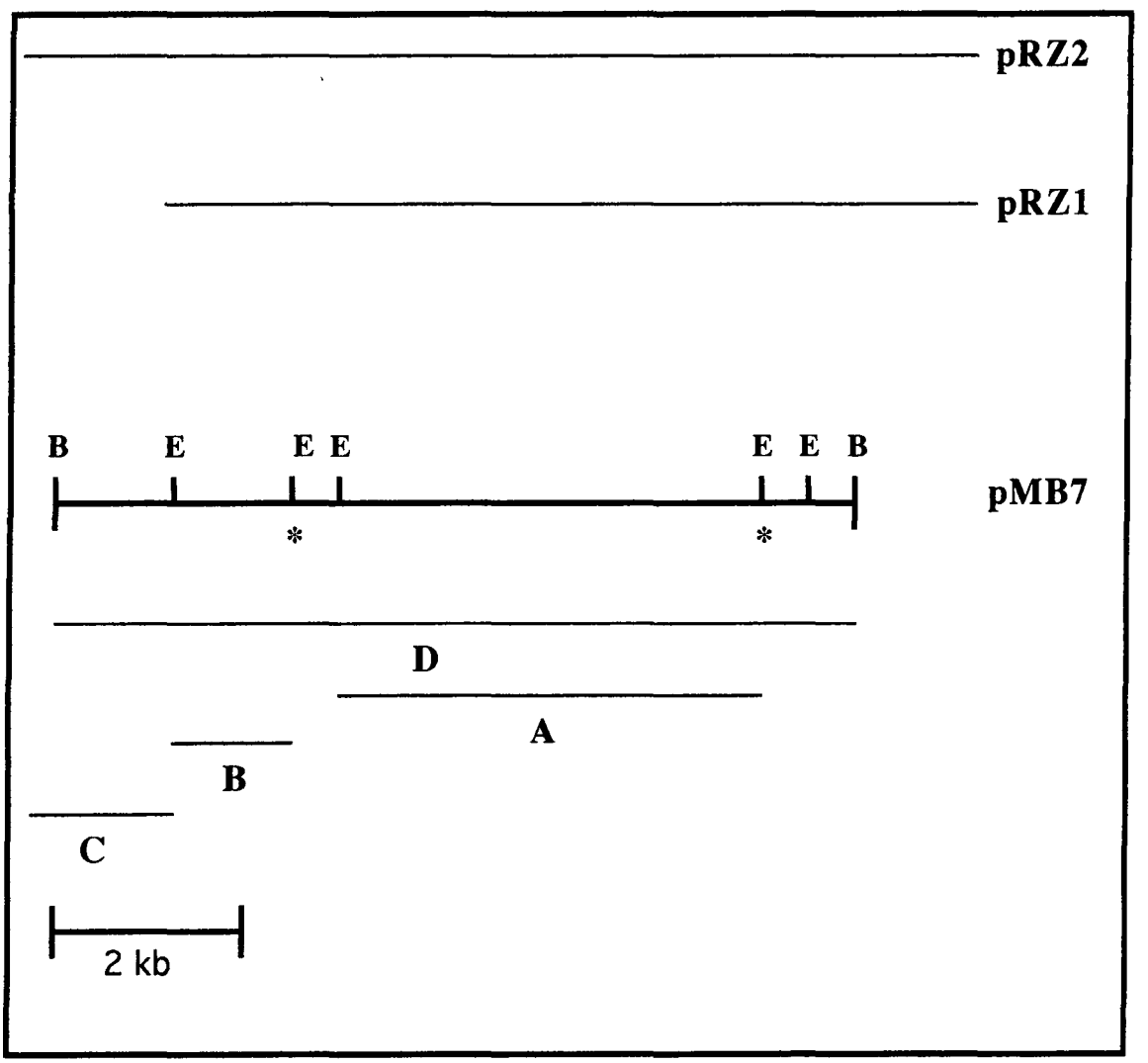


Figure 4.2.3 pMB7 plasmid DNA Southern blots probed with *orfX*, *flgF*, *fliI* and *orfZ* radiolabelled fragments.

Blot A ***orfX*** **probe on agarose gel (figure 4.2.1)**

LANE 1	pMB7	*Bam*HI
LANE 2	*Pst*I cut λ DNA	
LANE 3	pMB7	*Eco*RI

Blot B ***flgF*** **probe on agarose gel (figure 4.2.1)**

LANE 5	pMB7	*Bam*HI
LANE 6	*Pst*I cut λ DNA	
LANE 7	pMB7	*Eco*RI

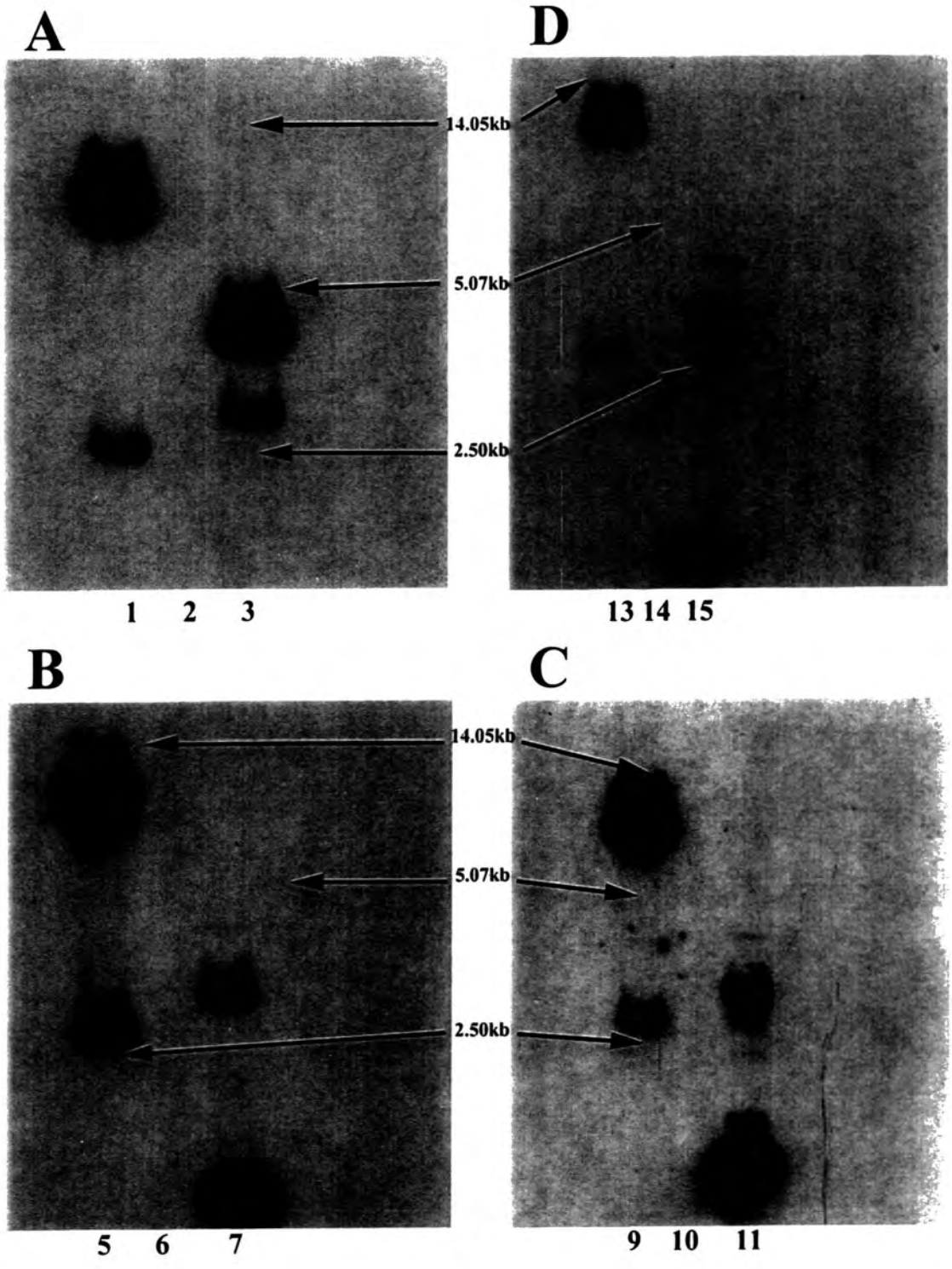
Blot C ***fliI*** **probe on agarose gel (figure 4.2.1)**

LANE 9	pMB7	*Bam*HI
LANE 10	*Pst*I cut λ DNA	
LANE 11	pMB7	*Eco*RI

Blot D ***orfZ*** **probe on agarose gel (figure 4.2.1)**

LANE 12	pMB7	*Bam*HI
LANE 13	*Pst*I cut λ DNA	
LANE 14	pMB7	*Eco*RI

The films were exposed to the blots for 48 hrs.


A
D
14.05kb
5.07kb
2.50kb
1 2 3
13 14 15
B
C
14.05kb
5.07kb
2.50kb
5 6 7
9 10 11

In the *Eco*RI digestion of pMB7, the *orfZ* probe showed strongest homology to a band of approximately 1 kb. This 1 kb band can be seen on the map of pMB7 in figure 4.2.2, to the left of the *flgF* and *fliI* *Eco*RI band, labelled C. It slightly overlaps the vector/insert site. The results of the *orfZ* blot are shown in figure 4.2.3.

From the results of the four blots a schematic diagram of pMB7 was constructed (figure 4.2.4) showing the postulated order of the four open reading frames.

4.3 Summation of the results of the pRZ blots and pMB7 blots.

Both sets of blots demonstrated that the four open reading frame fragments had homologous DNA sequences within *R. meliloti*. The open reading frame homologues appeared to be in the same order in *R. meliloti* as in *A. tumefaciens* although the direction of transcription looked to be inverted. Recently, the *R. meliloti* sequence homologous to this region was entered into the databases, Genbank accession number SMAJ4445 (23.2.98). It is now possible to compare the results of the Southern blots with that of the actual DNA sequence.

4.4 Confirmation of Southern blots by comparison with the actual *R. meliloti* sequence.

The R. *meliloti* DNA sequence entered in the database spans a region of 14,949 base pairs, mapping many chemotaxis, flagellar and motility genes to a single region of the *R. meliloti* chromosome from *orf10* to *flgC*. Chapter 3 gave the alignment results of the *R. meliloti* genes to the four *A. tumefaciens* open reading frames. Contrary to the results of the Southern blots earlier in this chapter, the sequence alignment results show that the order of transcription, relative to *motA*, of the four open reading frames is identical in *R. meliloti* and *A. tumefaciens*, and not inverted as was first thought. A restriction map of the *R. meliloti* DNA sequence was drawn, with the *Eco*RI and *Bam*HI sites marked. This can be seen in figure 4.4.1. From the sizes of the fragments it is obvious that the *Bam*HI fragment containing the four *A. tumefaciens* open reading frames was earlier mapped in the wrong orientation in the pRZ cosmids and pMB7 plasmid maps. The positions of all the *Eco*RI sites in the sequence confirm the band sizes hybridised to by the probes earlier in this chapter and also that the postulated order of the genes was correct. The revised hybridisation position and order of the four *A. tumefaciens* homologues are marked on figure 4.4.1. As was hypothesised, pRZ1 must extend further than was first suggested and all four open reading frames contain homology to it.

Figure 4.2.4 Diagram showing the postulated positions of the flagellar gene homologues in the pMB7 plasmid.

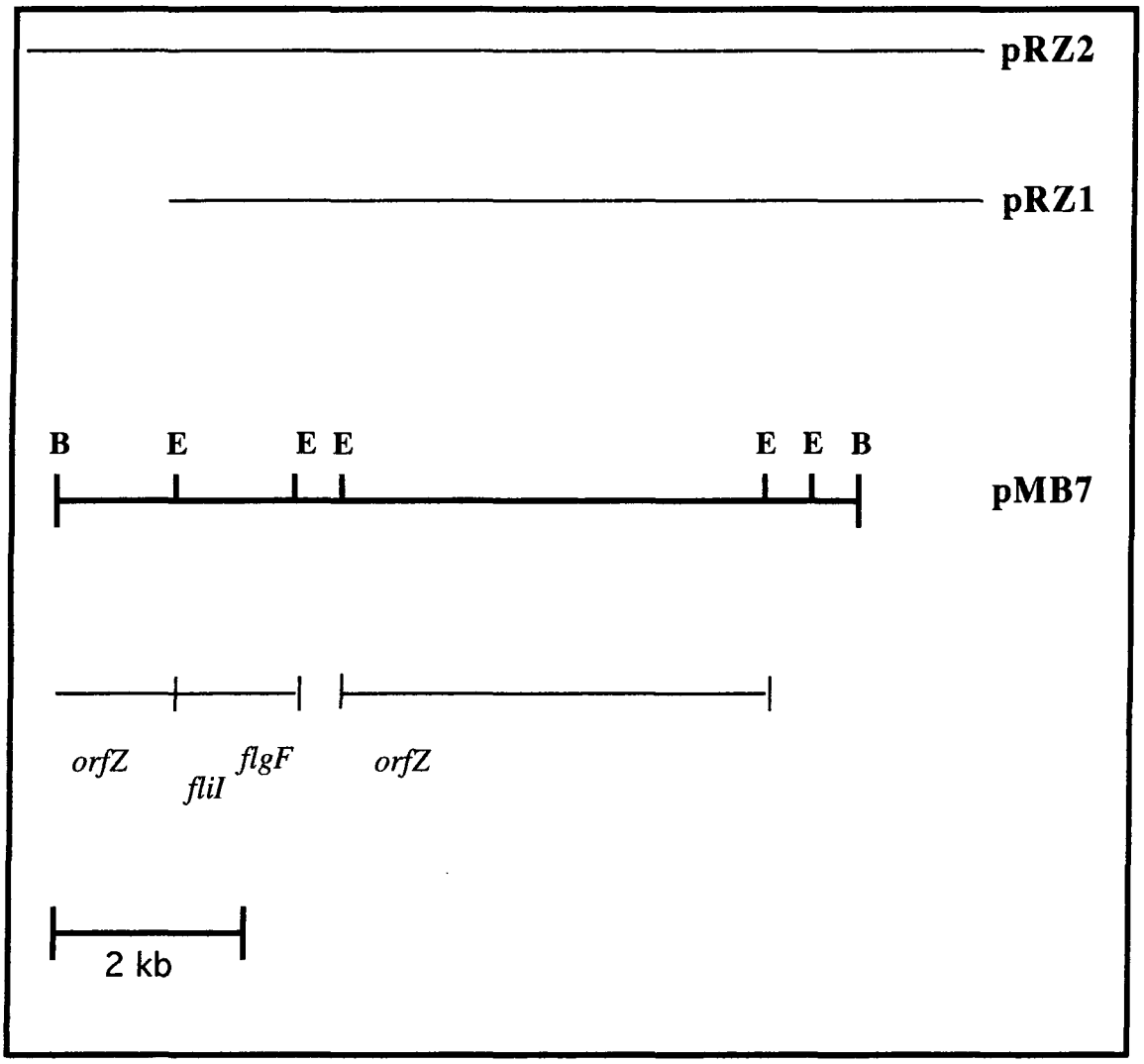
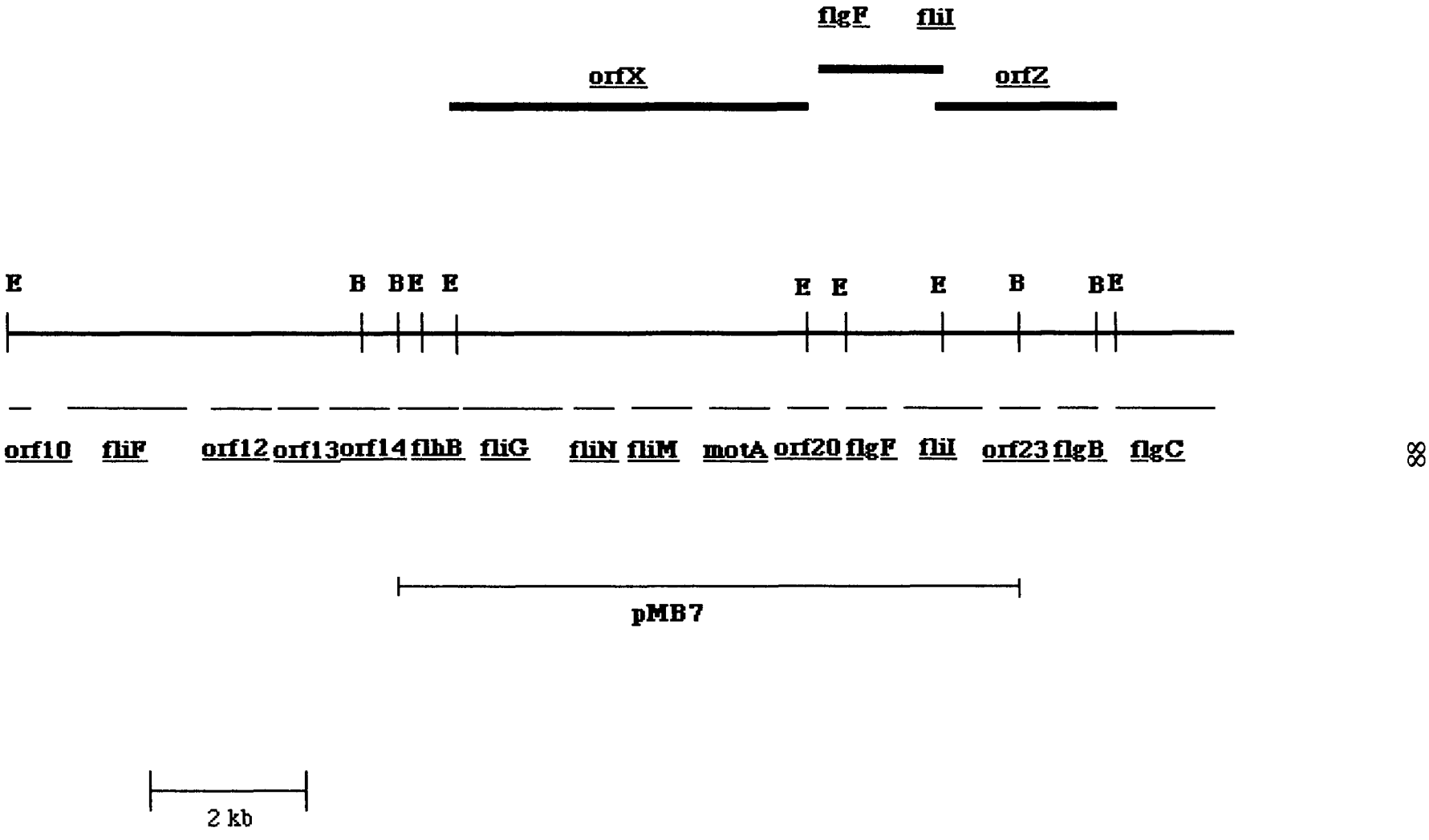


Figure 4.4.1 A diagram mapping almost 15 kb of *R. meliloti* DNA with regions showing homology to the four *A. tumefaciens* open reading frames.

The R. *meliloti* DNA sequence entered in the database spans a region of 14,949 base pairs, mapping several chemotaxis, flagellar and motility genes to a single region of the *R. meliloti* chromosome from *orf10* to *flgC*. This restriction map of the *R. meliloti* DNA sequence was drawn to show the positions of the *Eco*RI (E) and *Bam*HI (B) sites. Some of the identified genes and open reading frames within the *R. meliloti* sequence are shown below the map, with the revised pMB7 plasmid shown at the bottom of the diagram. The *Eco*RI hybridisation sites of the *A. tumefaciens* homologues are shown in bold at the top of the diagram.


flgF
fliI
orfX
orfZ
E
B
BE
E
E
E
E
B
BE
orf10
fliF
orf12
orf13
orf14
flhB
fliG
fliN
fliM
motA
orf20
flgF
fliI
orf23
flgB
flgC
pMB7
2 kb

Chapter 5

Mutagenesis

5.1 Creation of *orfX* and *orfZ* mutants

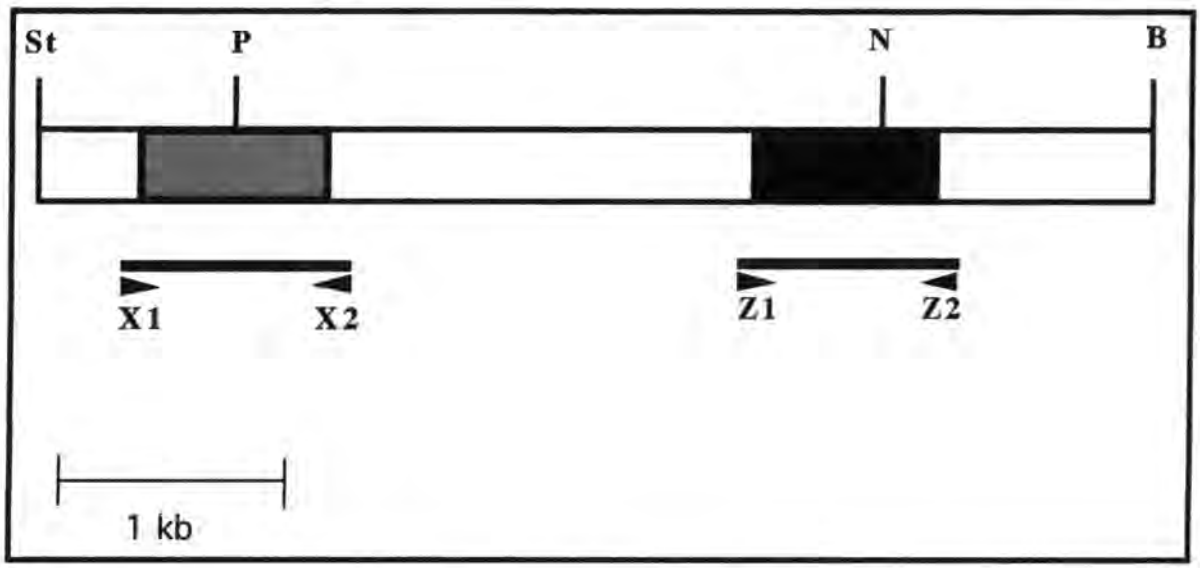
Mutations of the *orfX* and *orfZ* genes were created in an attempt to determine their phenotypes, that is, would disruption of the *orfX* and *orfZ* genes effect motility in *A. tumefaciens*? This procedure involved inserting a neomycin resistance cassette into a specific site within the chosen ORF in order to disrupt it. The insertion of a neomycin resistance cassette creates polar mutations, that is, because there is a termination of transcription, downstream ORFs are also affected. Thus, if *flgF* and *fliI* have similar functions in *A. tumefaciens* as in other bacteria, the insertion of a neomycin resistance cassette into *orfX* should produce a non-motile mutant. The insertion of a neomycin cassette into *orfZ* should determine if *orfZ* is necessary for motility and if downstream ORFs are affected by a termination of transcription within *orfZ*. The full experimental procedure is described in the methods section.

The first phase in the subcloning procedure involved the production of a plasmid. The plasmid used was pJQuc1 [Quandt and Hynes, 1993], a map of it can be seen in figure 5.1.2 (i). The second step was to clone fragments of DNA containing the *orfX* and *orfZ* genes into pJQuc1. Primers were designed, one from each end of both genes. The polymerase chain reaction was used to produce 1 kb fragments containing the genes. Figure 5.1.1 shows the size and location of the *orfX* and *orfZ* PCR fragments. The four primers are also shown on this diagram.

Primer X1 TCTCGAAGCAGCCAGACGT
Primer X2 GCGGAAGCCGACCGTGTTCAC

Primer Z1 CGGTGAAGCAGGTTCCGATC
Primer Z2 CCTGCATGACCGCTTCGAAG

Figure 5.1.1 Size and location of the *orfX* and *orfZ* PCR fragments



The unshaded box shows the region sequenced, with the bold lines showing the positions of the *orfX* and *orfZ* PCR fragments. The positions of the full length genes are also marked. *OrfX* is shown by the shaded region and *orfZ* by the black region. The four primers used in the PCR reactions are labelled:- X1, X2, Z1 and Z2. The restriction sites marked are St (*Stu*I), B (*Bam*HI), N (*Nco*I) and P (*Pst*I).

orfX

The *orfX* fragment was cloned into pJQuc1 at the *Sma*I site using a blunt-ended ligation procedure descibed in the methods section. The presence of β-galactosidase within pJQuc1 allowed blue/white selection for the insertion. The resulting plasmid was named pVEKX and is shown in figure 5.1.2 (ii)

The third phase was to disrupt the *orfX* gene by inserting the neomycin resistance cassette within it. The cassette was isolated from pDUB2033 with *Pst*I. This was subcloned into pVEKX at the *Pst*I site within *orfX* to create pVEKXn, figure 5.1.2 (iii). As pJQuc1 contains a gentamycin resistance gene, pVEKXn could be selected for due to its double resistance to neomycin and gentamycin.

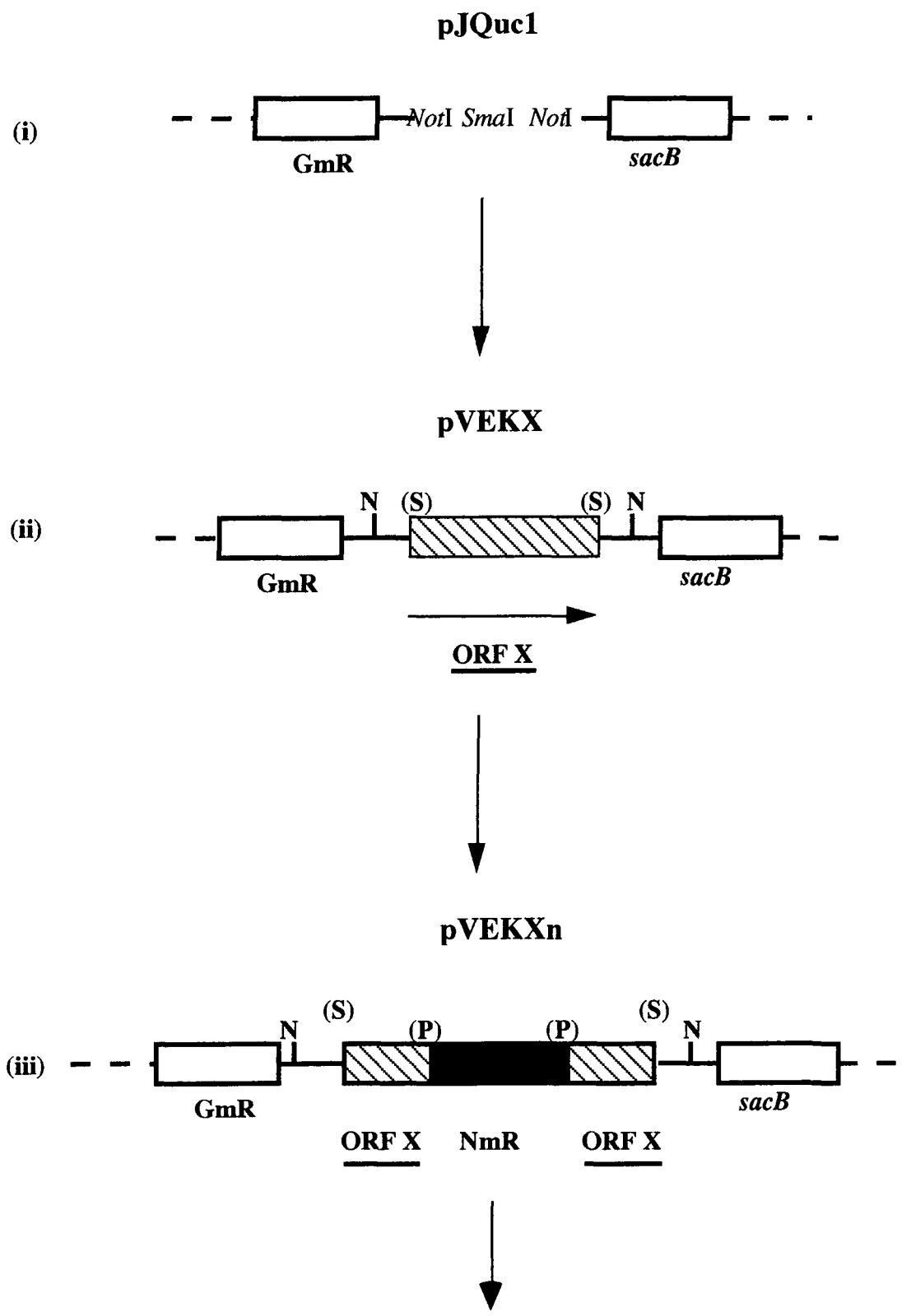
Figure 5.1.2 Diagrams showing the creation of the *orfX* mutants

(i) pJQuc1: The plasmid used was created by removing much of the multiple cloning site from pJQ200SK, a basic suicide vector, leaving only the *Not*I and *Sma*I restriction sites as shown. It has a functional *mob* site, the *sacB* gene and a gene for resistance to gentamycin.

(ii) pVEKX: The plasmid created by cloning the *orfX* fragment into pJQuc1 at the *Sma*I (S) site in a blunt-ended ligation. The shaded area indicates the *orfX* fragment. The (S) indicates the postion of the old *Sma*I site.

(iii) pVEKXn The third phase shows the disruption of the *orfX* gene by the insertion of a neomycin resistance cassette (solid black box) at the *Pst*I site (P) within *orfX*.

Finally pVEKXn was conjugated into *A. tumefaciens* and mutants resistant to neomycin and gentamycin selected for.



Conjugation into A. tumefaciens

orfZ

The *orfZ* fragment was cloned into pJQuc1 at the *Sma*I site in a blunt-ended ligation to create pVEKZ, as shown in figure 5.1.3 (ii).

The *orfZ* gene was disrupted with the insertion of a neomycin resistance cassette at the *Nco*I site. The resulting plasmid was named pVEKZn, figure 5.1.3 (iii). pVEKZn could again be selected for due to its double resistance to neomycin and gentamycin.

Finally, pVEKXn and pVEKZn were conjugated into *A. tumefaciens*. Double recombination allowed for strains to be selected for by their ability to grow on sucrose and neomycin plates. The mutant strains should have contained *orfX* and *orfZ* genes with a neomycin resistance cassette in the middle of them and hence an extra *Hin*dIII site.

5.2 Confirmation of an *orfZ* mutant

Six *orfZ* putative mutant colonies (Z MUT 1-5) which grew on neomycin, rifampicin and sucrose plates were grown up in culture. Rifampicin was included as it is a selectable marker for the *A. tumefaciens* strain, C58C1, used in this study. Chromosomal DNA was isolated and digested with *Hin*dIII. A sample of wild-type chromosomal DNA was also digested with *Hin*dIII and all restriction digests were then electrophoresed alongside each other. Figure 5.2.1 shows a photograph of the resulting agarose gel. The DNA was transferred to a nylon filter by Southern blotting and the *orfZ* specific probe (as described in Chapter 4) and neomycin resistance cassette specific probe were used in a hybridisation reaction. Figures 5.2.2 (i) and (ii) show the results of the two blots.

Blot A, figure 5.2.2 (i), was probed with the neomycin resistance cassette specific probe. As can be seen, no bands were hybridised to in the lanes containing wild-type chromosomal DNA as would be expected. The probe hybridised to three bands in all six mutant digests. Z MUT 5 in lane 6 gave the best results, as can be seen from the photograph of the agarose gel (figure 5.2.1) this DNA cut better than the rest of the mutants. The largest band hybridised to was approximately 11 kb in size and is assumed to be *Hin*dIII fragment A of pDUB1900 and the neomycin resistance cassette. It seems probable that the *Hin*dIII site in the neomycin cassette did not cut properly during the digestion due to an overloading of DNA or as a result of impurities in the genomic DNA isolation. This band is very faint in mutant Z MUT 5 and has been labelled **A** on figure 5.2.3. The second largest band hybridised to was approximately 9.05 kb in size. This band contains DNA from *Hin*dIII

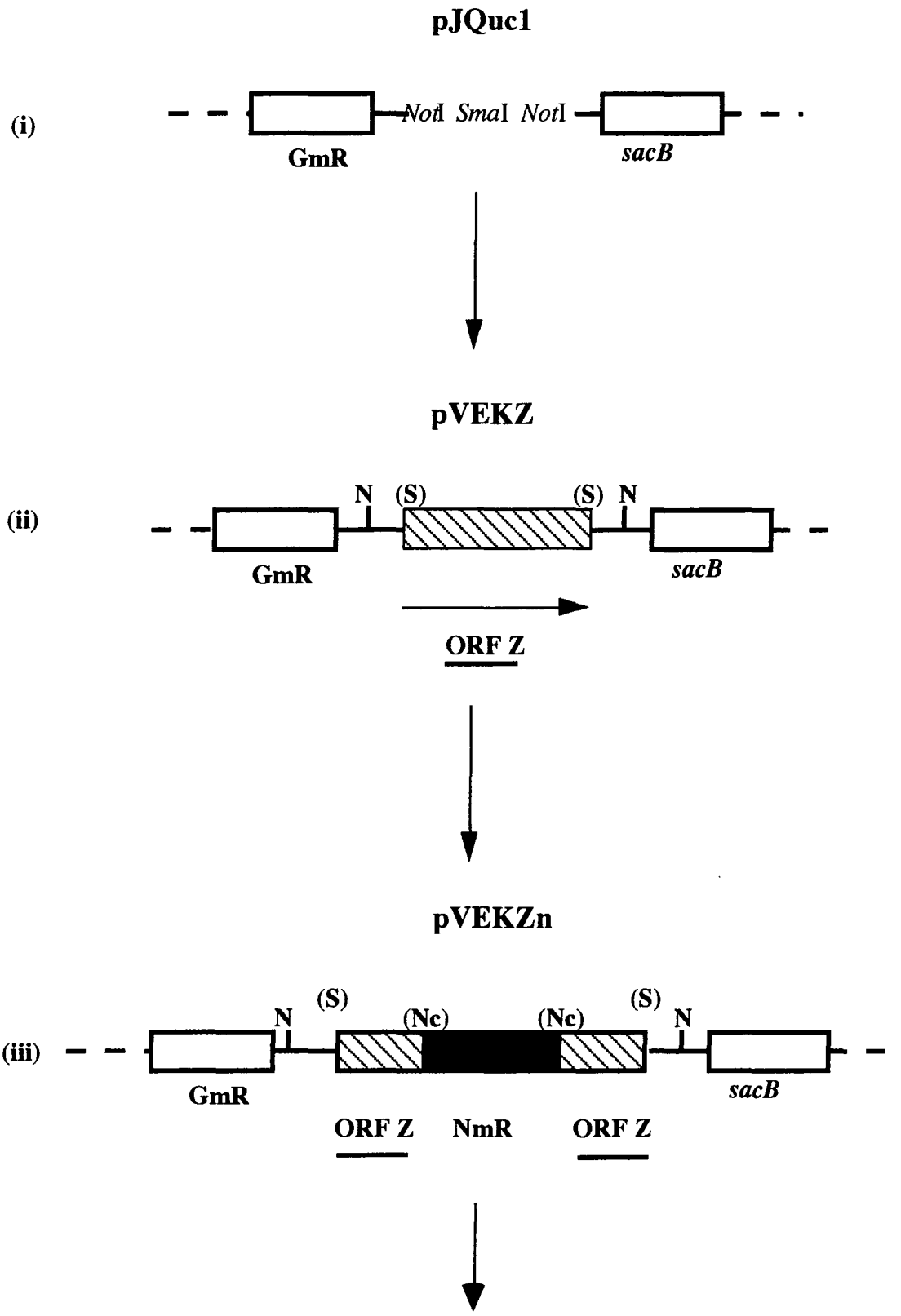
Figure 5.1.3 Diagrams showing the creation of the *orfZ* mutants

(i) pJQuc1: The plasmid used was created by removing much of the multiple cloning site from pJQ200SK, a basic suicide vector, leaving only the *Not*I and *Sma*I restriction sites as shown. It has a functional *mob* site, the *sacB* gene and a gene for resistance to gentamycin.

(ii) pVEKZ: The plasmid created by cloning the *orfZ* fragment into pJQuc1 at the *Sma*I (S) site in a blunt-ended ligation. The shaded area indicates the *orfZ* fragment. The (S) indicates the postion of the old *Sma*I site.

(iii) pVEKZn The third phase shows the disruption of the *orfZ* gene by the insertion of a neomycin resistance cassette (solid black box) at the *Nco*I site (N) within *orfZ*.

Finally pVEKZn was conjugated into *A. tumefaciens* and mutants resistant to neomycin and gentamycin selected for.


pJQuc1
(i)
NotI SmaI NotI
GmR
sacB
pVEKZ
(ii)
N (S) (S) N
GmR
sacB
ORF Z
pVEKZn
(iii)
N (S) (Nc) (Nc) (S) N
GmR
sacB
ORF Z NmR ORF Z
Conjugation into A. tumefaciens

fragment A, part of *orfZ* and part of the neomycin resistance cassette. It has been labelled **B** on figure 5.2.3. Z MUT 5 gave the brightest band. The smallest band hybridised to was approximately 1.95 kb in size and has been labelled **C** on figure 5.2.3. This contains DNA from the remainder of *Hin*dIII fragment A, *orfZ* and a small part of the neomycin resistance cassette. Again Z MUT 5 gave the brightest band as a result of the better digestion.

The same blot was then stripped of the neomycin resistance cassette specific probe and then re-probed using the *orfZ* specific probe. Blot B, figure 5.2.2 (ii), shows the results of this. This time a band was hybridised to in the lanes containing the wild-type chromosomal DNA as was expected. This band was approximately 10.1 kb in size which is the known size of *Hin*dIII fragment A of pDUB1900. In the lanes containing the six putative mutants, the same bands were hybridised to as for blot A, namely 11 kb, 9.05 kb and 1.95 kb (**A**, **B** and **C** on figure 5.2.3 respectively). Z MUT 5 again giving the best results.

These results show that the six colonies picked all contained the neomycin resistance cassette within *orfZ* and, hence, all were *orfZ* mutants. It had been hoped to repeat these procedures for the putative *orfX* mutants but unfortunately due to time restraints this was not possible.

The phenotypes of the six *orfZ* mutants were then analysed. Swarm plate analysis showed all six to be motile as did visualisation with a light microscope. Figure 5.2.4 shows a photograph of the swarm plates along with a wild-type swarm plate. The mutants were then observed with an electron microscope and all were found to have complete flagella. Figure 5.2.5 shows the electron micrographs produced.

If the two flagellar gene homologues and two unidentified open reading frames do infact form an operon, then it would be expected that polar effects, as a result of the neomycin resistance cassette insertion in *orfX* or *orfZ*, would also prevent transcription of *flgF* and *fliI*. In other bacteria, this has been shown to produce non-motile cells. However, the *orfZ* mutants were all found to be motile, suggesting that *orfZ* is not essential for flagella function. It may be a possibility that enough of *orfZ* remained before the insertion site to ensure it was still functional and thus, the putative operon was transcribed as normal. Unfortunately, time limitations prevented further analysis of the *orfZ* mutants and also of the possible *orfX* mutants.

Figure 5.2.1 Agarose gel electrophoresis of the six putative *orfZ* mutants' genomic DNA digested with *Hin*dIII.

LANE 1	*Pst*I cut λ DNA	
LANE 2	*A. tumefaciens* C58C1 genomic DNA	*Hin*dIII
LANE 3	Z MUT 1	*Hin*dIII
LANE 4	Z MUT 2	*Hin*dIII
LANE 5	Z MUT 3	*Hin*dIII
LANE 6	Z MUT 4	*Hin*dIII
LANE 7	Z MUT 5	*Hin*dIII
LANE 8	Z MUT 6	*Hin*dIII
LANE 9	*A. tumefaciens* C58C1 genomic DNA	*Hin*dIII
LANE 10	*Pst*I cut λ DNA	

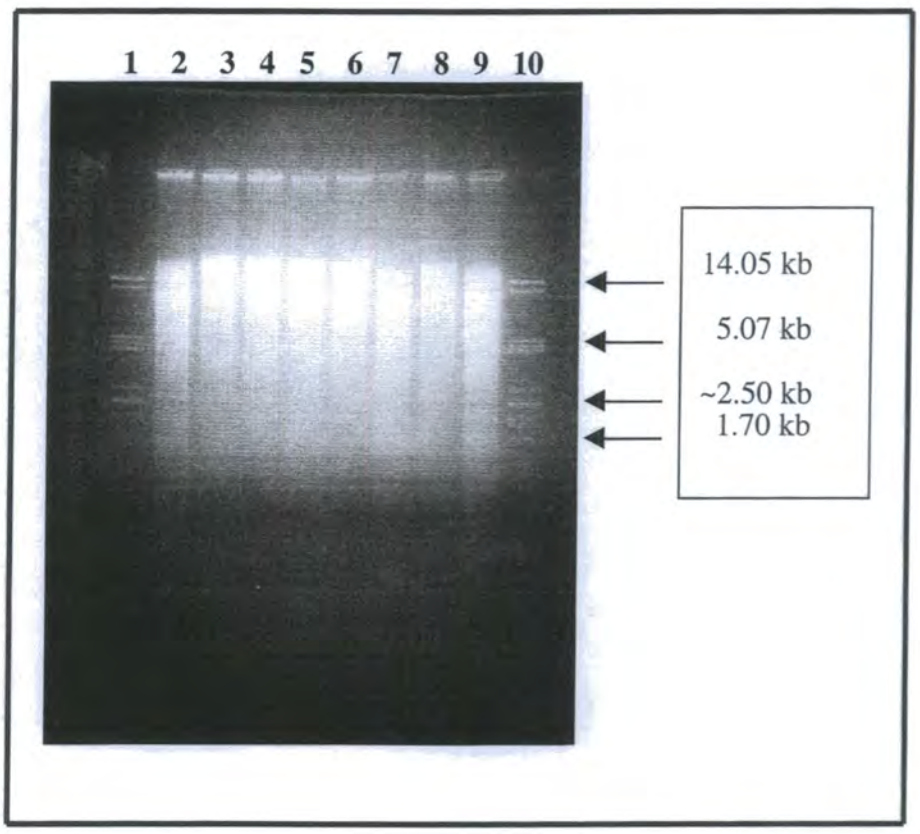
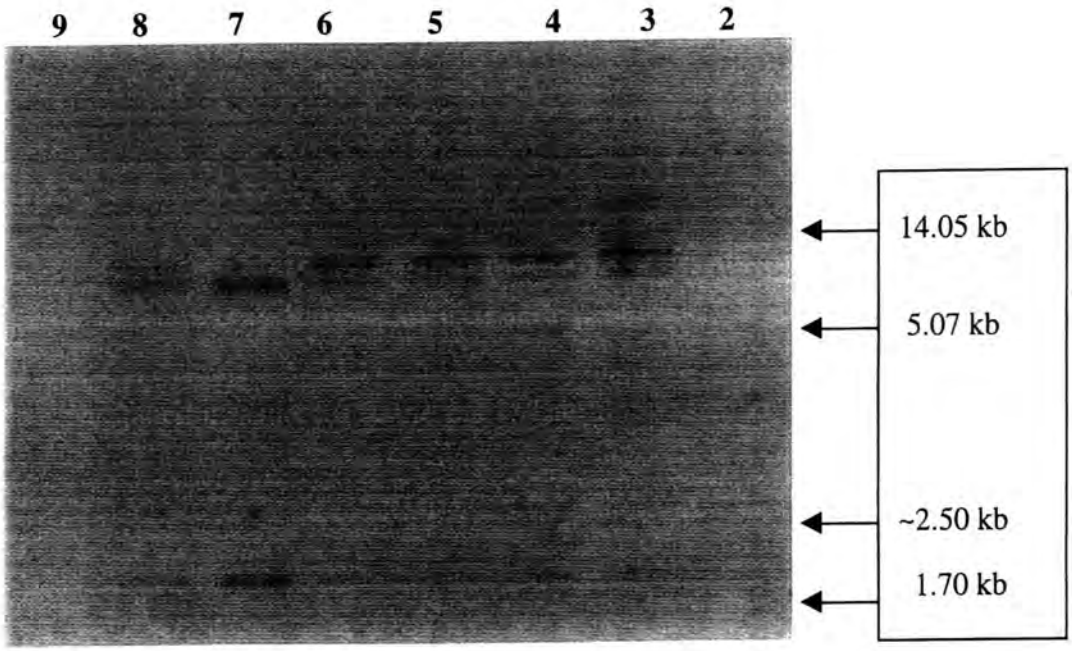


Figure 5.2.2 Southern blots of the putative *orfZ* mutants probed with the *orfZ* specific probe and a neomycin resistance cassette probe.

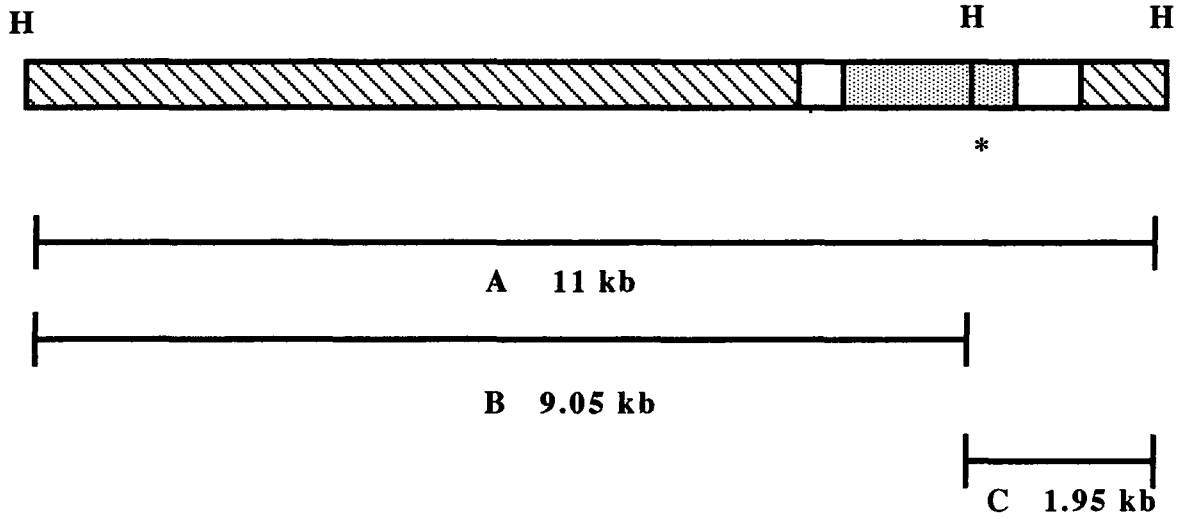
The lane labelling is as for figure 5.2.1. The films were exposed to the blots for 24 hours.

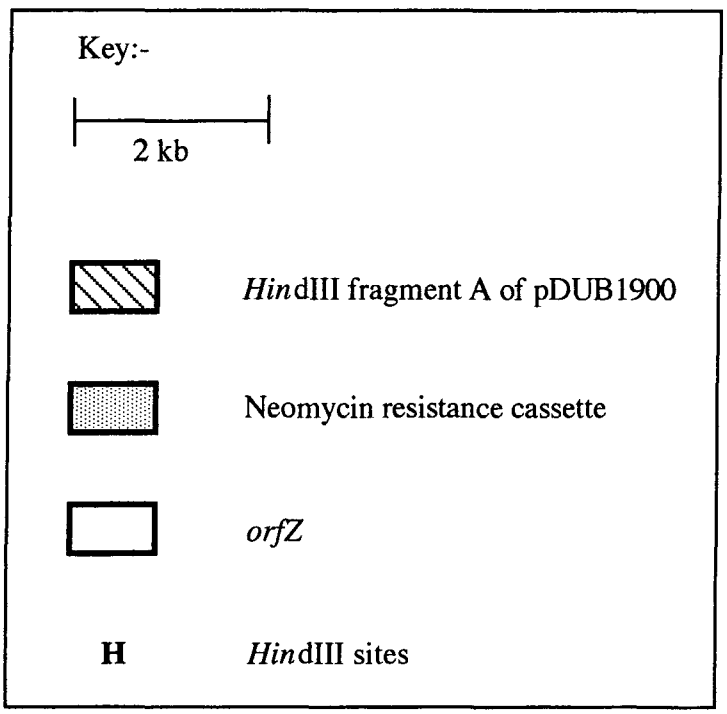
Blot A



Blot B

Figure 5.2.3 The *Hin*dIII fragments hybridised to by the radiolabelled DNA *orfZ* fragment and neomycin resistance cassette fragment in the mutant and wild-type genomic digests.





The * shows the extra *Hin*dIII site created by the insertion of the neomycin resistance cassette.

Figure 5.2.4 Photographs of swarm agar plates containing *orfZ* mutants and wild-type.

The bacteria were grown on 0.2 % swarm agar plates and incubated at 28 °C for 48 hours before being photographed.

1. Z MUT 1
2. Z MUT 2
3. Z MUT 3
4. Z MUT 4
5. Z MUT 5
6. Z MUT 6
WT. C58C1 (positive control)

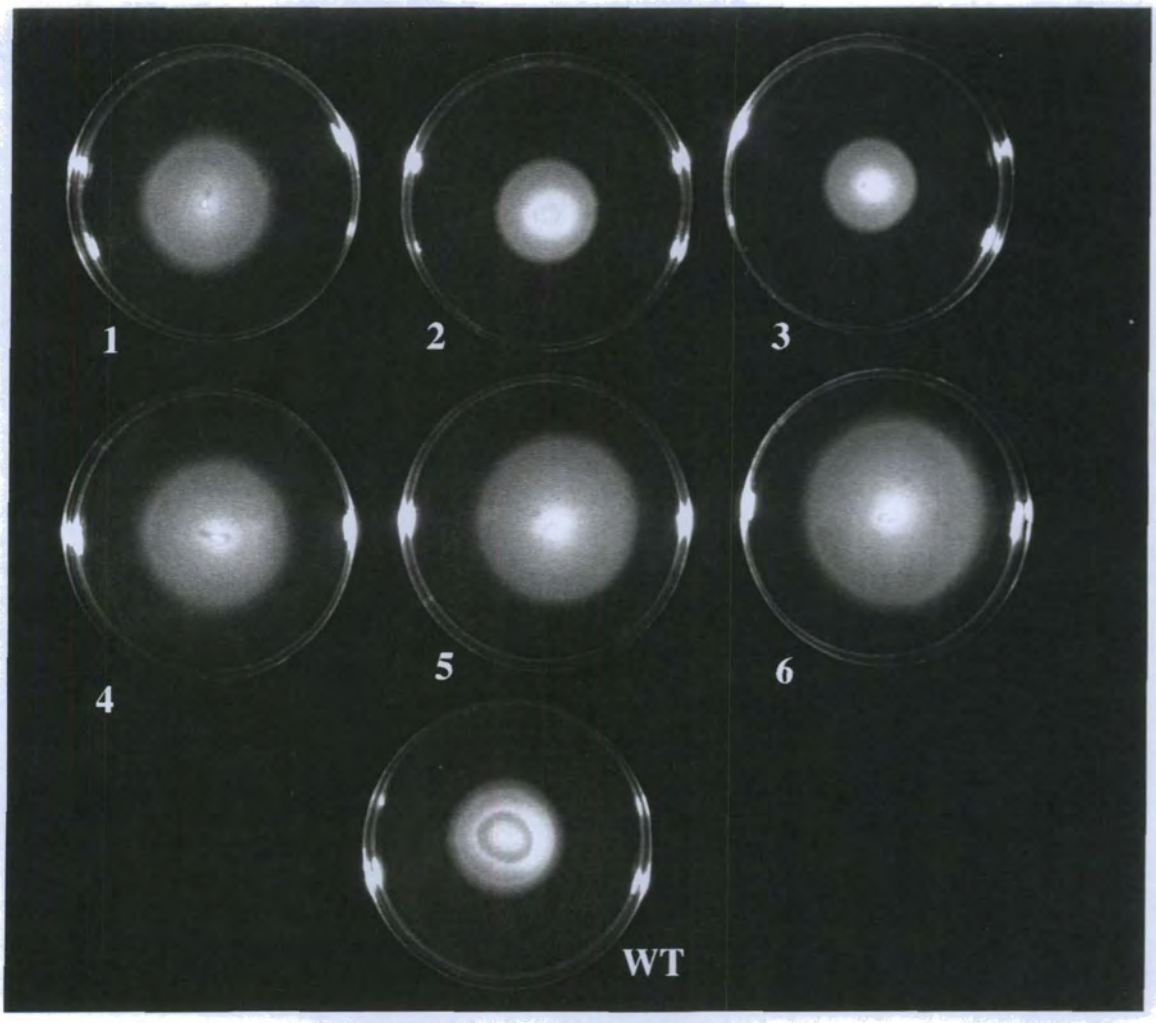
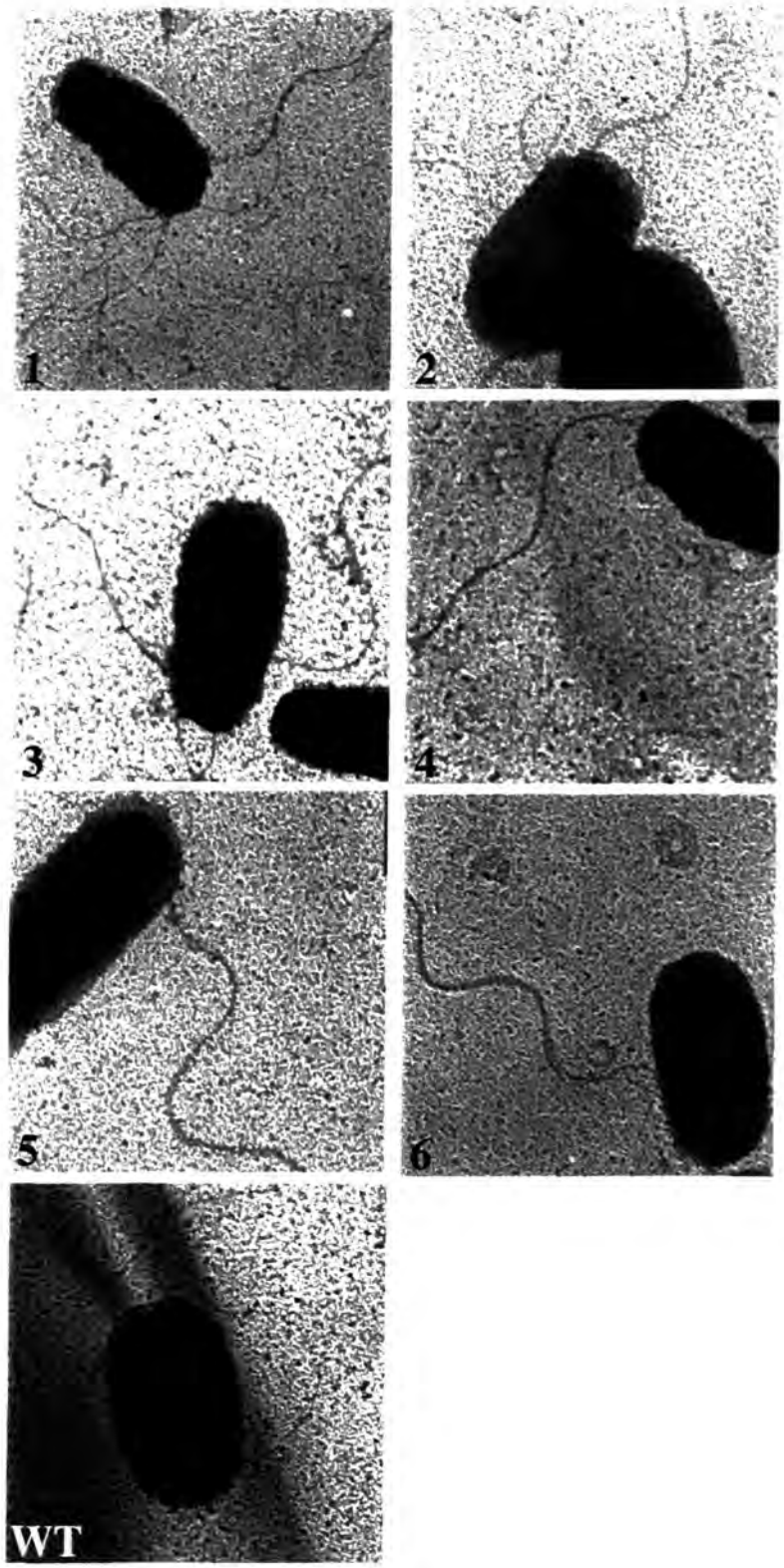

1
2
3
4
5
6
WT

Figure 5.2.5 Electron micrographs of *orfZ* mutants and wild-type.

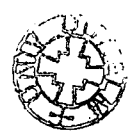
Electron micrographs showing the six *orfZ* mutants and a C58C1 positive control. The cells have been stained with 1 % uranyl acetate. The length of each cell body is approximately 1 μm.

1. Z MUT 1
2. Z MUT 2
3. Z MUT 3
4. Z MUT 4
5. Z MUT 5
6. Z MUT 6
WT. C58C1 (positive control)


1
2
3
4
5
6
WT

Chapter 6

Discussion



The data displayed describes the sequencing and consequent identification of two flagellar gene homologues and two open reading frames encoding potential proteins with no clear homologues in any database. These flagellar gene homologues and unidentified open reading frames were postulated to be transcribed together as an operon, independent of the two previously described operons up and downstream of the region sequenced [Deakin, *et al.*, 1997a; Deakin, *et al.*, 1997b].

TESTCODE predicted coding regions immediately upstream and downstream of *flgF* and *fliI*, and then DNA Strider identified two ORFs in these positions, namely *orfX* and *orfZ*. The position of the flagellar gene homologues, *orfX* and *orfZ* within *A. tumefaciens* suggest that they might be transcribed together. All four ORFs are in close proximity to one another with very little intergenic sequence which is characteristic of an operon. There were bigger regions of predicted non-coding sequence between *motA*/*orfX* and *orfZ*/*flgB*, again suggesting the presence of a separate transcription unit. It is impossible to state categorically where the putative operon starts. As *orfX* had no homology to previously identified genes, predictions of the correct start codon could not be made using comparisons with known genes. The promoter for the putative operon in the region upstream of *orfX* was unable to be located, mainly because the consensus sequence for flagellar operons in *A. tumefaciens* is not yet known and the *E. coli* consensus sequence may not have been sufficiently conserved to be identified in the sequence search. As there was a predicted non-coding region between *orfZ* and *flgB*, the sequence was searched for a possible transcription termination signal. A sequence encoding a potential hairpin-loop was found in the non-coding region that may fulfil this role. All of these findings suggested the presence of a small operon.

Further investigations are needed to determine the presence of this putative operon. Northern blotting experiments could confirm that all three operons, namely, the *motA* operon, *orfX* operon and *flgB* operon, are transcribed individually. Additionally, each operon could be mutated in polar fashion to determine if the others are still transcribed, although this may not work if a mutation in either *flgF* or *fliI* represses the *flgB* operon because FlgF or FliI is required before *flgB* can be transcribed.

The gene products of the *flgF* and *fliI* homologues were found to have significant sequence identity to proteins of several other bacteria following a search of the OWL database. Radiolabelled DNA fragments of *flgF* and *fliI* were found to hybridise to their homologous counterparts in *R. meliloti* within close proximity to one another. Furthermore, recent DNA sequence from *R. meliloti* revealed that they are found in the same order in *R. meliloti* as in *A. tumefaciens* relative to the *motA* and *flgB* operons.

The similarity of FliI in several bacteria to the β subunit of the *E. coli* ATP synthase complex is very interesting from both functional and evolutionary points of view. As was discussed in the introduction, the bacterium flagellum rotates and the energy source for this rotation is the transmembrane proton potential. Only under conditions of anaerobic glycolysis is ATP hydrolysis required, and this occurs via proton-linked membrane ATPase action [Albertini, *et al.*, 1991]. The *A. tumefaciens* FliI homologue also contains the nucleotide binding domains A and B (described by Walker, *et al.*, 1982), found in many other bacteria which in $F_1\beta$ are believed to make major contributions to the nucleotide-binding site of the β subunit. FliI has been shown to bind ATP [Dreyfus, *et al.*, 1993] in *S. typhimurium* and its copy number in the cell has been estimated to be about 1,500, which is much greater than the number of flagella. These both support the view that FliI is an ATPase which provides energy possibly for flagellar subunit export and assembly. The similarity between FliI and virulence proteins of mammalian and plant pathogens suggest the possibility of a superfamily of protein export systems. *spa47* of *Shigella flexneri* is 40 % identical to *fliI* from both *S. typhimurium* and *B. subtilis*. The *spa* region of *S. flexneri* is involved in the surface presentation of invasion plasmid antigens which are exported by a pathway that does not involve signal peptide cleavage [Venkatesan, *et al.*, 1992]. HrpB6 from the plant pathogen *X. campestris* is even more strongly related to FliI. The *hrp* region is needed for the basic pathogenic response of normal plants and the hypersensitive response of resistant plants [Fenselau, *et al.*, 1992]. Whatever the exact function of the FliI protein and its homologues amongst the virulence proteins, it seems possible to be able to conclude that they are components of a superfamily of systems that are used in a variety of ways to control and enable processes of export or assembly, or both, of proteins that lie beyond the cell membrane [Dreyfus, *et al.*, 1993]. A recent review [Yang He, 1998] examines the presence of a novel family of protein secretion systems classified as type III. Many bacterial pathogens responsible for diseases in higher plants, animals and humans contain type III secretion systems, a unique feature of the system being their ability to deliver bacterial virulence proteins directly into host cells. At least eight type III secretion proteins share sequence similarities with flagellar assembly components, therefore, Yang He [1998] speculates that various type III secretion systems may in fact have derived from the presumably more ancient flagellar assembly apparatus in adaptation to the evolution of eukaryotic organisms as hosts. FliI shows sequence similarities to the following type III secretion-associated proteins; YscN (*Yersinia*), HrcN (*Pseudomonas*), HrcN (*Erwinia*), HrcN (*Xanthomonas*), HrcN (*Rostonia*), InvC (*Salmonella*) and Spa47 (*Shigella*). HrcN and its flagellar counterpart, FliI, is a cytoplasmic protein that has several motifs characteristic of ATPases and is a candidate for energizing the assembly of flagella and the type III protein secretion apparatus [Yang He, 1998].

A recent study of FliI, after this work was completed, produced a site-specific mutant of *fliI* which resulted in a non-motile phenotype with no visible flagella [unpublished data]. Therefore, transcription of *fliI* is vital in *A. tumefaciens*, as in other bacteria, for the production of fully functional flagella.

As mentioned in the introduction, the basal body of *E. coli* and *S. typhimurium* consists of four rings and a rod. The proximal rod consists of three proteins, FlgB, FlgC and FlgF. The distal rod (FlgG) connects the proximal rod to the L and P ring proteins. In *E. coli* the axial flagellar proteins have localised consensus sequences at their amino-termini. FlgF from *A. tumefaciens* was also found to contain a high degree of conservation to this consensus sequence ANNLAN. Homma *et al.* [1990] suggested that this consensus does not form part of the recognition sequence of the flagellum-specific export pathway because not all flagella-associated proteins have this sequence. Instead it may be that the region determines a structural motif that is common to rod proteins, or in some way determines the temporal control of export [Zuberi, *et al.*, 1991].

Future work could involve producing a non-polar *flgF* mutant to confirm that its transcription is essential in *A. tumefaciens*, as it has been found to be in other bacteria, for the production of fully functional flagella.

In comparison with the flagellar gene homologues, it is impossible to infer functional roles for the gene products of *orfX* and *orfZ* from their homologies to previously identified proteins. However, they are conserved between species, as was found by probing *R. meliloti* DNA with radiolabelled fragments of *orfX* and *orfZ*. In addition, they are found in the same order in *R. meliloti* and *A. tumefaciens* relative to *flgF* and *fliI*. This would suggest that they are infact coding regions and, hence, genes as non-coding sequence is rarely conserved between two species.

Comparing the *R. meliloti* Orf20 sequence to OrfX highlighted some differences. Orf20 is approximately 100 amino acids bigger than OrfX as it has a GTG start codon upstream of the ATG start codon chosen for OrfX. This start codon was chosen for OrfX based upon the TESTCODE results which showed a non-coding region present between *motA* and *orfX* of approximately 330bp. This region should contain two promoters, one for the *motA* operon and one for the *orfX* operon. The position of the *orf20* start codon means that there is very little intergenic region between *motA* and *orf20*. If this was also true for *orfX*, the promoter for *motA* would have to be within the *orfX* coding region and *vice versa* for the *orfX* promoter. This is possible but seems unlikely given the TESTCODE predictions.

OrfZ and Orf23 are more similar in size, with Orf23 being only slightly bigger than OrfZ. The GTG start codon of OrfZ was chosen as it best fitted the results of the TESTCODE prediction. The mutagenesis work with *orfZ* failed to produce a non-motile phenotype. This may suggest that the gene product of *orfZ* is not essential for a motile phenotype. Alternatively, enough of *orfZ* may have remained before the neomycin resistance cassette insertion site to ensure *orfZ* remained functional.

As has been stated previously, it is impossible to infer functional roles for the gene products of *orfX* and *orfZ*, however, their position within a putative flagellar gene operon suggests the possibility that they may also be involved in flagellar assembly. The majority of previously identified flagellar proteins are all flagellar structural components and as there is a high degree of conservation between flagella structure in a great number of bacteria, it seems unlikely that the gene products of *orfX* and *orfZ* are structural components. Therefore, it seems more probable that OrfX and OrfZ have roles in regulating flagellar gene expression or in the export and assembly of flagellar components as with FliI. Firstly, protein expression studies will be required to show that the gene products of *orfX* and *orfZ* are expressed. Any effects on motility resulting from the over-expression of the *orfX* and *orfZ* gene products may also help define roles for these putative proteins. If either of the proteins is shown to be expressed, then immunocytochemistry could be carried out to localise its position in the cell and, thus, infer a possible insight into its function.

References

Adler, J. 1975. Chemotaxix in bacteria. Annu. Rev. Biochem. **44**: 341-356.

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

Aizawa, S.-I. Flagellar assembly in *Salmonella typhimurium.* Molecular Microbiol. **19:** 1-5.

Aizawa, S.-I., F. Vonderviszt, R. Ishima and K. Akasaka. 1990. Termini of *Salmonella* flagellin are disordered and become organised upon polymerisation into flagellar filament. J. Mol. Biol. **211:** 673-677.

Albertin, A.M., T. Caramori, W.D. Crabb, F. Scoffone and A. Galizzi. 1991. The *flaA* locus of *Bacillus subtilis* is part of a large operon coding for flagellar structures, motility functions, and an ATPase-like polypeptide. J. Bacteriol. **173**: 3573-3579.

Armitage, J. P. 1992. Behavioral responses in bacteria. Annu. Rev. Physiol. **54:** 683-714

Armitage, J. P. 1993. Methylation-independent behavioral responses in bacteria, pp. 43-65. *In* Kurjan and Taylor (ed.), Signal Transduction: Prokaryotic and Simple Eukaryotic Systems. Academic Press Inc., San Diego.

Armitage, J.P., W. A. Havelka and R. E. Sockett. 1990. Methylation-independent taxis in bacteria, pp. 177-197. *In* Armitage and Lackie (ed.), Molecular Biology of the Chemotactic Response. Cambridge University Press, Cambridge.

Ashby, A. M. 1988. Chemotaxis and crop protection. Ph.D. Thesis, University of Durham.

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

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

Atsumi, T. L., L. McCarter and Y Imae. 1992. Polar and lateral flagellar motors of marine *Vibrio* are driven by different ion-motive forces. Nature. **335:** 182-184

Berg, H. C., and R. A. Anderson. 1973. Bacteria swim by rotating their flagella filaments. Nature. **245:** 380-382

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

Bernaerts, M.J. and J. De Ley. 1963. A biochemical test for crown gall bacteria. Nature. **197**: 406-407.

Bhuvaneswari, T. V., A. Bhagwat and W. D. Bauer. 1981. Transient susceptibility of root cells in four common legumes to nodulation by rhizobia. Plant Physiol. **68:** 1114-1119

Binns, A. N. and M. F. Thomashow. 1988. Cell biology of *Agrobacterium* infection and transformation of plants. Annu. Rev. Microbiol. **42:** 575-606

Bischoff, D. S. and G. W. Ordal. 1992. Identification and characterization of FliY, a novel component of the *Bacillus subtilis* flagellar switch complex. Mol. Microbiol. **6:** 2715-2723

Block, S. M., J. E. Segall and H. C. Berg. 1982. Impulse responses in bacterial chemotaxis. Cell. **31:** 215-226

Bollinger, J., C. Park, S. Harayama and G. L. Hazelbauer. 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

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

Brown, A. P. 1992. Molecular characterisation of behavioural functions in *Agrobacterium tumefaciens*. Ph.D. Thesis, University of Durham.

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

Carpenter, P.B., D.W. Hanlon and G.W. Ordal. 1992. *flhF*, a *Bacillus subtilis* flagellar gene that encodes a putative GTP-binding protein. Mol. Microbiol. **6**: 2705-2713.

Currier, W. W. and G. A. Strobel. 1976. Chemotaxis of *Rhizobium* spp. to plant exudates. Plant Physiol. **57:** 820-823

Deakin, W.J. 1994. Molecular characterisation of flagellar genes in *Agrobacterium tumefaciens*. Ph.D. Thesis, University of Durham.

Deakin, W.J., C.S. Furniss, V.E. Parker, and C.H. Shaw. 1997a. Isolation and characterisation of a linked cluster of genes from *Agrobacterium tumefaciens* encoding proteins involved in flagellar basal-body structure. Gene. **189**: 135-137.

Deakin, W.J., J.L.C.I. Sanderson, T. Goswami and C.H. Shaw. 1997b. The *Agrobacterium tumefaciens* motor gene, *motA* , is in a linked cluster of genes with the flagellar switch protein genes, *fliG*, *fliM* and *fliN*. Gene. **189**: 139-141.

de Weger, L. A., C. I. M. Van de Vlugt, A. H. M. Wijfjes, P. A. H. M. Bakker, B. Schippers and B. Lugtenberg. 1987. Flagella of a plant growth stimulating *Pseudomonas fluorescens* strain are required for colonization of plant roots. J. Bacteriol. **169:** 2769-2773

Dhaese, P., H. De Greve, H. Decraemer, J. Schell and M. Van Montagu. 1979. Rapid mapping of transposon insertion and deletion mutants in the large Ti-plasmids of *Agrobacterium tumefaciens*. Nucl. Acids Res. **7**: 1837-1849.

Ditta, G., S. Stanfield, D. Corbin and D.R. Helsinki. 1980. A broad host range DNA cloning system for Gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA. **77**: 7347-7351.

Dreyfus, G., A. W. Williams, I. Kawagishi and R. M. Macnab. 1993. Genetic and biochemical analysis of *Salmonell typhimurium* FliI, a flagellar protein related to the catalytic subunit of the F_OF_1 ATPase and to virulence proteins of mammalian and plant pathogens. J. Bacteriol. **175:** 3131-3138

Driks, A., R. Bryan, L. Shapiro and D. J. DeRosier. 1989. The organization of the *Caulobacter crescentus* flagellar filament. J. Mol. Biol. **206:** 627-636

Eisenbach, M. 1990. Functions of the flagellar modes of rotation in bacterial motility and chemotaxis. Mol. Microbiol. **4:** 161-167

Francis, N. R., G. E. Sosinsky, D. Thomas and D. J. DeRosier. 1994. Isolation, characterization and structure of bacterial flagella motors containing the switch complex. J. Mol. Biol. **235:** 1261-1270

Galan, J. E., C. Ginocchio and P. Costeas. 1992. Molecular and functional characterization of the *Salmonella* invasion gene *invA*: homology of InvA to members of a new protein family. J. Bacteriol. **174:** 4338-4349

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

Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman and C.I. Kado. 1985. Positive selection for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. **164:** 918-921.

Gillen, K. L. and K. T. Hughes. 1991. Molecular characterization of *flgM*, a gene encoding a negative regulator of flagellin synthesis in *Salmonella typhimurium.* J. Bacteriol. **173:** 6453-6459

Gray, J., J. Wang and S. B. Gelvin. 1992. Mutation of the *miaA* gene of *Agrobacterium tumefaciens* results in reduced *vir* gene expression. J. Bacteriol. **174:**1086-1098

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

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**: 557-580.

Hawes, M. C. 1990. Living plant cells released from the root cap: a regulator of microbial populations in the rhizoshere? Plant Soil. **129:** 19-27

Hawes, M. C. and L. Y. Smith. 1989. Requirement for chemotaxis in pathogenicity of *Agrobacterium tumefaciens* on roots of soil-grown pea plants. J. Bacteriol. **171:** 5668-5671

Hawes, M. C., L. Y. Smith and A. J. Howarth. 1988. *Agrobacterium tumefaciens* mutants deficient in chemotaxis to root exudates. Mol. Plant-Microbe Interact. **1:** 182-186

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

Herrera-Estrella, A., M. Van Montagu and K. Wang. 1990. A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of *Agrobacterium* VirD2 protein directs ß-galactosidase fusion protein into tobacco nuclei. Proc. Nat. Acad. Sci. USA. **87:** 9534-9537

Higgins, D. G. and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene. **73:** 237-244

Homma, M., T. Iino, K. Kutsukake and S. Yamaguchi. 1986. *In vitro* reconstitution of flagellar filaments onto hooks of filamentless mutants of *Salmonella typhimurium* by addition of hook-associated proteins. Proc. Natl. Acad. Sci. USA. **83:** 6169-6173

Hooykaas, P.J.J. and A.G.M. Beijersbergen. 1994. The virulence system of *Agrobacterium tumefaciens.* Annual Review of Phytopathology. **32:** 157-179

Huang, M. W., G. A. Cangelosi, W. Halperin and E. W. Nester. 1990. A chromosomal *Agrobacterium tumefaciens* gene required for effective plan signal transduction. J. Bacteriol. **172:** 1814-1824

Hughes, K. T., K. L. Gillen, M. J. Semon and J. E. Karlinsey. 1993. Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator. Science. **262:** 1277-1280

Ikeda, T., K. Oosawa and H. Hotani. 1996. Self-assembly of the filamentcapping protein, FliD, of bacterial flagella into an annual structure. J. Mol. Biol. **259**: 679-686

Irikura, V. M., M. Kihara, S. Yamaguchi, H. Sockett and R. M. Macnab. 1993. *Salmonella typhimurium fliG* and *fliN* mutations causing defects in assembly, rotation, and switchin of the flagellar motor. J. Bacteriol. **175:** 802-810

Jones, C. J., M. Homma and R. M. Macnab. 1989. L-, P-, and M-ring proteins of the flagellar basal body of *Salmonella typhimurium:* gene sequences and deduced protein sequences. J. Bacteriol. **171:** 3890-3900

Jones, C. J. and R. M. Macnab. 1990. Flagellar assembly in *Salmonella typhimurium:* analysis with temperature-sensitive mutants. J. Bacteriol. **172:** 1327-1339

Kawagishi, I., V. Muller, A. W. Williams, V. M. Irikura and R. M. Macnab. 1992. Subdivision of flagellar region III of the *Escherichia coli* and *Salmonella typhimurium* chromosomes and identification of two additional flagellar genes. J. Gen. Microbiol. **138:** 1051-1065

Kawagishi, I., M. Homma, A.W. Williams and R.M. Macnab. 1996. Characterization of the Flagellar Hook Length Control Protein FliK of *Salmonella typhimurium* and *Escherichia coli.* J. Bacteriol. **178**: 2954-2959.

Khan, I. H., T. S. Reese and S. Khan. 1992. The cytoplasmic component of the bacterial flagellar motor. Proc. Natl. Acad. Sci. USA. **89:** 5956-5960

Koncz, C., N. Martini, R. Mayerhofer, Z. Koncz-Kalman, H. Korber, G. P. Redei and J. Schell. 1989. High-frequency T-DNA mediated gene tagging in plants. Proc. Natl. Acad. Sci. USA. **86:** 8467-8471

Kubori, T., N. Shimamoto, S. Yamaguchi, K. Namba and S.-I. Aizawa. 1992. Morphological pathway of flagellar assembly in *Salmonella typhimurium.* J. Mol. Biol. **226:** 433-446

Kutsukake, K., Y. Ohya and T. Iino. 1990. Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium.* J. Bacteriol. **172:** 741-747

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

Lapidus, I. R., M. Welch and M. Eisenbach. 1988. Pausing of flagellar rotation is a component of bacterial motility and chemotaxis. J. Bacteriol. **170:** 3627-3632

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

Lippincott, B. B. and J. A. Lippincott. 1969. Bacterial attachment to a specific wound site is an essential stage in tumour initiation by *Agrobacterium tumefaciens.* J. Bacteriol. **97:** 620-628

Loake, G. J. 1989. The genetic dissection of chemotaxis in *Agrobacterium tumefaciens.* Ph.D. Thesis, University of Durham.

Loake, G.J., A. M. Ashby and C. H. Shaw. 1988. Attraction of *Agrobacterium tumefaciens* C58C1 towards sugars involves a highly sensitive chemotaxis system. J. Gen. Microbiol. **134:** 1427-1432

Lukat, G. S., A. M. Stock and J. B. Stock. 1990. Divalent metal ion binding to the CheY protein and its significance to phosphotransfer in bacterial chemotaxis. Biochem. **29:** 5436-5442

Macnab, R. M. 1987. Flagella, pp. 70-83. *In* Neidhart, Ingraham, Low, Magasanik, Schaechter and Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology. American Society for Microbiology, Washington, D.C.

Macnab, R. M. 1987. Motility and Chemotaxis, pp. 732-759. *In* Neidhart, Ingraham, Low, Magasanik, Schaechter and Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology. American Society for Microbiology, Washington, D.C.

Macnab, R. M. 1990. Genetics, structure, and assembly of the bacterial flagellum, pp. 77-106. *In* Armitage and Lackie (ed.), Biology of the Chemotactic Response. Cambridge University Press, Cambridge.

Macnab, R. M. 1992. Genetics and biogenesis of bacterial flagella. Annu. Rev. Genet. **26:** 131-158

Macnab, R. M. and M. K. Ornston. 1977. Normal to curly flagellar transitions and their role in bacterial tumbling: stabilization of an alternative quarternary structure by mechanical force. J. Mol. Biol. **112:** 1-30

Manson, M.D., J.P. Armitage, J.A. Hoch and R.M. Macnab. 1998. Bacterial locomotion and signal transduction. J. Bacteriol. 1009-1022.

Manson, M. D., P. Tedesco, H. C. Berg, F. M. Harold and C. Van der Drift. 1977. A proton motive force drives bacterial flagella. Proc. Acad. Sci. USA. **74:** 3060-3064

Matthysse, A. G. 1983. Role of bacterial cellulose fibrils in *Agrobacterium tumefaciens* infection. J. Bacteriol. **154:** 906-915

McCarter, L. M. and M. E. Wright. 1993. Identification of genes encoding components of the swarmer cell flagellar motor and propeller and a sigma factor controlling differentiation of *Vibrio parahaemolyticu.* J. Bacteriol. **175:** 3361-3371

Melchers, L. S. and P. J. J. Hooykaas. 1987. Virulence of *Agrobacterium,* pp. 167-220. *In* Miflin (ed.), Oxford Surveys of Plant Molecular and Cell Biology. Oxford University Press, Oxford.

Morgan, D. G., R. M. Macnab, N. R. Francis and D. J. DeRosier. 1993. Domain organization of the subunit of the *Salmonella typhimurium* flagellar hook. J. Mol. Biol. **229:** 79-84

Muller, V., C. J. Jones, I. Kawagishi, S. -I. Aizawa and R. M. Macnab. 1992. Characterization of the *fliE* genes of *Escherichia coli* and *Salmonella typhimurium* and identification of the FliE protein as a component of the flagellar hook-basal body complex. J. Bacteriol. **174:** 2298-2304

Nester, E. W. and T. Kosuge. 1981. Plasmids specifying plant hyperplasias. Annu. Rev. Microbiol. **35:** 531-565

New, P. B. and A. Kerr. 1972. Biological control of crown gall: field measurements and glasshouse experiments. J. Appl. Bacteriol. **35:** 279-285

Ohnishi, K., Y. Ohto, S. -I. Aizawa, R. M. Macnab and T. Iino. 1994. FlgD is a scaffolding protein needed for flagellar hook assembly in *Salmonella typhimurium.* J. Bacteriol. **176:** 2272-2281

Ordal, G. W., L. Marquez-Magana and M. J. Chamberlin. 1993. Motility and chemotaxis, pp. 765-784. *In* Sonenshein, Hoch and Losick (ed.), *Bacillus subtilis* and other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics. American Society for Microbiology, Washington, D. C.

Packer, H. L. and J. P. Armitage. 1993. The unidirectional flagellar motor of *Rhodobacter sphaeroides* WS8 can rotate either clockwise or counterclockwise: characterization of the flagellum under both conditions by antibody decoration. J. Bacteriol. **175:** 6041-6045

Packer, H. L. and J. P. Armitage. 1994. The chemokinetic and chemotactic behaviour of *Rhodobacter sphaeroides:* two independent response. J. Bacteriol. **176:** 206-212

Packer, H.L., D.E. Gauden and J.P Armitage. 1996. The behavioural response of anaerobic *Rhodobacter sphaeroides* to temporal stimuli. Microbiol. **142**: 593-599.

Pugsley, A. P. 1993. The complete general secretory pathway in Gram-negative bacteria. Microbiol. Rev. **57:** 50-108

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

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. **239**: 487.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular Cloning: A laboratory manual. Cold Spring Harbour Laboratory Harbour Press, New York.

Sar, N., L. McCarter, M. Simon and M. Silverman. 1990. Chemotactic control of the two flagellar systems of *Vibrio parahaemolyticus.* J. Bacteriol. **172:** 334-341

Schroth, M. N., A. R. Weinhold, A. H. McCan, D. C. Hildebrand and N. Ross. 1971. Biology and control of *Agrobacterium tumefaciens.* Hilgardia. **40:** 537-552

Shaw, C. H. 1990. Swimming against the tide: chemotaxis in *Agrobacterium.* Bioessays. **13:** 25-29

Shaw, C. H., G. J. Loake, A. P. Brown, C. S. Garrett, W. Deakin, G. Alton, M. Hall, S. A. Jones, M. O'Leary and L. Primavesi. 1991. Isolation and characterisation of behavioural mutants and genes of *Agrobacterium tumefaciens*. J. Gen. Microbiol. **137:** 1939-1953

Sheng, J. and V. Citovsky. 1996. Agrobacterium-plant cell DNA transport: Have virulence proteins, will travel. The Plant Cell. **8:** 1699-1710

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

Smit, G., T.J.J. Logman, M.E.T.I. Boerrigter, J.W. Kijne and B.J.J. Lugtenberg. 1989 Purification and partial characterization of the *Rhizobium leguminosarum* biovar *viciae* Ca^{2+} dependent adhesin, which mediates the first step in attachment of cells of the family *Rhizobiaceae* to plant root hair tips. J. Bacteriol. **171**: 4054-5062.

Soby, S. and K. Bergman. 1983. Motility and chemotaxis of *Rhizobium meliloti* in soil. Appl. Env. Microboil. **46**: 995-998.

Sockett, H., S. Yamaguchi, M. Kihara, V. M. Irikura and R. M. Macnab. 1992. Molecular analysis of the flagellar switch protein FliM of *Salmonella typhimurium.* J. Bacteriol. **174**: 793-806.

Sosinsky, G.E., N.R. Francis, M.J.B. Stallmeyer and D.J. DeRosier. 1992. Substructure of the flagellar basal body of *Salmonella typhimurium.* J. Mol. Biol. **223**: 171-184.

Southern,. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**: 503.

Stephens, C.M. and L. Shapiro. 1993. An unusual promoter controls cell-cycle regulation and dependence on DNA replication of the *Caulobacter fliLM* early flagellar operon. Mol. Microbiol. **9**: 1169-1179.

Stolz, B. and H.C. Berg. 1991. Evidence for interactions between MotA and MotB, torque-generating elements of the flagellar motor of *Escherichia coli.* J. Bacteriol. **173**: 7033-7037.

Taylor, B. L. 1983. Role of proton motive force in sensory transduction in bacteria. Annu. Rev. Microbiol. **37**: 551-573.

Taylor, B. L. and M.S. Johnson. 1993. Universal themes of signal transduction in bacteria, pp. 3-15. *In* Kurjan and Taylor (ed.), Signal Transduction: Prokaryotic and Simple Eukaryotic Systems. Academic Press, Inc., San Diego.

Taylor, B. L. and J.W. Lengeler. 1990. Transductive coupling by methylated transducing proteins and permeases of the phosphotransferase system in bacterial chemotaxis, pp. 69-90. *in* Aloia, Curtain and Gordon (ed.), Membrane Transport and Information Storage. Alan R. Liss Inc., New York.

Taylor, B.L. and I.B. Zhulin. 1998. In search of higher energy: metabolism-dependent behaviour in bacteria. Mol. Microbiol. **28**: 683-690.

Tempe, J. and A. Goldman. 1982. Occurence and biosynthesis of opines, pp. 428-450. *In* Kahl and Schell (ed.), Molecular Biology of Plant Tumours. Academic Press Inc., New York.

Tisa, L.S. and J. Adler. 1992. Calcium ions are involved in *Escherichia coli* chemotaxis. Proc. Natl. Acad. Sci. USA. **89**: 11804-11808.

Tisa, L.S., B.M. Olivera and J. Adler. 1993. Inhibition of *Escherichia coli* chemotaxis by ω-conotoxin, a calcium ion channel blocker. J. Bacteriol. **175**: 1235-1238.

Venkatesan, M.M., J.M. Buysse and E.V. Oaks. 1992. Surface presentation of *Shigella flexneri* invasion plasmid antigens requires the products of the *spa* locus. J. Bacteriol. **174**: 1990-2001.

Vogler, A.P., M. Homma, V.M. Irikura and R.M. Macnab. 1991. *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F_0F_1, vacuolar, and archaebacterial ATPase subunits. J. Bacteriol. **173**: 3564-3572.

Walker, J.E., M. Saraste, M.J. Runswick and N.J. Gay. 1982. Distantly related sequences of the alpha and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**: 945-951

Willmitzer, L., P. Dhaese, P.H. Schreier, W. Schmalenbach, M. Van Montagu and J. Schell. 1983. Size, location and polarity of T-DNA encoded transcripts in nopaline crown gall tumours: common transcripts in octopine and nopaline tumours. Cell. **32**: 1045-1056.

Yamamoto, K. and Y. Imae. 1993. Cloning and characterisation of the *Salmonella typhimurium*-specific chemoreceptor Tcp for taxis to citrate and from phenol. Proc. Natl. Acad. Sci. USA. **90**: 217-221.

Yang He, S. 1998. Type III protein secretion systems in plant and animal pathogenic bacteria. Annu. Rev. Phytopathol. **36**: 363-92.

Ziegler, R.J., C. Peirce and K. Bergman. 1986. Mapping and cloning of a *fla-che* region of the *Rhizobium meliloti* chromosome. J. Bacteriol. **168**: 785-790.