Durham University

Durham E-Theses

Molecular characterisation of flagellar genes from agrobacterium tumefaciens

William James Deakin

How to cite:

Deakin, William James (1994) Molecular characterisation of flagellar genes from agrobacterium tumefaciens. Doctoral thesis, Durham University.

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a https://etheses.durham.ac.uk/id/eprint/5858/ is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
https://etheses.durham.ac.uk

Molecular characterisation of flagellar genes from *Agrobacterium tumefaciens*

by

William James Deakin.

A thesis submitted to the Department of Biological Sciences

University of Durham

In accordance with the requirements for the degree of

Doctor of Philosophy

September 1994

The copyright of this thesis rests with the author.
No quotation from it should be published without
his prior written consent and information derived
from it should be acknowledged.

13 JAN 1995

For My Parents

Declaration

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

The copyright of this thesis rests with the author. No quotation from it should be published without prior written consent and the information derived from it should be acknowledged.

Molecular characterisation of flagellar genes from *Agrobacterium tumefaciens.*

William James Deakin

PhD 1994

Abstract

Three behavioural mutants of *A. tumefaciens* C58C1 (*mot-1*, *mot-12* and *fla-15*) generated by transposon (Tn5) mutagenesis were studied. Analysis was initially at the molecular level, as a cosmid, pDUB1900, from a representative genomic library of C58C1 had been isolated that complemented the mutants. A region of 8624 nucleotides to which the Tn5 insertion sites of the three mutants had been mapped was sequenced completely in both directions. The comparison of this sequence with sequence databases and other computer analyses revealed six flagellar gene homologues (*flgI*, *flgH*, *fliP*, *flaA*, *flaB*, *flaC*), three open reading frames (ORFA, B and C) with no significant sequence identity to any open reading frames in the databases and the partial sequence of the flagellar gene homologue *flgG*. Computer analysis also showed that the *flgH*, *flgI* and *fliP* homologues, and ORFs A, B and C, could form the downstream region of a larger operon involved in chemotactic and motility functions. However putative transcription signals were also found within the operon.

A new mutant (MAN1) was created in the last gene (*fliP*) of the putative operon to investigate the function of possible transcription signals in the open reading frame immediately upstream of it (ORFC). The *mot-12* mutant phenotype of fully synthesised but paralysed flagella is brought about by the insertion of Tn5 in ORFC. ORFC contains a possible promoter for *fliP*. The Tn5 insertion in ORFC should have polar effects upon the expression of *fliP*, unless the putative promoter can cause expression of *fliP*. The MAN1 mutant had a flagella-less phenotype. FliP in other bacteria is required early in the synthesis of flagella and the null phenotype is *fla-*. Thus for flagella to be present in *mot-12* suggests *fliP* must have a promoter. The ORFC sequence is highly conserved in *R. meliloti* and the overall regulation of these flagellar gene homologues may be as an operon with other regulatory signals. Evidence from other operons (including motility operons) with multiple transcription signals is discussed.

The *flaABC* homologues were multiple copies of the gene encoding the flagellin protein of the flagellum. The *mot-1* phenotype of severely truncated filaments was caused by a Tn5 insertion in *flaA*. Analysis of the sequence showed *flaABC* to each have

transcription signals that could lead to separate transcription. Transcription analysis by Northern blotting showed *flaA* to be transcribed monocistronically. Flagella were isolated from *A. tumefaciens* and the flagellins separated by SDS-PAGE. The migrated distances (relative to those of markers) was not as predicted from the nucleotide sequence. This anomaly could be caused by unequivalent binding of SDS or post-translational modification of FlaA. The *A. tumefaciens* flagellar genes were most similar to those of *R. meliloti.* However *A. tumefaciens* flagella do not exhibit the characteristic cross-hatching of the complex flagella of *R. meliloti.* This study also showed *A. tumefaciens* flagella not to be dependent on divalent cations for subunit associations unlike *R. meliloti.* These properties of *A. tumefaciens* flagella were similar to those of *R. leguminosarum.*

The open reading frames found were isolated, radiolabelled and used as probes against Southern blots containing chromosomal DNA from a variety of soil bacteria, and cosmids known to contain motility genes in *R. meliloti.* Hybridisation revealed homologous DNA sequences in a number of these bacteria. All the *A. tumefaciens* open reading frames hybridised to homologous DNA in *R. meliloti* and are found in the same order in both species. This suggests that there are similarities at the molecular level in motility and chemotaxis functions between *R. meliloti* and *A. tumefaciens* as well as in the patterns of chemotaxis and motility observed previously.

Acknowledgements

I thank my supervisor Dr. Charlie Shaw for his support and encouragement during my PhD. I also thank SERC for providing funding during the three years of my study.

I am grateful to Dr. Adrian Brown and Dr. Elaine Broomfield for proof reading this thesis and encouragement whilst writing up.

Thanks also to Mrs Julia Bartley for operating the DNA Sequencer and much helpful advice, Mrs Christine Richardson for help with the electron microscopy, Terry Gibbons for technical support and assistance and to Dr. Mike Richardson for useful discussions concerning the abnormal SDS-PAGE migration patterns.

I consider myself lucky to have worked in the chemotaxis and motility research field and I thank the following members of this research community for their help: Drs. Kostia Bergman and Turlough Finan for helpful discussions and communicating unpublished results; Dr. Rudy Schmitt for providing strains and advice on their use; Drs. Judy Armitage and Jayne Robinson for many useful discussions; Professor Bob Macnab for communicating unpublished results, advice and searching his chemotaxis and motility gene database with ORFs A, B and C.

I am also grateful for the gift of the pJQ200 plasmids from Dr. Michael Hynes used in the gene replacement mutagenesis and for advice about their use.

Thanks to all the members of the Playgroup past and present, Mark, Caroline, Eleanor, Mohamed, Adrian, Adrian, Martin, Martin, Bert, Martin, Ivan and Donna for their help, encouragement, general abuse and for helping to pass the time. I am very grateful to Miss Caroline Furniss for too many things to list.

Finally I must thank my mum and dad for all their support (and a "bit" of money!) during my education and especially this PhD, without whom it would never have been started let alone finished.

Abbreviations

A_{260}	=	absorbance at 260nm
A_{280}	=	absorbance at 280nm
Amp	=	ampicillin
ATP	=	adenosine triphosphate
bp	=	base pairs
BSA	=	bovine serum albumin
cAMP	=	cyclic adenosine monophosphate
CAP	=	catabolite activator protein
CCW	=	counterclockwise
CTP	=	cytosine triphosphate
CW	=	clockwise
dNTP	=	deoxyribonucleoside triphosphate
DTT	=	dithiothreitol
EDTA	=	ethylenediaminetetraacetic acid
Gm	=	gentamyicn
kb	=	kilobase pairs
kDa	=	kilodaltons
MOPS	=	3-(*N*-morpholino)propane sulphonic acid
M_r	=	molecular weight
PEG	=	polyethylene glycol
Rif	=	rifampicin
RNAase	=	ribonuclease
SDS	=	sodium dodecyl sulphate
Sm	=	streptomycin
Tc	=	tetracycline
TEMED	=	*N,N,N',N'*,-tetramethylethylenediamine
Tris	=	tris(hydroxymethyl)aminomethane
X-gal	=	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
UV	=	ultraviolet
wrt	=	with respect to
5'	=	5' terminal phosphate of DNA molecule
3'	=	3' terminal hydroxyl of DNA molecule

Contents

Figures

1. Introduction

UNIV LOUNGE

Chemotaxis in *Escherichia coli* - a model system

1.1 Motility and behaviour of *E. coli*

The motility of *E. coli* cells is brought about by the rotation of flagella [24, 272]. A cell possesses between 0-15 peritrichous flagella (typically 8) each consisting of a helical filament that is rotated at its base to drive the cell forward. The length of the filament can be between 0-20μm (upto ~10 cell body lengths) but is usually 5-10μm [177, 178]. The motor is anchored in the cell envelope and uses the transmembrane proton potential as the energy source for its rotation [163, 185]. The MotA protein, a component of the motor, has been shown to be an ion channel that allows protons to cross membranes [34, 35]. The structure and assembly of the *E. coli* flagellum (and that of the closely related bacterium *Salmonella typhimurium*) will be discussed in section 1.2. The rotation of the flagellar motor is reversible and it continuously switches between the clockwise (CW) and counterclockwise (CCW) directions of rotation [272]. It is this continuous switching that accounts for the behaviour of *E. coli* in a uniform medium. Chemotaxis is caused by changes in the probability for either CW or CCW rotation [164].

CCW rotation of the flagella motor causes the individual filaments to adopt a left-handed structure before they coalesce to form a rotating bundle that drives the bacterium forward in a "run" [164, 178]. In the absence of any chemotactic gradients a run lasts for 1-2 seconds and is punctuated by a "tumble" lasting 0.1-0.2 seconds which re-orientates the cell randomly [25]. A tumble is brought about by the switching of some or all of the flagella to CW rotation. The switch to CW rotation causes a wave of conformational change from the cell along the filament to form a right-handed helical structure. The flagellar bundle is disrupted by the conformational change and individual filaments can be seen in transition between left- and right-handed helices. It is this transition between helical forms that results in the tumble on CW rotation. The brevity of CW rotation prevents full transformation to right-handed filaments - a less stable configuration. Full conversion to right-handed filaments can result in a run also, although because of the inherent instability of this form the run is slower and not as smooth [78, 178, 182].

The flagellar motor can also pause resulting in a stationary flagellum [272]. The frequency and duration of pausing has been shown to be affected by chemotactic stimuli and is hence postulated to be a component of the chemotactic process [162]. Pausing is possibly the result of incomplete switching events and so could play some part in the tumbling mechanism. A pause may also prevent the complete conformational change to the less stable structure of right-handed helices [78, 79, 162].

In the absence of chemical stimuli the continuous switching and pausing of the flagella motor results in a three dimensional random walk [25]. If the cell is exposed to a chemoeffector gradient, the suppression of tumbling when the cell is moving in a favourable direction leads to a biased random walk with net progress [26].

Chemoattractants for *E. coli* include certain amino acids, sugars, carboxylic acids and oligopeptides [3]. As well as oxygen, fumarate and nitrates which can act as electron acceptors in the respiratory chain [298]. Examples of repellents are hydrophobic amino acids, indole, ethanol, glycol and inorganic ions such as Co^{2+}, Ni^{2+} and S^{2-} [178]. Chemotaxis also occurs to variations in pH and temperature [178].

E. coli uses a temporal sensing of chemical stimuli to bring about chemotaxis [181, 262]. The bacterial cell is too small to detect spatial gradients of chemoeffectors accurately. The temporal sensing mechanism uses a simple memory to control the probability of CW rotation. Thus by continuously measuring chemoeffector concentrations and comparing them to concentrations in the immediate past, if the present concentrations are more favourable, CW rotation will be suppressed allowing the run to continue. However, the bacteria will eventually adapt to the new environment and return to the normal probability of tumbling. Thus the bacteria will be able to respond optimally to any subsequent changes in environmental conditions [298].

There are at least three systems for chemoreception in *E. coli* that must link together in the central signalling pathway of chemotaxis, these are the aerotactic response and taxis to other electron acceptors (section 1.2), the methylation independent behavioural response to carbohydrates transported by the phosphoenolpyruvate dependent phosphotransferase system (PTS) (section 1.3) and methylation dependent chemotaxis (section 1.4).

1.2 The aerotactic response

A behavioural response to oxygen was one of the earliest recognised [297]. The movement towards oxygen and alternative electron acceptors such as fumarate and nitrate is likely to be very important in the natural environment of bacteria. The electron transport chain which these compounds affect, provides the underlying regulation of the electrochemical proton gradient - essential for general cell integrity as well as motility [13]. Although there are specific binding proteins for these chemoattractants, the tactic response (at least for oxygen) is brought about by the metabolic use of the electron acceptor within the electron transport chain. Since proton translocation across the cell membrane is a direct result of electron transport, the actual stimulus is probably the perturbation of the proton

motive force [178]. The means by which the stimulus is signalled to the flagellar motor has yet to be identified [270]. For more detailed reviews of aerotaxis in *E. coli* see [13, 14, 297].

Although *E. coli* does not exhibit physiologically significant phototaxis, this response in photosynthetic bacteria is similarly stimulated by the proton motive force as a result of light-driven electron transport [178]. Phototaxis has been most thoroughly investigated in the photosynthetic bacterium *Rhodobacter sphaeroides*, for reviews see [12, 13, 14, 178].

1.3 Chemotaxis to PTS carbohydrates

The PTS is responsible for the uptake of a large number of carbohydrates, which can also act as chemoattractants [238]. The response is not mediated by the methyl-accepting chemotaxis proteins (see section 1.5) and is thus methylation-independent [299]. Instead a family of proteins, enzyme IIs (EII) recognise specific carbohydrates and transport them into the cell [13]. Extensive metabolism of the carbohydrate is not required for the chemotactic response, only its uptake and phosphorylation (whilst being translocated into the cell) are necessary to generate the signal [166, 299]. The EII protein acts in conjunction with the protein, enzyme III (EIII). A subclass of larger EII proteins also exists with an enzyme III-like carboxy-terminal domain. The EII-EIII protein complex must itself be phosphorylated to bring about the tranlocation, and phosphorylation, of the carbohydrates [299]. Phosphorylation of the EII-EIII complex results from the sequential transfer of a phosphoryl group from phosphoenolpyruvate to the protein, enzyme I (EI) and then to histidine-protein (HPr) which in turn phosphorylates EII-EIII [13]. The mechanism by which the chemotactic signal is relayed to the flagellar motor is still under investigation. However the phosphorylation levels of HPr are postulated to affect those of the CheA-CheY pathway [105, 299] whose interactions with the flagellar motor are understood (see section 1.7). The link between the two systems may be brought about by a, as yet hypothetical, phosphoryl-chemotaxis-protein (PCP) that responds to the phosphorylation levels of a PTS intermediate (possibly HPr) and relays these signals to either CheA or CheY [13, 166]. For more detailed reviews of chemotaxis towards PTS carbohydrates see [12, 13, 299].

Despite its importance in the natural environment of bacteria, the chemotactic responses to PTS attractants are much weaker than those elicited by the methylation-dependent pathway [317].

1.4 Methylation-dependent chemotaxis - an overview

A chemotactic response involves the recognition of a chemoeffector by a receptor at the cell surface, which generates a rapid excitation phase (affecting motor direction) and then an adaptive phase, to allow the resumption of a random walk. For methylation-dependent chemotaxis, the receptors are involved in the initiation of the excitation phase *via* the cytoplasmic signalling pathway (see section 1.7) and are responsible for adaptation, which is brought about by their methylation (see section 1.6).

The cell surface receptors are a group of proteins, the methyl accepting chemotaxis proteins (MCPs), which continuously monitor the cell's environment. Any stimulus is transferred to a protein kinase CheA, *via* an auxiliary protein CheW which phosphorylates a response regulator CheY. The phosphorylated form of CheY interacts with the flagellar motor proteins to control motility. Phosphorylated CheY is deactivated by the phosphatase CheZ. The methylation of the MCPs is catalysed by the methyltransferase CheR. Methyl groups are removed by the methylesterase CheB, once it has been phosphorylated by the kinase CheA [288]. Methylation-dependent chemotaxis is reviewed in more detail in [41, 178, 226, 288, 291].

1.5 The methyl-accepting chemotaxis proteins (MCPs)

E. coli possesses four MCPs; Tar, Tap, Tsr and Trg. They are all integral membrane proteins and approximately 60kDa in size. The sequences of all four proteins are closely related and thus they are all thought to have the same structural domains [38, 44, 154]. The amino-terminal portion of the protein is the receptor region, with two hydrophobic transmembrane sequences surrounding a hydrophilic ligand-binding domain that extends into the periplasm. A large hydrophilic carboxy-terminal domain is found in the cytoplasm and is responsible for intracellular signalling and adaptation after methylation [114]. MCPs exist as homodimers in the bacterial membrane [176, 198]. Recently the MCPs have been shown to aggregate predominantly at the cell poles in ternary complexes (see section 1.7) with the CheA and CheW proteins [183].

A response is initiated by MCPs binding compounds directly, or indirectly *via* interactions with periplasmic binding proteins that are the primary receptors for some compounds [114]. Tar binds aspartate directly and mediates maltose taxis by interaction with the maltose binding protein (MBP). The binding sites of aspartate and the MBP have been shown, by genetic studies, to be adjacent and partially overlapping on Tar [92]. Trg is responsible for taxis to galactose and ribose after they have been bound by their respective

periplasmic binding proteins and Tsr is required for movement towards serine. Tap mediates the response to dipeptides, which involves interaction with a dipeptide binding protein [1]. Each of the transducers is also responsible for movement away from one or more repellents [335]. The receptors are very sensitive and although there are several hundred receptors per cell a change in occupancy of only four generates a response [37].

A fifth MCP, Tcp (which responds to the attractant citrate and the repellent phenol) has been found in the closely related bacterium *S. typhimurium* [334]. Tcp is thought to be specific to *S. typhimurium* as citrate is not a carbon source for *E. coli* and detection of phenol is carried out by Trg in *E. coli* [334, 335].

In the unstimulated state, the ligand-binding sites are unoccupied and the methylation sites are continuously methylated and demethylated to give a steady state level of one methyl group per protein [114]. On binding of an attractant, the excitatory signal is a conformational change that decreases the interaction between the MCP cytoplasmic domain and the intracellular signalling system which suppresses tumbling. Signal transduction and changes in the demethylation rate result in increased levels of methylation of the MCPs, which cancels out the excitatory signal and allows the protein to cause tumbling again, restoring the random walk of the cell. Repellent binding causes conformational changes in the MCPs that reverse its interaction with the intracellular signalling system (relative to attractant binding) and bias the flagellar motors in a CW direction [288].

1.6 Adaptation

MCPs are reversibly methylated at four, five or six glutamate residues in their cytoplasmic domains during adaptation facilitate the temporal sensing of compounds, see the reviews [114, 193, 287]. Two of the glutamate residues are produced by the irreversible deamidation of glutamine residues by CheB [244]. S-adenosylmethionine is the methyl donor and the methyltransferase CheR catalyses the formation of glutamyl esters [280, 308]. CheR appears to work at a constant rate, whereas the activity of CheB is regulated by phosphorylation (see section 1.7).

Attractants increase the methylation levels of the MCPs and a highly methylated transducer signals the motor to bring about CW rotation. Thus, following the excitatory signal, increased levels of methylation return the MCP to a null signalling state. The time taken for adaptation and return to a random walk instead of straight runs varies from seconds to minutes depending on the degree of the chemotactic stimulus. Conversely repellent adaptation is accomplished by the loss of methyl groups, with unmethylated

receptors sending signals for CCW rotation. The changes in methylation states are specific to the individual receptor that bound the stimulating ligand [287].

The temporal sensing of the bacterium's environment is a result of the time difference between fast MCP conformational changes and relatively slow methylation changes upon binding of ligands. The state of methylation of the transducers corresponds to the chemical stimuli a few seconds previously, whilst the conformation of the protein reflects the occupancy of the ligand binding sites at that moment. The balancing of information from ligand occupancy and covalent modification achieved by the protein during signal output to the flagellar motor allows the comparison between current and past concentrations of compounds. Thus if a bacterium with transducers in the adapted null state moves in a long run down an attractant gradient, the still methylated but no longer ligand bound MCPs bias the flagella towards CW rotation and induce tumbling. Only if no more attractant molecules bind, do the proteins become demethylated and return to the unstimulated state.

1.7 The intracellular signalling pathway

Six cytoplasmic proteins: CheA, B, R, W, Y and Z, relay signals from the MCPs to the flagella during chemotaxis. Intra- and intermolecular phosphorylation reactions are responsible for intracellular signalling by modifying the activity of CheB and CheY causing behavioural responses and adaptation [219, 225, 226]. Phosphorylation of CheB activates it and phospho-CheY promotes CW rotation of the flagella by interacting with the FliG, FliM and FliN motor-switch complex [157, 323]. Recently the CheY protein has been shown to bind *in vitro* to the FliM protein of the flagellar motor-switch complex. The extent of CheY binding is dependent upon its level of phosphorylation and is unaffected by either FliN or FliG [323]. CheZ increases the rate of phospho-CheY hydrolysis to attenuate the tumbling response while CheR works antagonistically to phospho-CheB. Both CheR and CheZ appear to work at a constant rate and no covalent modifications of them have been reported [193, 226, 286].

CheA is able to autophosphorylate by transferring a phosphate group from ATP onto a histidine residue. CheA then transfers these phosphate groups to CheB and CheY [42, 117, 118]. The CheA, B and Y proteins are members of a family of two-component regulatory systems in bacteria that use phosphorylation for signalling [41, 227, 290]. An autophosphorylating kinase sensor (CheA) detects an event and causes a response by phosphorylating the response regulatory proteins (CheB and Y). The MCPs bias flagellar rotation by altering the rate at which CheA autophosphorylates and thus the production of

phospho-CheY. This process requires the presence of CheW [169, 212]. The CheA and CheW proteins have been shown to form a ternary complex with an MCP receptor (predominantly at a cell pole) leading to an increased rate of autophosphorylation and affinity for CheY [95, 96, 170, 183].

Binding of attractants by MCPs decreases the rate at which CheA autophosphorylates [40] and there is a bias towards straight runs due to decreased levels of phospho-CheY. Increased levels of methylation of the transducers, resulting from less phospho-CheB, reverses this effect of the conformational change that occurs on excitation allowing the tumble frequency to return to previous levels.

Calcium ions (Ca^{2+}) have also been suggested to be involved in *E. coli* chemotaxis [304, 305]. As *E. coli* maintains intracellular free Ca^{2+} at about 90nM, caged compounds which would release either Ca^{2+} or Ca^{2+} chelators, upon exposure to specific light wavelengths, were introduced into *E. coli*. This showed that increased intracellular Ca^{2+} levels resulted in cells tumbling and a decrease resulted in smooth swimming cells. The authors postulate this may be because Ca^{2+} maintains the phosphorylated state of CheY to prolong the tumble signal of the system, as there is other evidence that Ca^{2+} inhibits the dephosphorylation of phospho-CheY [175].

1.8 Flagella structure, regulation of synthesis and biogenesis

As for the chemotactic signalling pathways, the greatest understanding of bacterial flagella and their associated functions is in the closely related enteric bacteria *E. coli* and *S. typhimurium*, whose flagellar gene systems are almost identical [180]. The flagellum is one of the most complex structures in the bacterial cell, with almost 40 genes (identified so far) involved in its assembly and function. The nomenclature used for these genes and other chemotactic genes [131] is based upon their null phenotypes and chromosomal location (see section 1.10). Three categories of phenotype are recognised; non-flagellate (Fla^-), non-motile (Mot^-) and non-chemotactic (Che^-). The Fla^- category has by far the largest number of genes and has been sub-divided into three further classes (*flg*, *flh* and *fli*) in *E. coli* depending on chromosomal location and gene order. *S. typhimurium* has an additional class (*flj*) which is involved in phase variation of the flagellar filament proteins. The Mot^- class is the smallest with only two examples, *motA* and *motB*, where a complete flagellum appears to be synthesised and yet there is no rotation. The non-chemotactic phenotype, Che^-, can be used to describe defects in receptors, components of the intracellular signalling pathway and flagellar switch proteins. However, the gene symbol *che* is only given to genes whose protein products are involved in the intracellular signalling pathway as the transducers have

separate symbols and the switch proteins' genes have a Fla$^-$ null phenotype. The generation of mutants and the characterisation of their phenotypes has been a powerful tool in the study of bacterial flagella. Morphological and biochemical studies upon mutants have shown steps in flagella assembly and allowed the visualisation of intermediate structures during this process. Flagella structure and assembly has been extensively reviewed [150, 156, 177, 179, 180].

1.9 Flagella structure

A diagram showing the generalised flagellum and the genes which encode the various structural components is shown in figure 1.9.1. Electron microscopy has shown the flagellum to consist of a basal body imbedded in the cell surface, a hook and the helical filament. However other components are associated with the flagellum. The motor complex is thought to be located in the inner membrane adjacent to parts of the basal body, and a structure (the C-ring) postulated to be the switch complex has been shown to be associated with the cytoplasmic face of the basal body [89, 149]. The position of the export apparatus has not been shown and its function is poorly understood. It will be discussed in more detail in section 1.12.

The filament is composed of thousands of copies of a single protein, flagellin, encoded by the *fliC* gene. Some of the lysine residues of this protein are modified by methylation [9] but this appears to have no functional significance, as mutants lacking this modification function normally [177]. The filament is known to have a central channel [206] through which flagellin monomers can be exported for assembly at the distal tip of the filament [179]. To prevent loss during assembly the filament is capped with the gene product of *fliD* [180]. At the proximal end of the filament are two junction proteins FlgK and FlgL which are thought to be necessary for proper joining of the filament to the hook [123].

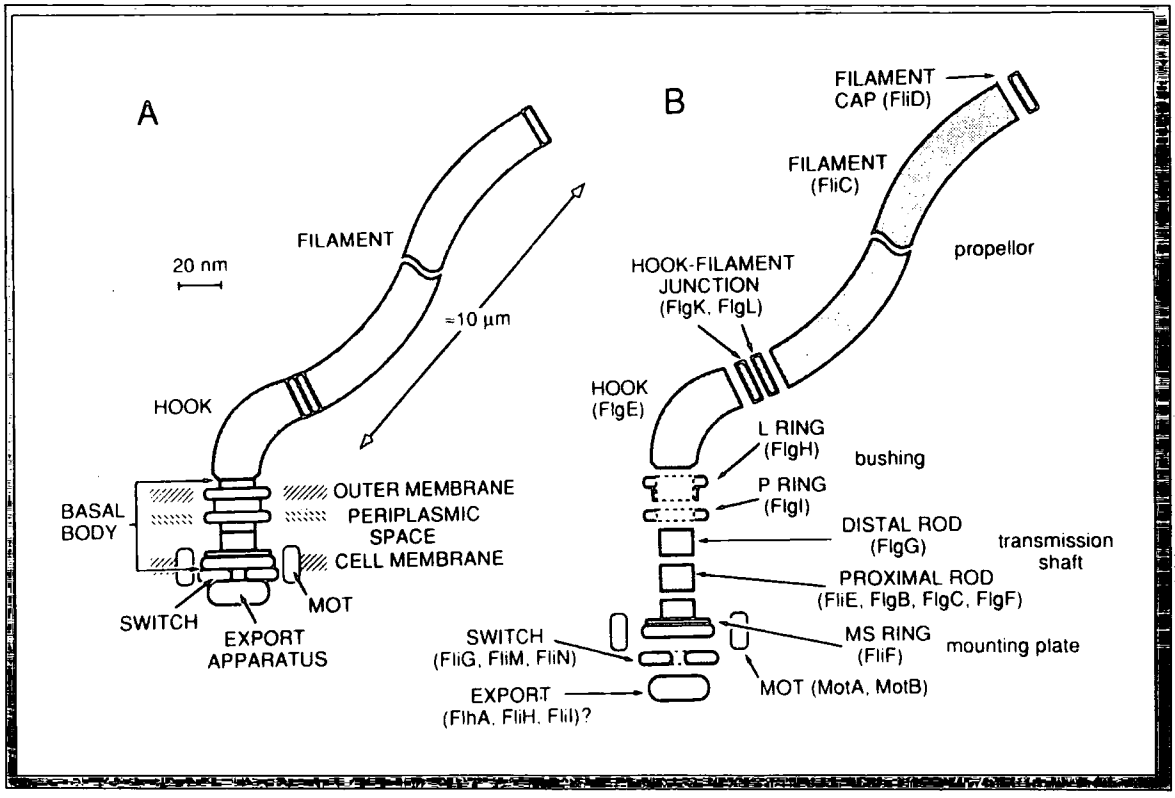
The hook is thought to act as a universal joint (or flexible coupling) between the filament and the cell [177]. It is composed solely of the *flgE* gene product. Like the filament it has a central channel (of comparable size) presumably to allow the export and assembly of hook and filament subunits [202]. This is not unsurprising since the hook and filament have been shown to have the same subunit packing [321]. There are no obvious similarities in the protein sequences between individual hook and flagellin proteins, apart from the heptad repeats which will be discussed later. However the hook proteins have been

Figure 1.9.1 The general structure of the flagellum of *E. coli* and *S. typhimurium*.

Diagram (A) shows the overall flagellum structure and the location of the components relative to the cell envelope. The location of the export apparatus has not been demonstrated but it is assumed to be at the flagellar base (see text).

Diagram (B) is an expanded version showing the substructures and the proteins encoding them, these are discussed in more detail in the text. The order of the proximal rod components FlgB, C, F and FliE within the rod is not known.

The figure was reproduced (with permission) from [180].



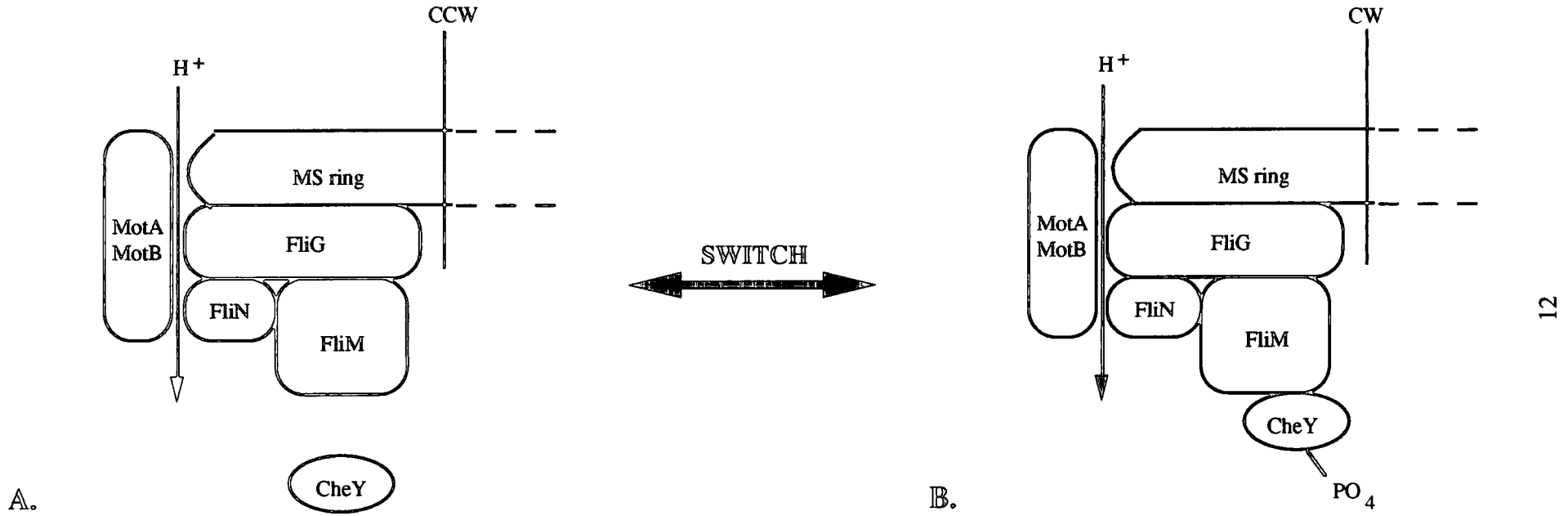
shown to possess three different domains, the innermost of which is a rod-like domain similar to that of the flagellins [202]. For a more detailed review of flagellin and hook protein organisation and packing see [6, 202, 206, 318, 319].

Figure 1.9.1 shows the basal body to be composed of a number of proteins. There are pairs of rings at the inner cell membrane and at the periplasmic space/outer membrane. There is also a central rod that passes through the outer rings and connects the inner pair of rings, hook and filament [180]. The functions of the basal body are usually compared to those of a mechanical motor [156, 177]. The inner membrane rings (MS ring) are considered to be the rotor or a structure onto which the rotor is mounted, the rod to be the transmission shaft and the outer membrane rings (L- and P rings) to be the bushing. The MS ring and part of the proximal rod are composed of a single protein FliF [307]. The FliE protein is thought to link the MS ring to the other proximal rod proteins FlgB, C and F [205]. The distal portion of the rod is composed of the FlgG protein [180]. The L- and P rings (FlgH and FlgI respectively) form a large pore in the periplasmic space and outer membrane, through which the rod passes [135]. The subunit stoichiometry of the basal body has been determined [277]. Each ring contains ~26 subunits, the rod contains ~6 copies each of FlgB, C and F and ~26 copies of FlgG.

The axial proteins FlgBCEFGKL and FliCD are all located extracellularly or within the cell envelope, although none have cleavable signal sequences [179, 239]. They are all thought to be exported by the flagellum-specific export pathway, see section 1.12. The axial proteins do all share a common feature of heptad repeats of hydrophobic amino acids at the amino- and carboxy-termini [121]. These regions are characteristic of sequences that fold into α helical coils and it has been suggested that the amino-terminus of one subunit together with the axially adjacent protein's carboxy-terminal region could form part of a coiled-coil domain. This interlocking organisation could be the common structural motif by which the nine axial proteins form a continuous structure [121, 202].

The MotA and MotB proteins are responsible for the formation of intramembrane particulate rings that surround the MS ring at the flagellar base [151]. These proteins are believed to be important components of the flagellar motor. MotA is a proton channel thought to be responsible for the formation of the proton gradient that drives the flagellum [34]. The presence of MotB is necessary for MotA to function, and they are believed to form a complex with MotB acting as a linker connecting MotA to the cell wall [292].

The switch complex is composed of three proteins: FliG, FliM and FliN, which are responsible for a bell-like structure (the C ring) at the base of the flagella [89, 149]. The


CCW
H+
MS ring
MotA
MotB
FliG
FliN
FliM
CheY
SWITCH
CW
H+
MS ring
MotA
MotB
FliG
FliN
FliM
CheY
PO4

A.

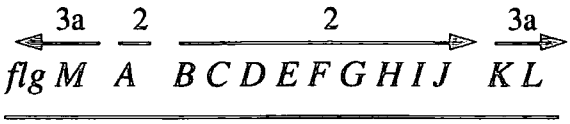
B.

switch proteins also play an important role in motor rotation (as well as direction of rotation) and mutations within the switch proteins can lead to Mot$^-$ phenotypes. Analysis of the phenotypes of a large number of switch protein mutants and in particular mutants that suppressed defects in the CheY and CheZ proteins has allowed a model of the structure and functional regions of the switch to be proposed [88, 132, 276] see figure 1.9.2. The FliG component of the switch complex is thought to be located at the cytoplasmic face of the MS ring, since mutants synthesising a fusion protein of FliF and FliG had almost normal function [88]. Immunoelectron microscopy showed the FliG component of the fusion protein to be located at the cytoplasmic face of the MS ring [88]. FliM is thought to be involved predominantly in the switching process [276] and thus to have a position more towards the cytoplasm, where it presumably interacts with phospho-CheY. As discussed earlier Welch *et al.* have shown phospho-CheY to bind FliM *in vitro* [323]. Upon binding of phospho-CheY, FliM probably directly interacts with FliG to cause switching [132]. On the basis of genetic studies FliN and FliG are thought to be involved in the mechanism of rotation of the flagella [132]. It is thus postulated that they interface with the MotA and MotB proteins and by some, as yet, unknown mechanism bring about rotation. This model of the switch-motor complex is discussed in more detail by Irikura *et al.* [132].

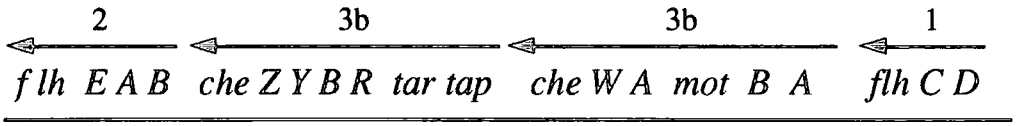
1.10 Regulation of transcription of the flagellar and chemotaxis genes

The flagellar and chemotaxis genes of *E. coli* and *S. typhimurium* are highly clustered on their respective chromosomes [179] see figure 1.10.1. There are four main regions labelled I, II, IIIa and IIIb. Each region contains several operons, which in turn contain between 1-9 genes. The expression of these operons is as a regulatory hierarchy, or regulon [159, 180]. The expression of one class of operon is required for the expression of lower operon classes [180]. Four classes (levels of hierarchy) have been identified; class 1, 2, 3a and 3b, see figure 1.10.1. The ultimate level of the hierarchy is class 1, which contains only one operon (the master operon) whose expression is required before that of any other classes. The seven class 2 operons are under the direct control of the master operon and contain genes whose products are required in the early to middle stages of flagellar synthesis, see section 1.11. One of the class 2 genes, *fliA*, has been found to encode a flagellum-specific sigma factor for classes 3a and 3b [16, 215]. The class 3b operons are thought to be solely under the control of FliA, whereas the class 3a operons are controlled by the master operon as well as FliA [180]. The mechanism of flagellar gene regulation is described in more detail by Macnab [180] as are the potential advantages of the control mechanisms.

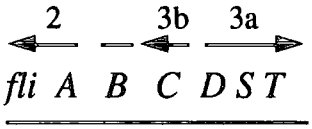
Region I :



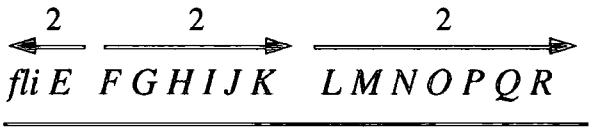
Region II :



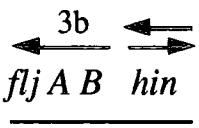
Region IIIa :

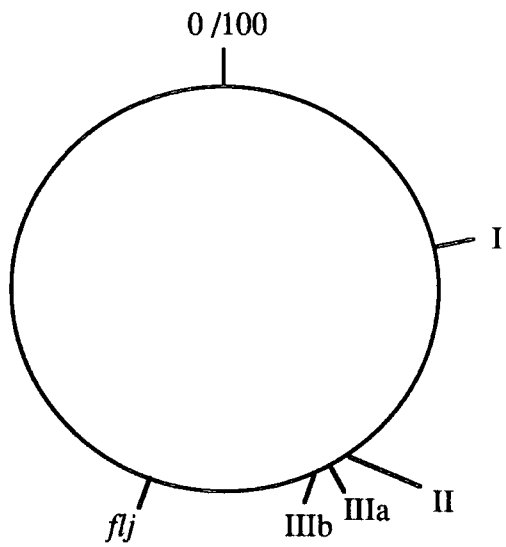


Region IIIb :



flj region :





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

The class 2 and 3 operons have flagellum-specific promoter consensus sequences, see figure 1.10.2 below, which was adapted with permission from [180].

Figure 1.10.2 A table showing the consensus sequences of the promoters for the classes of flagellar operons.

Operon class	-35 consensus		-10 consensus
Primary σ^{70}	TTGACA	N17	TATAAT
Flagellar class 1	CATC	N15	ACCGCTAA
Flagellar class 2	none evident		GCCGATAA
Flagellar classes 3a and 3b	TAAA	N15	GCCGATAA

As a comparison the consensus of the primary cellular sigma factor (σ^{70}) is also shown [113]. The class 1 consensus has little similarity to any other consensus sequences and the class 2 consensus appears to have no -35 consensus [180]. The presence of these consensus sequences infers that the expression of these operons is mediated by flagellum-specific sigma (σ) factors of RNA polymerase - such as FliA [115]. The nomenclature for the σ factors will be that of Macnab [180] σ^{C1}, σ^{C2} and σ^{C3} - where σ^{C1} represents the σ factor for the class 1 operon, σ^{C2} the undiscovered σ factor for class 2 operons and σ^{C3} to be FliA. FliA is also called σ^{F} [16, 115].

The class 1 (master) operon is under the control of cAMP *via* the regulatory protein CAP [271]. The nature of σ^{C1} is unknown, the operon may be transcribed using the primary cellular initiation factor (σ^{70}) or an alternative specific sigma factor. The promoter has no recognisable consensus to either the primary or flagellum-specific factors, see figure 1.10.2. The class 1 proteins FlhC and FlhD are positive regulators of the class 2 operons, although the mechanism through which this occurs (i.e. the identity of σ^{C2}) is unknown [159].

As mentioned before, σ^{C3} has been identified as the gene product of *fliA* [16, 215]. FliA can bring about expression from both class 3a and 3b operons, their consensus sequences appear indistinguishable, see figure 1.10.2. However, class 3a operons are also expressed in the absence of FliA, providing the master operon is expressed. Presumably the mechanism of transcriptional activation for 3a operons in this situation is the same as for class 2 operons - the -10 regions of the promoters have the same consensus. However the means by which σ^{C2} distinguishes between 3a and 3b operons is unknown [180].

In addition to the hierarchy of positive regulation, there are negative regulatory systems. The negative regulation of the *fliDST* operon and that of the FlgM protein. The observation that mutations in the *fliDST* operon result in elevated expression of class 3a and 3b operons [159] suggests that in wild-type cells the operon exerts some form of negative regulation. The specific mechanism of this negative regulation, called RflA activity, is not known. The *fliS* and *fliT* genes have only recently been identified as members of this operon and indirect evidence suggests they may have a regulatory role, although experimental evidence has yet to be obtained [139].

The FlgM protein blocks FliA-mediated expression of the class 3a and 3b operons [99] but is itself controlled by the state of flagellar assembly. This dual regulatory activity explains an essential feature of the hierarchical control. Any defect in a gene of a higher class will result in a large reduction in the expression of operons in lower classes, and hence a reduction in wasted energy expenditure. For example a mutation in a class 2 gene will result in non-expression of class 3b operons and only weak expression of class 3a operons (via σ^{C2}). In a wild-type cell the FlgM operon is transcribed from two promoters [100]. A class 3a promoter initiates transcription upstream of *flgM* and transcription is also initiated *via* the promoter of the upstream *flgA* operon (class 2). The *flgM* operon could thus be said to be a member of both regulatory classes [100]. The gene product of *flgM* has been shown to bind the σ factor FliA (σ^{C3}) with a 1:1 stoichiometry, and inactivate it [216]. Thus with FliA inactivated there can be no expression of class 3b operons and only weak expression of 3a operons. FlgM is therefore also auto-regulatory and in this situation *flgM* is predominantly transcribed from the *flgA* promoter [100]. The class 2 operons are transcribed normally and an immature flagellum, upto the completion of the hook (see section 1.11) will be assembled. It is at this point that the gene products of the class 3a and 3b operons are required and thus some mechanism must exist to free FliA from the FliA-FlgM complexes. This is achieved using the immature flagellum structure. The FlgM proteins are themselves exported through the central channel, which is later used for the export of the flagellin subunits [129, 158]. The FlgM anti-sigma factors are thus exported out of the cell and can be detected in the external medium [129, 158]. FliA is now free to activate the class 3a and 3b operons and complete the synthesis of the flagellum and the other chemotactic proteins. If any of the class 2 operons are defective, the structure of the immature flagellum will fail to assemble and the FlgM protein will not be transported out of the cell and no further flagellar synthesis will occur. Presumably the export of the FlgM proteins will diminish and eventually stop as the filament grows in length [129, 158, 174], thus inactivating the class 3a and 3b operons again by binding to any free FliA. However should the filament be damaged the export channel will again be open to the environment,

FlgM will be exported and activation of transcription of the class 3a and 3b operons will lead to the repair of the damaged filament [129, 158, 174].

There may also be other mechanisms involved in the regulation of flagellar and chemotactic gene expression. These may be general cellular mechanisms or may involve some of the products of genes essential for flagellar formation with an, as yet, unknown function *i.e.* FlgJ, FlhB, FlhE and FliJLOPQRSTUV [72, 139, 180]. Alternatively the proteins encoded by these genes may be involved in the assembly of flagella (for example as scaffolding proteins) or the flagellum-specific export pathway, see the next two sections.

1.11 Biogenesis of the bacterial flagellum

The biogenesis of the bacterial flagellum is on the whole linear, with assembly proceeding from proximal to cell distal structures and the synthesis of individual structures also occurring distally [180]. It is usually analysed and described according to the morphology of successive stages prior to complete assembly, see figure 1.11.1 [156, 180]. However such an approach does not account for any of the regulation (described in section 1.10) controlling the synthesis of a structure as complicated as the flagellum, for example the role of FlgM. Nor does it describe the mechanisms involved for the export and assembly of the individual components, following their synthesis in the cytoplasm, to the five different cellular locations used: peripheral to the inner cell membrane/cytoplasmic, integral to the inner cell membrane, periplasmic, integral to the outer cell membrane and extracellular.

The proteins peripheral to the cytoplasmic cell membrane probably self-assemble onto a membranous structure upon contacting it. The Sec-dependent pathway [239] is responsible for the export of some of the integral membrane and periplasmic proteins. The flagellum also has a specific export apparatus (see section 1.12) for the export of the axial proteins. The axial proteins form the rod, hook and filament structures which assemble distally and possess a central channel through which successive components can be exported [202]. The channel has not yet been observed in the rod [284]. However given the similarities in amino acid sequence between the axial proteins it is postulated they form a continuous structure and hence the rod is assumed to have a channel [202]. The morphogenetic pathway shown in figure 1.11.1 is now discussed in more detail, it is also reviewed in [156, 179, 180].

As described previously the MS ring is composed of a single protein FliF. Its insertion into the cytoplasmic membrane is thought to be the initiating event of flagellar

biogenesis [136, 156]. The assembly of the switch proteins FliMNG is assumed to proceed next, in part because of their location at the cytoplasmic face of the MS ring [136, 156]. A putative model of the organisation of the switch proteins was described in section 1.10. The FliF-FliG fusion protein which identified the putative location of FliG at the cytoplasmic face of the MS ring [88] must be assembled at the same time, inferring sequential assembly of the switch complex. The assembly of the switch complex is necessary for the construction of the rod - the next observable morphological stage, and is assumed to precede its assembly [156].

Rod formation is prevented in *flhA*, *fliH* and *fliI* mutants [156]. The products of these genes are believed to be involved in flagellum-specific export [316] see next section. The prevention is necessary as the rod is composed of axial proteins which the flagellum-specific export pathway transports, and thus assembly of the flagellum-specific export apparatus must precede its use in exporting the axial proteins. There are a number of other mutants which inhibit rod formation, see figure 1.11.1. Some, or all, of the products of these mutated genes may also be involved in the flagellum-specific export pathway.

The FliE protein is also required for the assembly of the rod [180]. It has been found to be a component of the basal body and may act as an adapter between the MS ring and rod substructure [205].

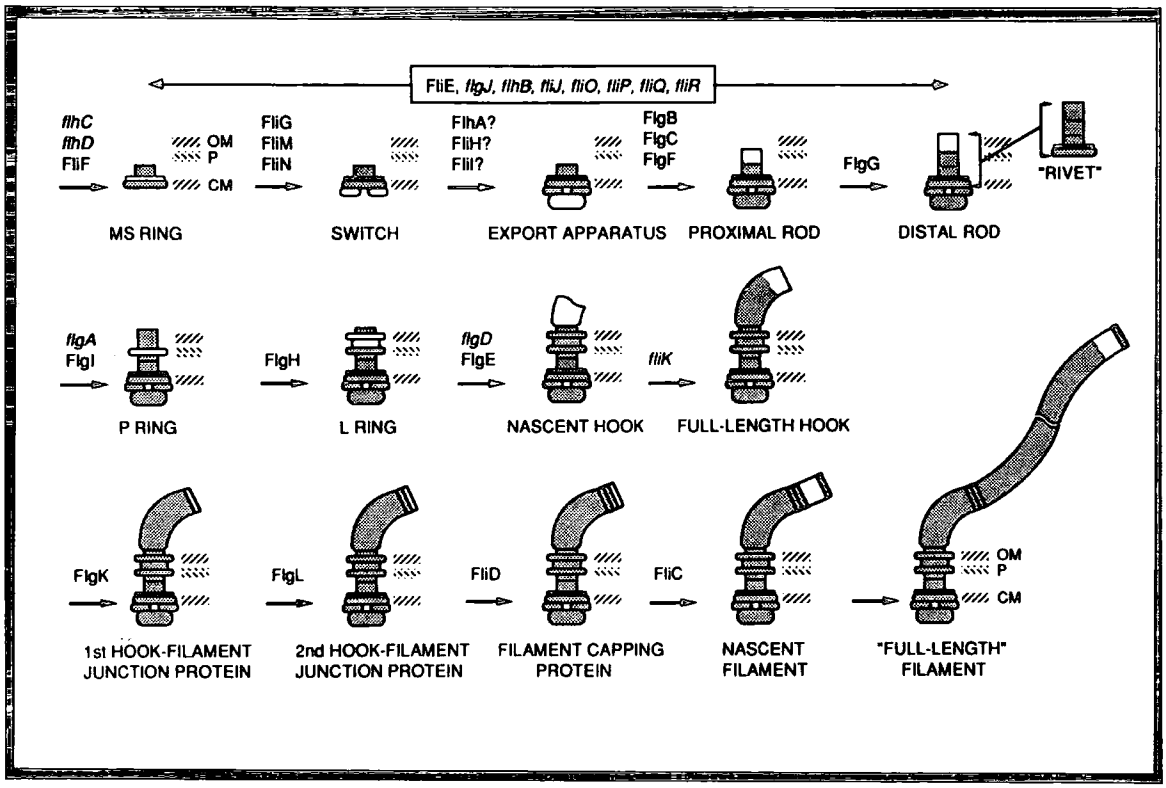
The rod proteins FlgB, C, F and G are thought to assemble cooperatively with FlgG found at the distal end [136, 156]. The FlgD protein has recently been shown to be added to the complete rod structure and to remain at the distal tip of the immature flagellum whilst the hook is synthesised [217]. Upon completion of the hook the FlgD protein is lost, thus its presence is only transient and so it is thought to be a scaffolding protein [217]. The role it plays will be described later in conjunction with hook assembly.

Assembly of the P- and L rings is not necessary for the initiation of hook synthesis, but a completed hook is only made if the outer cylinder has assembled [156]. The P- and L rings have cleavable signal peptides and are assumed to be exported to the periplasm and outer membrane (respectively) by the Sec-dependent pathway [239], where they encounter a completely synthesised rod upon which they assemble [180]. Whether this process occurs by diffusion or requires additional molecular chaperones is not known.

Hook synthesis begins only if FlgD has been inserted onto the distal end of the rod, implying FlgD is also a member of the axial proteins [217]. FlgD has been shown to act as a scaffolding protein and its function in this role is analogous to FliD, the filament capping

Figure 1.11.1 The biogenesis of the flagellum from *E.coli* and *S. typhimurium*.

The figure was reproduced (with permission) from [180]. The assembly stages shown are discussed in more detail in the text. Succeeding stages in the assembly of the flagellum are shown, along with the genes or proteins required for the next stage. When the protein is known to assemble into the structure it is written in roman letters, if the function of a gene product is not known, the gene is labelled in italics. *flgD* is shown in italics, although the function of FlgD is now known (see text). Each incremental feature is shown in white and all the preceding structure is shaded. The positions of the cell membrane (CM), periplasmic space (P) and outer membrane (OM) are shown. The "rivet" is the simplest structure detected by electron microscopy after the MS ring. The switch and export complexes (and maybe others) thought to be assembled between the MS ring and rivet are lost during the isolation procedure. FliE and the genes labelled within the box are required prior to distal rod assembly and approximately at the positions shown. The term "full length" when describing the filament is taken to mean a filament long enough to provide propulsion. The length of the "mature" filament is not defined, see text.



protein, to prevent the loss of hook subunits into the external environment prior to distal assembly [217]. Hook synthesis proceeds to a defined length under the influence of FliK [295]. When hook assembly is completed, the FlgD protein is lost to the environment (unlike FliD) and replaced by the hook-associated protein FlgK by an active displacement process [217]. The other hook-associated protein FlgL and the filament cap protein FliD are now added and the filament synthesised [217]. With the loss of FlgD the anti-sigma factor FlgM will be exported from the cell through the central channel (see section 1.11) to bring about full activation of the 3a and 3b operons *via* σ^{C3} (FliA). The presence of the FlgK, L and FliD proteins before the release of FliA can be accounted for by their transcription using σ^{C2} as they are all class 3a flagellar proteins. These proteins are synthesised and exported through the central channel continuously [122], presumably in case of filament breakage, the filament can be recapped with FliD and re-synthesised.

The flagellin subunits are exported through the central channel and assemble at the distal end of the growing filament. Whether FliD aids this assembly (as a scaffolding protein) or merely prevents the loss of the flagellins is unknown [179]. It has not yet been shown if the flagellin subunits are exported in a folded or unfolded state, since there is some debate over the diameter of the channel [202]. X-ray diffraction studies of the flagellar filament calculate the channel size to be ~60Å [206] and thus capable of allowing the passage of fully folded flagellin subunits [202]. Whereas electron microscopic studies show the channel diameter to be ~25Å which would necessitate at least the partial unfolding of the flagellin subunits [306]. Flagellin subunits have been shown to self-assemble in *in vitro* systems [17]. The extent of filament synthesis is not controlled like hook length. The rate of elongation decreases as the filament lengthens presumably as the passage of flagellin subunits down the channel slows due to of frictional forces [180]. As filaments lengthen they become more susceptible to damage, and breakage will reset the flagellar length.

As MotA and MotB are not required for the assembly of the flagellar structure, it is not known at what stage they assemble. Given their functional interactions with the flagellar switch (see section 1.10) it is possible that they assemble early in the pathway [180]. MotB has also been proposed to be the anchor of the overall structure in the cell membrane [56].

1.12 The flagellum-specific export pathway

The location of the flagellum-specific export apparatus is assumed to be at the flagellar base since the pathway for the exported products is through the nascent structure [316]. Very little is known about the export apparatus, whether energy is required or the

nature of the target sequences/motifs on the proteins to be exported that distinguish them from other proteins. Flagellar regrowth experiments with temperature sensitive mutants identified four flagellar genes which when mutated failed to reassemble flagella. Their gene products are all required prior to rod formation, and thus they were proposed to be possible components of the flagellum-specific export apparatus [316]. Of the four genes identified; *flhA, fliH, I* and *N*, three of the gene products had unknown functions. The fourth, FliN, was known to be an integral part of the switch-motor complex [132]. Thus its role in the flagellum-specific export pathway is thought to be indirect, possibly because of its position it may interact with the export apparatus and if mutated may affect the structure of the apparatus [316].

The function of the FliH protein remains unknown. The FlhA protein has several putative transmembrane domains inferring it may be integrated into the membrane [75]. The FliI protein was observed to have sequence similarity to subunits of the proton-translocating F_0F_1 ATPases and thus may be involved in supplying any energy that is necessary for export [75]. Furthermore FliI was shown to have a high copy number per cell (~1500) and probably interacts transiently with the flagella (~8 per cell) remaining free in the cytoplasm at other times [75].

Both the FliI and FlhA proteins were also found to have sequence similarity to proteins found in the virulence protein-export systems of mammalian and plant pathogens [75]. FliI was found to have sequence similarity to Spa47 of *Shigella flexneri* [314] and to HrpB6 of *Xanthomonas campestris* [83]. FlhA has sequence similarity to VirH of *S. flexneri* [314], HrpC2 of *X. campestris* [83], LcrD of *Yersinia pestis* [234] and to the InvA protein of *S. typhimurium* [91] - all of which are thought to be involved in the translocation of specific proteins across cell membranes [91]. A further flagellar protein (FliP) has been proposed to be involved in the flagellum-specific export apparatus on the basis of sequence similarity to proteins involved in the export systems of other bacteria. The FliP protein has sequence similarity to Spa24 of *S. flexneri* [314] and to ORF2 of *X. campestris* [130].

The homologies seen between the above proteins in this range of bacteria has led to the conclusion that they are all components of a superfamily of export, and possibly assembly, proteins [75]. However further components have still to be found, as well as the functions of those already identified.

1.13 Chemotaxis and motility systems in other bacteria

The chemotaxis and motility systems of *E. coli* and *S. typhimurium* have been the most studied and thus are used as models to which other micro-organisms are compared. It should not, however, be taken as an absolute, since a number of differences have been reported amongst other bacteria. Some of these differences from the more extensively studied bacteria are described below.

The Gram-positive bacterium *Bacillus subtilis* possesses homologues of many of the chemotaxis and motility genes found in *E. coli* and *S. typhimurium*, but some of their gene products have been shown to function in different roles [31]. These genes, and other chemotaxis and motility genes thought to be unique to *B. subtilis*, are found in five regions on the *B. subtilis* chromosome [220]. The majority of these are found in a very large operon of over 25kb containing genes encoding proteins involved in intracellular chemotaxis, flagellar structure and assembly as well as some of unknown function [220]. The transcription of these genes together infers a different regulatory mechanism to that described (in section 1.10) for *E. coli* and *S. typhimurium*. Furthermore flagellar synthesis in *B. subtilis* is not affected by cAMP/CAP and no genes homologous to *flhC* and *flhD* have been found [220]. Perhaps the most striking differences though are those in the MCP-dependent intracellular signalling system. The situation is a reversal of that described for *E. coli* and *S. typhimurium*. In *B. subtilis* attractants induce CheA to phosphorylate CheY which then brings about CCW rotation (and hence smooth swimming) of the flagellum. Adaptation is brought about by the transfer of methyl groups from the receptors to another protein, which is in turn thought to be demethylated at the flagellum producing methanol. No homologue of the enteric protein CheZ has been found in *B. subtilis* and the mechanism of phospho-CheY dephosphorylation has not been elucidated [220]. There are also differences at the switch complex, no *B. subtilis* homologue of *fliN* has been found, however the *B. subtilis* FliY protein has some similarity to FliM and FliN of *E. coli* [32]. The differences in the switch proteins probably reflect the differences in the intracellular signalling pathways. Despite these differences most of the chemotaxis and motility genes sequenced in *B. subtilis* have enteric homologues suggesting there must be some parallels in the structures and controls of the two systems.

The temporal regulation of flagellar biogenesis within the cell cycle has been studied in *Caulobacter crescentus*. An order of transcription has been proposed by observing morphological changes in flagella structure and transcription studies [69]. A hierarchial control, very similar to that seen in *E. coli* and *S. typhimurium*, is thought to operate with potentially four classes in *C. crescentus* [285]. However, no class 1 homologues have been found in *C. crescentus* and so the first flagellar genes known to be expressed are class 2

genes, whose transcription is coupled to chromosomal replication [285]. The majority of the structural components of the flagellar and chemotaxis machinery are class 3 and 4, which are transcribed with specific sigma factors and *trans*-acting factors [285]. *C. crescentus* has a single polar flagellum composed of three different types of flagellin monomers, which are organised in a defined manner [76]. The McpA chemoreceptor (an MCP) is located at the same pole [8].

Vibrio parahaemolyticus possesses two distinct motility systems which function separately depending upon the bacterium's environment. In a liquid environment, cells possess a single polar flagellum and are capable of swimming. However when placed on a surface the swimmer cell differentiates into a swarmer cell which possesses many lateral flagella. The differentiation is brought about, in part, by the polar flagellum acting as a tactile sensor, sensing forces that restrict its movement [191]. The flagellar systems are different, the polar flagellum is composed of two different types of flagellin subunits and is sheathed by an extension from the outer membrane [192]. In comparison the lateral filaments, which are arranged peritrichously, resemble those of *E. coli* and *S. typhimurium* and are composed of only one flagellin subunit LafA [192]. Furthermore the lateral flagella are driven by the proton motive force (like *E. coli* and *S. typhimurium*) but the polar flagellum is driven by the sodium ion motive force [21]. More components of the lateral flagella system have been identified, including motor components, a hook-associated protein and a flagella sigma factor [192]. Despite the differences in flagellar systems, there is thought to be a common chemosensory apparatus [254].

R. sphaeroides possesses a single flagellum composed of one type of flagellin. Rotation is unidirectional (clockwise) and changes in the direction of swimming are brought about by pauses in flagellar rotation and subsequent cell body reorientation by Brownian motion [15]. A variant of *R. sphaeroides* has been isolated that rotates its flagellum CCW but exhibits normal motility and chemotactic behaviour [221]. *R. sphaeroides* is thought not to possess MCP chemoreceptors and only a limited PTS-dependent chemotaxis system [12]. It exhibits two distinct chemoresponses - chemokinesis and chemotaxis [12, 14]. Chemotaxis occurs when *R. sphaeroides* detects a nutrient gradient and results in an accumulation of the bacteria in a region of increased concentration [14]. This behaviour is characterised by a change in the stopping frequency of the flagellum [12]. Chemotaxis requires the transport and metabolism of the chemoattractant [237]. Chemokinesis is brought about by an increase in the concentration of the limiting metabolite and is demonstrated by a sustained increase in the swimming speed of *R. sphaeroides* [12]. The chemokinetic response results in a dispersal of the bacterial population [14]. Chemotaxis and chemokinesis have been shown to be independent responses (although some

chemoeffectors elicit both) and that transport of the chemoeffector is essential for the response [222].

The rhizosphere bacterium *Azospirillum brasilense* has also been shown to exhibit similar chemoresponses as those of *R. sphaeroides*, and is not thought to possess an MCP-dependent chemotactic signalling system [340]. The chemotaxis and motility systems of other Rhizobiaceae will be discussed later, in particular those of *Agrobacterium tumefaciens* the organism predominantly studied in this work.

1.14 *Agrobacterium tumefaciens*

A. tumefaciens is a Gram-negative bacterium, whose cells have rod-shaped morphology. These rods are 0.6-1.0μm wide by 1.5-3.0μm long. The bacteria are motile, possessing between one and six peritrichously arranged flagella (see figure 1.14.1). *A. tumefaciens* is mesophilic with 28°C being the optimum temperature for growth [144].

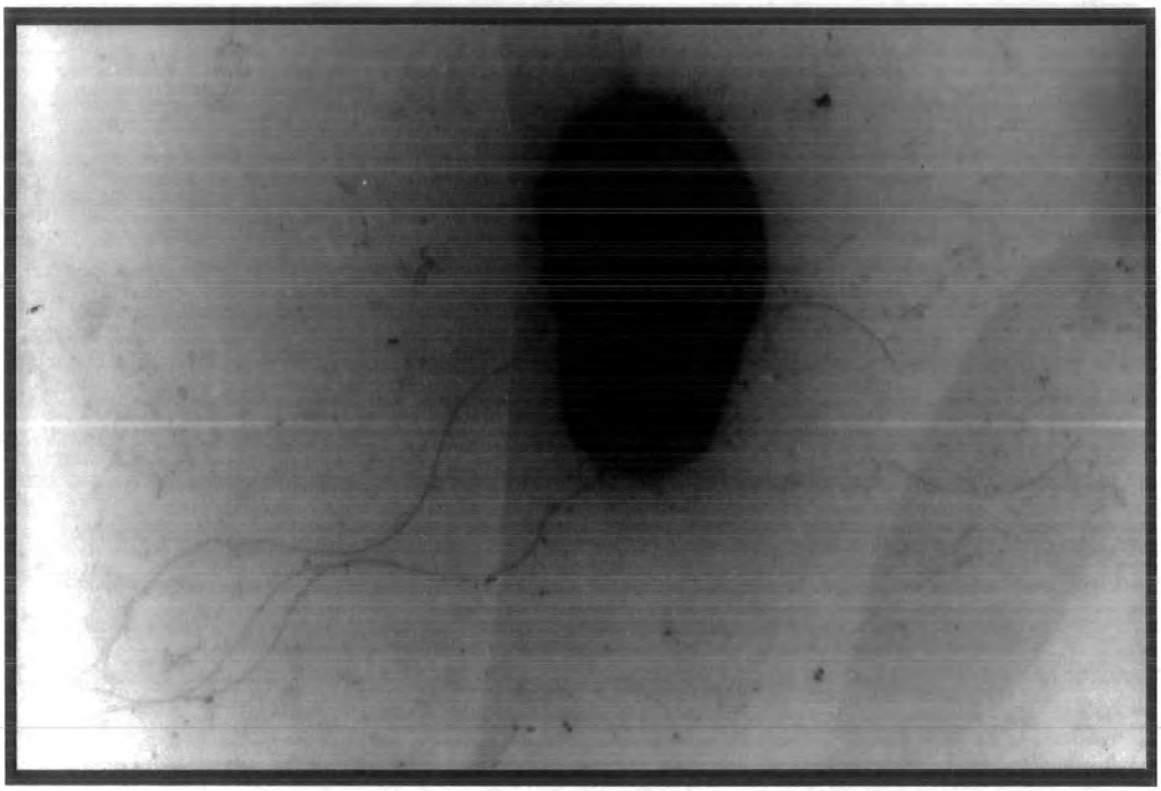
The genus *Agrobacterium* is a member of the family Rhizobiaceae and is closely related to the genus *Rhizobium* [147]. The genus traditionally consists of four species: *A. tumefaciens* (which causes Crown Gall Tumour); *A. rhizogenes* (the causative agent of Hairy Root); *A. rubi* (which causes Cane Gall Tumour); *A. radiobacter* (which is avirulent) [209].

The phytopathogenicity has been shown to be unstable and is thus unsuitable for species identification. The behaviour is dependent upon large tumour-inducing (Ti) or root-inducing (Ri) plasmids that a species may or may not contain [200, 309, 322, 324]. These plasmids can be transferred from one strain to another [97, 145] resulting in the recipient strains acquiring the pathogenic properties of the donor strain [302, 310].

Thus an alternative classification has been proposed that divides the genus into three biovars [146, 147] according to various chromosomally encoded biochemical and physiological differences. The three biovars are genetically and phenotypically distinct and contain both pathogenic and non-pathogenic strains. Despite this classification, the type strain remains the virulent *A. tumefaciens* [137] and thus the old nomenclature is the one that is still in common use (and will be used here). All the strains used in this study are derivatives of *Agrobacterium tumefaciens* C58 (which is a biovar 1 strain) that contains the Ti plasmid pTiC58 [309].

Figure 1.14.1 ***Agrobacterium tumefaciens*** **C58C1.**

An electron micrograph of a wild-type *A. tumefaciens* cell showing the two polar flagella, and two lateral flagella. The cell body is approximately 1.2μm in length. The photograph was taken by Dr. C. H. Shaw.



1.15 Soil ecology and plant transformation

A. tumefaciens is an opportunistic pathogen of most dicotyledonous plants and some monocots [59, 74]. It is a common component of the soil microflora world-wide, especially within the rhizosphere, where the number of bacteria associated with the roots can be approximately one thousand times that found in nearby soil [211]. Non-pathogenic bacteria greatly outnumber pathogenic strains, which are usually only detected around infected roots and in galls, since the Ti and Ri plasmids are easily lost without selective pressure [43, 142]. Pathogenic agrobacteria cause neoplastic overgrowths on plant roots (and sometimes stems) which stunt the growth of infected plants and are responsible for significant agricultural losses [141, 143]. These neoplastic overgrowths of Crown Gall and Hairy Root Disease result from the transformation of plant cells into tumourous cells which undergo uncontrolled proliferation. The transformed cells are characterised by the synthesis and secretion of opines, unusual sugar and amino acid conjugates, which are not normally present in plant cells [300].

The type of tumour that develops and the type of opine that is synthesised are dependent on the oncogenic plasmid present in the infecting bacteria [322, 324]. This is because to transform plant cells a small piece of DNA from the Ti or Ri plasmid is transferred into the plant genome [53, 54, 325]. The expression of this DNA is responsible for the tumourous phenotype and opine synthesis (see section 1.18).

The opines are used as a source of nitrogen and carbon by the infecting *Agrobacterium* strain. This catabolism is also a trait encoded by the oncogenic plasmid [39]. The host plant and the majority of other micro-organisms are unable to use the opines [23, 252]. Opines also induce the conjugal transfer of self-transmissible Ti plasmids to plasmid-less agrobacteria [80], thus enabling non-virulent strains to become tumourigenic and metabolise opines. This induction could give agrobacteria a competitive advantage over other soil-micro-organisms unable to catabolise opines. However field investigations suggest opine production does not greatly affect soil colonisation and survival in the rhizosphere of agrobacteria, since most species isolated are non-pathogenic and other micro-organisms can utilise opines. The production of opines however, probably stimulates growth of the tumourigenic *Agrobacterium* species within a Crown Gall [255].

1.16 The *A. tumefaciens* : plant interaction - an overview

A. tumefaciens is able to detect and localise itself at wounded sites on plants (see section 1.21). This is essential for infection as wounding of the plant is a pre-requisite for

tumour development [30]. The bacteria will now attach themselves to the plant cells and expression of the Ti plasmid-encoded *vir*ulence (*vir*) genes is induced by wound-specific chemicals secreted by these cells. The Vir proteins bring about the transfer of a piece of bacterial DNA, the transfer or T-DNA, to the plant. The T-DNA becomes incorporated into the plant chromosomal DNA and the genes upon it expressed. These genes encode plant growth regulators, which cause tumour production, and also the enzymes responsible for opine synthesis.

1.17 Chromosomal genes involved in virulence

As well as the *vir* genes, that process and transfer the T-DNA, there are many chromosomal genes (*chv*) required for tumour production. The *chv* (*ch*romosomal *v*irulence) genes encode proteins involved in the attachment of agrobacteria to the plant cells and in *vir* gene induction. *chv* genes necessary for *vir* gene induction include, *chvD*, *chvE*, *Ivr* -221, -223, -225 and *miaA* [103, 127, 196]. The function of the ChvD and ChvE proteins will be discussed in section 1.20.

A. tumefaciens must attach itself to the surface of the wounded plant cells to cause Crown Gall tumours [168]. The attachment process occurs in two stages, firstly a loose reversible attachment mediated by a bacterial protein called rhicadhesin [274] and then a tight irreversible binding. Since the binding of agrobacteria to plant cells is saturable, and unrelated bacteria are unable to compete for binding, whereas related do, the presence of specific binding sites has been proposed [168]. However no plant cell receptor has yet been found, although there is evidence to support its existence [30].

The irreversible binding of *A. tumefaciens* to plant cells is brought about by the synthesis of cellulose fibrils [188, 189]. The fibrils anchor the bacteria to the plant and entrap other agrobacteria forming aggregates which increase the chances of successful tumour production [188]. The *cel* genes responsible for cellulose fibril production are also chromosomally encoded [245].

The active form of rhicadhesin requires endogenous β-1,2-glucan either for its correct processing or for anchoring the molecule to the membrane [296]. The products of the *chvA*, *chvB* and *pscA* genes are all essential for virulence and involved in the biosynthesis or processing of β-1,2-glucan [51, 301].

Cyclic β-1,2-glucans are only found in *Agrobacterium* and *Rhizobium* species [294] and the *chvA*, *chvB* and *pscA* genes have homologous and functionally interchangeable

counterparts in *Rhizobium meliloti* [77]. This conservation and the ability to render *R. leguminosarum* strains tumourigenic by the introduction of a Ti plasmid to it [124] suggests there is a general chromosomally-encoded plant-attachment process among Rhizobiaceae prior to more specific interactions occurring.

1.18 The Ti plasmid

Ti plasmids are classed into four groups dependent on the opines produced by the tumour they induce. They range in size from 190-240kb and have a low copy number (1-3) in the cell [194, 329]. They consist of five distinct regions: the T-DNA, the *vir* region, the opine metabolism region, a region for bacterial conjugation and one for replication and incompatibility (see figure 1.18.1) [30]. The regions required for pathogenicity are the *vir* genes and the T-DNA (including its flanking border sequences). The features of these regions will be discussed. The Ti plasmid also contains the genes responsible for opine-stimulated-conjugal transfer [233].

The opine metabolism region encodes proteins that enable the infecting *Agrobacterium* strain to take up and catabolise opines produced by the tumour [208]. The region is under the control of an opine-inducible promoter [329, 339].

Agrobacteria also possess cryptic "Megaplasmids" whose functions are unknown [311]. Curing *A. tumefaciens* C58 of the 410kb cryptic plasmid pAtC58 has no effect on the virulence of the bacterium [251].

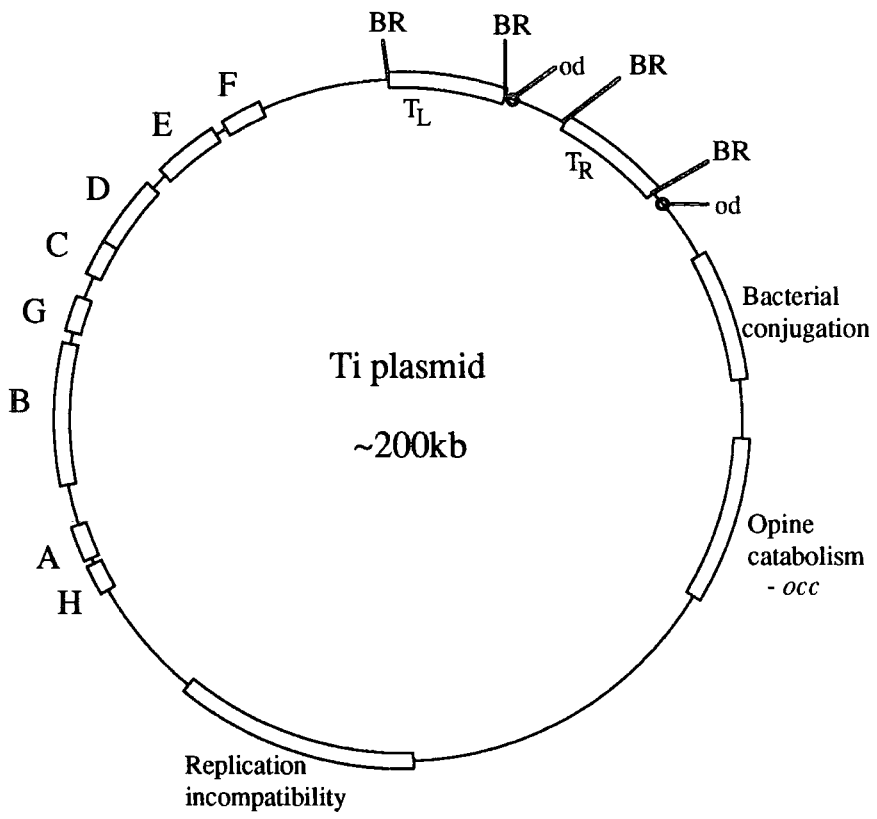
The T-DNA can be found in two separate regions, such as the octopine-type Ti plasmid in figure 1.18.1, or as a continuous region in other opine producing Ti plasmids. In both cases a highly conserved set of genes are present which are responsible for phytohormone production and hence the tumourous phenotype in transformed plant cells [326]. A new auxin biosynthetic pathway is established by the T-DNA genes *iaaM* and *iaaH* [260, 312] and *iptZ* expression brings about cytokinin production [48]. The overproduction of these plant growth regulators leads to the unorganised and uncontrolled growth of plant cells. The T-DNA also encodes opine biosynthetic enzymes and a gene encoding a transport protein allowing excretion of opines from the plant cells [195].

T-DNA regions are flanked by 25bp direct repeat sequences, called the border repeats [120, 333]. The right border repeat is essential for T-DNA transfer [268]. The overdrive sequence, a 24bp T-DNA transfer enhancer is found to the right of the right

Figure 1.18.1 A genetic map of an octopine-type Ti plasmid.

The figure was adapted, with permission, from Dr. M. H. Levesley [167] and shows the approximate locations of the *vir* genes, the T-DNA and border repeats. The relative positions of the (approximate) locations of the loci for plasmid replication and incompatibility, bacterial conjugation and opine catabolism are also shown.

In octopine-type plasmids, the 13kb left T-DNA (TL) encodes the genes needed for plant transformation and octopine synthesis, whilst different opine biosynthetic genes are encoded by the 7.8kb right T-DNA (TR) [243]. Nopaline-type Ti plasmids have a single 22kb continuous T-DNA [243]. The *vir* regions of Ti plasmids are very similar, although nopaline-type Ti plasmids don't contain the *virF* or *virH* loci, see text.



The figure is not to scale. BR = border repeat; od = overdrive sequence.

border repeat in octopine Ti plasmids [231]. The products of the *vir* genes bring about T-DNA transfer to the plant cell (see section 1.19), the early stages of which are thought to be analogous to bacterial conjugation [329]. The target sites in the plant genome for T-DNA integration appear to share significant homology to the T-DNA border repeats and active transcriptional regions are preferred [153, 190]. T-DNA gene expression is performed by the plant cell and their promoters possess typical features of plant promoters (CAAT and TATA boxes) [329].

T-DNA structure and function and the roles of the *vir* gene products are described in more detail in the following reviews [30, 328, 329, 337, 338].

1.19 The *vir* region and functions of the Vir proteins

The *vir* region is (approximately) 35kb in length and encodes over twenty proteins which are required for T-DNA transfer [30]. Only *virA* and *virG* are monocistronic. The other loci consist of operons encoding numerous proteins. The gene products of the *virA*, *virB*, *virD* and *virG* loci are absolutely required for tumour formation, whilst the *virC* and *virE* (and *virF* and *virH* in octopine Ti plasmids) are the "host range loci" that are necessary for normal tumour development on certain plants only [30]. The *virA* and *virG* genes are constitutively expressed and activate transcription of the other *vir* loci (and themselves) if plant-derived inducing factors are present [329]. The Vir proteins bring about the copying and transfer of the T-DNA.

The border repeats of the T-DNA are nicked by VirD2 after the DNA has been relaxed by the topoisomerase VirD1 [338]. These nicks allow the synthesis of a single stranded section of DNA (T-strand) from the T-DNA that has been unwound [282]. The unwinding is possibly aided by the VirD1 and VirD2 proteins, as well as the new synthesis itself [282, 338]. In octopine Ti plasmids the VirC1 and VirC2 proteins bind to overdrive and enhance the formation of the T-strand [329].

The T-strand enters a plant cell as part of a protein/DNA structure called the T-DNA transfer complex [30]. The single stranded DNA of the T-strand has a VirD2 protein bound at its 5' end and is coated with VirE2 [125]. The VirE2, by binding to the T-strand, is thought to protect it from bacterial and plant cell nucleases [57]. The bound VirD2 may function as a pilot protein for the T-DNA transfer complex within the plant cell, as it has two nuclear targeting signals [116]. The T-DNA transfer complex is thought to leave the bacterial cell through a channel formed by the eleven VirB proteins [303].

Many of the *vir* gene loci contain open reading frames whose products have no clear functions. The *virF* and *virH* loci are only found on octopine Ti plasmids, and affect T-DNA transfer depending on the host plant species [329]. The function of VirF is unclear, although it may be responsible for the production of a secreted factor or it may function within the plant cell [203, 264]. The *virH* loci (also called *pinF*) encodes gene products which may function to counteract bactericidal plant products encountered during the infection process [338].

1.20 *vir* gene induction and regulation of expression

Several different environmental factors determine the extent of induction of the *vir* genes: certain low-molecular weight phenolic compounds, pH, some monosaccharides, phosphate levels and some stress stimuli [186]. The level of expression of the *vir* genes is dependent on the pool size of VirG [329]. VirG is a transcriptional activator that positively regulates the *vir* genes [283, 331]. Two promoters, P1 and P2, control the amount of VirG in the cell [327]. The upstream P1 promoter is induced by low phosphate levels and phenolic compounds, whereas P2 is induced by acidic media [327]. The P2 promoter is also induced by other stimuli stressful to cell growth, such as heavy metal ions, and its sequence is similar to the consensus heat shock promoter of *E. coli* [186].

Phosphate levels are often low in soil and this will induce the P1 promoter [327]. The chemotaxis of *A. tumefaciens* towards wound sites (see section 1.21) will activate the P2 promoter, since wound sites are acidic [138]. A functional ChvD protein is required for maximal induction from both these stimuli [332]. The P2 promoter may be further activated by phytoalexins, antimicrobial compounds produced by some wounded plants [186]. The activation of these two promoters should provide sufficient VirG to allow successful induction, when the main signals for the virulence process are detected.

The VirA protein is a transmembrane environmental sensor that detects signals released from the plant [329]. Upon detection, VirA autophosphorylates and then phosphorylates VirG, which causes an increase in *vir* gene expression (see section 1.21). The VirA and VirG proteins are members of the two component sensor-regulator family of control systems, described earlier, that use phosphorylation in intracellular signalling [250]. VirA detects the presence of specific monosaccharides, most of which are plant metabolites or monomers of plant cell wall polysaccharides [11]. Wound-specific phenolic compounds synthesised and secreted by the plant are also detected. The first such *vir*-inducer to be detected was acetosyringone, a monocyclic aromatic hydrocarbon, but a large number of related compounds have now been shown to have inducing activity [279, 281]. Some of

these compounds are intermediates for cell repair biosynthetic pathways and for cell wall repair [104, 108]. Thus *A. tumefaciens* has evolved a system for detection and targeting those cells wounded and undergoing repair [279].

VirA does not bind directly to the extracellular stimuli. ChvE, a periplasmic binding protein, interacts with the *vir*-inducing monosaccharides and then the periplasmic region of VirA [50]. The phenolic compounds may be detected by the p10 and p21 periplasmic proteins which possibly then interact with VirA [52, 165]. The presence of opines potentates the induction of the *vir* genes by acetosyringone [313].

VirA autophosphorylates following the putative interactions of the phenolic-bound p10 and p21 proteins [128, 134]. The autophosphorylation is enhanced by the interactions of monosaccharide-bound ChvE with VirA [50]. The kinase properties of VirA transfer the phosphate group to VirG which then brings about an increase in transcription of the *vir* genes [133, 134]. Phosphorylated VirG binds to consensus sequences in the *vir* gene promoters (*vir* boxes) to activate transcription [228, 229]. Each *vir* gene possesses upto five *vir* boxes, although expression of the gene does not require all of them [228]. VirG is able to bind to *vir* boxes in either its unphosphorylated or phosphorylated form, however only the phosphorylated form brings about transcriptional activation possibly by converting the closed RNA polymerase-promoter complex to its open form to allow transcription [133]. VirG also activates transcription of the *virA* and *virG* genes to form a positive regulatory loop, committing the bacterium to the infection process [329].

The chromosomally-encoded Ros protein negatively regulates the *virC* and *virD* loci [58]. It does this by binding to the *virC* and *virD* promoters that overlap the *vir* boxes [60].

vir gene induction involves the interaction of several signalling pathways and is a complex process. A plant wound site provides the environmental stimuli necessary for induction and the availability of *vir* inducers is a reason why the wounding of plant cells is a pre-requisite for *A. tumefaciens* infection. The importance of chemotaxis towards the plant wound site (and *vir*-inducing molecules) and the possible involvement of VirA, VirG and ChvE in this process is discussed in the next section.

1.21 The importance of chemotaxis in *Agrobacterium*

The molecular events of the *Agrobacterium* infection process, in particular T-DNA transfer and tumour formation, have been extensively studied. However the chemotactic ability of *Agrobacterium* to locate itself initially within the rhizosphere and then at plant wound sites has been relatively poorly characterised. A study of *Agrobacterium* chemotaxis has been initiated, and results have led to the proposal of a mechanism by which *Agrobacterium* detects and becomes concentrated at wound sites in sufficiently high numbers to cause infection [19, 20, 172, 223, 265].

Loake *et al.* showed the avirulent *A. tumefaciens* strain C58C1 to have a highly sensitive chemotaxis system, responding to a number of sugars abundant in plant extracts [172]. The most intense and sensitive response was to sucrose, the commonest translocated plant sugar. The monosaccharides studied were detected with equivalent sensitivity, but with a lower response magnitude, and in general oligosaccharides were better attractants than their component monosaccharides [172]. This greater response to oligosaccharides may be a result of their breakdown in the periplasm and then the synergistic effects of the component monosaccharides [172]. The inability to metabolise carbohydrates in the periplasm or the impermeability of the outer membrane could be reasons for some compounds being non-attractants [172]. The only amino acids that generated a chemotactic response were valine, arginine and glutamic acid, and in general amino acids were found to be relatively poor chemoattractants [172].

ChvE has been shown to be required for chemotaxis towards some sugars [50]. As these responses were observed in the Ti plasmid-cured strain C58C1, VirA cannot be the signalling molecule for this chemotaxis [172]. It has been suggested instead that the sugar-bound ChvE protein interacts with a Trg-like protein [50].

The release of fixed carbon molecules into the rhizosphere by plant cells can affect the microbial population in the vicinity of the root [110]. Agrobacteria have been shown to accumulate around roots [261] and the maintenance of agrobacteria within the rhizosphere is probably a result of its sensitive chemotactic responses to plant sugars.

A. tumefaciens chemotaxis to plant wound exudate-specific phenolic compounds was also determined [18, 19]. Two classes of chemoattractants were identified, as well as compounds that gave no chemoattraction. The first group required the presence of a Ti plasmid and these compounds; acetosyringone, sinapinic acid and syringic acid elicit the strongest response [18, 19, 20]. The second group of chemoattractants had to be present at much higher concentrations to elicit a response and this response did not require a Ti

plasmid [18, 19]. It was proposed that two types of chemotaxis occur in response to wound-specific compounds - one that is weak and chromosomally-encoded, and another which is much more sensitive and specified by the Ti plasmid [18, 19].

The acetosyringone concentration required for chemotaxis is one hundred times lower than that giving maximal *vir* gene induction [19, 281]. Thus strains containing Ti plasmids could be attracted from the rhizosphere and localised at a wound site by phenolic compounds, which would then induce the *vir* genes.

The Ti plasmid loci required for chemotaxis towards acetosyringone were shown to be *virA* and *virG*, suggesting a multifunctional role for these two proteins [223, 265]. Furthermore, phosphorylation of VirA and VirG has been shown to be necessary for this chemotaxis towards acetosyringone [223]. Presumably this chemotactic response involves interactions with the chromosomally-encoded chemotaxis system. It has been postulated that phosphorylated VirG has a greater affinity for part of the chemotaxis pathway and thus the small amounts of phosphorylated VirG produced at low concentrations of phenolics would be preferentially involved in chemotactic signalling [223]. As the concentration of phenolic-inducers increases, the interaction with the chemotactic component becomes saturated and phosphorylated VirG now activates the *vir* genes [223]. VirA has homology to CheA and VirG to CheB and CheY [289] but the exact interactions and function in chemotaxis have yet to be elucidated.

Another group using a less sensitive assay, were unable to demonstrate that acetosyringone is a chemoattractant for *A. tumefaciens* and suggested that chemotaxis to other *vir* inducers assayed is a chromosomally encoded phenomenon [224]. The difference in results could be explained by the differences in *A. tumefaciens* strains used - the strain used in the above work has been observed to be less vigorously motile than C58C1 [19].

Despite these results and speculation over the exact mechanism of chemoattraction to phenolic compounds, chemotaxis is undoubtedly an important part of the infection process. Root and shoot extracts of both monocotyledonous and dicotyledonous plants have been shown to be chemoattractants for *A. tumefaciens* [19]. A requirement for chemotaxis in the pathogenicity of *A. tumefaciens* on pea roots in soil has been shown [111]. However environmental conditions, such as soil type, are also important for the chemotactic ability of *A. tumefaciens* [329]. Mutants non-chemotactic to root exudates were avirulent on soil-grown plants, but fully virulent on plants grown in sand [111]. A mutant attracted to wounded cells but non-chemotactic towards root exudates was also avirulent in soil, suggesting movement towards wound products is of secondary importance to attraction by naturally released carbohydrates [111]. The probable major role of chemotaxis in

pathogenicity is to maintain high numbers of bacteria within the rhizosphere, with any attraction to wounded cells likely to be important only on a micrometre scale [266].

1.22 Motility and chemotaxis of rhizosphere bacteria

Numerous plant pathogenic and soil borne bacteria have been shown to be attracted to plant roots, plant extracts or components of plant extracts in laboratory experiments [112]. However studies to show the biological significance of these attractions in soil are limited, and thus the importance of chemotaxis in the movement of bacteria in soil and the colonisation of plant roots is unclear. Chemotaxis and motility have been shown to be necessary for the efficient spreading of *R. meliloti* in soil [275] and flagella were shown to be required for potato colonisation by *Pseudomonas fluorescens* [63]. Another study [126] showed wheat roots could be colonised by non-motile *P. fluorescens*. No correlation between motility and root colonization for *Pseudomonas* and *Serratia* spp. has also been demonstrated in another study [256]. The requirement for chemotaxis may have been abolished in this latter study as the bacteria were inoculated directly onto the seeds. Differences in experimental design, inoculation methods, soil type and estimation of bacterial numbers in the rhizosphere may account for some of the contrasting data.

The chemotactic involvement in *Agrobacterium* infection (see section 1.21) and *Rhizobium* nodulation events have been studied and it does appear to be important under field conditions. Non-chemotactic mutants of *Rhizobium* spp. can nodulate host roots, but their nodulation efficiency compared with wild-type strains is reduced [10, 49].

The postulated chemotactic events leading upto nodulation for rhizobia are parallel with those previously described for *A. tumefaciens* infection. Rhizobia become localised at specific regions of roots as a result of chemoattractants within plant root exudates [61, 94, 107]. The chemoattractants are secreted by cells in the zone of elongation and thus rhizobia aggregate at the regions of the root most susceptible to infection [29]. Chemoattractants include amino acids and sugars as well as *nod*-gene inducing flavones, such as luteolin and naringenin [4, 49]. Chemotaxis towards luteolin secreted from roots is thought to occur at low concentrations, and as rhizobia move towards the roots and the luteolin concentration increases, chemotaxis diminishes and the *nod* genes are rapidly induced [232].

1.23 General features of *A. tumefaciens* motility and chemotaxis

As discussed in section 1.14 (and shown in figure 1.14.1) *A. tumefaciens* has an unusual flagellation pattern of two polar flagella and 2-4 lateral flagella. All the filaments show a typical sinusoidal curvature in electron microscopes. *A. tumefaciens* is a motile bacterium, presumably as a result of these flagella, and will swarm in semi-solid (swarm) agar plates. Enrichment of motile populations can be achieved by successive sub-culturing on swarm agar plates, see section 2.5, for the analysis of motile cells.

The strain C58C1 was observed microscopically to be an active swimmer, in comparison to other strains (such as LBA4301 and A136) which were relatively poorly motile [19, 172]. Subsequent investigations on *A. tumefaciens* chemotaxis and motility used strain C58C1 [171] and all the *A. tumefaciens* strains used in this work are derivatives there of.

The motility of C58C1 is characterised by long straight or curved runs rarely interrupted by sudden changes in direction and tumbling motions (as are seen in *E. coli*). Furthermore the direction of rotation of tethered cells, attached by their flagella to glass coverslips or slides, suggest that the flagella rotate unidirectionally in a clockwise direction [171, 172]. These behavioural features and others observed for *A. tumefaciens* strongly resemble those of *R. meliloti* [102] and are distinct from those of *E. coli*.

Chemoattractants for *A. tumefaciens* abundant in plant extracts were described earlier. The use of avirulent C58C1 demonstrated the chromosomally encoded chemotactic system to certain sugars.

1.24 Analysis of chemotactic and motility genes in *A. tumefaciens*

The MCP chemotactic signalling system found in enteric bacteria has been suggested to be used by most motile bacteria [214]. Several pieces of experimental evidence imply that *A. tumefaciens* also possesses MCP-like proteins. Loake demonstrated the involvement of methylation for chemotaxis of *A. tumefaciens* towards sucrose [171]. Antibodies to the *E. coli* MCPs Trg and Tar both cross-reacted with protein extracts from *A. tumefaciens* [47, 201]. The proteins detected in *A. tumefaciens* were approximately the same size as the MCPs found in *E. coli* [47]. Recently Farrand and his co-workers have sequenced an open reading frame that could encode a protein with 75% identity to McpA of *C. crescentus* [82].

A number of chemotactic mutants were created using the transposon Tn5 to study in more detail the genes involved in *A. tumefaciens* chemotaxis [171, 267]. The creation of the mutants, their initial characterisation and the strategies used to clone wild-type DNA complementing the mutants are described in more detail in [47, 171, 267].

Twenty chemotactic and motility mutants were isolated and these were analysed on swarm agar plates and by light and electron microscopy. Preliminary phenotypes were assigned to the mutants. There were seven *mot* and seven *fla* mutants. The only *che* mutant isolated (designated *che-2*) tumbled continuously under microscopic examination. It is likely to be a general non-chemotactic mutant as the swarm plates used during its isolation contained a large number of attractants. Four mutants demonstrated normal chemotactic behaviour to L-broth in chemotactic assays, but only produced intermediate sized swarms on swarm-agar plates. The final mutant had a phenotype of paired cells - extracellularly joined at the cell poles.

All twenty mutants were shown to be virulent after the introduction of a Ti plasmid. The bacteria were inoculated directly onto sunflower seedlings to circumvent any need for chemotaxis and motility in the infection process. The inability of the *mot-1* mutant to colonise newly grown roots was demonstrated using the method of de Weger *et al.* [63] which supports the idea that chemotaxis is required for rhizosphere colonisation [111].

A representative genomic library of C58C1 genomic DNA was constructed in the cosmid pLAFR3 [267]. Each of the cosmid clones had a genomic insert of 20-25kb and 1200 separate clones were maintained separately [267]. The library, in conjunction with the chemotactic and motility mutants created, was used to isolate DNA encoding chemotactic and motility functions. The insertional inactivation of a behavioural gene by Tn5 allows flanking sequences corresponding to the mutated gene to be isolated. This "molecular tagging" is possible as Tn5 encodes resistance to the antibiotic kanamycin. The procedure used is described by Shaw [263] and the isolated flanking sequences can be radiolabelled and used as probes to screen the genomic library to recover cosmids containing homologous DNA sequences [267]. The cosmid clone identified can then be used in complementation tests on the mutant, and any others found to hybridise to the same cosmid. A cosmid clone, pDUB1900, was isolated in this manner [267]. pDUB1900 either hybridised to or complemented nine of the mutants: *mot-1, mot-4, mot-9, fla-10, fla-11, mot-12, mot-14* and *fla-15* (see figure 1.24.1). Originally *mot-12* and *mot-14* were thought to have two separate Tn5 insertion sites, mapping very close together, that gave slightly different phenotypes [267]. DNA sequencing around both insertion sites showed them both to have the same Tn5 insertion site and thus the *mot-14* nomenclature was dropped (this study).

Figure 1.24.1 The cosmid clone pDUB1900.

The top line represents the *A. tumefaciens* chromosomal DNA from which the chromosomal DNA of pDUB1900 was taken, and the *Eco*RI (E) and *Hpa*I (H) sites upon it. The boxes below this show the positions of *Bam*HI, *Eco*RI and *Hind*III sites within the chromosomal insert of pDUB1900. The lines below depict the extents of the cloned fragments from the various mutants, the Tn5 insertion in each is denoted with an arrow head. The horizontal arrows and labels represent the known extents of putative flagellar genes found, see text. The diagram was redrawn with permission from Dr. C. H. Shaw.

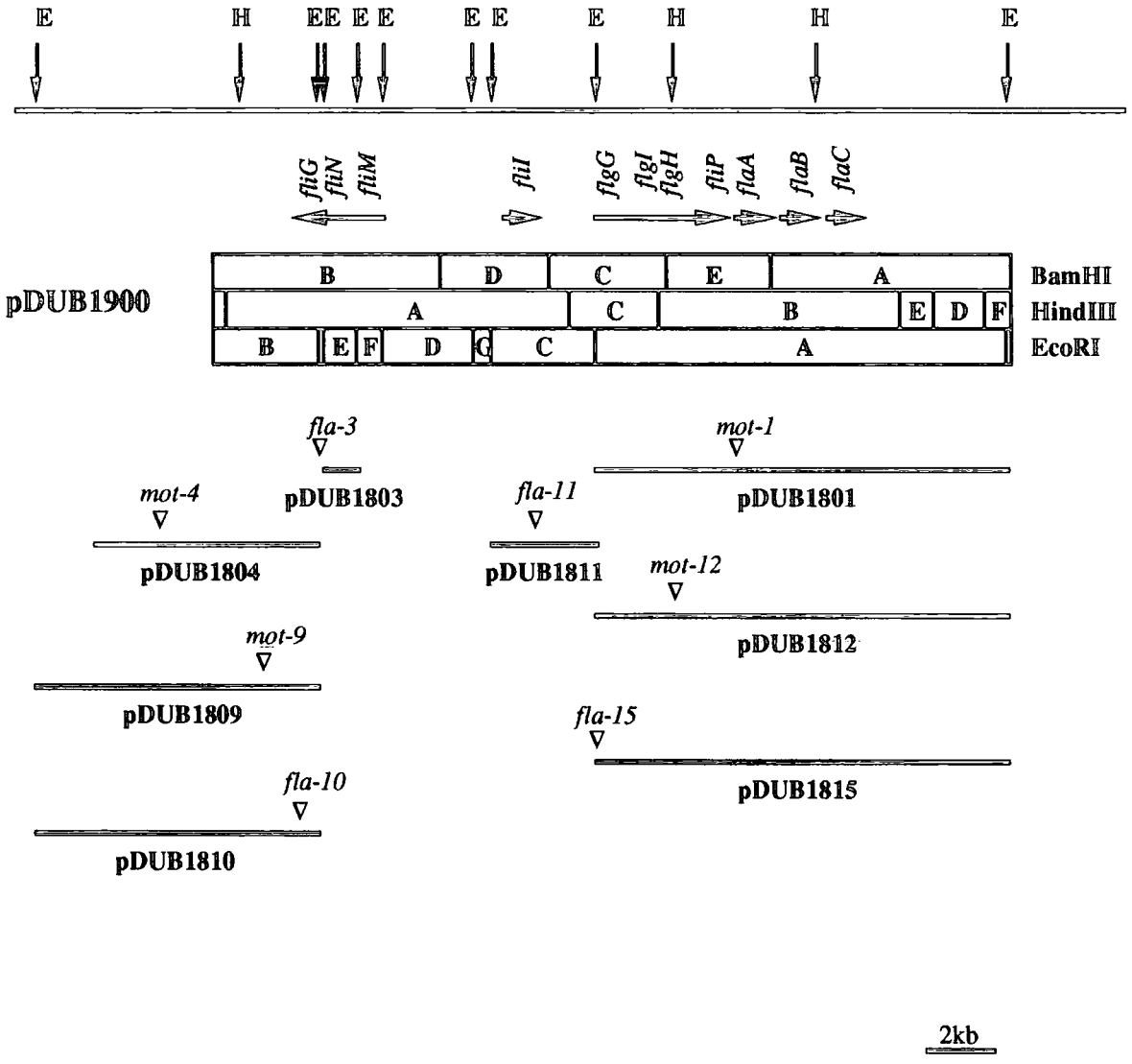


Figure 1.24.1 also shows the regions of pDUB1900 that have been sequenced and the flagellar gene homologues identified. Homologues of the flagellar switch genes *fliG, fliM* and *fliN* were identified around the insertion sites of the *fla-3* and *fla-10* mutants (C. H. Shaw, unpublished results). Partial DNA sequence homologous to the *fliI* gene of *E. coli*, that encodes a putative ATPase thought to be involved in the flagellum-specific export pathway, was also identified near the *fla-11* Tn5 insertion site (W. J. Deakin & C. H. Shaw, unpublished data). Other flagellar gene homologues were identified during the course of this study and will be discussed in chapters 3, 4 and 5.

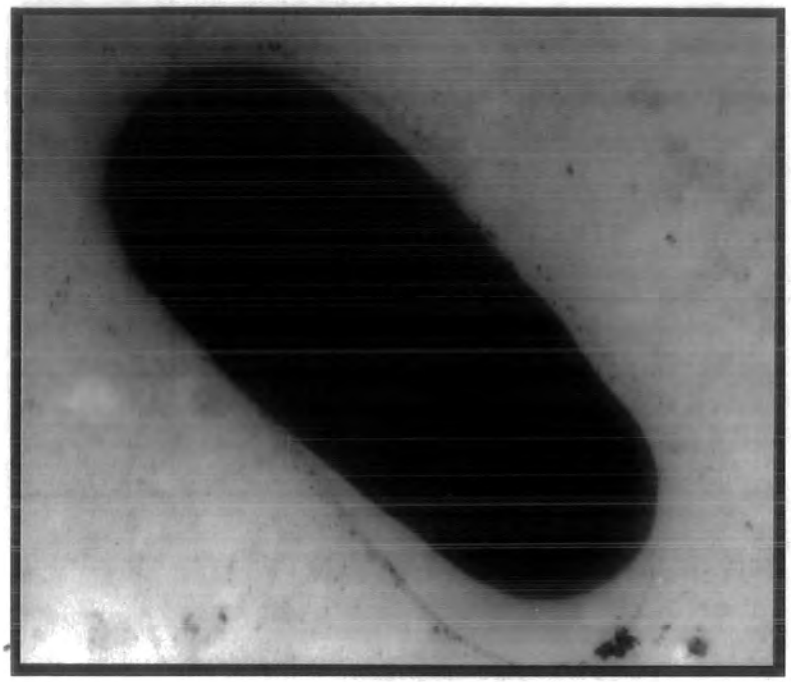
Three more of the mutants, *che-2, mot-6* and *fla-8*, were shown to map onto and be complemented by a separate cosmid, pDUB1905 [47]. Partial DNA sequencing around these Tn5 insertion sites revealed DNA homologous to *fliR* of *E. coli* for the *mot-6* and *fla-8* mutants (which map close to each other) and to *cheL* of *C. crescentus* for the *che-2* mutant (S. P. de Martino, A. P. Brown, W. J. Deakin and C. H. Shaw, unpublished results). Both *fliR* (of *E. coli*) and *cheL* (of *C. crescentus*) encode proteins of unknown functions in the chemotaxis and motility systems of their respective bacteria [148, 184].

1.25 Aims of this study

The aims of this work were to characterise three of the behavioural mutants: *mot-1*, *mot-12* and *fla-15*. The Tn5 insertion sites of these mutants all mapped to the pDUB1900 cosmid, within 5kb of each other, see figure 1.24.1. Electron micrographs of the three different mutants are shown in figures 1.25.1 and 1.25.2. The phenotypes of the three mutants are distinct. *fla-15* is non-flagellate, *mot-1* has short, straight, truncated filaments and *mot-12* possesses (apparently) fully synthesised flagella that are paralysed.

As discussed earlier a fourth mutant (*mot-14*) originally thought to be phenotypically distinct and to map within this region close to the *mot-12* Tn5 insertion site was described [267]. Electron micrographs of *mot-12* and *mot-14* apparently showed them to have different phenotypes. *mot-12* appeared to lack the polar tuft pair of flagella and possess only lateral ones and *mot-14* to lack lateral flagella and only possess the polar pair of flagella [267]. An original goal of this project was to investigate this putative non-equivalence in lateral and polar flagella, however DNA sequencing showed the "two" mutants to share a common Tn5 insertion site. A single mutant was presumably duplicated at some stage of the isolation procedure and the apparent difference in phenotypes were due to electron micrograph artefacts. Despite this, non-motile mutants with fully synthesised but paralysed flagella are thought to be unusual, especially in the enteric bacteria [150], where such phenotypes are only due to mutations in switch or motor proteins.

Figure 1.25.1 Electron micrographs of the *fla-15* and *mot-1* mutants.



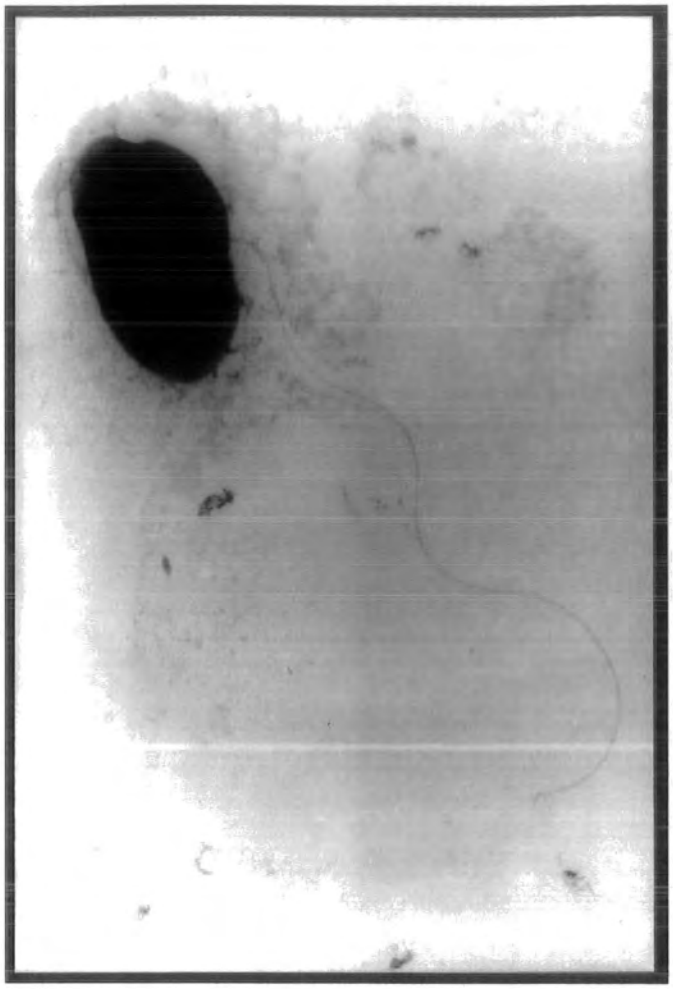
1μm

fla-15

1μm

mot-1

Figure 1.25.2 Electron micrograph of the *mot-12* mutant.



1μm

mot-12

The region of pDUB1900 to which the Tn5 insertion sites mapped was initially to be sequenced and depending upon the results of this and any flagellar gene homologues identified, further characterisations would be carried out. The proximity of the Tn5 insertion sites on pDUB1900 meant any chemotaxis and motility genes identified in this region could be regulated together as part of an operon. As described earlier chemotaxis and motility genes are often found within large operons. Thus it is possible other chemotaxis and motility genes could be identified and then subjected to further characterisation.

Phenotypes resembling the *mot-1* mutant had been observed before in *R. meliloti* and *E. coli* and had been found to be caused by mutations within the gene encoding the flagellin protein [27, 36, 235]. In the case of *R. meliloti* the mutation was in one of the two copies of the flagellin genes it possesses (see chapter 5). Sequencing around the region of the *mot-1* Tn5 insertion site may thus reveal multiple copies of the flagellin genes. These could then be investigated further by genetic and biochemical methods to reveal whether the similarities in motile behaviour are reflected by similarities at the molecular level.

2. Materials and Methods

2.1 Materials

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

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

Caesium chloride was from Boehringer Mannheim (U.K.) Ltd., Lewes, U.K..

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

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

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

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

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..

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

Ficoll 400 and Sephadex G-50 were from Pharmacia Fine Chemicals, Uppsala, Sweden.

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

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

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

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

2.2 Bacterial strains and plasmids

2.2.1 *E. coli* strains

DH5α *supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1* (lab stock).
GM119 *dcm-6 dam-3 metBI thi-1 lacY1 galK2 galT22 mtl-2 tonA2 tsx-78 supE44* (lab stock).

2.2.2 *A. tumefaciens* strains

C58C1 rifampicin resistant (lab stock).
GM19023 cured of cryptic plasmid [251].
C58C1 *mot-1* non-motile Tn5 insertion mutant [171].
C58C1 *mot-12* non-motile Tn5 insertion mutant [171].
C58C1 *fla-15* Tn5 insertion mutant with no flagella [171].
MAN1 non-flagellated mutant of C58C1 due to the insertion of a neomycin cassette in the *fliP* homologue of *A. tumefaciens* (this study).

2.2.3 *Rhizobium* strains

***R. meliloti* 1021** a streptomycin resistant derivative of Su47 (lab stock).
Su47 *che-1* a non-chemotactic mutant of Su47 [341].
Su47 *che-3* a non-chemotactic mutant of Su47 [341].
Su47 *fla-1* a non-flagellated mutant of Su47 [341].
R. leguminosarum* biovar *viciae sym+ and sym- strains (lab stock).
***R. leguminosarum* biovar *phaseoli* 8401** (lab stock).
***R. lupini* H13-3** [155].

2.2.4 *Pseudomonas* strains

P. reactans (lab stock).
P. talassii (lab stock).
P. rhodos possesses both complex and plain flagella [259].

2.2.5 Plasmids

pUC19 general *E. coli* vector, Amp^R [336].
pRK2013 helper plasmid for tri-parental matings in *A. tumefaciens*, has the ColE replicon containing the transfer function of RK2, Km^R [85].

Recombinant plasmids containing cloned behavioural genes.

pRZ1, pRZ2 and pRZ4 pLAFR-1 broad host range vector containing cloned behavioural genes of *R. meliloti*, Tc^R [341].
pDUB1900 a C58C1 cosmid library clone containing a *Bam*HI chromosomal fragment in the vector pLAFR-3 (carries cloned behavioural genes) Tc^R [171, 267].
pWJD1 pUC19 containing *Bam*HI fragment E from pDUB1900, Amp^R (this study, see figure 3.2).
pWJD2 pUC19 containing the ~2kb *Bam*HI-*Eco*RI fragment from *Bam*HI fragment C of pDUB1900, Amp^R (this study, see figure 3.2).
pWJD3 pUC19 containing the ~4kb *Bam*HI-*Hind*III fragment from *Bam*HI fragment A of pDUB1900, Amp^R (this study, see figure 3.2).
pWJD4 pUC19 containing the ~2kb *Bam*HI-*Dra*I fragment from *Bam*HI fragment E of pDUB1900, Amp^R (this study, see figure 3.2).

Recombinant pUC plasmids containing Tn5 plus flanking sequences from the mutants *mot-1*, *mot-12* and *fla-15*, respectively.

pDUB1801 Amp^R, Km^R [171].
pDUB1812 Amp^R, Km^R [171].
pDUB1815 Amp^R, Km^R [171].

Plasmids used for the gene replacement mutagenesis of *A. tumefaciens fliP*.

pJQ200SK basic vector, possesses *sacB* and Gm^R [240].
pUTD26 pJQ200SK with *Pst*I to *Xho*I of the multiple cloning site removed (this study).
pUTD27 pUTD26 carrying the *Bam*HI-*Dra*I fragment of pWJD4 (this study).
pDUB1100 neomycin resistance cassette of Tn903 cloned into the *Eco*RI site of pUC1813 (lab stock).
pUTD29 pUTD27 with the neomycin cassette of pDUB1100 inserted at the *Sal*I site of *fliP* from *A. tumefaciens* (this study).

2.3 Bacterial growth media, conditions and procedures

The following media were used in this work.

LB broth

10g.l^{-1} trypticase peptone, 5g.l^{-1} yeast extract, 5g.l^{-1} NaCl.

Lab M nutrient broth no.2 (LM broth)

25g made up to 1 litre with distilled water gives final concentrations of 10g.l^{-1} beef extract, 10g.l^{-1} balanced peptone no.1, 5g.l^{-1} NaCl, pH 7.5 ± 0.2.

Chemotaxis media [2]

0.1mM EDTA pH 7.0, 10mM phosphate buffer pH 7.0. (Phosphate buffer is 1M KH_2PO_4 adjusted to pH 7.0 with 1M KOH.)

Rhizobium initiation media [4]

5g.l^{-1} trypticase peptone, 3g.l^{-1} yeast extract, 1.3g.l^{-1} $CaCl_2.2H_2O$.

MinA media [197]

20ml 5x MinA salts, 1ml 20% D(+) galactose, 0.1ml 1M $MgSO_4$ made up to 100ml with distilled water. (5x MinA salts contain 105g.l^{-1} K_2HPO_4, 45g.l^{-1} KH_2PO_4, 10g.l^{-1} $(NH_4)_2SO_4$, 5g.l^{-1} Na.citrate.$2H_2O$.)

If necessary, agar was added at a concentration of 1% for solid plates or 0.16% for swarm plates, unless ready-made Lab M nutrient agar (LM agar) was used.

Lab M nutrient agar (LM agar)

28g made up to 1 litre with distilled water, gives final concentrations of 5g.l^{-1} peptone, 3g.l^{-1} beef extract, 8g.l^{-1} NaCl, 12g.l^{-1} agar no.2, pH 7.3 ± 0.2.

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

For *E. coli*; ampicillin 50μg.ml^{-1}, gentamycin 15μg.ml^{-1}, kanamycin 50μg.ml^{-1}, neomycin 50μg.ml^{-1} and tetracycline 15μg.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, 40μl of 20mg.ml^{-1} X-gal (in DMF) were spread over the surface of agar plates.

For *R. meliloti*; as *A. tumefaciens* and streptomycin 100μg.ml^{-1}.

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

Liquid bacterial cultures were incubated on an orbital shaker at 200rpm at temperatures of 37°C for *E.coli* and 28°C for *Agrobacterium*, *Rhizobium* and *Pseudomonas* strains. *Rhizobium* strains were grown in *Rhizobium* initiation media and *Pseudomonas* strains in MinA media. Short term (1-2 months) stocks of cultures were kept at 4°C on solid agar plates. Long term stocks were kept in 50% glycerol at -80°C.

Liquid cultures were inoculated with a flamed loop or a sterile cocktail stick. Solutions and bacterial cultures were spread onto agar plates using a glass spreader which had been sterilised in 70% ethanol.

Aseptic technique was used throughout with bacterial cultures. All glassware, plasticware and other equipment were sterilised by autoclaving at 121°C, 15p.s.i. for 15 minutes. Generally solutions were prepared according to Sambrook *et al.* [253], and autoclaved as above if possible. Otherwise solutions were filter-sterilised through a 0.22μm nitrocellulose filter (Sartorius) into a sterile container.

2.4 Bacterial growth measurement

The optical density of bacteria was measured at 600nm on a Beckman DU7500 spectrophotometer. The optical density was compared against standard growth curves of the changes in optical density over time and changes in cell number over time (Dr. M. H. Levesley, personal communication) to estimate the number of bacteria in the sample.

2.5 Preparation of motile bacteria

Motile cells within a bacterial strain were obtained periodically by inoculating a loopful of bacteria onto the centre of a MinA swarm agar plate. The plate was incubated at the optimum growth temperature for the bacteria. Motile bacteria moved towards the edge of the plate, thus by the inoculation of another swarm plate with a loopful of bacteria from the edge of the swarm, motile bacteria were selected. The swarm plates were usually incubated for two days, although this varied depending on the strain and how quickly it swarmed. New plates were always inoculated before the swarm reached the edge of the plate. The procedure was repeated at least three times before the bacteria were streaked to

single colonies on a master plate with antibiotic selection. Single colonies were taken from this plate and used for later procedures such as RNA or flagella isolation.

2.6 The 3-keto-lactose assay for *Agrobacterium*

The method used was that of Bernaerts and De Ley [28]. Putative *Agrobacterium* colonies were streaked onto plates containing 2% $CaCO_3$, 2% glucose, 1% yeast extract and 1% agar and incubated overnight at 28°C. A loopful of the bacteria was streaked onto lactose plates, (1% lactose, 0.1% yeast extract and 1% agar) and incubated for two days at 28°C. The plates were then flooded with Benedict's reagent (173g sodium citrate, 100g sodium carbonate and 17.3g copper sulphate per litre) and left at room temperature for up to two hours. A yellow zone appears around colonies capable of converting lactose into 3-keto-lactose (a reducing sugar) in this case *A. tumefaciens*.

2.7 Conjugation of plasmids into *Agrobacterium*

Triparental matings based upon the method of Ditta *et al.* [70] were used to mobilise plasmids into *Agrobacterium* with pRK2013 as a helper plasmid. Cultures of the recipient *Agrobacterium* strain, and the *E.coli* plasmid donor and helper strains were grown to mid log phase. 100µl of each culture were mixed together in an eppendorf tube and the resulting 300µl pipetted onto a 0.22µm nitrocellulose filter disc on the surface of a LM-agar plate. This was incubated at 28°C overnight, and the disc then transferred to a universal bottle containing 10ml of 10mM $MgSO_4$. This was vortexed vigorously to wash the bacteria off the disc. Dilutions were made from the resulting cell suspension which were plated onto agar plates with antibiotic selection.

2.8 Complementation analysis of motility mutants

Wild-type copies of mutated genes (on plasmids or library cosmids) were introduced into various motility mutants by conjugation. The presence of the plasmid was checked via DNA minipreps (see section 2.10.1). When the motility mutant was an *Agrobacterium* strain, the 3-keto-lactose test (section 2.6) was also performed to confirm this. A sterile metal dissection needle was used to inoculate the strain onto a MinA swarm agar plate, as well as the corresponding plasmid-free mutant. This was to try and ensure approximately equal numbers of cells were inoculated. Overnight broth cultures of both strains were grown, and the needle, having been flamed and cooled, was dipped into the culture before

being stabbed to the bottom of the swarm agar plate, for each strain. The swarm plates were wrapped in Nesco film and incubated at the optimum growth temperature for the bacteria.

2.9 Microscopy

2.9.1 Light microscopy

A loopful of bacteria was resuspended in chemotaxis media and observed under phase contrast optics (using a Nikon Optiphot microscope.)

2.9.2 Electron microscopy

A 30µl sample of resuspended bacteria was allowed to settle on a Formvar-coated grid for 30 minutes before the excess chemotaxis media was blotted off. Grids were not allowed to dry out completely at this stage. The grid was then placed upside down in a 50µl drop of 1% uranyl acetate (w/v in 70% ethanol) for 30 seconds, washed by transferring to several successive drops of distilled water and then finally air dried. A Philips EM400 was used to view and photograph the grids.

Isolated flagellar filaments were negatively stained using a slightly modified method. A 20µl sample was pipetted onto a Formvar-coated grid and the excess media immediately blotted off. 20µl of 1% uranyl acetate was then pipetted onto the grid which was left for up to one minute, before the excess uranyl acetate was blotted off and the grid allowed to air dry. The grids were viewed and photographed as before.

2.10 Isolation of DNA

2.10.1 Alkaline lysis plasmid minipreps

This method, used to prepare small amounts of relatively pure plasmid DNA was according to Sambrook *et al.* [253].

A single colony of bacteria was grown overnight in 5ml of LM-broth containing the appropriate antibiotic selection. 1.5ml of this culture was pipetted into a sterile eppendorf tube and the cells harvested by centrifugation for 1 minute in a microfuge (MSE MicroCentaur, approximately 10,000g). The supernatant was discarded and the pellet resuspended in 100µl of ice-cold solution 1 (1% glucose, 10mM EDTA pH8.0, 25mM

Tris.HCl pH8.0) and left at room temperature for 5 minutes. 200µl of solution 2 (0.2M NaOH, 1% SDS) was then added and the contents of the tube mixed by gentle inversion. After 5 minutes on ice, 150µl of ice-cold solution 3 was added to the mixture and the tube vortexed briefly before being placed on ice for a further 5 minutes. (Solution 3 was prepared by adding 11.5ml of glacial acetic acid to 28.5ml of distilled water and then adding 60ml of 5M potassium acetate. The solution has an overall pH of 4.8 and is 3M wrt potassium and 5M wrt acetate.) The tube was then microfuged as before for 5 minutes to remove bacterial debris. The supernatant was transferred to a fresh tube and extracted with an equal volume of TE (10mM Tris.HCl pH8.0, 1mM EDTA pH8.0) - saturated phenol:chloroform:isoamyl alcohol (25:24:1), see section 2.11.2. After a 2 minute spin in the microfuge, the aqueous phase was transferred to a fresh tube and the DNA precipitated by the addition of 2 volumes of ethanol. The tube was left for 5 minutes at room temperature and the DNA collected by centrifugation for 5 minutes. The DNA pellet was washed in 70% ethanol and dried under vacuum. The final pellet was resuspended in 50µl of TE buffer with RNAase A added to a concentration of 20µg.ml^{-1} (for the preparation of the stock solution of RNAase A see section 2.11.1).

2.10.2 "Quick" plasmid minipreps

When a large number of plasmid minipreps had to be performed from *E.coli* strains, this quicker method was used. Plasmid DNA of lower yield and purity was obtained, relative to the alkaline-lysis miniprep. However it could always be digested by restriction enzymes and hence once the correct plasmid containing strain had been identified by this method, alkaline lysis minipreps were then carried out to obtain larger/purer amounts of the plasmid DNA for further work.

1.5ml of a 5ml overnight culture was pipetted into an eppendorf tube and spun for 1 minute in the micofuge to pellet the cells. The pellet was resuspended in 100µl of TE, before 400µl of TE-saturated phenol:chloroform:isoamyl alcohol (25:24:1) was added, and the tube was vortexed for 5 seconds. The tube was then microfuged for 3 minutes, the aqueous layer transferred to a fresh eppendorf tube and 400µl of chloroform:isoamyl alcohol (24:1) added. Again the tube was vortexed for 5 seconds and spun for 3 minutes in the microfuge. The aqueous layer, containing the plasmid DNA, was transferred to a fresh tube.

2.10.3 Large scale plasmid preparation

A single colony was inoculated into 5ml of LM-broth with antibiotic selection and grown overnight. The entire culture was used to inoculate 500ml of LM-broth (with antibiotic selection), which was incubated overnight with shaking. The culture was transferred to two 250ml centrifuge bottles and the bacterial cells harvested by centrifugation at 4000g/4°C for 10 minutes in a MSE High Speed 18 centrifuge. The supernatant was discarded, the cells washed in 20ml of STE (TE buffer with 0.1M NaCl added) and this solution transferred to two 50ml Oakridge centrifuge tubes. These were re-centrifuged for 10 minutes at 4000g/4°C. The supernatants were again removed, the pellets resuspended in 10ml of ice-cold solution 1 and the cell suspension left at room temperature for 5 minutes. (Solutions 1, 2 and 3 were prepared according to section 2.10.1.) 20ml of solution 2 was added to each tube, which were gently mixed and placed on ice for 20 minutes. 15ml of ice-cold solution 3 was added to each tube, the tubes were mixed well and then put back on ice for 10 more minutes. The tubes were then centrifuged for 20 minutes at 4°C and 15000g in the MSE 18 to remove cell debris. The supernatants were transferred to sterile 30ml Corex tubes, 0.6 volumes of isopropanol were added and the DNA left to precipitate at room temperature for 15 minutes. The DNA was pelleted by centrifugation for 20 minutes at 4°C and 12000g in the MSE 18. The pellet was washed in 70% ethanol, dried and resuspended in 5ml TE. Plasmid DNA was purified by caesium chloride density gradient centrifugation.

2.10.4 Caesium chloride/ethidium bromide density gradient centrifugation of DNA

The DNA sample, 20.6g of caesium chloride and 0.3ml of 10mg.ml^{-1} aqueous ethidium bromide were added to a clean measuring cylinder and the volume made up to 27ml with TE buffer. This solution was transferred to a medium Beckman Quickseal tube using a 10ml syringe and wide bore needle. If necessary for balancing the tube was filled with a top up solution - 0.76g CsCl and 11µl of 10mg.ml^{-1} ethidium bromide per ml. After the tube was heat sealed, it was centrifuged for 17-20 hours at 250,000g and 15°C in a Du Pont Instruments Sorvall OTD 65B ultracentrifuge with a Beckman 70Ti rotor. Following centrifugation the DNA within the tube could be visualized under UV light because of the presence of the ethidium bromide. Usually two bands were visible, the upper one consisting of chromosomal DNA and open circular plasmids, and the lower supercoiled plasmid DNA. The top of the tube was punctured with a needle and the relevant DNA band collected with a second wide bore needle into a 5ml syringe. The DNA solution was transferred to eppendorf tubes and the ethidium bromide removed by exhaustive extraction with butan-2-ol saturated with distilled water and CsCl. (An equal volume of the saturated butan-2-ol was added to the DNA solution and vortexed well to mix the two phases. The

tubes were spun for 3 minutes in a microfuge, the top [aqueous] layer removed to a fresh tube and the process repeated until the pink colour, of the ethidium bromide, disappeared from both the organic and aqueous phases.) 300µl aliquots of the DNA solution were put into eppendorf tubes, and the DNA precipitated by the addition of 600µl 0.45M sodium acetate (pH 4.8) and 540µl isopropanol. The tubes were left at 4°C for 15 minutes and the DNA pelleted by centrifugation for 10 minutes in the microfuge. The pellet was washed with 70% ethanol, dried and resuspended in TE buffer.

2.10.5 Large scale preparation of bacterial chromosomal DNA

This was a modified method based upon that of Dhaese *et al.* [66].

A single bacterial colony was inoculated into 5ml of LM-broth with antibiotic selection and grown up overnight. 100µl of the resulting culture was used to inoculate 40ml of LM-broth (with selection) which was incubated overnight. The bacterial cells were pelleted in 2 McCartney bottles at 4000g for ten minutes in a Wifug 500E bench-top centrifuge. The supernatants were removed and the cell pellets resuspended in a 5ml solution of 0.08% SDS, 2mM EDTA pH8.0, 50mM Tris.HCl pH8.0 and 1mg.ml^{-1} Proteinase K (stock solution of Proteinase K is 20mg.ml^{-1} in distilled water [253]). The solutions were incubated at 37°C for 1 hour, or until the solution had become clear and viscous. The chromosomal DNA was now sheared by pipetting the solutions up and down in a Pasteur pipette. The solutions were then extracted twice with equal volumes of phenol:chloroform:isoamyl alcohol (25:24:1), the aqueous layers being transferred to fresh bottles each time. A single chloroform:isoamyl alcohol (24:1) extraction was then performed, the aqueous layers removed to fresh bottles and the DNA precipitated as in section 2.11.4. The final pellet was resuspended in 1ml of TE buffer. The DNA was further purified by density gradient centrifugation (see section 2.10.4).

2.10.6 Small scale preparation of bacterial chromosomal DNA

This method produced chromosomal DNA of sufficient purity to be cleavable by restriction endonucleases and was a personal communication from Dr. C. O'Reilly.

A 5ml LM-broth culture of bacteria was grown to stationary phase and 400µl of this culture transferred to an eppendorf tube. The cells were spun down for one minute in a microfuge, the supernatant removed and the pellet left at -20°C for 30 minutes. The pellet was then resuspended in 200µl of TE buffer, and the cell suspension incubated with 8µl of lysozyme (10mg.ml^{-1} aqueous stock [253]) for 30 minutes at 37°C. The cells were then

lysed by the addition of 40µl 4M sodium perchlorate, 24µl of 10% SDS and 8µl of Proteinase K (20mg.ml^{-1} aqueous stock) and incubated at 45°C for 2 hours. The DNA was now precipitated by the addition of 2 volumes of ethanol and then pelleted by centrifugation in the microfuge for 5 minutes. The pellet was washed with 70% ethanol, dried and resuspended in 500µl of TE buffer. The DNA solution was extracted twice with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), the aqueous layer being transferred to new eppendorf tubes each time, and then extracted once with chloroform:isoamyl alcohol (24:1). This final aqueous layer was precipitated with ethanol as above and the pellet finally resuspended in 50µl of TE buffer with RNAase A added to a final concentration of 20µg.ml^{-1}.

2.11 DNA manipulations

2.11.1 RNAase treatment of DNA solutions

A stock solution of RNAase was prepared as follows. Pancreatic RNAase A was dissolved at a concentration of 10mg.ml^{-1} in 15mM NaCl, 10mM Tris.HCl pH7.5. Any DNAases present were inactivated by boiling the solution for 15 minutes and allowing it to cool slowly to room temperature, before aliquots were stored at -20°C.

Contaminating RNA was removed from a DNA solution by the addition of RNAase A (to a final concentration of 20µg.ml^{-1}) and incubation at 37°C for an appropriate length of time. Usually for RNAase treatment of plasmid DNA this incubation was carried out along with restriction endonucleases, see section 2.11.6. For larger amounts of plasmid DNA or chromosomal DNA preparations, contaminating RNA was removed by digestion with 50µg.ml^{-1} RNAase A for 1 hour at 37°C. The enzyme was then removed by phenol-chloroform extraction (see section 2.11.2) before precipitation of the DNA.

2.11.2 Phenol-chloroform extraction of DNA

A 25:24:1 solution of phenol:chloroform:isoamyl alcohol was equilibrated 3 times with TE buffer and stored under TE in a light-proof bottle at 4°C. To remove proteins from DNA solutions an equal volume of phenol:chloroform:isoamyl alcohol was added, the solutions mixed by vortexing for 30 seconds and the phases separated by centrifugation for 2 minutes in a microfuge. The aqueous phase was transferred to a fresh tube. This was repeated until no further protein was visible (as a white precipitate) at the boundary of the two phases. Finally a similar extraction with chloroform:isoamyl alcohol (24:1) was carried out to remove any traces of phenol from the DNA solution.

2.11.3 Removal of proteins from DNA solutions using silica fines

This method was a personal communication from Dr. N. J. Robinson. The main advantage of this procedure was that highly corrosive phenol was not used. Two volumes of sodium iodide solution (90.8g NaI and 1.5g Na_2CO_3 dissolved in distilled water, filter sterilised and saturated with 0.5g Na_2SO_3 - stored at 4°C in a light-proof bottle) were added to the DNA solution (minimum 200µl volume) in an eppendorf tube. 5µl of silica fines were added, mixed and the solution left for 10 minutes at room temperature with occasional shaking. The fines were spun down for 15 seconds in a microfuge, the supernatant removed and the fines washed with 70% ethanol. The silica fine pellet, having been dried under vacuum, was resuspended in 50µl of TE buffer and incubated at 37°C for 10 minutes with occasional shaking. The fines were spun down again in the microfuge for 15 seconds and the supernatant, containing the DNA, collected.

Preparation of the silica fines:

250ml of silica 325 mesh powder was resuspended in distilled water to give a total volume of 500ml. The suspension was stirred for 1 hour and left to settle for a further hour. The suspension was then centrifuged at 5000g in a Beckman J2-HS centrifuge using a JA-14 rotor. The pellet was resuspended in 150ml of distilled water plus 150ml nitric acid. The suspension was then heated to 98°C and allowed to cool to room temperature. The silica fines were then repeatedly washed with sterile distilled water until the pH was greater than 5.5. Silica fines were stored at 4°C as a 50% slurry in sterile distilled water.

2.11.4 Ethanol precipitation of DNA

0.1 volumes of 3M sodium acetate (pH4.8) and 2 volumes of ethanol were added to the DNA solution and mixed by vortexing. Plasmid DNA solutions were placed at -80°C for at least 30 minutes, whilst chromosomal DNA solutions were left for 15 minutes at room temperature. The DNA was pelleted by centrifugation in a microfuge for 10 minutes. After which the supernatant was removed, the pellet washed in 70% ethanol, dried under vacuum and finally resuspended in TE buffer or sterile distilled water.

2.11.5 Spectrophotometric quantification of DNA solutions

The absorbance of a 1:50 dilution of the DNA sample in sterile distilled water was read at 260nm and 280nm on a Beckman DU7500 spectrophotometer, using sterile distilled water as a blank. As an A_{260nm} of 1.0 is equivalent to a concentration of $50\mu g.ml^{-1}$ of double stranded DNA or $\sim 33\mu g.ml^{-1}$ of single stranded oligonucleotides, sample DNA concentrations could be calculated.

2.11.6 Restriction endonuclease digestions

Digestions were carried out according to the enzyme manufacturer's instructions. Generally plasmid DNA was digested in a total volume of 10-30µl, with 5 units of restriction endonuclease, 0.1 volumes of the supplied 10x concentrated enzyme buffer and sterile distilled water to make up the volume. The reaction was incubated at the recommended temperature (usually 37°C) for 1-2 hours. If more than one restriction enzyme was to be used in the same reaction and the buffers supplied differed, the reaction was buffered using one-phor-all buffer PLUS (Pharmacia). Chromosomal DNA was digested in a larger volume, 100-200µl, with 10 units of restriction enzyme added for every microgram of DNA, as well as the appropriate amounts of buffer and sterile distilled water. The reaction mixture was covered with a layer of mineral oil to prevent evaporation and maintain the buffering conditions, before being incubated overnight at the required temperature.

If the digestions were to be analysed by gel electrophoresis, 0.2 volumes of 6x gel-loading buffer were added. 6x gel-loading buffer [253], contains 0.25% bromophenol blue, 0.25% xylene cyanol FF and 40% sucrose in distilled water. This was filter sterilised and stored at 4°C.

If the digested DNA was to be used in further subcloning steps, the digestion was stopped by removing the restriction enzyme(s) using silica fines (see section 2.11.3).

2.11.7 Agarose gel electrophoresis

Gel electrophoresis of DNA samples was carried out with large 180x150mm maxigels (volume 200ml), 100x80mm (volume 70ml) or 77x55mm minigels (volume 50ml). Minigels were run in Pharmacia gel apparatus GNA-100 electrophoresis tanks. The concentration of agarose within a gel could be varied depending on the size of DNA to be separated [253]. Usually a 0.7% agarose gel was used, which efficiently separated linear DNA between 10-0.8kb. The required amounts of agarose and 1x TAE buffer (50x stock-

242g Tris, 100ml EDTA pH8.0, 57.1ml glacial acetic acid per litre [253]) were mixed and the agarose dissolved by microwaving the mixture. The solution was cooled to about 60°C, 10mg.ml^{-1} ethidium bromide was added to a final concentration of 0.2µg.ml^{-1}, and the agarose poured into the gel mould with a well comb in place. Once the agarose had set the gel was put in a tank and covered with 1x TAE buffer containing 0.2µg.ml^{-1} ethidium bromide. The DNA samples (and size markers) were loaded and electrophoresis carried out at 5-10 V.cm^{-1} for the required amount of time. The size markers used were either λ-DNA digested with *Pst*I and/or λ-DNA digested with *Hind*III.

***Pst*I digested λ-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...

The fragments enclosed in brackets run together on an agarose gel. Smaller fragments are also produced, but were rarely seen in this work.

***Hind*III digested λ-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.

DNA within the gel was visualised on a transilluminator (UVP Inc.), and photographed with a Polaroid RP4 Land camera (using a red filter) onto Polaroid 667 film.

2.11.8 DNA fragment isolations from agarose gels using silica fines

This method was a personal communication from Dr. N. J. Robinson, and uses the same solutions as described in section 2.11.3.

The required DNA band was excised from an agarose gel with a sterile scalpel blade and placed into an eppendorf tube. 1ml of NaI solution was added and the tube incubated at 70°C until the agarose fragment had melted, usually about 5 minutes. The tube was then left at room temperature to cool for 5 minutes before 5µl of silica fines were added and a silica fines extraction performed as in section 2.11.3.

2.11.9 Filling in 3'-recessed termini

The DNA fragment (maximum of 500ng) was resuspended in 10-15µl of sterile distilled water following isolation from an agarose gel. A 1µl solution containing all 4 dNTPs (each at 1mM) was added to the DNA. 2µl of Klenow buffer (10x) was added (10x buffer is 0.5M Tris.HCl pH 7.6, 0.1M $MgCl_2$) and the reaction buffer made upto 20µl with

sterile distilled water plus 1µl (1 unit) of Klenow fragment. The reaction mixture was left at room temperature for 30 minutes and then the Klenow fragment removed using silica fines or inactivated by incubating at 70°C for 5 minutes.

2.11.10 Ligation of DNA

T4 DNA ligase was used to ligate DNA fragments with compatible cohesive or blunt termini. The fragments of insert and vector DNA were usually mixed at a ratio of 3:1 (insert:vector) with a maximum of 300ng DNA. 0.1 volume of 10x ligase buffer (0.66M Tris.HCl pH7.5, 50mM $MgCl_2$, 50mM DTT, 10mM ATP - made fresh regularly and stored at -20°C) was added and for cohesive termini 1 unit of DNA ligase added. This was then incubated overnight at 4°C. For blunt-ended termini, 3 units of ligase were added and the reaction incubated at 15°C overnight. The ligation mix was then used immediately to transform competent *E.coli* cells.

2.12 Transformation of *E.coli*

The transformation procedure used was a modified version of that described by Hanahan, D. [109].

2.12.1 Preparation of competent cells

5ml of LM broth was inoculated and grown overnight at 37°C, subcultured 1:100 into fresh LM broth and the cells grown to an OD_{600} of 0.3-0.35. The culture was chilled for 5 minutes on ice before being spun down in pre-chilled centrifuge tubes at 4000g, 4°C for 7 minutes. The supernatants were poured off and the cell pellets resuspended in $^2/_5$ of the original culture volume of solution A (30mM potassium acetate, 100mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride and 15% glycerol. The solution was adjusted to pH 5.8 with 0.2M acetic acid and filter sterilised.) The tubes were held on ice for 5 minutes and spun down as before. The supernatants were poured off and the pellets resuspended in $^1/_{25}$ of the original culture volume of solution B (10mM MOPS, 75mM $CaCl_2$, 10mM $RbCl_2$ and 15% glycerol. The solution was adjusted to pH 6.5 with KOH and filter sterilised.) The tubes were left on ice for 15 minutes, before suitable volumes (200µl) were added to pre-chilled eppendorfs, frozen in liquid nitrogen and stored at -80°C.

2.12.2 Transformation procedure

The cells were thawed by hand and then placed on ice for 10 minutes. The DNA was added, up to 250ng/200µl cells, and the tube held on ice for at least 15 minutes (up to 45 minutes). The cells were heat shocked at 42°C for 90 seconds, held on ice for 3 minutes and then 800µl of pre-warmed (to 37°C) LM broth were added. The tube was incubated for 1 hour at 37°C, with occasional shaking. Finally appropriate aliquots, usually $^{1}/_{10}$th and $^{9}/_{10}$ths of the tube, were spread onto selective agar plates.

2.13 DNA hybridisation procedures

2.13.1 Radio-labelling of DNA fragments

DNA fragments were labelled with [α-^{32}P] dCTP by the random primer labelling method using an Amersham Multiprime kit. 30-50ng of the DNA to be labelled, in a total volume of 28µl, were boiled for 5 minutes and then held on ice for 2 minutes. 10µl of labelling buffer, 5µl of random hexanucleotide primers, 5µl ^{32}P-dCTP (equivalent to 50µCi) were added, followed by 2µl of Klenow enzyme. The labelling reaction was now left to proceed either at room temperature overnight or for 2-3 hours at 37°C. The labelled DNA was boiled for 5 minutes immediately before use.

2.13.2 Southern blotting

DNA was transferred to Hybond-N (Amersham) nylon membranes, according to the manufacturer's instructions. Firstly the agarose gel containing the DNA samples was photographed with a ruler down its side. Gels known to have DNA fragments greater than 10kb in size would first be soaked in 0.25M HCl for 15 minutes, to partially depurinate the DNA, and rinsed twice with distilled water. For blotting the gel was soaked in denaturation buffer (1.5M NaCl, 0.5M NaOH) with occasional shaking for 30 minutes. The gel was then rinsed twice with distilled water and soaked in neutralisation buffer (1.5M NaCl, 0.5M Tris.HCl pH7.2, 0.001M EDTA). After rinsing the gel twice with distilled water, the blot was set up. For single sided (one-way) blots a reservoir of 10x SSC was set up (20x SSC - 3.0M NaCl, 0.3M Na.citrate pH7.0). A platform was placed over this reservoir and a long piece of Whatman 3MM paper (presoaked in 10x SSC) put on this platform with its ends dipping into the reservoir. The gel was placed, wells uppermost, on the 3MM paper, and a piece of Hybond-N nylon membrane cut to the same size as the gel placed on top of it. Any air bubbles were carefully removed before 3 sheets of Whatman 3MM paper, cut to the gel size and presoaked in 10x SSC, were placed on top. Finally 2 layers of disposable nappies

(also cut to gel size) were placed on top, the stack covered with a glass plate and a 1kg weight placed on top. Double sided (two-way) blots were created by sandwiching the agarose gel between two equivalent piles of a glassplate, 2 layers of disposable nappies, 3 pieces of Whatman 3MM paper (presoaked in 10x SSC) and the Hybond-N nylon membrane. A 1kg weight was placed on the top glass plate. Liquid retained in the gel transferred the DNA onto both membranes. Both types of blot were left for at least 16 hours to allow DNA transfer. After which time the apparatus was dismantled, the positions of the wells marked on the nylon and the DNA fixed onto the membrane according to the manufacturer's instructions. The filter was allowed to air dry for 1 hour before wrapping it in clingfilm and exposing it to UV light for two minutes to fix the DNA to the membrane.

2.13.3 Hybridisation of radio-labelled probes to Southern blots

Hybridisation reactions were carried out using Techne Hybridisation tubes in a Techne Hybridiser HB-1 oven. The nylon filter was put inside the hybridisation tube and 200µl of pre-hybridisation solution (5x SSC, 5x Denhardt's solution [50x Denhardt's solution is 1% ficoll, 1% polyvinylpyrrolidone, 1% BSA fraction V], 0.5% SDS and 100µg.ml^{-1} of denatured salmon sperm DNA) added per cm^2 of filter. Any air bubbles were carefully removed and the tube incubated at 65°C, whilst being rotated in the oven. After two hours the labelled probe was denatured, by boiling for 5 minutes, and added to the tube contents. The tube was replaced in the oven and incubated at 65°C for at least 12 hours. The hybridisation temperature could be varied depending on the desired stringency, but usually it was at 65°C. After incubation the pre-hybridisation solution (containing the probe) would be poured off or transferred to a glass bottle, if the probe was needed again.

2.13.4 Washing of probed Southern blots

The nylon filter was washed within the hybridisation tube and was never allowed to dry out. For high stringency washing, the filter was washed twice in 2x SSC, 0.1% SDS for ten minutes at room temperature, followed by one wash in 0.1x SSC, 0.1% SDS for 15 minutes at 65°C. Lower stringency washes were done with 2x SSC alone at 42°C. After each washing solution was removed, the filter was checked with a Geiger counter and the washing continued until sufficient (apparent) non-specific radio-labelled probe was removed. Finally the filters were wrapped in clingfilm.

2.13.5 Detection of hybridising probes

The wrapped filter was taped onto a larger piece of Whatman 3MM paper and radioactive ink spotted onto the edges of the 3MM paper. Radioactive bands could be

detected on the filter by exposing Fuji RX-100 X-ray film to the filter. The film sheets were pre-flashed once to sensitise the film, and the exposure carried out at -80°C, for varying amounts of time. Exposed films were developed with Ilford Phenisol developer for up to 4 minutes and fixed with Kodak Unifix fixer for 2 minutes. Aligning the radioactive ink spots on the film to those on the 3MM paper allowed the position of the wells on the filter to be marked on the film. Hence the size of any hybridising fragments could be calculated using the original gel photograph (with ruler included).

2.13.6 Removal of radioactive probes from nylon filters

To strip radioactive probes from nylon filters for reprobing, the filter was washed with 0.4M NaOH at 45°C for 30 minutes. This was followed by a wash with 0.1x SSC, 0.1% SDS and 0.2M Tris.HCl pH7.5 for 30 minutes at 45°C. Successful removal of probes from nylon filters only occurred if the filter had never been allowed to dry out completely after hybridisation/washing. The stripped filter was checked for probe removal by re-exposing X-ray film.

2.14 Gene replacement mutagenesis

This technique involves the specific replacement of a gene, with a copy of the gene's DNA sequence containing an insert of a selectable marker, to prevent transcription of a functional wild-type gene product. The actual protocol used in this work was that of Quandt, J. and Hynes, M. F. [240], the vectors described therein were a gift from the above authors. A diagrammatic explanation of the constructs used is presented in section 4.10.

The basic vector used in this work was pJQ200SK, which carries a gene for resistance to gentamycin and a functional *mob* site allowing *A. tumefaciens* conjugation. To prepare the gene replacement constructs, the *A. tumefaciens* genes to be mutated were subcloned into pJQ200SK. Positive subclones were selected by inactivation of β-galactosidase. Each *A. tumefaciens* gene was then disrupted by the introduction of a neomycin-resistance cassette from pDUB1100. Transformants were in this case selected through resistance to gentamycin and neomycin. The resultant plasmid was then conjugated into *A. tumefaciens* (see section 2.7.)

pJQ200SK cannot replicate outside of enterobacteria, and thus acts as a suicide vector in *A. tumefaciens*. This vector can only be maintained in *A. tumefaciens* following integration into the chromosomal DNA by homologous recombination. For gene replacement to be achieved, the integration event must be followed by the excision of the

plasmid and functional gene to leave only the mutated gene. The replacement event is rare, and cannot be directly selected for by neomycin resistance, as both integrated and replacement recombinants carry this selection. Nevertheless pJQ200SK carries an alternative selectable marker, the *sacB* gene of *Bacillus subtilis*. *sacB* is inducible by sucrose and encodes the enzyme levansucrase, which is lethal in Gram negative bacteria. Therefore *A. tumefaciens* cells containing the mutated gene alone, *i.e.* with gene replacement, could be selected for on LM-agar plates containing 5% sucrose and neomycin, whilst integrated recombinants would be killed.

To confirm gene replacement, the chromosomal DNA from putative mutants/recombinants was isolated and digested with a restriction enzyme. This should produce a different sized DNA fragment than wild-type genomic DNA cut with the same enzyme, because of the insertion of the neomycin-resistance cassette. Southern hybridisation of the digests with a radiolabelled copy of the wild-type gene was used to detect this size difference and thus confirm gene replacement.

2.15 DNA sequencing

DNA sequencing was always carried out with an Applied Biosystems 373A DNA Sequencer using double stranded DNA templates. Usually fluorescently labelled universal M13 primers from the Applied Biosystems PRISM™ Ready Reaction Dye Primer Cycle Sequencing Kit were used, according to the manufacturer's instructions. Occasionally custom-synthesised primers (synthesised on an Applied Biosystems 381A DNA Synthesiser) were used in conjunction with the Applied Biosystems Taq DyeDeoxy™ Terminator Cycle Sequencing Kit.

2.15.1 Preparation of double stranded DNA templates

For optimum sequence data the purity of the plasmid DNA template was critical. This protocol is the one recommended by Applied Biosystems and is listed in their sequencing kits. It is a modified alkaline-lysis miniprep (see section 2.10.1) with a polyethylene glycol precipitation step at the end to purify the DNA.

A 5ml volume of LM broth, with the appropriate antibiotic selection, was inoculated with the relevant bacterial strain and incubated overnight at 37°C with shaking. 1.5ml of the resulting culture was transferred to an eppendorf tube and the cells spun down for 1 minute in a microfuge. The supernatant was poured off, the pellet resuspended in 100µl of ice-cold solution 1 and the tube left for 5 minutes at room temperature. 200µl of freshly

prepared solution 2 was now added, the sample mixed by inversion and incubated on ice for 5 minutes. 150µl of ice-cold solution 3 was then added, the solution mixed by inversion for 10 seconds and incubated on ice for 5 minutes. The sample was microfuged for 5 minutes and the supernatant carefully transferred to a fresh tube avoiding any transfer of the precipitated cell debris. This tube was microfuged for 5 more minutes, the supernatant transferred to a fresh tube and RNAase A added to a final concentration of 20µg.ml^{-1}. The sample was incubated for 20 minutes at 37°C, after which an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the tube vortexed for 30 seconds before being microfuged for 1 minute. The aqueous phase was transferred to a fresh tube and precipitated with 2 volumes of ethanol, as in section 2.10.1. The DNA pellet was dissolved in 16µl of sterile distilled water and combined with 4µl of 4M NaCl and 20µl of 13% polyethylene glycol (M_r 8000). The tube contents were mixed well and incubated on ice for 20 minutes. The tube was now microfuged for 10 minutes, the supernatant carefully removed and the pellet washed twice with chilled 70% ethanol before being vacuum dried. Finally the pellet was dissolved in 10µl of sterile distilled water. 1µl of this was digested by a restriction endonuclease and the digest products run out on a gel to check the isolated plasmid's quantity and quality. Another 1µl was used to spectrophotometrically determine the DNA concentration (see section 2.11.5) accurately prior to sequencing. For dye primer reactions 7µl of DNA at 250ng.µl^{-1} were used, and for dye terminator reactions 5µl at 200ng.µl^{-1}.

2.16 RNA procedures

2.16.1 General techniques

Due to the high stability of ribonucleases, care was taken to create a ribonuclease-free environment whenever possible. Sterile, disposable plasticware (considered to be RNAase-free) was used. Non-sterile eppendorf tubes and pipette tips were used from unopened bags and autoclaved immediately. Gloves were always worn whilst performing RNA-related work to prevent contamination from RNAases. All aqueous solutions which were used with RNA were first treated with diethyl pyrocarbonate (DEPC) which is a non-specific inhibitor of RNAases. DEPC was added to a final concentration of 0.1%, the solution shaken vigorously to disperse the DEPC and then left overnight at room temperature. The solution would then be autoclaved to remove the DEPC. Solutions containing Tris could not be treated directly with DEPC, as the two compounds react to form a stable complex. Instead Tris buffers were made with autoclaved DEPC-treated distilled water and then re-autoclaved. Corex tubes used for extractions were soaked

overnight in 0.1% DEPC, before being rinsed with DEPC-treated distilled water and then autoclaved.

2.16.2 Extraction of total RNA from *A. tumefaciens*

When broth cultures were used as the source of bacterial cells for the extraction procedures, the cultures were grown to mid-log phase in MinA media with the appropriate selection. Occasionally cells were harvested from agar plates, in this case, 1ml of a *A. tumefaciens* culture grown overnight in MinA media was spread onto a MinA plate (with selection). The plate was incubated for 16 hours at 28°C and then at room temperature for 30 hours. Finally the cells were washed off the plate with 3ml of 0.15M NaCl and collected.

a) Quick minipreparations of total RNA

This method was a modified version of that of Barry *et al.* [22]. 1.5ml of culture, grown as described in section 2.16.2, was spun in a microfuge for 1 minute. The supernatant was poured off, the pellet resuspended in 20µl of DEPC-treated distilled water and 1mg of lysozyme added. 3µl of 0.5% DEPC (diluted with distilled water) was added, followed by 200µl of ice-cold acetone and the tube mixed by inversion. The tube was microfuged for two minutes, the supernatant carefully removed and the pellet resuspended in 30µl of DEPC-treated distilled water. 1µl of 100µg.ml^{-1} Proteinase K was added and the sample incubated on ice for 10 minutes. 3.5µl of 0.5% DEPC was added, followed by 200µl of Tris-saturated phenol (pre-heated to 70°C) and 150µl of chloroform:isoamyl alcohol (24:1). The tube was mixed by inversion, 120µl DEPC-treated distilled water were added and the tube microfuged for 5 minutes. The aqueous phase was removed to a fresh tube containing 1ml of ethanol and the RNA precipitated at -20°C for 1 hour. The RNA was pelleted by microfuging for 10 minutes, the pellet was resuspended in 50µl of DEPC-treated distilled water and stored at -80°C.

b) RNA extraction utilising diethyl pyrocarbonate

This method was adapted from that of Summers, W. C. [293]. 10ml aliquots of *A. tumefaciens* cell culture were placed in 15ml Corex tubes, and the cells pelleted by centrifugation at 4000g, 4°C for 5 minutes in a Beckman J2-HS centrifuge using a JA-14 rotor. The supernatants were removed, the pellets resuspended in 5ml of protoplasting medium (0.015M Tris.HCl pH8.0, 0.45M sucrose, 0.008M EDTA and 80µg.ml^{-1} lysozyme) and incubated on ice for 5 minutes. The tubes were spun again (as above), the supernatants removed and the pellets resuspended in 0.5ml of lysing buffer (0.01M Tris.HCl pH8.0,

0.01M NaCl, 0.001M Na.citrate and 1.5% SDS). 15µl of DEPC were added, the tubes incubated at 37°C for 5 minutes and 300µl of chilled (saturated) NaCl solution were to precipitate cell debris. The tubes were spun at 10,000g, 4°C for 10 minutes and the supernatants transferred to fresh tubes. 2 volumes of ethanol were added and the tubes incubated at -20°C for 4 hours. The RNA precipitate was collected by centrifugation at 10,000g, 4°C and the pellets resuspended in 100µl of DEPC-treated distilled water.

2.16.3 Quality and quantity assessment of extracted RNA

After extraction, an aliquot of the RNA was run out on an agarose gel. Basically this was performed as described in section 2.11.7. However all the gel apparatus was thoroughly rinsed with water, and then DEPC-treated distilled water before use. The gel and gel buffer were also made using DEPC-treated distilled water. A 1.2% agarose gel was prepared and run for half an hour to minimise exposure of the samples to RNAases. An RNA extraction produced a characteristic pattern of ribosomal RNA bands at (approximately) 2900bp, 1540bp and 120bp as well as a smear of transfer RNA fragments, about 75bp in size.

The concentration of RNA was estimated using a spectrophotometer, as in section 2.11.5, as an A_{260} of 1.0 is equivalent to an RNA concentration of ~40µg.ml^{-1}.

If necessary RNA solutions were more concentrated by precipitation and resuspension in a smaller volume. Precipitation was brought about by the addition of 0.1 volumes of 8M LiCl and 3 volumes of ethanol. The mixture was left on ice for at least 2 hours, before the RNA was collected by centrifugation at 15,000g, 4°C for 20 minutes in a Beckman J2-HS centrifuge using a JA-14 rotor.

2.16.4 Formaldehyde-containing agarose gels

RNA samples that were to be transferred to nylon filters were electrophoresed in 1.2% or 1.4% agarose gels containing formaldehyde. Either a GNA200 electrophoresis system (Pharmacia) kept specifically for RNA samples or a thoroughly washed GNA100 tank was used. The gel was made up with 1xMOPS buffer (10xMOPS buffer - 0.2M MOPS, 0.05M Na.citrate, 0.01M EDTA - this was adjusted to pH7.2 and DEPC-treated) diluted with DEPC-treated distilled water. The agarose was microwaved to dissolve it, the solution cooled to about 60°C and 37% formaldehyde (12.3M) added to a final concentration of 2.2M, before gel casting. 3 volumes of a denaturing solution (500µl formamide, 100µl 10xMOPS buffer, 150µl formaldehyde, 250µl DEPC-treated distilled water) were added to the RNA samples. The samples were incubated at 60°C for 5 minutes,

then chilled on ice before 0.1 volume of 10x loading buffer was added, (10x formaldehyde gel loading buffer, 50% glycerol, 1mM EDTA pH8.0, 0.25% bromophenol blue and 0.25% xylene cyanol). The samples were loaded and the gel run at $5V.cm^{-1}$, using 1xMOPS as the running buffer. After electrophoresis the RNA within the gel was stained with $0.5\mu g.ml^{-1}$ ethidium bromide and viewed as described in 2.11.7. Often RNA bands were poorly visible because of the lower affinity of the stain for RNA and "background" stain due to the gel itself binding ethidium bromide. In which case the gel was destained by gentle shaking in distilled water until resolution was clear. The gel was photographed with a ruler along its side. The highly conserved ribosomal RNA bands were used as approximate size markers on the gel.

2.16.5 Northern blotting

RNA was transferred to Hybond-N nylon membranes essentially as described in section 2.13.2. The gel was destained overnight to remove the formaldehyde and ethidium bromide and then soaked in 10xSSPE for 1 hour, (20xSSPE is 3.6M NaCl, 0.2M sodium phosphate and 0.02M EDTA pH7.7). A one-way capillary blot was then set up as described, using 10xSSPE in the reservoir and to soak the components.

The radiolabelled probe was prepared as described in 2.13.1, but unincorporated nucleotides were removed using a column to reduce background hybridisation [253]. 10µl of a stop-dye mix (60mM EDTA pH8.0, $100mg.ml^{-1}$ blue dextran and $1mg.ml^{-1}$ xylene cyanol) was added to the labelling reaction after the desired incubation. The contents were mixed and then pipetted into a glass pasteur pipette containing Sephadex G-50 (equilibrated with TE buffer). The DNA was washed through with TE buffer. The probe came off with the first dye front (blue dextran) in a volume of about 500µl, whilst the unincorporated nucleotides elute with the xylene cyanol dye.

Pre-hybridisation and hybridisation were carried out as described in 2.13.3. Except SSPE was used instead of SSC (at the same concentration) and formamide was added to a final concentration of 50%. The incubation temperature was always 42°C.

2.16.6 Washing of Northern hybridisation blots

The conditions used to wash Northern blots varied considerably. Occasionally a wash in 2xSSPE, 0.1% SDS for 5 minutes at room temperature was sufficient. Usually further washing had to be carried out with 1xSSPE, 0.1% at 42°C for 5 minute intervals. After each the blot would be checked using the Geiger counter and this step repeated until

excess probe had been removed. Hybridising bands were visualised as described in 2.13.5. The films were exposed initially for 48 hours at -80°C.

2.16.7 Removal of hybridising probes from Northern blots

A boiling solution of 1xSSPE and 0.1% SDS was poured over the blot and then removed. The blot was allowed to cool (but not dry out) before a repeated wash with boiling 0.1xSSPE, 0.1% SDS. Probe removal was now checked by re-exposing an X-ray film to the blot.

2.17 Protein procedures

2.17.1 Flagellar filament isolation

The method used to detach and isolate flagellar filaments from bacterial cells was a modified version of those of Robinson *et al.* [249] and Krupski *et al.* [155].

25ml of MinA media was inoculated with a loopful of *A. tumefaciens* cells taken from the edge of a swarm on a swarm plate and incubated at 28°C for 25 hours. 1ml aliquots of this culture were then spread onto 20 MinA agar plates, which were all incubated at 28°C for 16 hours and then at room temperature for 30 hours. The cells were washed off the plates using 3ml of 0.15M NaCl, pooled and collected in 50ml polypropylene centrifuge tubes. The flagella were detached from the cells by vortexing for 15 seconds. The cell bodies were pelleted by centrifuging the tubes at 12,000g, 4°C for 10 minutes in a Beckman J2-HS centrifuge using a JA-14 rotor. The supernatants were removed to fresh centrifuge tubes and centrifuged again for 10 minutes at 15,000g, 4°C. The supernatants were now transferred to medium Beckman Quickseal tubes and the flagella filaments pelleted by centrifuging at 100,000g, 4°C for 2 hours in a Sorvall OTD 65B ultracentrifuge with a Beckman 70Ti rotor. The supernatants were removed and the flagellar filaments resuspended in 1ml of HEPES buffer (10mM HEPES, 10μM EDTA pH8.0, 200μM $CaCl_2$). The quantity and quality of the isolated filaments were now assessed by SDS-PAGE analysis, section 2.17.3.

2.17.2 Extraction of cell protein fractions

The method used to extract protein fractions from bacterial cells was a modified version of that of Neu *et al.* [210].

1.5 ml of cell culture was taken and pelleted in an eppendorf tube at low speed (3000g) in a microfuge (MSE MicroCentaur) for 10 minutes. The supernatant was carefully removed and stored on ice as the extracellular fraction. The pellet was resuspended in 1.5 ml of a solution containing 20% sucrose and 0.03M Tris.HCl pH 8.0, and 3µl of 500mM EDTA pH 8.0 were then added. The cell suspension was shaken at room temperature for 10 minutes at low speed on a Scientific Industries Vortex Genie II mixer and then centrifuged for 10 minutes at low speed in a microfuge. The supernatant was saved as the "wash" fraction and the pellet resuspended in 1.5ml of ice cold distilled water, vortexed for 15 seconds and then shaken on ice at low speed for 10 minutes. The suspension was spun for 10 minutes at low speed in the microfuge and the supernatant was saved as the periplasmic fraction. The pellet was resuspended in 1.5ml of distilled water and sonicated for 1 minute at 14µ amplitude using a MSE Soniprep 150. The resulting suspension was spun for 5 minutes at high speed in a microfuge. The supernatant was saved as the cytoplasmic fraction. The pellet was resuspended (by vortex mixing) in 1% Triton X-100 and this was treated as the membrane fraction.

2.17.3 SDS polyacrylamide gel electrophoresis

SDS polyacrylamide gel electrophoresis (SDS-PAGE) for proteins was performed in either the Bio-Rad Mini Protean II or Protean II xi equipment using the method of Laemmli, U. K. [161].

Running gels were made using 10% acrylamide, 0.1% SDS, 0.07% ammonium persulphate and 0.375M Tris.HCl pH 8.8. Stacker gels were made using 5% acrylamide, 0.1% SDS, 0.07% ammonium persulphate and 0.125M Tris.HCl pH 6.8. The gels were set using TEMED (N,N,N',N'-tetramethylethylenediamine) - 1 $\mu l.ml^{-1}$ of gel. The ammonium persulphate was usually made fresh and kept for a week at 4°C. Before loading, the samples were mixed with an equal volume of 2x sample buffer (4% SDS, 20% glycerol, 120mM Tris.HCl pH 6.8 and 0.005% bromophenol blue, with 10% β-mercaptoethanol added immediately before use), then boiled for 5 minutes and spun down in the microfuge before loading with Bio-Rad Prot/Elec tips. Molecular size markers were obtained from Sigma either pre-stained (cat. no. SDS 7) or unstained (cat. no. SDS 7B). Gels were run in reservoir buffer containing 0.025M Tris.HCl pH 8.8, 0.19M glycine and 0.1% SDS, at 150V until the blue marker was 1cm from the bottom of the gel.

The acrylamide used in the gels was diluted from a 30% acrylamide stock which was made by the addition of 29g acrylamide and 1g N,N'-methylenebisacrylamide to 60ml distilled water. The solution was heated to 37°C to dissolve the chemicals and the volume then adjusted to 100ml with distilled water.

2.17.4 Non-denaturing (native) polyacrylamide gel electrophoresis

Native polyacrylamide gel electrophoresis (PAGE) was carried out with thoroughly washed (to remove any residual SDS) mini-Protean II gel equipment. The same solutions were used as for SDS-PAGE except any denaturing agents, (SDS and β-mercaptoethanol) were not added. Lower percentages of acrylamide were also used, 3% in the stacker and 7.5% in the running gel. The samples were also not boiled before loading. Molecular weight markers for non-denaturing gels were obtained from Sigma, cat. no. MW-ND-500.

2.17.5 Staining polyacrylamide gels with Coomassie Brilliant Blue

The stain was prepared by dissolving 0.25g of Coomassie Brilliant Blue in 90ml of methanol:distilled water (1:1 v/v) and 10ml of glacial acetic acid. The solution was filtered through Whatman no. 1 filter paper. This solution was poured over the gel in a plastic tray, a lid put on the tray, and the tray microwaved for 30 seconds at medium power. (The stain solution should become hot - but NOT boil.) The stain was poured off and kept for future use. The gel was then destained by covering it in 90ml of methanol:distilled water (1:1 v/v) and 10ml acetic acid and microwaving as before. This step was repeated until protein bands were visible in the gel. Destaining was further accelerated by placing a small piece of sponge in the tray which absorbed the stain as it leached from the gel. The gel was now photographed using FP4 black and white film or Kodak Kodacolour Gold 2 colour film.

2.17.6 Silver staining polyacrylamide gels

The method used to silver stain polyacrylamide gels was a personal communication from Dr. K. Elborough.

Gels were carefully removed from the glass plates of the electrophoresis apparatus and placed in a polythene box containing enough of a solution of 5% formaldehyde and 40% ethanol solution to cover the gel. After 30 minutes, this solution was replaced with distilled water and the gel left for 30 minutes before being placed in 50% methanol overnight. The gel was then soaked in distilled water containing a small amount of dithiothreitol for 30 minutes followed by soaking in enough 0.1% silver nitrate to just cover the gel. After a further 30 minutes, the gel was washed 3 times with distilled water, before the developer (3% Na_2CO_3, 0.0185% formaldehyde) was added. This was left for 2 to 5 minutes. Once the required intensity of the bands had been reached 20ml of 50% citric acid was added to stop the reaction and the gel was washed with distilled water. The gel was then photographed as above (2.17.5.)

2.17.7 Densitometry

Densitometry was performed on Coomassie stained polyacrylamide gels using a LKB 2222-020 Ultroscan XL Enhanced Laser Densitometer, (Bromma).

3. DNA Sequencing of pDUB1900

Almost 9kb of DNA from the largest *Eco*RI fragment of pDUB1900 was sequenced. This region was known to contain DNA complementing three phenotypically distinct motility mutants (*mot-1*, *mot-12* and *fla-15*) and is shown in figure 3.1.

Generally, DNA fragments to be sequenced were subcloned into pUC19 and the universal M13 forward and reverse primers used as described in section 2.15. The sequencing strategy used and also a partial restriction map of the region sequenced is given in figure 3.2. Two custom-synthesised oligonucleotides were used as primers in a number of the sequencing reactions, which are also shown in figure 3.2.

Oligonucleotide T

G A A A A C G G G A A A G G T T C C G T (tcaggacgctacttgtgtataagagtcag).

The oligonucleotide synthesised is given in bold, block, capitals. It is complementary to a region of the inverted repeats of the transposon Tn5 29 nucleotides from each termini (these are listed in lower case afterwards). Using this oligonucleotide in a sequencing reaction allowed sequence to be obtained out of one of the Tn5 termini and directly into the "chromosomal" DNA of a cloned mutant.

Oligonucleotide F

G C A T A A A G G T C C G A A A G G C T G

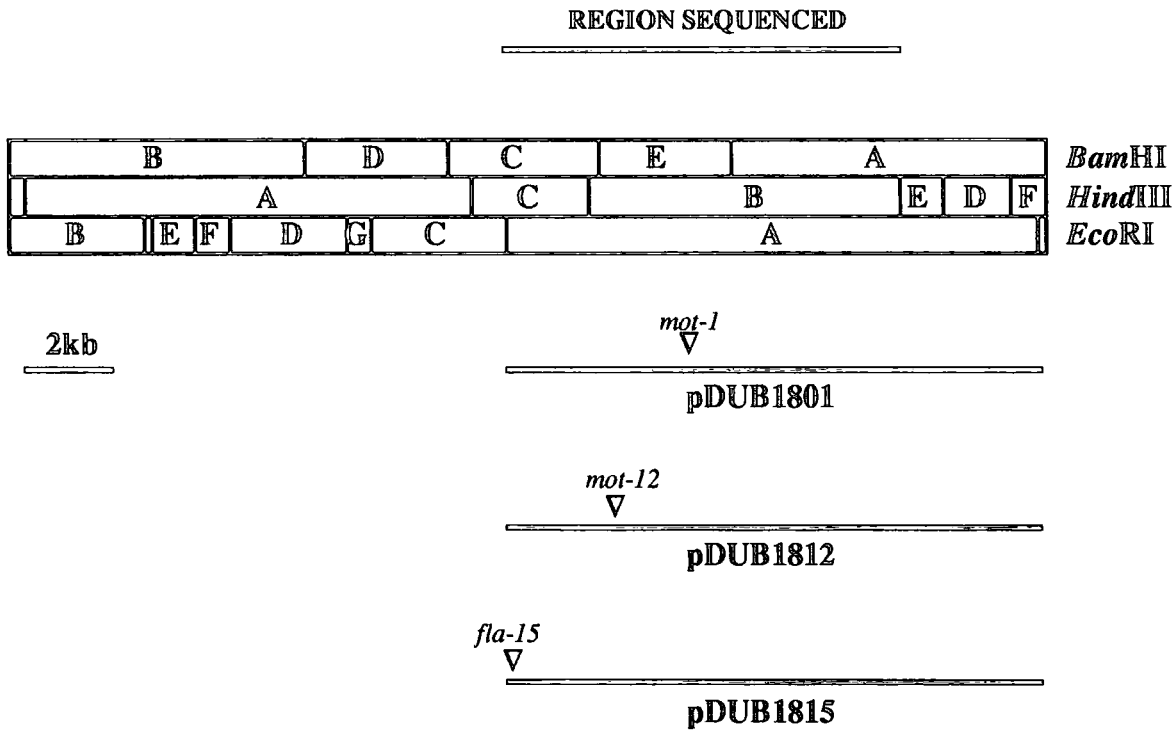
This oligonucleotide was complementary to a DNA sequence of pDUB1900 approximately 20 base pairs upstream from a region that created band-compression whilst trying to sequence through it with the Universal primers. The use of this oligonucleotide allowed the clear resolution of this region.

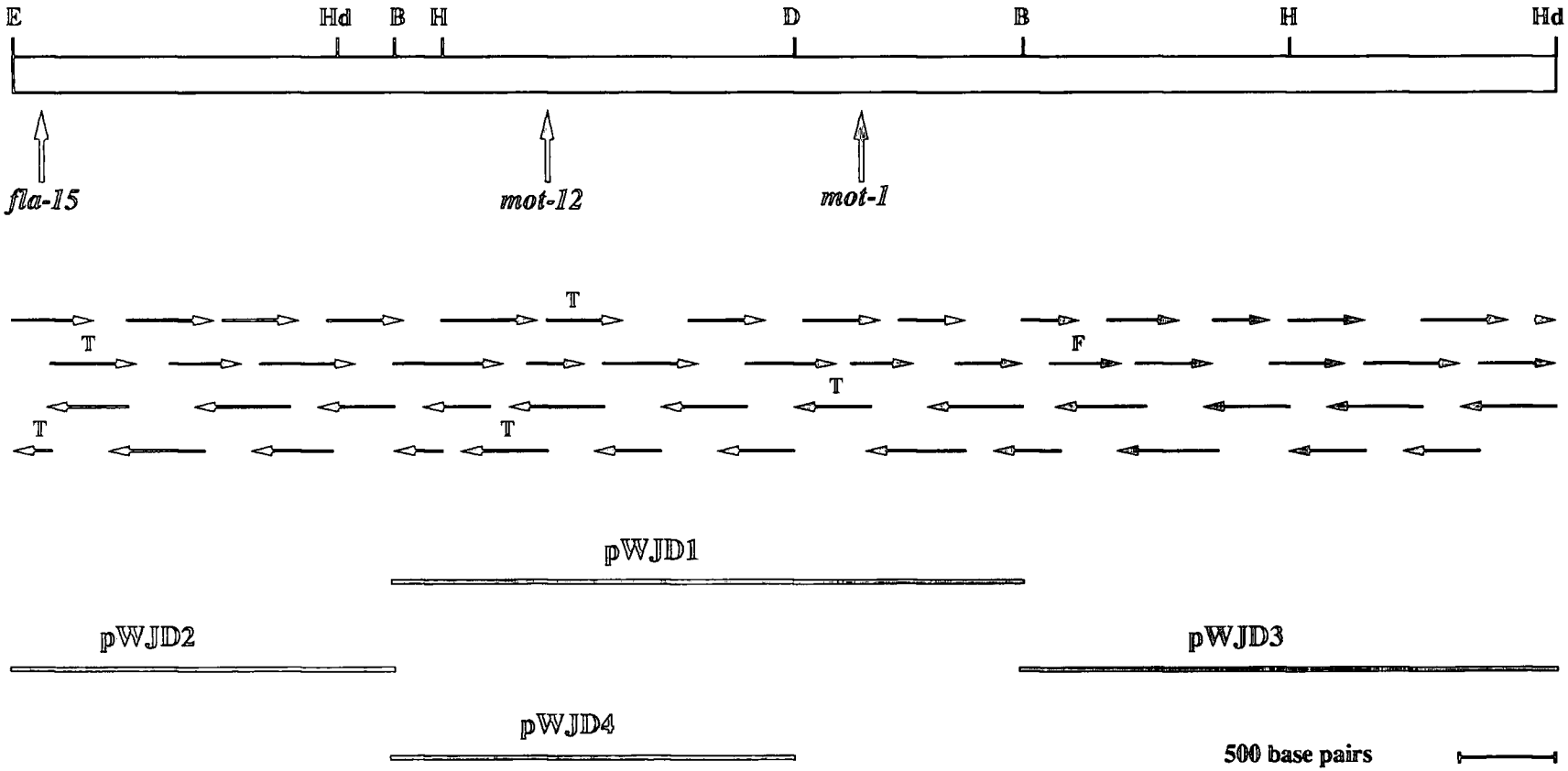
The DNA sequences generated were initially analysed and aligned using the Macintosh computer programs DNA Strider™ 1.2 [187] and DNAid. The entire region was sequenced completely in both directions with overlapping DNA fragments. Some of the predicted restriction enzyme sites were confirmed by experimental digests of the DNA, and the sizes of the generated fragments compared with those predicted by the computer.

The sense strand of the complete sequence (8624 nucleotides) is listed in its entirety in figure 3.6 at the end of this chapter.

Figure 3.1 A partial restriction map of pDUB1900 showing the region that was sequenced.

The boxes in the centre of the figure represent the chromosomal insert of pDUB1900 and the *Bam*HI, *Eco*RI and *Hind*III sites within it. The line above shows the extent of *Eco*RI fragment A that was sequenced. The three lines below depict the cloned fragments isolated from the (labelled) mutants, with the Tn5 insertion site in each case denoted by ∇. The diagram was adapted from one provided by Dr. C. H. Shaw.



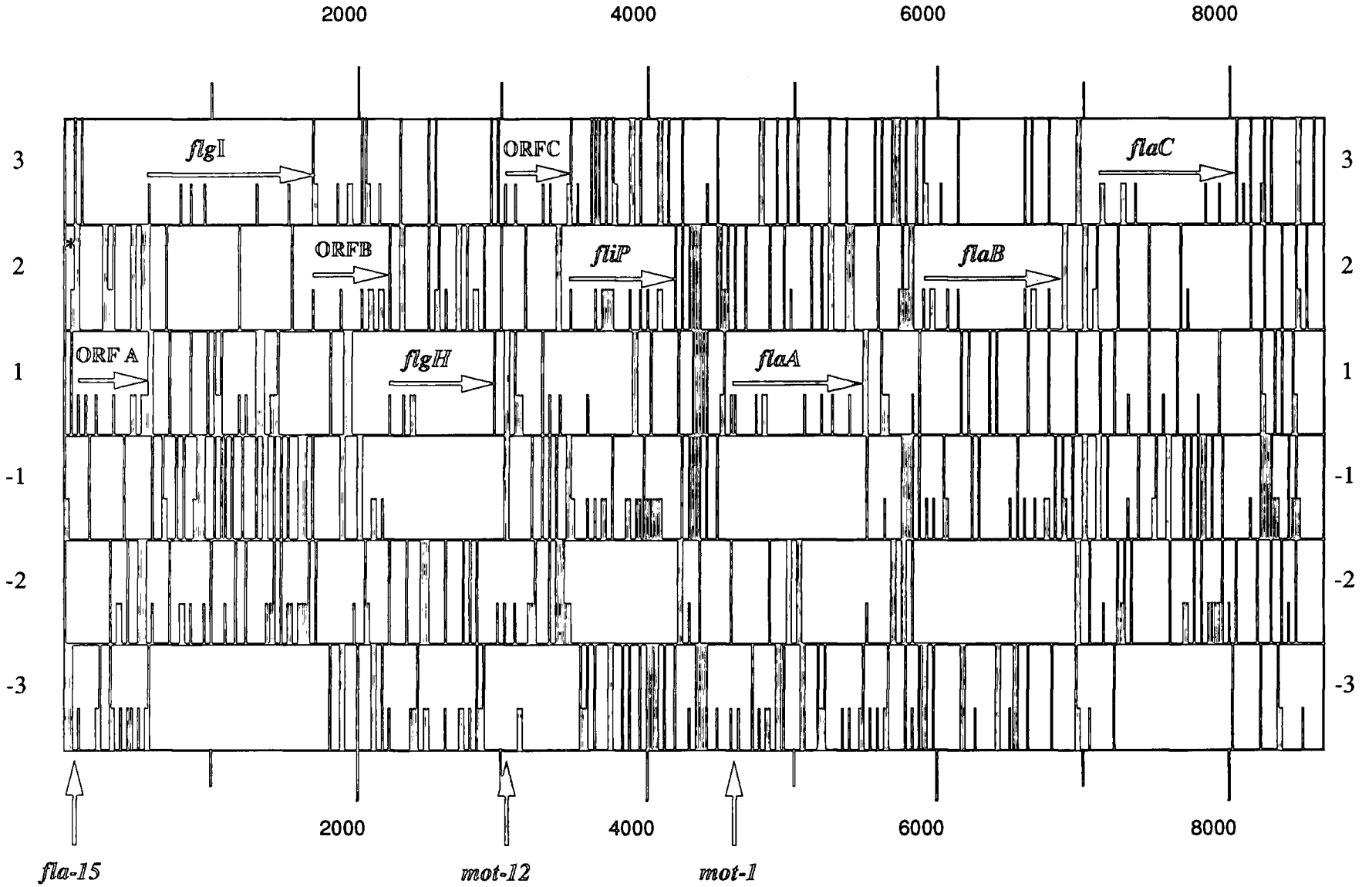

E
Hd
B
H
D
B
H
Hd
fla-15
mot-12
mot-1
T
T
T
T
F
pWJD1
pWJD2
pWJD3
pWJD4
500 base pairs

The sequence was compared against nucleic acid and protein databases to identify homologous regions. Sections of the DNA were manipulated using the University of Wisconsin Genetics Computer Group (UWGCG) Sequence Analysis Software Package [65] mounted on the SEQNET VAX 3600 at SERC, Daresbury, U.K., and then compared against the GenBank and EMBL databases using the FASTA program [230]. Alternatively open reading frames (ORFs) in the DNA sequence were identified using DNA Strider, translated into their protein sequence and compared against the Owl database using the SEQNET programme SWEEP.

As a result of the database searches several open reading frames with sequence identity to previously identified genes/proteins involved in flagellar structure and assembly were located. The position and sizes of the open reading frames containing the homologues are given in figure 3.3. More detail concerning these homologies will be given in the next two chapters. Where possible the new unified flagella nomenclature according to Iino *et al.* [131] has been used for the *A. tumefaciens* homologues.

Initially the homologues were divided into two groups according to their position (or involvement) with flagella. Four of the homologues identified (*flgG*, *flgH*, *flgI* and *fliP*) are associated with the basal body of the flagella in *E. coli*. The FlgI and FlgH proteins make up the P- and L-rings, respectively, of the basal body. FliP is known to be required at an early stage for flagellar assembly. FlgG is found at the distal end of the rod [179]. Only the region encoding the carboxy-terminal section of the *A. tumefaciens* FlgG homologue was sequenced in this study.

The second group of homologues represent three copies of the gene encoding flagellin, which is the main structural protein of the flagellar filament itself. The homologues were arranged in tandem, the furthest upstream being mutated by the Tn5 insertion site in the mutant *mot-1*. Since there are multiple copies of the flagellin homologues, the nomenclature used for them differs from that of Iino *et al.* [131] which has only one name for the flagellin gene (*fliC*). Instead the nomenclature used is that of the previously identified flagellin genes of *R. meliloti* [27, 235] *flaA*, *flaB* and *flaC*, to which the *A. tumefaciens* homologues were most similar.


2000
4000
6000
8000
3
2
1
-1
-2
-3
flgI
ORFC
flaC
ORFB
fliP
flaB
ORF A
flgH
flaA
2000
4000
6000
8000
fla-15
mot-12
mot-1

As figure 3.3 shows, between each of the flagellar gene homologues (that is the partial *flgG* homologue and the *flgI*, *flgH* and *fliP* homologues) are open reading frames (ORFs A, B and C) which hold no significant homology to any previously sequenced genes. ORFs A and C are mutated by Tn5 insertion sites in the *fla-15* and *mot-12* mutants, respectively. Initially these open reading frames were considered to be putative coding regions, because no potential promoters or transcription terminators could be found for any of the four flagellar gene homologues. When the entire region of sequenced DNA was analysed for potential protein coding sequences using the UWGCG program TESTCODE [84], ORFs A, B and C, and all the flagellar gene homologues, were predicted to be protein coding regions. The results of the TESTCODE analysis are shown in figure 3.4, as well as a brief explanation of them.

The results of the TESTCODE analysis and the absence of any recognisable promoter or terminator sequences within the DNA region from the partial *flgG* homologue to the *fliP* homologue, led to the preliminary conclusion that this entire region might be under the same transcriptional control, *i.e.* an operon. The promoter region is presumably further upstream on pDUB1900. Flagellar genes (including these homologues) have been found in large operons in other bacteria (see Introduction). The *A. tumefaciens* flagellin gene homologues however all appear to be monocistronic, from sequence analysis. A summary of this is presented in figure 3.5. Subsequent examination of the DNA region from the partial *flgG* homologue to the *fliP* homologue revealed three possible promoter sequences. The presence, and possible functional roles, of which will be discussed in Chapter 4.

Further analysis of all the open reading frames in the putative operon is presented in Chapter 4. In particular, the properties of the unidentified open reading frames (A, B and C) are discussed as well as arguments for and against the presence of the putative operon. The flagellin gene homologues, their transcriptional control and their contribution to flagellin structure are discussed in Chapter 5.

ORFA
flgI
ORFB
flgH
ORFC
fliP
flaA
flaB
flaC
TestCode
14
12
10
8
6
4
0
2,000
4,000
6,000
8,000

Figure 3.5 The putative transcriptional arrangement of the flagellar gene homologues and unidentified open reading frames.

The region of DNA sequenced is represented by the box. Below it are the positions of selected restriction enzyme sites, B=*Bam*HI, E=*Eco*RI, H=*Hpa*I and Hd=*Hind*III. The positions of the Tn5 insertion sites of the three mutants are shown beneath the box. For reasons of scale the *flgG* homologue is not labelled.

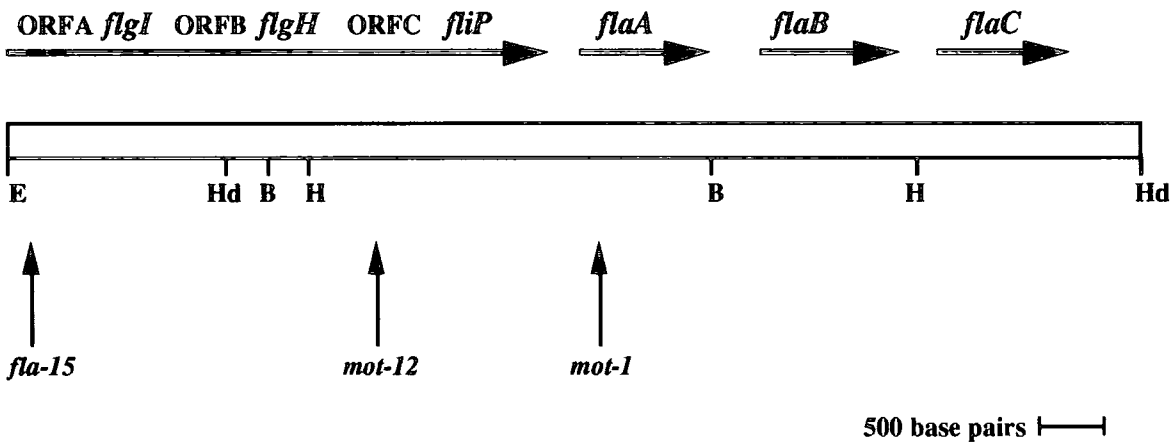


Figure 3.6 The complete listing of the sense strand of DNA sequenced in pDUB1900.

```
        |   10      |   20      |   30      |   40      |   50      |   60
   1 GAATTCCAAG GTCATCACCA CCGCTGACGA  AATGGCTTCG ATTGTCAGCA AGAACCTGAA 60
  61 GTAAGAACGG AAAGGCAGGC GAACATGAGG  TTTGGCCGGA ACAATAGCAG CTGCAGAACG 120
 121 GCACTCGTCC GTATGTGTCT CGCGTCTGCC  TTTTCACTGG GCGCCCTGGC GCCCGCTCTT 180
 181 GCGCAGGCGC CAATGGCGCT GGTTCCCGTG  CGCACCATCT ATCCGGGCGA GGCGATCTCG 240
 241 CCCGAACAGG TGAAGTCGGT GGAAGTGACC  AACCCGAATA TTTCCGCTGG TTATGCGAGC 300
 301 GATATCAGCG AAGTGGAAGG CATGATCTCC  AAGCAGACAT TGCTTCCCGG CCGCACGATC 360
 361 CCCATTGCGG CGCTTCGCGA ACCATCACTG  GTGGTGCGCG GTACAAGCGT CAAACTCGTT 420
 421 TTCCACATTG GCAACATGAC GCTGATGGCT  TCCGGAACGC CAATGAGCGA CGGCTCGCTC 480
 481 GGCGAGGTCG TCAGGGTACG CAACATCGAT  TCCGGCGTGA TGGTCAGCGG CACGGTCATG 540
 541 AAGGACGGAA CCATTCAGGT GATGGCGAAA  TGAGAGTGCT TCGTATCATA GCCGCGGCTT 600
 601 TGGTCTTTTC GGCCCTGCCC TTCCTCTCCA  CGCCGCCCGC GCAGGCCGAC ACGTCGCGAA 660
 661 TCAAGGATAT CGCATCCCTG CAGGCCGGAC  GTGATAACCA GCTGATCGGT TACGGTCTCG 720
 721 TCGTGGGTCT GCAGGGTACC GGCGACAGCC  TGCGCTCCTC GCCTTTCACC GAACAGTCCA 780
 781 TGCGCGCCAT GCTGCAGAAC CTCGGCATCA  CCACGCAGGG TGGCCAGTCC AACGCCAAGA 840
 841 ACATCGCCGC CGTGATGGTG ACGGCCAACC  TGCCGCCATT CGCCAGCCCC GGCAGCCGTG 900
 901 TGGACGTCAC GGTCAGCTCG CTTGGCGACG  CCACCTCGCT TCGTGGCGGC AACCTCATTA 960
 961 TGACTTCGCT TTCCGGCGCG GACGGGCAGA  TTTACGCCGT CGCGCAAGGG GCGCTGATCG 1020
1021 TCAACGGTTT TTCGGCCCAG GGCGATGCCG  CGACGCTGAC GCAGGGCGTC ACCACGTCGG 1080
1081 CGCGCGTGCC GAACGGCGCG ATCATCGAGC  GGGAACTGCC GTCGAAGTTC AAGGATTCGG 1140
1141 TCAATCTGGT CCTTCAGCTG CGCAACCCGG  ATTTTTCGAC AGCCGTGCGT GTGGCTGATG 1200
1201 TCGTCAACGC CTTTGCCCGC GCCCGTTATG  GCGACCCGAT CGCCGAACCG CGCGATTCGC 1260
1261 AGGAAATCGC GGTGCAGAAA CCCCGCGTCG  CAGACCTCAC CCGGCTGATG GCGGAAATCG 1320
1321 AAAATCTGAC GGTCGAGACG GATACGCCTG  CCAAGGTCGT GATCAACGAG CGCACCGGCA 1380
1381 CGATCGTCAT CGGAGCGGAT GTGCGCATTT  CCCGCGTCGC CGTCAGCTAT GGAACGCTCA 1440
1441 CCGTGCAGGT GACGGAATCG CCGCAGGTCA  TTCAGCCCGC ACCCTTCTCG CGCGGCCAGA 1500
1501 CGGCAGTGCA GCCGCAGACG GATATCATGG  CCATGCAGGA AGGTAGCAAG GTCGCCATCG 1560
1561 TCGAAGGCCC GGACCTTCGC ACGCTGGTTG  CCGGTCTCAA CAGCATCGGG CTCAAGGCGG 1620
1621 ACGGCATCAT CGCCATCCTG CAAGGCATCA  AATCCGCCGG CGCCCTTCAG GCGGAGCTCG 1680
1681 TGCTGCAATG ATGGAACGTC AGACCAAAAA  TCCGCTCTCA AATGGCCTCG TCCGCTTTGC 1740
1741 GGCCGTCGCG TCGCTGCTGT TCCTGTTGCC  GGTTGCCGGC GCCGAAAGCC AGCAGAACGT 1800
1801 GGTGTCTGAA CTCAGCACGC AGGACGAAAT  CCAGAAATTC TGCACCAATA TTGCCGATGC 1860
1861 GGCCCCGCGAC CAGCGTTACC TCATGCAGAA  GCAGGACCTT GAAAAGCTTC AGGCCGACGT 1920
1921 CAATGAGCGT ATTTCCGTGC TGGAAAACCG  CAAGGCGGAA TATGAGGACT GGCTGGCGCG 1980
1981 CCGCGAACAT TTCCTCAATC AGGCCAAAAG  CAACCTCGTC GACATCTACA AGACGATGAA 2040
2041 GGCGGACGCC GCAGCTCCCC AGCTAGAAAA  GATGCATGTG GAAATCGCGG CCGCCATCAT 2100
2101 CATGCAGCTG CCACCGCGCC AATCGGGCCT  CATCCTCAGC GAAATGGATG CGCAGAAAGC 2160
2161 CGCAACCGTC GCCGGCATCA TGTCGCAGGC  AATCGACAAA AACACTTCGA AGGATCCTTC 2220
2221 ATGAGCACGC GTCGTCTTCC GGCCCTCCTC  CTGCCGCTCG CTCTCCTGGC CGGCTGCCAG 2280
2281 AATAACCAGA CCCTGAAGGA AATCGGCAAC  GCACCCGCGA TGAGCCCGAT CGGCAGCGGC 2340
2341 CTGCAATTCA GCCAGACACC GCAGATGGGC  ATGTATCCCA AGCAGCCGAA ACACATGGCA 2400
2401 AGCGGCTATT CGCTGTGGAG CGACAGCCAG  GGTGCGCTGT TCAAGGATCT GCGCGCACTC 2460
2461 AATATCGGCG ACATCCTGAC CGTTAACATC  CAGATCAACG ACAAGGCCGA TTTCGACAAC 2520
2521 GAGACCGAGC GTAACCGCAC CAATGCCAGT  GGCCTGAACT GGAAGGCAAA AGCCCAGATC 2580
2581 CTGGGCTGGA CGCCGGATGC GGATTCCAGC  ATCAAATACG GCTCCGACAC CGACACACAG 2640
2641 GCCAAGGGCA AGACCAAGCG CTCCGAAAAG  CTGACGCTGC TGGTGGCTGC CGTCGTGACC 2700
2701 GGCATTCTGG AAAACGGCAA CCTCATCATC  AGCGGCTCGC AGGAAGTGCG TGTGAACCAC 2760
2761 GAAATCCGCA TCCTCAACGT CGGCGGTATC  GTCCGTCCGC AGGATGTCGA TGCCCAGAAC 2820
2821 ATCATCTCCT ACGAGCGCAT CGCCGAAGCA  CGCATCTCCT ACGGCGGTCG TGGCCGTCTG 2880
2881 ACGGAAGTGC AGCAGCCCCC GGTCGGGCAG  CAGGTCGTTG ACCTGTTCTC GCCGCTCTGA 2940
2941 CGCCGCACAC GGGAACGGAC TGACCCATGG  AAAACGAACA GGCTGAGGGC AAAAAAAAAT 3000
3001 CCTCCCCTCT GGTCATGACC ATCGCAGGCG  TTGTGATCCT CACCCTGCTC GGTGCGGGCG 3060
3061 GCGGCTGGCT CGTGGGCGGC ATGATCGCCC  CGAAAGTTGC CGCAACGGAG GCCCATGCCA 3120
3121 CCGCTGCGGC CGGCGGCCAT GGCGAGAAGA  AGGGCGAAGG CCTCGATAAA ATCGACGCCG 3180
3181 AAGCGAACGG CATCGTCCAG CTCGATCCCA  TCACCACCAA TCTCGCCTAC CCTTCGACCA 3240
3241 ACTGGGTGCG GCTGGAAGTG GCGCTGATGT  TCAAGGGGCC GGTGGAAGTC GGGCTGGCGG 3300
3301 AGGATATCCA CCAGGACATC ATGGCCTATG  TCCGGACCGT TTCCCTCCAG CAGCTGGAAG 3360
3361 GCCCGCGCGG CTTTCAATAT CTTAAGGATG  ACATTCAGGA ACGAGTTGAC CTGCGCTCTC 3420
3421 AAGGGCGCGT ATCCAAGGTC ATGTTCCGGA  CCTTTGTCAT CGAATGATTC GATTTCTTGT 3480
3481 AACCATCGCC GTCCTCCTCG CTTTGCCGGG  CCTTGCCAAT GCCCAGCAAT TTCCGTCCGA 3540
3541 TCTGTTCAAC ACGCAGATCG ACGGCTCCGT  CGCGGCATGG ATCATCCGCA CCTTCGGCCT 3600
3601 GTTGACGGTT CTCTCCGTCG CGCCCGGCAT  CCTGATCATG GTCACGAGCT TTCCGCGCTT 3660
3661 CGTGATCGCC TTTTCGATCC TGCGCTCCGG  CATGGGACTT GCGTCGACAC CGTCGAACAT 3720
```

```
3721 GATCCTGCTG TCGATGGCGA TGTTCATGAC  CTTTTACGTC ATGTCCCCCA CCTTCGACAA 3780
3781 AGCCTGGACG GATGGCGTGC AGCCGTTGCT  GCAGAACCAG ATCAACGAAC AGCAGGCCGT 3840
3841 GCAACGCATT GCCGAACCCT TCCGCACCTT  CATGAACGCC AATACGCGTG ACAAGGACCT 3900
3901 GAAGCTGTTC GTTGACATCG CCCGCGAGCG  CGGACAGGTG GTCATGACCG ACAATGTCGT 3960
3961 GGACTATCGC GTACTGGTTC CTGCCTTCAT  GCTCTCGGAA ATCCGGCGCG GTTTTGAAAT 4020
4021 TGGTTTCCTC ATCATCCTGC CGTTCCTCGT  CATCGATCTG ATCGTCGCCA CCATTACCAT 4080
4081 GGCGATGGGC ATGATGATGC TGCCGCCCAC  CTCGATTTCG CTGCCGTTCA AGATCCTGTT 4140
4141 TTTCGTGCTG ATCGACGGCT GGAACCTGCT  CGTGGGAAGC CTCGTCAGAT CGTTCAACTG 4200
4201 ACGGCCCTTA TACCCCGCTG ATACAGATTA  ACCCTGTTTT TTAACCCCGC ACGCCCAAGG 4260
4261 CCGCGGGGTT TTCGCATTTT AGTGCTTTGG  TTACCATTTT CACTAGCGAC TGGGTAACTA 4320
4321 TATTTTTAAA TAAAAAAGTT ACTATTAAGC  TAAATCAAAA AGTAACCATA AGTATTTTAC 4380
4381 AGTATTTCTT AACGCTCAAA TTAATGATTG  CGCAATTATT CGCTGCGAAA TTCTTCCCAC 4440
4441 ACGCAGCGGG TTTGGTTTCT GGAACTAAGC  GTTTCTGAAC CGAGTGGCAT GAAGCCGAAA 4500
4501 CGTTGCGACC GGTAATATCC AACCCGGTAT  GTCCCCACTC CGTATCTCGT AAAAACAAGG 4560
4561 GACACATTTA TTATGGCAAG CATTCTGACC  AACAACAACG CAATGGCCGC TCTCTCGACC 4620
4621 CTGCGTTCCA TCGCTTCCGA CCTGTCGACC  ACGCAGGACC GTATTTCTTC CGGCCTGAAG 4680
4681 GTTGGCTCGG CTTCGGACAA CGCTGCTTAC  TGGTCGATCG CGACCACCAT GCGCTCCGAC 4740
4741 AACAAGGCTC TCGGCGCAGT TTCTGACGCG  CTGGGCATGG GCGCTGCCAA GGTTGACACC 4800
4801 GCTTCCGCCG GTATGGACGC CGCCATCAAG  GTCGTGACCG ACATCAAGGC AAAGGTCGTT 4860
4861 GCCGCCAAGG AACAGGGCGT TGACAAGACC  AAGGTTCAGG AAGAAGTTTC TCAGCTTCTC 4920
4921 GATCAGCTGA AGTCGATCGG CACGAGCGCG  TCTTTCAACG GTGAAAACTG GCTCGTTTCC 4980
4981 AGCGCCAATG CTACCAAGAC CGTTGTATCC  GGCTTTGTTC GTGACGCTGG CGGCACCGTT 5040
5041 AGCGTCAAGA CGACGGACTA CGCGCTGGAC  GCCAACTCCA TGCTTTACAC GGAAGGTACA 5100
5101 CCGGGCACGA TCGACGCAAA CTCCGGTATC  CTGAACGCGA CCGGTGCAAC GACCACCGTC 5160
5161 GGCGCCAAGA CCTACACACA GATCTCCGTG  CTCGACATGA ACGTCGGCAC CGACGACCTC 5220
5221 GACAACGCTC TCTACTCCGT TGAAACCGCT  CTGACGAAGA TGACCAGCGC TGGCGCCAAG 5280
5281 CTCGGTTCGC TGTCTGCACG TATCGACCTG  CAGAGCGGCT TCGCCGACAA GCTGTCCGAC 5340
5341 ACCATCGAAA AGGGTGTAGG CCGTCTCGTC  GACGCTGACA TGAACGAAGA GTCCACCAAG 5400
5401 CTGAAGGCTC TGCAGACGCA GCAGCAGCTG  GCAATCCAGG CTCTGTCGAT CGCCAACAGC 5460
5461 GACTCGCAGA ACATCCTGTC GCTCTTCCGC  TAAGAGCCGA GCTTCTCATA GCCCTTCGGG 5520
5521 GCTGGGGGAT CCATTTCTTC CCACAAGCAA  TCGGGGGCTG ACCGAACGCA GATTGGGATA 5580
5581 GAAAGCATAA AGGTCCGAAA GGCTGAATCG  GACTGAATGC ATCCACGAGG CCGCATCCGG 5640
5641 AAACGGATGC GGCCTTATCT TTTTTTTAAA  CTTTTGCAGA TACGCGCGCG TGCGCGTGAG 5700
5701 GGTTGCAATT TTAATTTTGC CGGGAATGAA  TTTCTTTCCG CGCTTATCAA TTCATTAACC 5760
5761 AAACCCGCGT TACCTAATGC TCATCGAAAC  GGCGAGCTAA CGTTAAAAAT CAGCCAGTTA 5820
5821 GCAGGCATGA TGCTGTGCGC CGTTCCCGGT  GAAACCGGAA TGTCCCTTCT TTAATCAGCC 5880
5881 ATTCAAGGGG CACACTACTA TGACGAGCAT  TATCACGAAT GTCGCAGCAA TGTCTGCGCT 5940
5941 CCAGACCCTG CGCTCTATCG GCCAGAACAT  GGAATCCACC CAGGCACGCG TCTCCTCCGG 6000
6001 CCTTCGCGTC GGCGATGCTT CCGACAACGC  TGCCTACTGG TCGATCGCAA CCACCATGCG 6060
6061 CTCCGACAAC ATGGCTCTCT CTTCCGTTTC  CGACGCTCTC GGCCTCGGCG CCGCAAAGGT 6120
6121 GGACACTGCT TCCGCCGGTA TGAGCTCGGC  AATTGACGTC GTTAAGGAAA TCAAGGCAAA 6180
6181 GCTGGTCACC GCGACTGAAG AAGGCGTCGA  CCGCACCAAG GTTCAGGAAG AAATCGGCCA 6240
6241 GCTGCAGAAG CAGCTCGCAT CGATCTCGCA  GGGCGCTTCC TTCTACGGCG AAAACTGGCT 6300
6301 CGTCGGCGTC TCGACGCTCG GCGCAGCAAC  GCCCGGCACT GACCCGGACA AGTCTGTCGT 6360
6361 CGCCGGCTTC GTCCGCGCTT CCGGCGGTGC  AGTTAGCGTT ACGACCACGA AATACGCCCT 6420
6421 CGACAACACC GCTACCGGCA ACGTTCTGTT  CGGTTCGGTC GACGGTACCG GCACTCCCGA 6480
6481 CGCATCCGGC ATTCTCGGCA CTGTTGGCAC  GTTCACGACG GTCGCTGCCC AGTCTGTCTA 6540
6541 CACACTGGAT ATCACCCAGT ACGCCGCCGC  AGACCGCGCC ACCAACATGG CAGAAGCCCT 6600
6601 GACGCTGGTT GAAAACTCGC TTAAGGCAAT  GACCAGCGCC GCTGCAAAGC TCGGCTCGCT 6660
6661 CTCCATGCGT ATCGGCCTGC AGGAAGACTT  CGCTTCCAAG CTGTCCGACT CCGTCGAAAA 6720
6721 GGGCATCGGC CGCCTCGTGG ACGCTGACAT  GAACGAAGAG TCCACCCGCC TCAAGGCTCT 6780
6781 GCAGACACAG CAGCAGCTCG GCGTTCAGGC  TCTCTCGATC GCCAACAGCA ACTCCGAAAG 6840
6841 CATCCTGTCG CTCTTCCGTT AATCGAAAGC  TGCAGCCGAA AACATGTAAC CGCGCCAGAA 6900
6901 AGGCGCCGGT TTTTGTTTGT GCTCTTACCT  GCGAAATTCA CGAGATTTTT TGATTAAGAG 6960
6961 TTCATTAACC AAATCACGGC TAAATCAGCG  CATCGAAACG ACTCGTTAAC CAAGACGAAA 7020
7021 AAGAGGTTAA CAGGCATCAA GCCGCTCAGT  CGTTTCCGGC GATCCCGGAA TGTCCCTTCC 7080
7081 CATTTAGCCA GCTCAGAGGG GCAATTATGA  CAAGTATTCT GACGAACACC GCCGCCATGT 7140
7141 CGGCGCTCCA GACCTTGCGC GCAATCAGCG  GCCAGCTCGA AGACACACAA TCCAGAGTTT 7200
7201 CTTCTGGATT ACGCGTAAAA TCGGCCTCCG  ATAACGCCGC CTATTGGTCG ATCGCGACGA 7260
7261 CCATGCGTTC CGACAACATG GCTCTCTCCG  CCGTTCAGGA TGCATTGGGT CTCGGCGCGG 7320
7321 CCAAGGTCGA CGTCGCTTAT TCCGCGATGG  AAAGTACCGT CGAAGTCGTC AAGGAAATCA 7380
7381 AGTCGAAAAT CGTCGCGGCA ACCGAAGAGG  GTGTCGACAA GACCAAAATC CAGGAAGAAA 7440
7441 TCGACCAGCT GAAAAAACAG CTTGAATCGA  TTGCACAAGG TGCATCTTTC AGTGGCGAGA 7500
7501 ACTGGCTGCT GGGCACGGGC GCCAAGACCG  TCGTTTCCGG TTTTGTGCGC GATGGCGGCG 7560
7561 GCACCGTCAG CGTTACGAAA ACGGACTATA  CGCTTATCGA CACCACCGCC GGCACAACAA 7620
7621 CCGCCAACGT GCTTTTCGGC CTGACCGGCT  CGCCCGCAAC GCTTGACACA ACAAAGGGTA 7680
7681 TCATCGGCCA GCCCGGCACG GCTTCCGGCA  TTTCCGTATG GGATATCGAT CTCAAGCTGT 7740
7741 TCAATTCTGC AACGCCGCCG ACCTACACCA  TCGGGAACCT TCTCACCGAT GTCGAAACGG 7800
7801 CATTTCAGTC GATCAGCAGT GCATCCGCCG  CTCTCGGTTC GATCAAGATG CGCATTGGCC 7860
```

```
7861 TGCAGGAAGA CTTCGTCTCG AAACTCACCG  ACTCCATCGA CAAGGGCATC GGCCGTCTCG 7920
7921 TCGATGCGGA CATGAACGAA GAATCCACCA  AGCTCAAGGC TCTTCAGACA CAGCAGCAGC 7980
7981 TGGGTATCCA GTCGCTCTCC ATCGCCAATA  CCAGTTCCGA AAATATCCTG TCGCTATTCC 8040
8041 GCCAGTAAGC GGCGCCCGTT GGCTACGGGC  GGTTTCGTCC CGCTCTGGCA GCAACAGAAT 8100
8101 ATTCGGATAT GGAGACCGCG CTTCGCAAGA  GGCGCGGTTT TCCATTTGAC GCGGCCTGAT 8160
8161 GCAACGCGTG CGAACCGAAG CGCGACGACG  CCTTCGGCGT TTGCGATTAT CTAAGTATCT 8220
8221 GTTTTATGGG AGAAAATGGT GCTGCTAGAG  AGATTTGAAC TCTCGGCCTC TCCCTTACCA 8280
8281 AGGGAGTGCT CTACCCCTGA GCTATAGCAG  CATCCGGTGC CGAAGCGTCT GCTTCAGCAT 8340
8341 CAAGCGTGGC GGCCTATTGC CATAGGTTTT  TGACGAGCGC AAGTCGCAAA ACGATATTCT 8400
8401 TCATCCAGTG GGGCAAAAAA GCGTGCCCCC  TGTTGAAAAA CGCATTTTTC AGATATTGCT 8460
8461 GAATCTATGA ACGAACAACA TGACAAACAG  GCGAAGGCGG CGGTTTCCGT GGACATTTCG 8520
8521 AAGGAACCGT CGGCCGTAGT GGCAAACGGC  AACACGCAGG CCGGCCCCGG TGAAAAGGCG 8580
8581 CGCCAGCGTG AGGCGGAGGC AAGGCGCGAG  CGGCGAAGAA GCTT                  8624
```

4. Further Analysis of the Putative Flagellar Gene Operon of *A. tumefaciens*

This chapter presents a more detailed analysis of the sequenced flagellar homologues and unidentified open reading frames within the putative operon. Hybridisation of these homologues to chromosomal DNA digests was used to assess the degree of sequence conservation in other bacteria. More detailed mapping of the positions of the homologues (and unidentified open reading frames) was performed in the same way against the pRZ cosmids, known to complement motility mutants (and contain flagellar genes) from the closely related bacterium *R. meliloti* [27, 341]. The construction of the *fliP*- motility mutant is also described in more detail. Finally the experimental data will be discussed with respect to the existence of the flagellar operon.

4.1 The *flgG* homologue

This open reading frame was only 63 base pairs in length. When translated, the 21 amino acids had 55% identity to the C-terminal amino acids of FlgG from *B. subtilis* [342]. The position of the sequence identity is shown in the gap alignment below. The alignment was produced by the UWGCG program GAP, which uses the algorithm of Needleman and Wunsch [207].

Figure 4.1.1 A gap alignment between the partial *A. tumefaciens* FlgG homologue and the carboxyl-terminal amino acids of FlgG from *B. subtilis*.

```
 Percent Similarity: 70.000   Percent Identity: 55.000

  1 .........................................NSKVIT 6
                                              |||:||
201 TASAANQPGDGGTGALKSGFLEMSNVDLTDEFTEMIVAQRGFQSNSKIIT 250


  7 TADEMASIVSKNLK 20
    |.||:   :  |||
251 TSDEILQELV.NLKR 264
```

The amino acid sequence on the top line is that of the *A. tumefaciens* FlgG homologue, and the C-terminal amino acid sequence of FlgG from *B. subtilis* is shown below. In *B. subtilis* the complete FlgG protein is 264 amino acids, and in *S. typhimurium* the FlgG protein is 261 amino acids [342]. Thus if this is the FlgG homologue and it is similarly conserved, the open reading frame would be expected to extend at least 700 base

pairs upstream. Completion of this open reading frame sequence will be necessary before its role in the *A. tumefaciens* flagellum can be determined.

Twenty base pairs downstream of the stop codon of the partial *flgG* homologue is the start codon of ORFA. Approximately one hundred base pairs downstream (within ORFA) are a number of sequences that could form potential stem-loop structures. The position of these structures is shown in the section describing ORFA (4.5). Their possible regulatory roles are considered in the discussion section of this chapter.

In the enteric bacteria FlgG forms the distal part of the rod. It is a member of the axial family of proteins, exported by the flagellum-specific export pathway [179, 180]. FlgG in *B. subtilis* is believed to have a similar structural role and export mechanism [7, 342]. As described in the introduction the rod is assembled cooperatively from the proteins FlgB, C, F, G and possibly FliE [156, 205]. FlgG is the major protein of the rod with c. 26 copies per basal body [278].

4.2 The *flgI* homologue

The gene product encoded by this 1122bp open reading frame has 373 amino acids (predicted M_r 38,943Da) with a putative signal peptide at its amino-terminus. This predicted protein was found to have sequence similarity to the FlgI protein from several bacteria; 52% identity (69% similarity) over 370 residues to FlgI of *C. crescentus*, 44% identity (63% similarity) over 365 residues to FlgI of *S. typhimurium* and 47% identity (66% similarity) over 367 residues to FlgI of *Pseudomonas putida*. The extent of the similarity is shown in a multiple alignment, produced using the SEQNET program CLUSTAL V [119] of the four FlgI homologues in figure 4.2.1. The alignment shows the similarity between the proteins to be throughout their entire lengths, the greatest divergence between them being within their putative signal peptides at the amino-termini. The similarity is also shown in figure 4.2.2 which gives hydropathy plots of the four proteins using the Kyte and Doolittle analysis [160]. All four proteins have comparable plots. This similarity will be discussed later in relation to the function of the FlgI proteins.

The predicted cleavage site of the FlgI signal peptide is shown on figure 4.2.3. Removal of this 26 amino acid long signal peptide would leave a polypeptide of 347 amino acids with an molecular weight of 36,209Da. The 26 residues meet the criteria described for signal peptides [218]. There are two positively charged residues close to the amino-terminus, followed by 21 predominantly hydrophobic amino acids and an Ala-Gln-Ala sequence that could serve as a cleavage site.

The nucleotide sequence of the *flgI* homologue in *A. tumefaciens* has a higher guanine and cytosine content, 64%, than regions of the DNA sequence thought to be non-coding. The TESTCODE program predicts the largest non-coding regions to be between each of the *fliP*, *flaA*, *B* and *C* homologues, and the average guanine and cytosine content of these regions is 46.9%. This is reflected in the *flgI* DNA sequence by a preference in the third position of codons for guanine and cytosine nucleotides, indicative of a coding region [320]. The *flgI* homologue is preceded by a potential ribosome binding site, AGGT [269] nine base pairs from the start codon. This is shown in figure 4.2.3, which also shows a possible class II flagellar gene promoter approximately 100 base pairs upstream of the start codon.

Consensus class II -10 promoter GCCGATAA
A. tumefaciens putative promoter GCCAATGA

As described in the introduction flagellar gene expression is highly regulated, often using flagellum-specific promoter sequences [180]. In the enteric bacterial model the *flgI* gene is a class II structural gene and hence comes under the control of the -10 consensus described above. Whether the *A. tumefaciens* promoter is used or not, is unknown. However its presence within the putative operon will be considered later in the discussion section of this chapter.

Downstream of the *flgI* coding region there is no evidence of any transcription termination signals. In fact the translation termination codon TGA overlaps with the start codon (ATG) of ORF B immediately downstream, to give "ATGA". This configuration also occurs with the stop codon of ORF A and the start codon of *flgI*.

The FlgI homologue was highly similar to a number of FlgI proteins from other bacteria. The hydropathy profiles of the FlgI homologue also matched those of the other FlgI proteins, and all possessed putative signal peptides. The *Sac*II-*Sac*I fragment of the *flgI* homologue's DNA sequence (see figure 4.2.3) was radiolabelled and used as a probe against *A. tumefaciens* genomic DNA cut with *Bam*HI, *Eco*RI and *Hind*III. For each separate digest the size of the band to which the probe hybridised was equivalent to that predicted by the restriction map of pDUB1900 (data not shown). This implies that this is the only copy of this DNA sequence within the *A. tumefaciens* genome. This evidence suggests that this *flgI* homologue is the gene encoding the FlgI protein in *A. tumefaciens*. However as this homologue is potentially part of an operon, a mutation of it (leading to a *fla*- phenotype) would not necessarily show that its expression is necessary for flagellar formation, since the mutation would probably effect transcription of the homologues (*flgH* and *fliP*) downstream,

the absence of either of which also leads to a *fla-* phenotype in *S. typhimurium*. Protein expression studies will be required to prove that this homologue is FlgI and is expressed.

The *flgI* gene encodes the protein that forms the P ring of the basal body in *S. typhimurium* and *C. crescentus* [135, 148]. The P ring lies within the periplasmic space, possibly in association with the peptidoglycan layer [135]. The P ring protein subunits interact with themselves and probably the L ring protein subunits to form a cylinder embedded in the outer membrane. The rod passes through this cylinder and is thought to be free to rotate within it. The L- and P rings act as a bushing to help anchor and stabilise the basal body in the cell envelope [135, 148, 156]. The conservation shown between the various FlgI amino acid sequences is probably a result of the various structural components the protein must interact with [179]. These factors also probably determine the number of subunits (c. 26) in the P ring of *S. typhimurium*, as well as the diameter of the rod around which they must assemble [179, 278]. The FlgI proteins are believed to be exported by the *sec*-dependent general secretory pathway [179, 239]. Accordingly, before assembly the FlgI proteins are presumed (in *S. typhimurium* it has been shown experimentally) to be processed to remove their signal peptides [135, 156]. The mature *S. typhimurium* FlgI protein is weakly hydrophilic as expected from its position within the aqueous periplasmic space [135].

Despite the similarities of the amino acid sequences, there does appear to be some differences between P rings of the enteric bacteria (*E. coli* and *S. typhimurium*) and those of *A. tumefaciens*. Mutants have been isolated in *E. coli* which fail to assemble functional flagella. The defect appears to be due to the inability to form a disulphide bond in the P ring, which the authors state is essential for flagella assembly [62]. However the predicted protein sequence from the *A. tumefaciens flgI* homologue contains no cysteine residues to form disulphide bonds. This points to a slightly different form of intramolecular interaction between the FlgI subunits in *A. tumefaciens*. The amino acid sequence of FlgI from *C. crescentus* contains two cysteine residues, but one of these is within the predicted signal peptide sequence. Since this sequence is probably lost before the mature proteins assemble, only one cysteine residue will remain. This does not rule out the possibility of intermolecular disulphide bonds between P ring subunits in *C. crescentus*. Furthermore the predicted FlgI sequence of *P. putida* contains only one cysteine residue within the signal sequence. Thus the mature form of this protein also probably does not have the ability to form disulphide bonds, as in the *A. tumefaciens* FlgI.

Figure 4.2.1 Multiple alignment of the *A. tumefaciens* FlgI homologue to those of *C. crescentus*, *P. putida* and *S. typhimurium*.

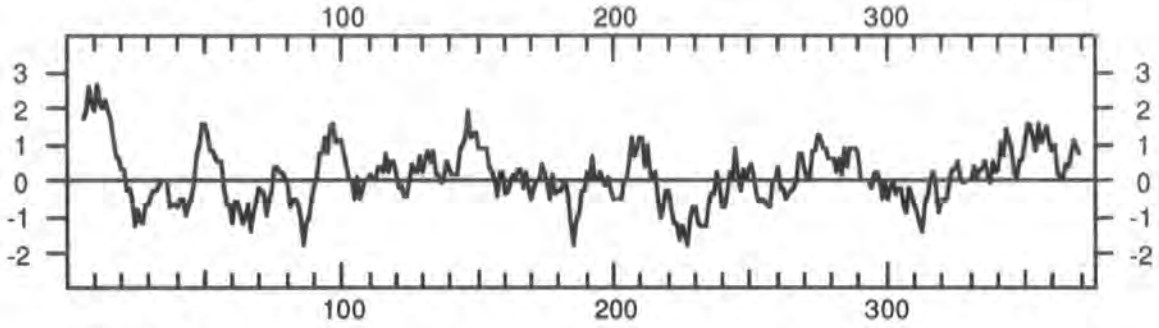
Identities are indicated by asterisks, and conservative substitutions by dots. The abbreviations are, *At* = *A. tumefaciens*, *Cc* = *C. crescentus*, *Pp* = *P. putida* and *St* = *S. typhimurium*. The alignment was performed using the SEQNET program CLUSTAL V [119].

```
FlgI At  MR--VLRIIAAALVFSALPFLSTPPAQADTSRIKDIASLQAGRDNQLIGYGLVVGLQGTG
FlgI Cc  MPRSLFRTVLTALV--AASCLIAAPALAK-SRIKDIVSFEGVRENQLIGYGIVVGLNGTG
FlgI Pp  MFN--VRQLIATTL------LLSCAFAAQAERLKDIASISGVRSNQLIGYGLVVGLNGTG
FlgI St  MF----KALAGIVL------ALVATLA-HAERIRDLTSVQGVRENSLIGYGLVVGLDGTG
         *     . .    .                    *..*. *  . * ******.****.***

FlgI At  DSLRSSPFTEQSMRAMLQNLGITTQGG--QSNAKNIAAVMVTANLPPFASPGSRVDVTVS
FlgI Cc  DSLRNAPMTKQSLEAMLERQGVNVRDN--NLNTKNTAAVMVTANLPPFSASGSKVDVTVS
FlgI Pp  DQTTQTPFTLQTFNNMLSQFGIKVPAGSGNVQLKNVAAVSVHADLPPFAKPGQVVDITVS
FlgI St  DQTTQTPFTTQTLNNMLSQLGITVPTGT-NMQLKNVAAVMVTASYPPFARQGQTIDVVVS
         *    .* * *.   **   *.       . . ** *** * *  ***.  *  .*. **

FlgI At  SLGDATSLRGGNLIMTSLSGADGQIYAVAQGALIVNGFSAQGDAAT-LTQGVTTSARVPN
FlgI Cc  TLGDAKSLLGGTLLVTSLQGADGQTYAVAQGTVQTGSVSAGGASGSSVTKGVPTAGRIAG
FlgI Pp  SIGNSKSLRGGSLLMTPLKGIDGNVYAIAQGNLVVGGFDAQGRDGSKITVNVPSAGRIPG
FlgI St  SMGNAKSLRGGTLLMTPLKGVDSQVYALAQGNILVGG-AGASAGGSSVQVNQLNGGRITN
         ..*.. ** ** *..*.* * *.. **.*** .   .  . .  .. .      ..*.

FlgI At  GAIIERELPSKFKDSVNLVLQLRNPDFSTAVRVADVVNAFARARYGDPIAEPRDSQEIAV
FlgI Cc  GGVIERETGFQMVNMDIMRLTLRNPDFTTARRVADAINA----KFPG-CAQAQNPTIIAT
FlgI Pp  GASVERAVPSGFNQGNTLTLNLNRPDFTTAKRIVDKVNDL----LGPGVAQAVHGGSVRV
FlgI St  GAIIERELPTQFGAGNTINLQLNDEDFTMAQQITDAINRA----RGYGSATALDARTVQV
         *. .**           . * *   **. * .. * .*          * . .   .

FlgI At  QKPRVA-DLTRLMAEIENLTVETDTP-AKVVINERTGTIVIGADVRISRVAVSYGTLTVQ
FlgI Cc  RPPPGM-DMISFMTNIENLMVEPDGP-AKVVIDEVAGVIVMGDDVRISQVAIAQGNLTIT
FlgI Pp  SAPMDPSQRVDYLSILENLEIDPGQAVAKVIINSRTGTIVIGQNVKVSP-AVTHGSLTVT
FlgI St  RVPSGNSSQVRFLADIQNMEVNVTPQDAKVVINSRTGSVVMNREVTLDSCAVAQGNLSVT
           *         .. ..*. ..     ***.*.  .* .*.  .* .   *.. * *..

FlgI At  VTESPQVIQP-APFSRGQTAVQPQTDIMAMQEGSK--VAIVEGPDLRTLVAGLNSIGLKA
FlgI Cc  VQENPAVSQP-APFSQGQTAVVPQSTVNVEEEKGKQLLTLGGAPSLKGLIGGLNALGVTP
FlgI Pp  ITEDPIVSQPGA-FSNGQTAVVPRSRVNAEQE-AKPMFKFGPGTTLDEIVRAVNQVGAAP
FlgI St  VNRQLNVNQPNTPFGGGQTVVTPQTQIDLRQS-GGSLQSVRSSANLNSVVRALNALGATP
         .     * ** . *. *** * *.. .   .  .        .  *  .. ..* .*  .

FlgI At  DGIIAILQGIKSAGALQAELVLQX
FlgI Cc  RDMISILQAVKAAGALQADIEVMX
FlgI Pp  GNLMAILEALKHRP-LQADLIV-I
FlgI St  MDLMSILQSMQSAGCLRAKLEI-I
           ...**....    *.* . .
```

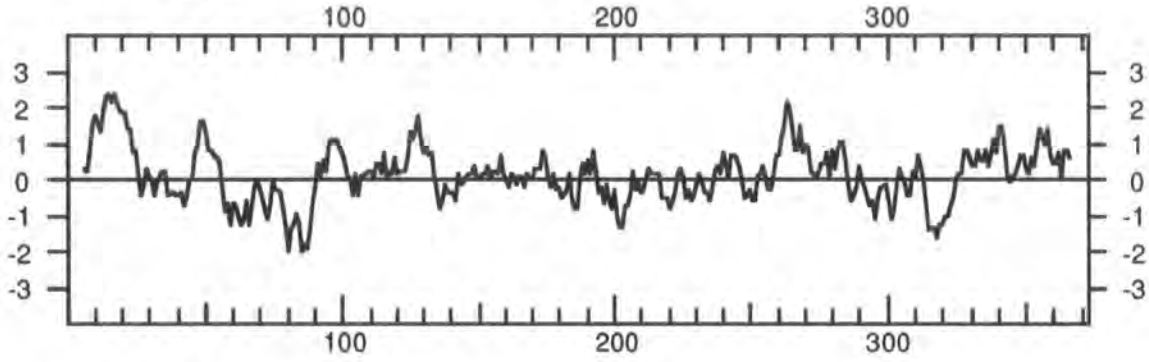
Figure 4.2.2 Hydropathy profiles of the *A. tumefaciens* FlgI homologue and the FlgI proteins of *C. crescentus*, *P. putida* and *S. typhimurium*.

The profiles were drawn with DNA Strider, using the Kyte and Doolittle analysis. The abbreviations are given in figure 4.2.1.

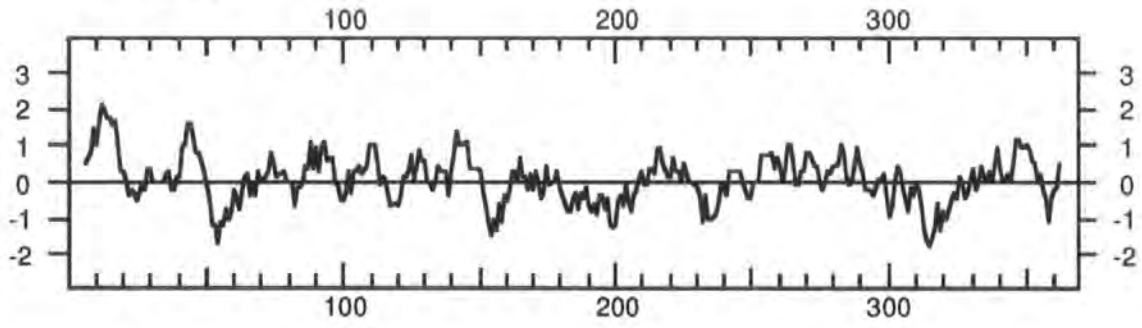
a) FlgI *At*



b) FlgI *Cc*



c) FlgI *Pp*



d) FlgI *St*

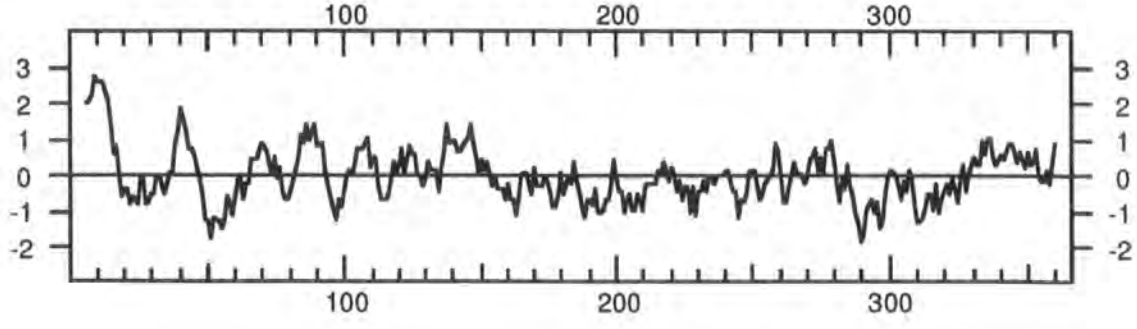


Figure 4.2.3 The entire coding and untranslated regions of the *flgI* homologue and the predicted translation product.

The FlgI coding region begins at nucleotide position 220 and finishes at 1341. Nine base pairs upstream of it a possible ribosome binding site (AGGT) is shown and labelled "rbs". Approximately 100 base pairs upstream of the start codon, the putative flagellar class II promoter is shown in bold, block capitals (nucleotide positions 109-116).

The cleavage site of the potential signal peptide is marked with a double asterisk, after amino acid residue 26.

The *Sac*II (nucleotide position 242-247) and *Sac*I (nucleotide position 1324-1329) restriction enzyme sites used to excise the *flgI* specific probe for radiolabelling are shown within the coding region in bold.

The stop codon of ORFA is overlined and labelled with a "Z" at nucleotide position 221-3. The start codon of ORFB is also overlined and labelled with an "M" at nucleotide position 1338-40. These codons overlap with the start and stop codons (respectively) of the *flgI* homologue, giving the "ATGA" configuration found between most of the open reading frames within the putative operon.

```
   1 ccgcacgatccccattgcggcgcttcgcgaaccatcactggtggtgcgcggtacaagcgt 60

  61 caaactcgttttccacattggcaacatgacgctgatggcttccggaacGCCAATGAgcga 120

 121 cggctcgctcggcgaggtcgtcagggtacgcaacatcgattccggcgtgatggtcagcgg 180
                                                (ORFA)  Z
 181 cacggtcatgaaggacggaaccattcAGGTgatggcgaa ATG AGA GTG CTT CGT  234
   1                           rbs           M   R   V   L   R    5
                SacII
 235 ATC ATA GCC GCG GCT TTG GTC TTT TCG GCC CTG CCC TTC CTC TCC  279
   6 I   I   A   A   A   L   V   F   S   A   L   P   F   L   S   20

 280 ACG CCG CCC GCG CAG GCC GAC ACG TCG CGA ATC AAG GAT ATC GCA  324
  21 T   P   P   A   Q   A * D   T   S   R   I   K   D   I   A   35
                          *
 325 TCC CTG CAG GCC GGA CGT GAT AAC CAG CTG ATC GGT TAC GGT CTC  369
  36 S   L   Q   A   G   R   D   N   Q   L   I   G   Y   G   L   50

 370 GTC GTG GGT CTG CAG GGT ACC GGC GAC AGC CTG CGC TCC TCG CCT  414
  51 V   V   G   L   Q   G   T   G   D   S   L   R   S   S   P   65

 415 TTC ACC GAA CAG TCC ATG CGC GCC ATG CTG CAG AAC CTC GGC ATC  459
  66 F   T   E   Q   S   M   R   A   M   L   Q   N   L   G   I   80

 460 ACC ACG CAG GGT GGC CAG TCC AAC GCC AAG AAC ATC GCC GCC GTG  504
  81 T   T   Q   G   G   Q   S   N   A   K   N   I   A   A   V   95

 505 ATG GTG ACG GCC AAC CTG CCG CCA TTC GCC AGC CCC GGC AGC CGT  549
  96 M   V   T   A   N   L   P   P   F   A   S   P   G   S   R  110
```

```
550 GTG GAC GTC ACG GTC AGC TCG CTT GGC GAC GCC ACC TCG CTT CGT  594
111 V   D   V   T   V   S   S   L   G   D   A   T   S   L   R    125

595 GGC GGC AAC CTC ATT ATG ACT TCG CTT TCC GGC GCG GAC GGG CAG  639
126 G   G   N   L   I   M   T   S   L   S   G   A   D   G   Q    140

640 ATT TAC GCC GTC GCG CAA GGG GCG CTG ATC GTC AAC GGT TTT TCG  684
141 I   Y   A   V   A   Q   G   A   L   I   V   N   G   F   S    155

685 GCC CAG GGC GAT GCC GCG ACG CTG ACG CAG GGC GTC ACC ACG TCG  729
156 A   Q   G   D   A   A   T   L   T   Q   G   V   T   T   S    170

730 GCG CGC GTG CCG AAC GGC GCG ATC ATC GAG CGG GAA CTG CCG TCG  774
171 A   R   V   P   N   G   A   I   I   E   R   E   L   P   S    185

775 AAG TTC AAG GAT TCG GTC AAT CTG GTC CTT CAG CTG CGC AAC CCG  819
186 K   F   K   D   S   V   N   L   V   L   Q   L   R   N   P    200

820 GAT TTT TCG ACA GCC GTG CGT GTG GCT GAT GTC GTC AAC GCC TTT  864
201 D   F   S   T   A   V   R   V   A   D   V   V   N   A   F    215

865 GCC CGC GCC CGT TAT GGC GAC CCG ATC GCC GAA CCG CGC GAT TCG  909
216 A   R   A   R   Y   G   D   P   I   A   E   P   R   D   S    230

910 CAG GAA ATC GCG GTG CAG AAA CCC CGC GTC GCA GAC CTC ACC CGG  954
231 Q   E   I   A   V   Q   K   P   R   V   A   D   L   T   R    245

955 CTG ATG GCG GAA ATC GAA AAT CTG ACG GTC GAG ACG GAT ACG CCT  999
246 L   M   A   E   I   E   N   L   T   V   E   T   D   T   P    260

1000 GCC AAG GTC GTG ATC AAC GAG CGC ACC GGC ACG ATC GTC ATC GGA  1044
 261 A   K   V   V   I   N   E   R   T   G   T   I   V   I   G    275

1045 GCG GAT GTG CGC ATT TCC CGC GTC GCC GTC AGC TAT GGA ACG CTC  1089
 276 A   D   V   R   I   S   R   V   A   V   S   Y   G   T   L    290

1090 ACC GTG CAG GTG ACG GAA TCG CCG CAG GTC ATT CAG CCC GCA CCC  1134
 291 T   V   Q   V   T   E   S   P   Q   V   I   Q   P   A   P    305

1135 TTC TCG CGC GGC CAG ACG GCA GTG CAG CCG CAG ACG GAT ATC ATG  1179
 306 F   S   R   G   Q   T   A   V   Q   P   Q   T   D   I   M    320

1180 GCC ATG CAG GAA GGT AGC AAG GTC GCC ATC GTC GAA GGC CCG GAC  1224
 321 A   M   Q   E   G   S   K   V   A   I   V   E   G   P   D    335

1225 CTT CGC ACG CTG GTT GCC GGT CTC AAC AGC ATC GGG CTC AAG GCG  1269
 336 L   R   T   L   V   A   G   L   N   S   I   G   L   K   A    350

1270 GAC GGC ATC ATC GCC ATC CTG CAA GGC ATC AAA TCC GCC GGC GCC  1314
 351 D   G   I   I   A   I   L   Q   G   I   K   S   A   G   A    365
                 SacI                   M   (ORFB)
1315 CTT CAG GCG GAG CTC GTG CTG CAA TGA tggaacgtcagaccaaaaatccgc 1365
 366 L   Q   A   E   L   V   L   Q   *                           374

1366 tctcaaatggcctcgtccgctttgcggccgtcgcgtcgctgctgttcctgttgccggttg 1425

1426 ccggcgccgaaagccagcagaacgt                                    1450
```

4.3 The *flgH* homologue

The gene product encoded by this open reading frame has 239 amino acids and a predicted molecular weight of 25,994Da. This gene product was found to have sequence similarity to the FlgH protein of two bacteria. It has 40% identity (57% similarity) over 242 amino acids to FlgH of *C. crescentus* and 26% identity (40% similarity) over 232 amino acids to FlgH of *S. typhimurium.* The extent of the similarity is shown in a multiple alignment (figure 4.3.1) of the *A. tumefaciens* FlgH homologue and the two FlgH proteins named above. The hydropathy plots of the *A. tumefaciens* FlgH homologue and the other two FlgH proteins, using the Kyte and Doolittle algorithm [160], are roughly comparable (figure 4.3.2). All three contain several alternating hydrophilic and hydrophobic regions and a hydrophilic carboxy-terminus.

The amino-terminus of the *A. tumefaciens* FlgH homologue, like the FlgH proteins of *C. crescentus* and *S. typhimurium*, contains a putative signal peptide sequence. There are however two possible cleavage sites, Leu-Leu-Ala or Leu-Ala-Gly, as shown in figure 4.3.3. Cleavage at the first site would remove a 17 amino acid signal peptide and leave a mature protein of 222 amino acids (M_r 24,164Da). Cleavage at the second would remove an 18 amino acid signal peptide and leave a mature protein of 221 amino acids (M_r 24,107Da). The rest of the sequence meets the criteria described for signal peptides [218], with two positively charged residues close to the amino-terminus and then 12 or 13 predominantly hydrophobic amino acids. The signal sequence of the FlgH proteins of *C. crescentus* and *S. typhimurium* also have at least two potential cleavage sites [68, 135]. Only the signal sequence of the *S. typhimurium* FlgH has been shown to be removed experimentally, although the amino-terminus of the mature protein is thought to be blocked [135]. Jones *et al.* postulate that the block may be due to the attachment of lipid to a cysteine residue generated by cleavage at one of the sites since lipoprotein, the predominant outer membrane protein of Gram negative bacteria, has lipid covalently attached to its amino-terminal cysteine residue. The block in FlgH (also an outer membrane protein) of *S. typhimurium* may be due to a similar modification [45, 135]. At least one of the possible cleavage sites of FlgH from *A. tumefaciens* and *C. crescentus* would also generate a cysteine at the amino-terminus.

The putative start codon of the *flgH* homologue overlaps with the stop codon of ORFB, to give the "ATGA" configuration again. Six base pairs upstream of the start codon is a potential ribosome binding site, AGGA [269]. Further upstream (almost 200 base pairs) is a sequence with only two mismatches from the -10 consensus of a class II flagellar promoter (see figure 4.3.3).

Consensus class II -10 promoter	GCCGATAA
A. tumefaciens putative promoter (*flgH*)	GACGATGA
[*A. tumefaciens* putative promoter (*flgI*)	GCCAATGA]

Like *flgI*, *flgH* in the enteric bacteria is within an operon under (at least) the control of such a class II flagellar gene promoter [180]. Again the presence of this promoter-like sequence within the putative operon will be considered in the discussion section of this chapter.

The sequence downstream of the *flgH* homologue was searched for potential transcription stop signals using the UWGCG program, TERMINATOR, which uses the method of Brendel and Trifonov [46]. There was no indication of a "classical" rho-independent terminator. However, approximately 160 base pairs downstream (within ORFC) there was a potential stem-loop sequence that might constitute a transcription terminator (see figure 4.3.3). The start codon of ORFC does not overlap the stop codon of the *flgH* homologue with the "ATGA" configuration. In fact it starts 26 base pairs downstream, giving a possible short intergenic region.

The FlgH homologue of *A. tumefaciens* was significantly similar to the FlgH proteins from two other bacteria. All of these have putative signal peptide sequences. The hydropathy plot of the *A. tumefaciens* FlgH homologue was comparable to those of the other two FlgH proteins, see figure 4.3.2. The *flgH* open reading frame of *A. tumefaciens* was predicted, at the 95% confidence level, to be a coding region by the program TESTCODE, see figure 3.4. The guanine and cytosine content of the open reading frame was 61%. This is significantly higher than in regions thought to be non-coding, and again indicative of a coding region. The *Bam*HI-*Nco*I fragment of the *flgH* homologue's DNA sequence (see figure 4.3.3) was radiolabelled and used as a probe against *A. tumefaciens* genomic DNA cut with *Bam*HI, *Eco*RI and *Hind*III. For each separate digest the size of the band(s) to which the probe hybridised was that predicted by the restriction map of pDUB1900 (data not shown). This implies that this is the only copy of this DNA sequence within the *A. tumefaciens* genome. Again as was the case for the *flgI* homologue, all of this evidence suggests that this *flgH* homologue is the gene encoding the FlgH protein in *A. tumefaciens*. A mutation could be made in the *flgH* homologue, as described for the *fliP* homologue in sections 2.14 and 4.10, to show its expression is necessary for flagellar formation. This mutation would be expected to produce a *fla*-phenotype. The homologue is, however, potentially part of a flagellar gene operon, and unfortunately the mutation would also probably affect transcription of the *fliP* homologue downstream. A mutation in the *fliP* homologue has also been shown to result in a *fla*-phenotype. Hence protein expression

studies will be required to prove that this homologue is the gene encoding the L ring protein, *flgH*, and is expressed.

The *flgH* gene (originally called *flbN* in *C. crescentus*) encodes the protein that forms the L ring of the basal body in *S. typhimurium* and *C. crescentus* [68, 135]. The L ring lies within the outer membrane. As described in section 4.2, the L ring is thought to interact with the P ring, to form a cylinder embedded in the outer membrane. This cylinder probably has a minimal role in the motor mechanism and instead stabilises the basal body, as the flagellum rotates, in the cell envelope [135]. As in the case of the FlgI proteins, the conservation shown between the FlgH proteins is probably necessary for the various interactions with other structural components [179]. The FlgH proteins must interact with themselves as well as, potentially, the P ring, the outer membrane and the rod [135, 156]. There are thought to be c. 26 FlgH subunits in the L ring. Again this is thought to be a result of the interactions listed as well as the diameter of the rod [180, 277].

The FlgH subunits are exported to the outer membrane by the *sec*-dependent general secretory pathway [179, 239]. Thus the signal peptides are probably removed prior to assembly. In *S. typhimurium* this has been shown [135]. Although the FlgH proteins are found within the outer membrane, the hydropathy plots (figure 4.3.2) show no clear transmembrane domains. The signal peptide sequence could possibly form one, although this is probably removed in the mature protein. Jones *et al.* [135], observed that the mature form of FlgH in *S. typhimurium* was predicted to have extensive β-structure, 26% according to the Chou & Fasman algorithm [55], and 30% according to the Garnier *et al.* algorithm [93]. In some outer membrane proteins, for example porin (which also has no predicted hydrophobic membrane spanning domains), β-sheets potentially form the dominant trans-membrane structural motif [152, 315]. The predicted β-structures of the FlgH proteins of *A. tumefaciens* and *C. crescentus* were also fairly high, approximating that found in the *S. typhimurium* FlgH: for *A. tumefaciens* 27% according to the Chou & Fasman algorithm and 21% according to the Garnier *et al.* algorithm and for *C. crescentus* 23% and 30%, respectively.

Figure 4.3.1 Multiple alignment of the *A. tumefaciens* FlgH homologue to those of *S. typhimurium* and *C. crescentus.*

Identities are indicated by asterisks, and conservative substitutions are denoted by dots. The abbreviations are as follows, *At* = *A. tumefaciens*, *Cc* = *C. crescentus* and *St* = *S. typhimurium*. The alignment was performed using the SEQNET program CLUSTAL V [119].

```
FlgH At   MSTRRLPALLLPLALLAGCQNNQ--TLKEIGNAPAMS-PIGSGLQFSQTPQMGMYPKQPK
FlgH Cc   M---RRPAILAAAVLLAPLAACS--TVKEAVKGPDLA-PVGYPAPLAPMQQQYVSAREPA
FlgH St   MQKYALHAYPVMALMVATLTGCAWIPAKPLVQGATTAQPIPGPVPVANGSIF----QSAQ
          *      *      ..*          *   ...  . *.      .          . .

FlgH At   HMASGYSLWSDSQGALFKDLRALNIGDILTVNIQINDKADFDNETERNRT---NASGLNW
FlgH Cc   PQRFGQLAVAGRGPAFFNDQRASRVGDIVTVMIDIDDSARTKKPTNSSRTANMKAGVRHL
FlgH St   PINYGY-------QPLFEDRRPRNIGDTLTIVLQENVSASKSSSANASRDGKTSFGF---
              *          ..* * *.  .** .*. .. .  *     .. .*      .

FlgH At   KAKAQILG-WTPDA-DSSIK-YGSDTDTQAKGKT-KR-SEKLTLLVAAVVTGILENGNLI
FlgH Cc   LGMESSLGKFLPGGFDPASA-LETNSTTTNAGSGGSR-SEKISLTIAAVVSQLLPNGNMV
FlgH St   ----DTVPRYLQGLFGNSRADMEASGGNSFNGKGGANASNTFSGTLTVTVDQVLANGNLH
                .     .  . .     .       *      *. ..  ..  *  .* ***.

FlgH At   ISGSQEVRVNHEIRILNVGGIVRPQDVDAQNIISYERIAEARISYGGRGRLTEVQQPPVG
FlgH Cc   IQGTQEVRTNAELRQLTVAGIVRPEDISSANTIRHTQIAEARISYGGRGDISRVQKTPAG
FlgH St   VVGEKQIAINQGTEFIRFSGVVNPRTISGSNSVPSTQVADARIEYVGNGYINEAQNMGWL
          . * ...  *     .  .*.* *  . . * .   ..*.*** * * * .   *.

FlgH At   QQVVDLFSPLX
FlgH Cc   QSLVEKFSP-F
FlgH St   QRFFLNLSPM-
          *     .**
```

Figure 4.3.2 Hydropathy profiles of the *A. tumefaciens* FlgH homologue and the FlgH proteins of *C. crescentus* and *S. typhimurium*.

The profiles were drawn with DNA Strider, using the Kyte and Doolittle analysis. The abbreviations are shown in figure 4.3.1.

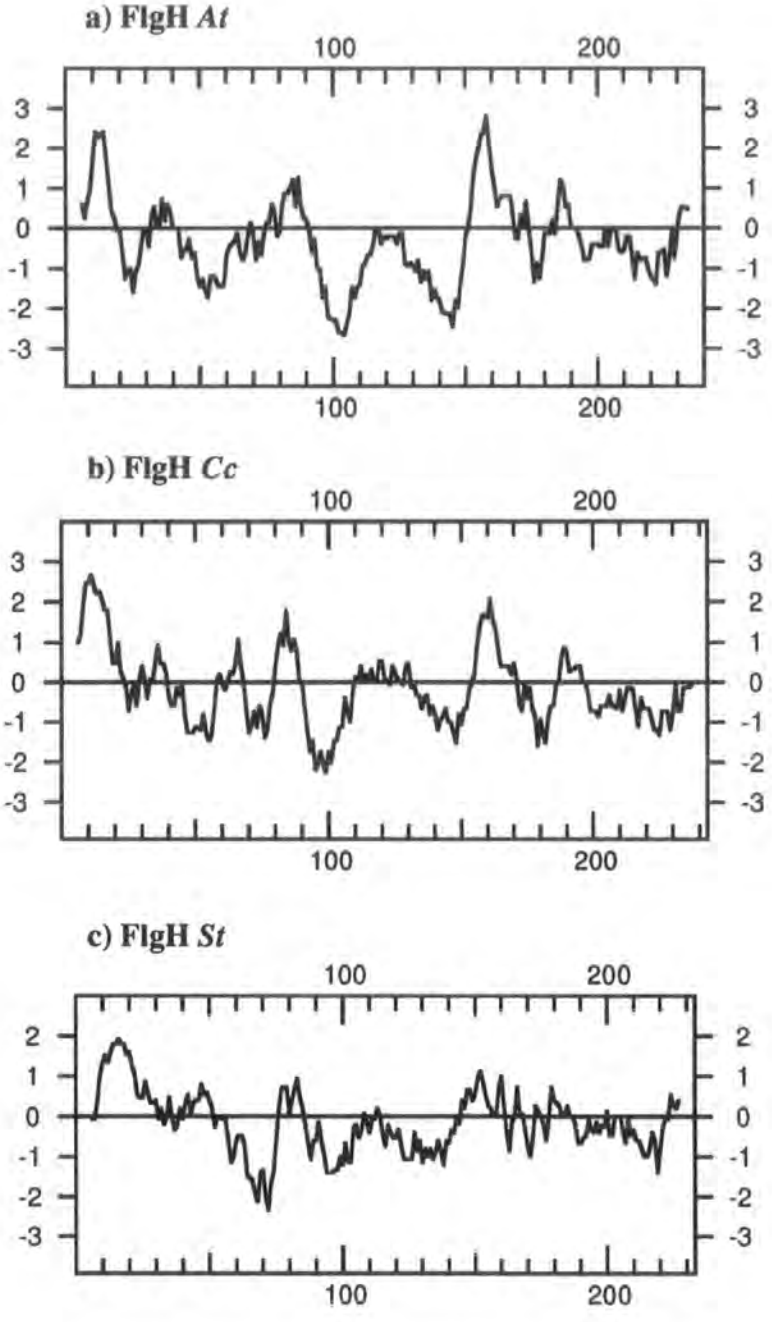


Figure 4.3.3 The complete coding and untranslated regions of the *flgH* homologue and predicted translation product.

The FlgH coding region begins at nucleotide position 321 and finishes at 1040. Six base pairs upstream of the start codon, there is a possible ribosome binding site, AGGA, shown in bold. Approximately 200 base pairs upstream is the putative flagellar class II promoter shown in bold, block capitals (nucleotide positions 132-139).

The two possible signal peptide cleavage sites are marked with double asterisks, after amino acid residues 17 and 18.

The *Bam*HI (nucleotide position 312-317) and *Nco*I (nucleotide position 1065-1070) restriction enzyme sites used to excise the *flgH* coding region as a specific probe for radiolabelling are shown.

The stop codon of ORFB is overlined and labelled with a "Z" at nucleotide positions 322-324. It overlaps the start codon of *flgH*, giving the "ATGA" configuration. The start codon of ORFC (nucleotide positions 1067-1069) is overlined and labelled with an "M" twenty six base pairs downstream of the termination codon of *flgH*. The potential stem-loop structure, which may be a transcription termination signal is also shown, between nucleotides 1199-1230 by staggered arrows.

```
   ---ORFB--->
  1 gaaaagcttcaggccgacgtcaatgagcgtatttccgtgctggaaaaccgcaaggcggaa 60

 61 tatgaggactggctggcgcgccgcgaacatttcctcaatcaggccaaaagcaacctcgtc 120

121 gacatctacaaGACGATGAaggcggacgccgcagctccccagctagaaaagatgcatgtg 180

181 gaaatcgcggccgccatcatcatgcagctgccaccgcgccaatcgggcctcatcctcagc 240

241 gaaatggatgcgcagaaagccgcaaccgtcgccggcatcatgtcgcaggcaatcgacaaa 300
                BamHI      Z
301 aacacttcgaAGGAtccttc ATG AGC ACG CGT CGT CTT CCG GCC CTC CTC 350
  1           rbs        M   S   T   R   R   L   P   A   L   L   10

351 CTG CCG CTC GCT CTC CTG GCC GGC TGC CAG AAT AAC CAG ACC CTG  395
 11 L   P   L   A   L   L   A * G * C   Q   N   N   Q   T   L    25
                             *   *
396 AAG GAA ATC GGC AAC GCA CCC GCG ATG AGC CCG ATC GGC AGC GGC  440
 26 K   E   I   G   N   A   P   A   M   S   P   I   G   S   G    40

441 CTG CAA TTC AGC CAG ACA CCG CAG ATG GGC ATG TAT CCC AAG CAG  485
 41 L   Q   F   S   Q   T   P   Q   M   G   M   Y   P   K   Q    55

486 CCG AAA CAC ATG GCA AGC GGC TAT TCG CTG TGG AGC GAC AGC CAG  530
 56 P   K   H   M   A   S   G   Y   S   L   W   S   D   S   Q    70

531 GGT GCG CTG TTC AAG GAT CTG CGC GCA CTC AAT ATC GGC GAC ATC  575
 71 G   A   L   F   K   D   L   R   A   L   N   I   G   D   I    85
```

```
576 CTG ACC GTT AAC ATC CAG ATC AAC GAC AAG GCC GAT TTC GAC AAC  620
 86 L   T   V   N   I   Q   I   N   D   K   A   D   F   D   N    100

621 GAG ACC GAG CGT AAC CGC ACC AAT GCC AGT GGC CTG AAC TGG AAG  665
101 E   T   E   R   N   R   T   N   A   S   G   L   N   W   K    115

666 GCA AAA GCC CAG ATC CTG GGC TGG ACG CCG GAT GCG GAT TCC AGC  710
116 A   K   A   Q   I   L   G   W   T   P   D   A   D   S   S    130

711 ATC AAA TAC GGC TCC GAC ACC GAC ACA CAG GCC AAG GGC AAG ACC  755
131 I   K   Y   G   S   D   T   D   T   Q   A   K   G   K   T    145

756 AAG CGC TCC GAA AAG CTG ACG CTG CTG GTG GCT GCC GTC GTG ACC  800
146 K   R   S   E   K   L   T   L   L   V   A   A   V   V   T    160

801 GGC ATT CTG GAA AAC GGC AAC CTC ATC ATC AGC GGC TCG CAG GAA  845
161 G   I   L   E   N   G   N   L   I   I   S   G   S   Q   E    175

846 GTG CGT GTG AAC CAC GAA ATC CGC ATC CTC AAC GTC GGC GGT ATC  890
176 V   R   V   N   H   E   I   R   I   L   N   V   G   G   I    190

891 GTC CGT CCG CAG GAT GTC GAT GCC CAG AAC ATC ATC TCC TAC GAG  935
191 V   R   P   Q   D   V   D   A   Q   N   I   I   S   Y   E    205

936 CGC ATC GCC GAA GCA CGC ATC TCC TAC GGC GGT CGT GGC CGT CTG  980
206 R   I   A   E   A   R   I   S   Y   G   G   R   G   R   L    220

981 ACG GAA GTG CAG CAG CCC CCG GTC GGG CAG CAG GTC GTT GAC CTG 1025
221 T   E   V   Q   Q   P   P   V   G   Q   Q   V   V   D   L    235
                                                  M --ORFC-->
1026 TTC TCG CCG CTC TGA cgccgcacacgggaacggactgacCCATGGaaaacgaaca 1080
 236 F   S   P   L   *                          NcoI              240

1081 ggctgagggcaaaaaaaaatcctcccctctggtcatgaccatcgcaggcgttgtgatcct 1140

1141 caccctgctcggtgcgggcggcggctggctcgtgggcggcatgatcgccccgaaagttgc 1200
                                                                --
1201 cgcaacggaggcccatgccaccgctgcggccggcggccatggcgagaagaagggcgaagg 1260
     ---- --- -->    <-- --- ------
1261 cctcgataaaatcgacgccgaagcgaacggcatcgtccagctcgatcccatcaccaccaa 1320

1321 tctcgcctacccttcgaccaactgggtgcg                              1350
```

4.4 The *fliP* homologue

The gene product encoded by this open reading frame has 245 amino acids and a predicted molecular weight of 27,349Da. This gene product was found to have sequence similarity to the FliP protein of four other bacteria. It has 78% identity (88% similarity) over 245 amino acids to FliP of *R. meliloti*, 44% identity (71% similarity) over 219 amino acids to FliP of *B. subtilis*, 43% identity (69% similarity) over 245 amino acids to MopC of *Erwinia carotovora* subspecies *atroseptica* and 42% identity (69% similarity) over 245 amino acids to FliP of *E. coli*. The MopC protein of *E. carotovora* was isolated because of a mutation downstream that caused a reduction in the virulence of this organism. It has over 50% sequence identity to the FliP protein of *B. subtilis* which is found within an operon containing other flagellar gene homologues [204]. The extent of this similarity is shown in a multiple alignment (figure 4.4.1) between the *A. tumefaciens* FliP homologue and the four FliP proteins named above. The hydropathy plots of the *A. tumefaciens* FliP homologue and the four FliP proteins, prepared using the Kyte and Doolittle algorithm [160], are also similar (figure 4.4.2). The hydropathy profiles show the FliP proteins to be predominantly hydrophobic and to possess several possible membrane spanning domains.

The amino-terminus of the *A. tumefaciens* FliP homologue, like the FliP proteins of *R. meliloti* and *E. coli*, contains a putative signal sequence. The FliP protein of *B. subtilis* and the MopC protein of *E. carotovora* do not appear to have any recognisable signal sequences. The predicted cleavage site of the FliP homologue's signal peptide is shown in figure 4.4.3. Removal of this 20 amino acid long signal peptide would leave a polypeptide of 225 amino acids and a molecular weight of 25,272Da. There is a short, positively charged amino-terminus, followed by 14 (predominantly) hydrophobic residues and an Ala-Asn-Ala sequence that could serve as a cleavage site [218].

The putative start codon of the *fliP* homologue overlaps with the stop codon of ORFC, to give an "ATGA" configuration. Thirteen base pairs upstream is a potential ribosome binding site, GGA [269]. Further upstream (almost 150 base pairs) is a sequence with three mismatches from the -10 consensus of a class II flagellar gene promoter.

Consensus class II -10 promoter	GCCGATAA
[*A. tumefaciens* putative promoter (*flgH*)	GACGATGA]
[*A. tumefaciens* putative promoter (*flgI*)	GCCAATGA]
A. tumefaciens putative promoter (*fliP*)	GCCTATGT

The *fliP* gene in *E. coli* is found within an operon under (at least) the control of a flagellar class II promoter [180]. Again the presence of this promoter-like sequence in *A.*

tumefaciens will be considered in the discussion section of this chapter. The transcription/translation sequences mentioned are all shown in figure 4.4.3.

The open reading frame encoding the FliP protein from the closely related bacterium *R. meliloti*, is preceded by similar transcription/translation sequences [86]. These are shown in figure 4.4.4. Upstream of the *R. meliloti fliP* open reading frame is a partial open reading frame with significant identity to ORFC (from *A. tumefaciens*), see section 4.5. The *R. meliloti* open reading frame does not overlap with the *fliP* coding region, it finishes 7 base pairs upstream. There is no apparent ribosome binding site preceding the *R. meliloti fliP* gene. However approximately 150 base pairs upstream there is a sequence with three mismatches from the -10 consensus of a flagellar class II promoter.

Consensus class II -10 promoter	GCCGATAA
A. tumefaciens putative promoter (*fliP*)	GCCTATGT
R. meliloti putative promoter (*fliP*)	GCCTATCT

The sequence downstream of the *A. tumefaciens fliP* homologue was searched for potential transcription stop signals using the UWGCG program TERMINATOR, which uses the method of Brendel and Trifonov [46]. There was no indication of a "classical" rho-independent terminator. However there were a number of (short) potential stem-loop sequences that could constitute transcription terminators. The presence of a transcription terminator is expected downstream of the *fliP* homologue, since it is presumed to be the last gene in the putative operon. Approximately 400 base pairs downstream of the stop codon of the *fliP* homologue is the first of the flagellin genes (*flaA*). This gene has a possible promoter sequence upstream and a rho-independent transcription terminator sequence downstream. Furthermore transcription studies have shown *flaA* to be monocistronic (see chapter five).

The FliP homologue of *A. tumefaciens* was significantly similar to the FliP proteins of four other bacteria. Figure 4.4.2 shows the hydropathy plot of the *A. tumefaciens* FliP homologue to be very similar to those of the four other FliP proteins. The *fliP* open reading frame of *A. tumefaciens* was predicted, at the 95% confidence level, to be a coding region by the program TESTCODE, see figure 3.4. The guanine and cytosine content of the open reading frame was 57.2%. This is significantly higher than in regions thought to be non-coding, and again indicative of a coding region. Section 4.10 of this chapter describes the creation of a mutation in the *A. tumefaciens fliP* open reading frame. This was carried out by inserting a kanamycin-resistance cassette at the *Sal*I site in the *fliP* coding region. This insertion would severely truncate the *fliP* transcript. This mutant was found to be non-flagellated. The same phenotype has also been obtained from mutations of the *E. coli, B.*

subtilis and *R. meliloti fliP* genes [33, 86, 184]. All of the above data implies that this open reading frame does encode the FliP protein in *A. tumefaciens*.

The exact function of the FliP protein in *E. coli* and *S. typhimurium* is not known. However it is known to be required early in flagellum synthesis, after the assembly of the MS-ring in the cytoplasmic membrane but before the rod is completed [75, 156, 180]. The signal peptide of the *E. coli* FliP protein has been shown to be removed experimentally [184]. The hydropathy plots (see figure 4.4.2) of the FliP proteins show that the mature forms of the proteins contain hydrophobic segments that could be membrane spanning regions. The overall hydrophobicity of the FliP sequence implies that it is an integral membrane protein. The *B. subtilis* FliP protein does not possess an amino-terminal signal peptide which might suggest that the FliP protein is located in the inner membrane [33]. On the basis of this Malakooti *et al.* [184], postulate that the FliP protein may be a structural component of the basal body, since unidentified structures in association with the MS-ring are composed of proteins with a similar molecular weight to that of the *E. coli* FliP protein. The authors also point out that there are significant homologies between FliP and proteins involved in the control of translocation across the outer membrane, as discussed in the Introduction. These observations suggest a role for FliP in the flagellum specific export pathway, which performs a similar role. *A. tumefaciens* FliP has 30.5% sequence identity over 203 amino acids to Spa24 of *Shigella flexneri* and 34.2% sequence identity over 202 amino acids to ORF2 of *Xanthomonas campestris* pv. glycines. The extent of this similarity is shown by the multiple alignment in figure 4.4.5. The Spa24 protein, in *S. flexneri*, is thought to have a role in the surface presentation of the invasion plasmid antigens [314]. The ORF2 gene product of *X. campestris* pv. glycines is required for pathogenicity of this plant pathogen [130]. Furthermore the pathogenicity genes of *X. campestris* pv. campestris may be involved in the secretion of proteins by controlling translocation across the outer membrane [73].

UNIV DUNDEE

Figure 4.4.1 Multiple alignment of the *A. tumefaciens* FliP homologue to the FliP proteins of *B. subtilis*, *E. coli*, *R. meliloti* and the MopC protein of *E. carotovora*.

Identities are indicated by asterisks, and conservative substitutions are denoted by dots. The following abbreviations are used, *At* = *A. tumefaciens*, *Bs* = *B. subtilis*, *Ec* = *E. coli*, *Er* = *E. carotovora* and *Rm* = *R. meliloti*. The alignment was performed using the SEQNET program CLUSTAL V [119].

```
FliP At  MIRFLVTIAVLLALP----GLANAQQFPS--------DLFNTQIDGSVAAWI--IRTFGL
FliP Bs  MNEFINIFS-----------------------------------SSDPENVSSTVKLLLL
FliP Ec  MRRL---LSVAPVL----------LWLITPLAFAQLPGITSQPLPGGGQSWSLPVQTLVF
MopC Er  MSALPNYFRITALLRTQRYGLAIGLLFMAPSVWAQLPGIVTQPLPNGGQSWTLSVQTLVL
FliP Rm  MLRFATFIIAMMAMS----GIAGAQSFPA--------DILNTPVDGSVASWI--IRTFGL
         *  .   .                                                .. . .

FliP At  LTVLSVAPGILIMVTSFPRFVIAFSILRSGMGLASTPSNMILLSMAMFMTFYVMSPTFDK
FliP Bs  LTVFSVAPGILILMTCFTRIVIVLSFVRTSLATQSMPPNQVLIGLALFLTFFIMAPTFSE
FliP Ec  ITSLTFIPAILLMMTSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDK
MopC Er  LTSLTFLPAALLMMTSFTRIIIVLSLLRNALGTPTAPPNQVLLGLTLFLTFFVMSPVLNR
FliP Rm  LTVLSVAPGILIMVTSFPRFVIAFAILRSGMGLATTPSNMIMVSLALFMTFYVMAPTFDR
         .* ..  *. *...* * *..* ....* ...  . *.* .......*.**..*.* .

FliP At  AWTDGVQPLLQNQINEQQAVQRIAEPFRTFMNANTRDKDLKLFVDIARERGQVVMTDNVV
FliP Bs  INKEALTPLMDNKISLDEAYTKAEEPIKEFMSKHTRQKDLALFMNYA--KMDKPESLKDI
FliP Ec  IYVDAYQPFSEEKISMQEALEKGAQPLREFMLRQTREADLGLFARLA--NTGPLQGPEAV
MopC Er  VYDEAYLPFSQDQISMEVAIERGAEPVREFMLRQTRETDLALFTRLA--EIPEIQGPEAV
FliP Rm  AWRDGIDPLLKNEISETDAMQRMSEPFREFMVANTRDKDLQLFIDIAREKGQTVVVDEKV
            ..  *.  . *.   *  .  .* . **  .**. ** **   *            .

FliP At  DYRVLVPAFMLSEIRRGFEIGFLIILPFLVIDLIVATITMAMGMMMLPPTSISLPFKILF
FliP Bs  PLTTMVPAFAISELKTAFQIGFMIFIPFLIIDMVVASVLMSMGMMMLPPVMISLPFKILL
FliP Ec  PMRILLPAYVTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLML
MopC Er  PMRVLLPAFVTSELKTAFQIGFTVFIPFLIIDLVVASVLMALGMMMVPPATISLPFKLML
FliP Rm  DLRAVVPAFMISEIRRGFEIGFLIMLPFLVIDLIVATITMAMGMMMLPPTAISLPFKILF
            ..**.  **.. .*.*** . .***.**...*.. *..****.**  *.****...

FliP At  FVLIDGWNLLVGSLVRSF-NX
FliP Bs  FVLVDGWYLIVKSLLQSFX--
FliP Ec  FVLVDGWQLLVGSLAQSFYSX
MopC Er  FVLVDGWQLLLGSLAQSFY-S
FliP Rm  FVLIDGWNLLVGSLVRSF-IX
         ***.*** *.. ** .**
```

Figure 4.4.2 Hydropathy profiles of the *A. tumefaciens* FliP homologue and the FliP proteins of *B. subtilis*, *E. coli*, *E. carotovora* and *R. meliloti*.

The profiles were drawn with DNA Strider, using the Kyte and Doolittle analysis. The abbreviations are as listed in figure 4.4.1.

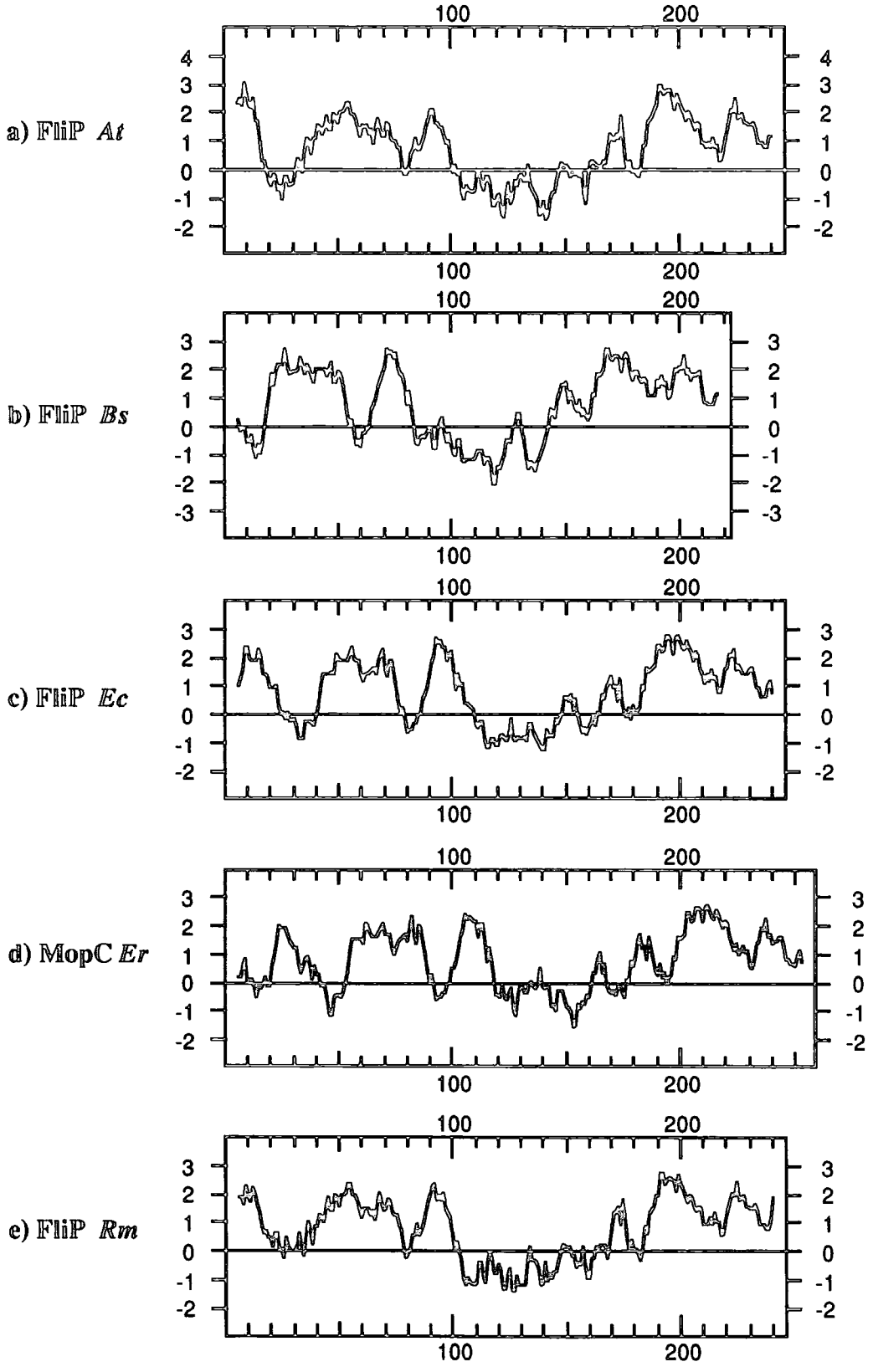


Figure 4.4.3 The coding and untranslated regions of the *fliP* homologue and the predicted translation product.

The FliP coding region begins at nucleotide position 201 and finishes at 938. Thirteen base pairs upstream of the start codon, there is a possible ribosome binding site, GGA, shown in bold capitals and labelled "rbs". Approximately 150 base pairs upstream the putative flagellar class II promoter is shown in bold, block capitals (nucleotide positions 61-68).

The potential signal peptide cleavage site is marked with a double asterisk, after amino acid residue 20.

The *Sal*I (nucleotide position 440-445) and *Dra*I (nucleotide position 1062-1067) restriction enzyme sites used to excise the *fliP* probe for radiolabelling are shown.

The stop codon of ORFC is overlined and labelled with a "Z" at nucleotide positions 202-204. It overlaps the start codon of *fliP*, giving the "ATGA" configuration. Approximately 400 base pairs downstream of the *fliP* stop codon the start codon of *flaA* is shown, nucleotide positions 1310-1312. Two potential stem-loop structures, which may be transcription termination signals are also shown with staggered arrows, between nucleotides 1051-1079 and 1140-1172.

```
   ---ORFC--->
  1 ctgatgttcaaggggccggtggaagtcgggctggcggaggatatccaccaggacatcatg 60

 61 GCCTATGTccggaccgtttccctccagcagctggaaggcccgcgcggctttcaatatctt 120

121 aaggatgacattcaggaacgagttgacctgcgctctcaagggcgcgtatccaaggtcatg 180
                              Z
181 ttccGGAcctttgtcatcga ATG ATT CGA TTT CTT GTA ACC ATC GCC GTC 230
  1     rbs               M   I   R   F   L   V   T   I   A   V  10

231 CTC CTC GCT TTG CCG GGC CTT GCC AAT GCC CAG CAA TTT CCG TCC  275
 11 L   L   A   L   P   G   L   A   N   A * Q   Q   F   P   S   25
                                         *
276 GAT CTG TTC AAC ACG CAG ATC GAC GGC TCC GTC GCG GCA TGG ATC  320
 26 D   L   F   N   T   Q   I   D   G   S   V   A   A   W   I   40

321 ATC CGC ACC TTC GGC CTG TTG ACG GTT CTC TCC GTC GCG CCC GGC  365
 41 I   R   T   F   G   L   L   T   V   L   S   V   A   P   G   55

366 ATC CTG ATC ATG GTC ACG AGC TTT CCG CGC TTC GTG ATC GCC TTT  410
 56 I   L   I   M   V   T   S   F   P   R   F   V   I   A   F   70
                                         SalI
411 TCG ATC CTG CGC TCC GGC ATG GGA CTT GCG TCG ACA CCG TCG AAC  455
 71 S   I   L   R   S   G   M   G   L   A   S   T   P   S   N   85

456 ATG ATC CTG CTG TCG ATG GCG ATG TTC ATG ACC TTT TAC GTC ATG  500
 86 M   I   L   L   S   M   A   M   F   M   T   F   Y   V   M  100

501 TCC CCC ACC TTC GAC AAA GCC TGG ACG GAT GGC GTG CAG CCG TTG  545
101 S   P   T   F   D   K   A   W   T   D   G   V   Q   P   L  115
```

```
546 CTG CAG AAC CAG ATC AAC GAA CAG CAG GCC GTG CAA CGC ATT GCC  590
116 L   Q   N   Q   I   N   E   Q   Q   A   V   Q   R   I   A  130

591 GAA CCC TTC CGC ACC TTC ATG AAC GCC AAT ACG CGT GAC AAG GAC  635
131 E   P   F   R   T   F   M   N   A   N   T   R   D   K   D  145

636 CTG AAG CTG TTC GTT GAC ATC GCC CGC GAG CGC GGA CAG GTG GTC  680
146 L   K   L   F   V   D   I   A   R   E   R   G   Q   V   V  160

681 ATG ACC GAC AAT GTC GTG GAC TAT CGC GTA CTG GTT CCT GCC TTC  725
161 M   T   D   N   V   V   D   Y   R   V   L   V   P   A   F  175

726 ATG CTC TCG GAA ATC CGG CGC GGT TTT GAA ATT GGT TTC CTC ATC  770
176 M   L   S   E   I   R   R   G   F   E   I   G   F   L   I  190

771 ATC CTG CCG TTC CTC GTC ATC GAT CTG ATC GTC GCC ACC ATT ACC  815
191 I   L   P   F   L   V   I   D   L   I   V   A   T   I   T  205

816 ATG GCG ATG GGC ATG ATG ATG CTG CCG CCC ACC TCG ATT TCG CTG  860
206 M   A   M   G   M   M   M   L   P   P   T   S   I   S   L  220

861 CCG TTC AAG ATC CTG TTT TTC GTG CTG ATC GAC GGC TGG AAC CTG  905
221 P   F   K   I   L   F   F   V   L   I   D   G   W   N   L  235

906 CTC GTG GGA AGC CTC GTC AGA TCG TTC AAC TGA cggcccttataccccg 954
236 L   V   G   S   L   V   R   S   F   N   *                  246

955 ctgatacagattaaccctgtttttaaccccgcacgcccaaggccgcggggttttcgcat 1014
                                          ------ - -- -> <- -- - -
1015 tttagtgctttggttaccattttcactagcgactgggtaactatattTTTAAAtaaaaaa 1074
     -----                                          DraI
1075 gttactattaagctaaatcaaaaagtaaccataagtattttacagtatttcttaacgctc 1134
          --  --- ---> <--- ---  --
1135 aaattaatgattgcgcaattattcgctgcgaaattcttcccacacgcagcgggtttggtt 1194

1195 tctggaactaagcgtttctgaaccgagtggcatgaagccgaaacgttgcgaccggtaata 1254

1255 tccaacccggtatgtccccactccgtatctcgtaaaaacaagggacacatttatt ATG  1312
                                                             M    1
```

Figure 4.4.4 FliP of *R. meliloti*, the upstream DNA sequence and a gap alignment with the *A. tumefaciens* FliP homologue.

Diagram (A) shows the upstream DNA sequence. The putative class II flagellar gene promoter discussed in the text is shown in bold, block capitals at nucleotide positions 73-80. The stop codon of the *R. meliloti* ORFC homologue is also shown in bold, block capitals and labelled with a "Z" at nucleotide positions 214-216. The possible cleavage site of the *R. meliloti* FliP signal sequence is marked (after residue 20) with a double asterisk.
Diagram (B) shows a gap alignment between the FliP proteins of *A. tumefaciens* (*At*) and *R. meliloti* (*Rm*). This was performed using the UWGCG program GAP.

(A)

```
   1 ctcgaggtcgcgctccagttcgatcgagctcccgacgttgcactggcggagacgatacac 60

  61 caggacattgcaGCCTATCTgaagaccgtgtcgctgcagcagattcaaggaccgcgcggc 120

 121 ttccaatatctccgggatgacatccaggagcgggttgacctgcgctccgagggccgcgta 180
                   --(Rm)-ORFC-->       Z
 181 acgaatgtgatgttccgcaccttcgtcatccaaTGAttcgctg ATG CTC CGG TTT  235
   1                                            M   L   R   F    4

 236 GCC ACC TTC ATA ATC GCC ATG ATG GCG ATG TCG GGA ATT GCC GGG  280
   5 A   T   F   I   I   A   M   M   A   M   S   G   I   A   G    19

 281 GCT CAG AGC TTC CCC GCC GAT ATC CTG AAC ACG CCG GTC GAC GGC  325
  20 A * Q   S   F   P   A   D   I   L   N   T   P   V   D   G    34
       *
 326 TCC GTC GCC TCG TGG ATC ATC CGC ACC TTC                      355
  35 S   V   A   S   W   I   I   R   T   F                       44
```

(B)

```
At    1 MIRFLVTIAVLLALPGLANAQQFPSDLFNTQIDGSVAAWIIRTFGLLTVL 50
        |:|| . | .::|:.|:|.|| ||.|::||.:|||||.||||||||||||
Rm    1 MLRFATFIIAMMAMSGIAGAQSFPADILNTPVDGSVASWIIRTFGLLTVL 50
                  .         .         .         .         .
At   51 SVAPGILIMVTSFPRFVIAFSILRSGMGLASTPSNMILLSMAMFMTFYVM 100
        ||||||||||||||||||||.|||||||||.||||||::|:|:|:|||||||
Rm   51 SVAPGILIMVTSFPRFVIAFAILRSGMGLATTPSNMIMVSLALFMTFYVM 100
                  .         .         .         .         .
At  101 SPTFDKAWTDGVQPLLQNQINEQQAVQRIAEPFRTFMNANTRDKDLKLFV 150
        .||||:|| ||::|||.|:|.| :|:||:.||||.|| ||||||||.||:
Rm  101 APTFDRAWRDGIDPLLKNEISETDAMQRMSEPFREFMVANTRDKDLQLFI 150
                  .         .         .         .         .
At  151 DIARERGQVVMTDNVVDYRVLVPAFMLSEIRRGFEIGFLIILPFLVIDLI 200
        |||||:||.|:.|: ||.|.:|||||:||||||||||||:|||||||||
Rm  151 DIAREKGQTVVVDEKVDLRAVVPAFMISEIRRGFEIGFLIMLPFLVIDLI 200
                  .         .         .         .
At  201 VATITMAMGMMMLPPTSISLPFKILFFVLIDGWNLLVGSLVRSFN 245
        ||||||||||||||||.|||||||||||||||||||||||||||
Rm  201 VATITMAMGMMMLPPTAISLPFKILFFVLIDGWNLLVGSLVRSFI 245
```

Figure 4.4.5 Multiple alignment of the *A. tumefaciens* and *E. coli* FliP proteins to the Spa24 protein of *S. flexneri* and to the gene product of ORF2 from *X. campestris* pv. glycines.

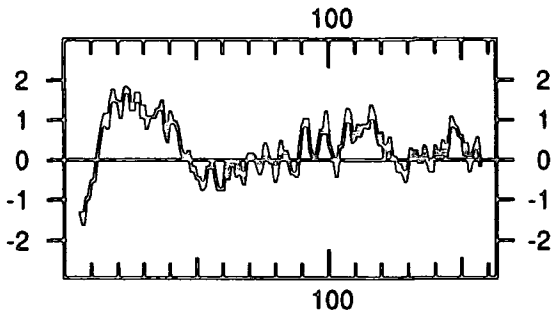
Identities are indicated by asterisks, and conservative substitutions are denoted by dots. The abbreviations are as used in figure 4.4.1, except for the Spa24 and ORF2 proteins. The alignment was performed using the SEQNET program, CLUSTAL V.

```
FliP At   MIRFLVTIAVLLALPGLANAQQFPSDLFNTQIDGSVAAWI--IRTFGLLTVLSVAPGILI
FliP Ec   MRRLLSVAPVLLWLITPLAFAQLPG-ITSQPLPGGGQSWSLPVQTLVFITSLTFIPAILL
Spa24     M---LSDM-SLI-----------------------------ATLSFFTLLPFLVAA--
ORF2      MQ--MPDVGSLL-----------------------------LVVIMLGLLPFAAMV--
          *   .     *.                                  .  *

FliP At   MVTSFPRFVIAFSILRSGMGLASTPSNMILLSMAMFMTFYVMSPTFD---KAWTDGVQPL
FliP Ec   MMTSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVID---KIYVDAYQPF
Spa24     -GTCYIKFSIVFVMVRNALGLQQVPSNMTLNGIALIMALFVMKPIIEAGYENYLNGPQKF
ORF2      -VTSYTKIVVVLGLLRNAIGVQQVPPNMVLNGVALLVSCFVMAPV---GMEAF-KAAQNY
            * . .. . . ..*...*    *.*  * ..*.... ..* *          . *

FliP At   --LQNQINEQQAVQRIAEPFRTFMNANTRDKDLKLFVDIARERGQVVMTDNVVDYRV--L
FliP Ec   --SEEKISMQEALEKGAQPLREFMLRQTREADLGLFARLAN--TGPLQGPEAVPMRI--L
Spa24     DTISD---IVRFSDSGLMEYKQYLKKHTDLELARFFQRSEEEN-ADLKSAENNDYSLFSL
ORF2      GAGSDNSRVVVLLDACREPFRQFLLKHTREREKAFFMRSAQQIWPKDKAATLKSDDLLVL
              .       .     . ..  .*     .*                    .  *

FliP At   VPAFMLSEIRRGFEIGFLIILPFLVIDLIVATITMAMGMMMLPPTSISLPFKILFFVLID
FliP Ec   LPAYVTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVD
Spa24     LPAYALSEIKDAFKIGFYLYLPFVVVDLVISSILLALGMMMMSPITISVPIKLVLFVALD
ORF2      APAFTLSELTEAFRIGFLLYLVFIVIDLVVANALMAMGLSQVTPTNVAIPFKLLLFVAMD
           **.  **.  .* *** . . *...**...   .*.*.  . *  ...*.*...** .*

FliP At   GWNLLVGSLVRSF-NX--
FliP Ec   GWQLLVGSLAQSFYSX--
Spa24     GWGILSKALIEQYINIPA
ORF2      GWSMLIHGLVLSYR----
          ** .*  .*   .
```

4.5 ORFA

This is a 489 base pair open reading frame between the partial *flgG* and the *flgI* homologues. ORFA contains the Tn5 insertion site in the mutant *fla-15*. The gene product of ORFA is 162 amino acids long (predicted M_r of 16,979Da) and has no significant sequence identity to any other proteins in the OWL database. The hydropathy profile, according to the Kyte and Doolittle algorithm [160], of the putative protein is shown below in figure 4.5.1. There is a hydrophobic region at the amino-terminus that could be a transmembrane domain.

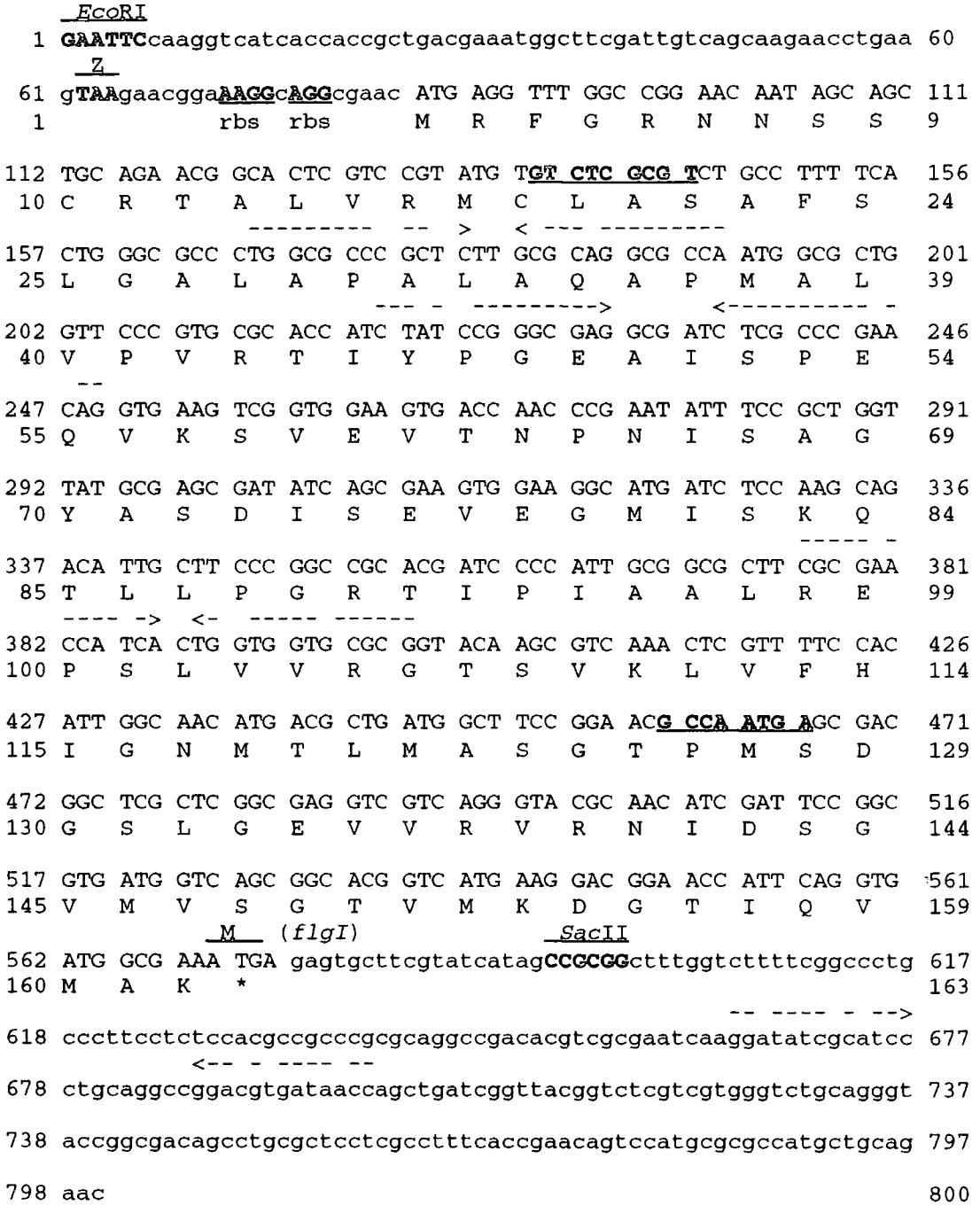
Figure 4.5.1 The hydropathy profile of ORFA from *A. tumefaciens* using the Kyte and Doolittle analysis.



Upstream of ORFA only 84 nucleotides have been sequenced. There are no promoter-like sequences within this region, although these could be present further upstream. There are two putative ribosome binding sites, "AAGG" nine base pairs upstream and "AGG" five base pairs upstream. There is a twenty base pair intergenic gap between the stop codon of the partial *flgG* homologue and the start codon of ORFA. The stop codon of ORFA does overlap with the start codon of *flgI* to give the "ATGA" configuration. These sequences, and the nine base pair Tn5 insertion site in *fla-15*, are all shown in figure 4.5.2.

The DNA sequence downstream of ORFA was searched for possible transcription termination sequences, using the UWGCG program TERMINATOR. No "classical" rho-independent termination sequences were found, but there was a large possible stem-loop structure that might constitute a transcription termination signal. Within ORFA itself are three large, possible, stem-loop sequences, as shown in figure 4.5.2. Their possible role in the regulation of the putative operon will be considered in the discussion section of this chapter.

The program TESTCODE predicted (at the 95% confidence level) that ORFA is a coding region. The guanine and cytosine content of ORFA was 59.5% and is significantly higher than that of non-coding regions. The flagella-less phenotype of the *fla-15* mutant is presumably a result of the insertion of Tn5 in ORFA. However the polar effects of the Tn5 insertion may have prevented the transcription of the other open reading frames downstream in the putative operon. The absence of this could also have led to a flagella-less phenotype. Protein expression studies will be required to show that the gene product of ORFA is expressed. Any effects on motility resulting from over-expression of the ORFA gene product may also help define a role for this putative protein. If the protein is shown to be expressed, immunocytochemistry could then be carried out to localise its position in the cell and hence give some insight into its function. To study the possible regulatory role of the stem-loop structures within ORFA, site-specific mutagenesis could be performed. Ideally this would only alter the stem-loop structures, and not substantially affect the coding region.


EcoRI
Z
rbs rbs
M R F G R N N S S
M A K *
(flgI)
SacII

4.6 ORFB

This is a 537 base pair open reading frame between the *flgI* and *flgH* genes. The 178 amino acid long gene product has a predicted molecular weight of 19,862Da. This putative protein has some sequence identity to the amino-terminal half of FliG from *S. typhimurium*. It has 25% sequence identity (47% sequence similarity) over 159 amino acids. A gap alignment between the two protein sequences is shown in figure 4.6.1. The FliG protein of *S. typhimurium* is actually 331 amino acids long and is involved in the switching of the flagellum (see introduction). FliG of *S. typhimurium* has been extensively studied by the analysis of a large number of mutations within the gene [132]. The amino-terminal portion was found to be relatively unimportant. One mutation which deleted 94 amino acids from the amino-terminus still allowed partial rotation and switching of the flagellum. Another *fliG* homologue has already been partially sequenced on pDUB1900 (see introduction). This encodes a 355 amino acid protein which has 21% sequence identity to the *S. typhimurium* FliG, and 21% sequence identity to the gene product of ORFB. The hydropathy profiles, using the Kyte and Doolittle algorithm [160], are shown in figure 4.6.2. The ORFB protein's hydropathy profile shows a possible hydrophobic transmembrane domain at the amino-terminus, whereas the other two FliG hydropathy profiles do not. The ORFB gene product also has 26% sequence identity (44% sequence similarity) over 94 amino acids to the product of an open reading frame within a *B. subtilis* flagellar operon. A gap alignment between the two putative proteins is shown in figure 4.6.3. The function of this open reading frame (Orf6) in *B. subtilis* is not known. It does have slight sequence identity to the carboxy-terminal region of the FrzCD protein of *Myxococcus xanthus*. FrzCD has sequence identity with the MCPs of the enteric bacteria [7]. The ORFB putative gene product does not have any significant similarity to the MCPs. The hydropathy profiles of the putative gene products of ORFB and Orf6 of *B. subtilis*, see figure 4.6.2, are similar. Both proteins have a possible hydrophobic membrane spanning domain at the amino-terminus with the remainder of the protein being predominantly hydrophilic.

Figure 4.6.4 shows that 11 base pairs upstream of ORFB is a putative ribosome binding site "GGAG". There are no promoter-like sequences upstream and no transcription termination sequences downstream. The start and stop codons overlap with the stop and start codons of *flgI* and *flgH*, respectively, to give the "ATGA" configuration.

ORFB is predicted at the 95% confidence level to be a coding region by the UWGCG program TESTCODE, see figure 3.4. The high guanine and cytosine content, 56.8%, relative to that of predicted non-coding sequences is again suggestive that ORFB is a coding region. The possible functions of the gene product of ORFB are unknown. The sequence identity to FliG is significant although the ORFB gene product may not perform a

similar role. It is much smaller and lacks the important carboxy-terminus sequence which is involved in switching and motor function [132]. The sequence identity to the unknown open reading frame of *B. subtilis* is perhaps more relevant. Both of these proteins are of a similar size and have comparable hydropathy profiles. It is unfortunate that the roles of both are unknown. Further studies, as described for ORFA, need to be performed on ORFB to show whether a protein is expressed and to elucidate its function.

Figure 4.6.1 A gap alignment between the gene product of ORFB and FliG of *S. typhimurium*.

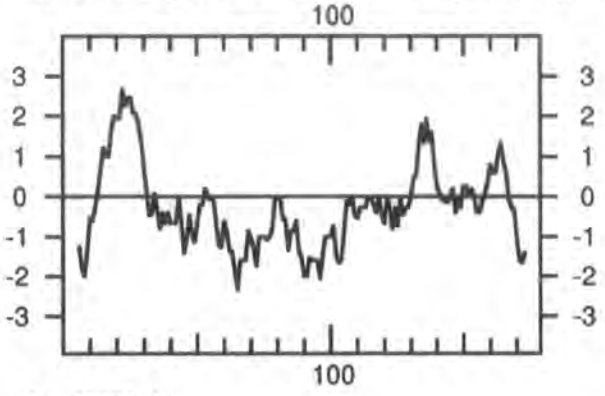
A gap alignment between the gene product of ORFB (labelled *At*) and FliG of *S. typhimurium* (labelled *St*). The UWGCG program GAP, which uses the algorithm of Needleman and Wunsch [207], was used to create it.

```
     Percent Similarity: 46.626   Percent Identity: 24.540

                .         .         .         .         .
At     1 MMERQTKNPLSNGLVRFAAVASLLFLLPVAGAESQQNVVSELSTQDEIQK 50
                    : .:.:... ::|| . |.:...:|...|||. |:|
St     1 ..............MSNLSGTDKSVILLMTIGEDRAAEVFKHLSTR.EVQA 36
                  .         .         .         .         .
At    51 FCTNIADAARDQRYLMQKQDLEKLQADVNERISVLENRKAEYEDWLARRE 100
         ::|.:|: .|: .    .: |..::.: .|.:..|: . .||  .:  :.
St    37 LSTAMAN.VRQISNKQLTDVLSEFEQE.AEQFAALNINANEYLRSVLVKA 84
                  .         .         .         .         .
At   101 HFLNQAKSNLVDIYKT............MKADAAAPQLEKMHVEIAAAII 138
            :.|.| | ||..|            |.::.||. : . |.:| |.|:
St    85 LGEERASSLLEDILETRDTTSGIETLNFMEPQSAADLIRDEHPQIIATIL 134
                  .         .         .         .         .
At   139 MQLPPRQSGLILSEMDAQ.......KAATVAGIMSQAIDKNTSKDPS. . 178
         ::|...|.: ||. :|..       : ||.:|: ..|:.. |.   :
St   135 VHLKRSQAADILALFDERLRHDVMLRIATFGGVQPAALAELTEVLNGLLD 184
```

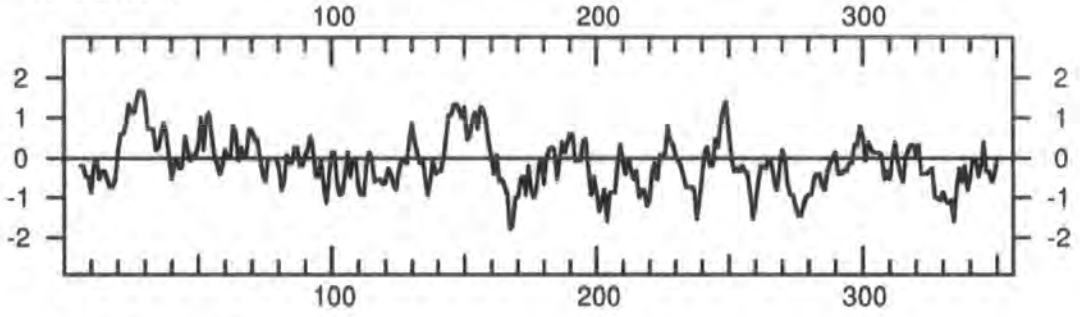
Figure 4.6.2 Hydropathy profiles of the protein encoded by ORFB, the larger *A. tumefaciens* FliG homologue, the *S. typhimurium* FliG and the protein encoded by Orf6 from *B. subtilis*.

The profiles were drawn with DNA Strider, using the Kyte and Doolittle analysis.

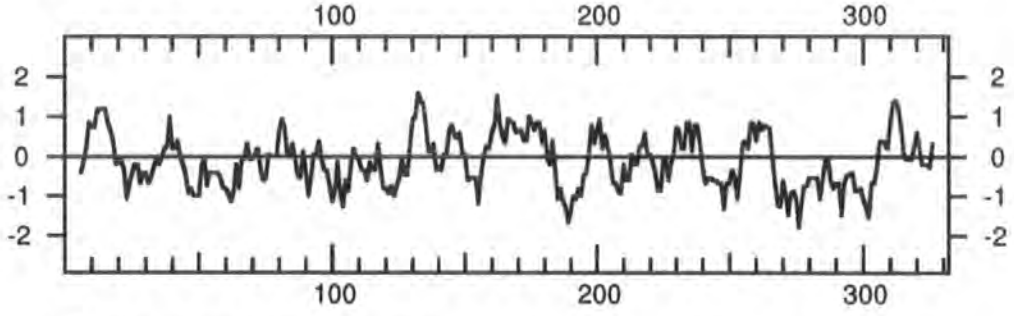
a) ORFB protein from *A. tumefaciens* (*At*)



b) FliG *At*



c) FliG *S. typhimurium*



d) Orf6 protein from *B. subtilis*

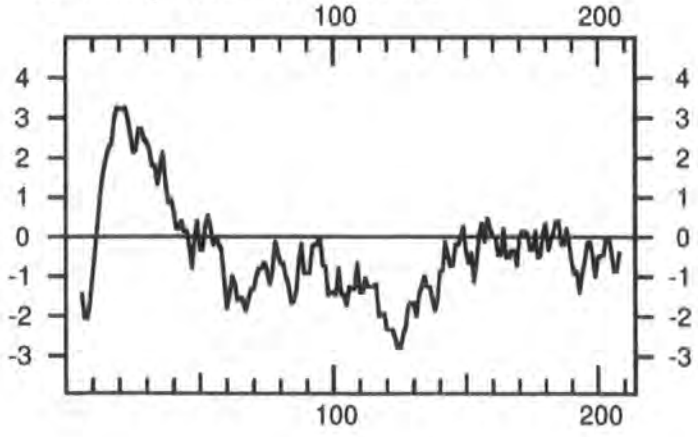


Figure 4.6.3 A gap alignment between the putative gene products of ORFB from *A. tumefaciens* and Orf6 of *B. subtilis*.

A gap alignment between the gene product of ORFB (labelled *At*) and Orf6 of *B. subtilis* (labelled *Bs*). The UWGCG program GAP, which uses the algorithm of Needleman and Wunsch [207], was used to create it.

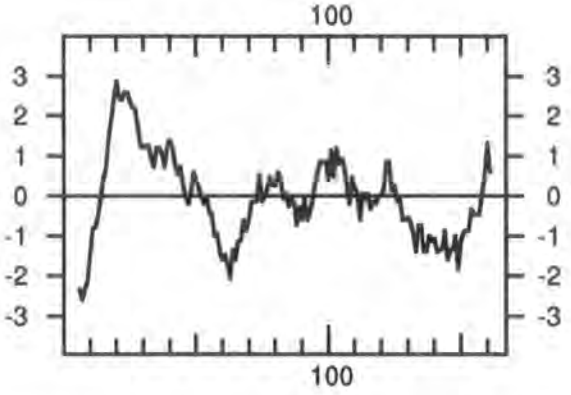
```
      Percent Similarity: 43.750   Percent Identity: 25.568

                    .          .          .          .          .
At    1 MMERQTKNPLSNGLVRFAAVASLLFLL.................PVAGAES 34
          |. ..|:. .   | :  : .|:|||                 |:.:|..
Bs    1 .MSGKKKESGKFRSVLLIIILPLMFLLIAGGIVLWAAGINVLKPIQDAAA 49
                    .          .          .          .          .
At   35 QQNVVSELSTQDEIQKFCTNIAD....AARDQRYLMQKQDLEKLQADV.. 78
          .  |:.|| .:.| .| ...  |    || :. .  || ::. |. |:
Bs   50 KTPVLKELVPETENKKGAASSKDSSNTAALEKTIKDQKSEISILNKDLET 99
                    .          .          .          .          .
At   79 ........NERISVLENRKAEYEDWLARREHFLNQAKS.......NLVDI 113
                 |::|. ||  |.. ::. .. :|  . |.|       .::.:
Bs  100 SKSEIDRLNQKIRSLE..KTAEDQKKSSEDHTEGSADSKASSENDKVISV 147
                    .          .          .          .          .
At  114 YKTMKADAAAPQLEKMHVEIAAAIIMQLPPRQSGLILSEMDAQKAATVAG 163
          ||.|..: ||. :..:. : |  |: .|..:| : ||..|.::.||| .:
Bs  148 YKSMDSGKAAKIIAQLKEQEALKILNGLSKKQLADILAKMTPEQAATYTE 197
                    .
At  164 IMSQAIDKNTSKDPS  178
          :...   .. |:...
Bs  198 KIAAQPRIGGMKGEAA 213
```

```
  1 acggcagtgcagccgcagacggatatcatggccatgcaggaaggtagcaaggtcgccatc 60

 61 gtcgaaggcccggaccttcgcacgctggttgccggtctcaacagcatcgggctcaaggcg 120

121 gacggcatcatcgccatcctgcaaggcatcaaatccgccggcgcccttcaggcGGAGctc 180
                                                          rbs
181 gtgctgca ATG ATG GAA CGT CAG ACC AAA AAT CCG CTC TCA AAT GGC 227
  1          M   M   E   R   Q   T   K   N   P   L   S   N   G  13

228 CTC GTC CGC TTT GCG GCC GTC GCG TCG CTG CTG TTC CTG TTG CCG 272
 14 L   V   R   F   A   A   V   A   S   L   L   F   L   L   P  28

273 GTT GCC GGC GCC GAA AGC CAG CAG AAC GTG GTG TCT GAA CTC AGC 317
 29 V   A   G   A   E   S   Q   Q   N   V   V   S   E   L   S  43

318 ACG CAG GAC GAA ATC CAG AAA TTC TGC ACC AAT ATT GCC GAT GCG 362
 44 T   Q   D   E   I   Q   K   F   C   T   N   I   A   D   A  58
                                                          Hd
363 GCC CGC GAC CAG CGT TAC CTC ATG CAG AAG CAG GAC CTT GAA AAG 407
 59 A   R   D   Q   R   Y   L   M   Q   K   Q   D   L   E   K  73
    III
408 CTT CAG GCC GAC GTC AAT GAG CGT ATT TCC GTG CTG GAA AAC CGC 452
 74 L   Q   A   D   V   N   E   R   I   S   V   L   E   N   R  88

453 AAG GCG GAA TAT GAG GAC TGG CTG GCG CGC CGC GAA CAT TTC CTC 497
 89 K   A   E   Y   E   D   W   L   A   R   R   E   H   F   L 103

498 AAT CAG GCC AAA AGC AAC CTC GTC GAC ATC TAC AAG ACG ATG AAG 542
104 N   Q   A   K   S   N   L   V   D   I   Y   K   T   M   K 118

543 GCG GAC GCC GCA GCT CCC CAG CTA GAA AAG ATG CAT GTG GAA ATC 587
119 A   D   A   A   A   P   Q   L   E   K   M   H   V   E   I 133

588 GCG GCC GCC ATC ATC ATG CAG CTG CCA CCG CGC CAA TCG GGC CTC 632
134 A   A   A   I   I   M   Q   L   P   P   R   Q   S   G   L 148

633 ATC CTC AGC GAA ATG GAT GCG CAG AAA GCC GCA ACC GTC GCC GGC 677
149 I   L   S   E   M   D   A   Q   K   A   A   T   V   A   G 163
                                                  BamHI
678 ATC ATG TCG CAG GCA ATC GAC AAA AAC ACT TCG AAG GAT CCT TCA 722
164 I   M   S   Q   A   I   D   K   N   T   S   K   D   P   S 178
    M       (flgH)
723 TGA gcacgcgtcgtcttccggccctcctcctgccgctcgctctcctggccggctgccag 781
179 *                                                         179

782 aataaccagaccctgaagga                                       801
```

4.7 ORFC

This is a 501 base pair open reading frame between the *flgH* and *fliP* homologues. ORFC contains the Tn5 insertion site in the mutant *mot-12*. The gene product of ORFC is 166 amino acids long (predicted M_r 17,834Da) and has no significant sequence identity to any other proteins in the OWL database. The hydropathy profile, according to the Kyte and Doolittle algorithm [160], of the putative protein is shown below in figure 4.7.1. There is a hydrophobic region at the amino-terminus that could be a transmembrane domain.

Figure 4.7.1 The hydropathy profile of ORFC from *A. tumefaciens* using the Kyte and Doolittle analysis.



Upstream of the start codon of ORFC there are two putative ribosome binding sites. Both have the sequence "GGA" and are 12 and 7 base pairs upstream. Further upstream no promoter-like sequences were identified. There is a 26 base pair intergenic gap between the stop codon of *flgH* and the start codon of ORFC. The stop codon of ORFC does overlap with the start codon of *fliP* to give the "ATGA" configuration. These sequences, and the nine base pair Tn5 insertion site in *mot-12* are shown in figure 4.7.2.

The DNA sequence downstream of ORFC was searched for possible transcription termination sequences, using the UWGCG program TERMINATOR. No "classical" rho-independent termination sequences were found, but there was a large possible stem-loop structure that might constitute a transcription termination signal. Within ORFC itself there was another large, possible, stem-loop sequence. The position of these two possible stem-loop sequences are shown in figure 4.7.2. The possible roles of these stem-loop structures in the regulation of the putative operon will be considered in the discussion section of this chapter.

The program TESTCODE predicted (at the 95% confidence level) that ORFC is a coding region. The guanine and cytosine content of ORFC is 60.2%, significantly higher than that of non-coding regions, and is again suggestive that ORFC is a coding region. Just over 200 base pairs of sequence upstream of the *R. meliloti fliP* homologue were obtained from T. M. Finan [86]. This sequence contained a partial open reading frame which translated into a protein which has 75% sequence identity (89% sequence similarity) to the predicted protein product of ORFC. A gap alignment between the two putative proteins is shown in figure 4.7.3(B). Downstream of the *R. meliloti* ORFC coding region is a possible stem-loop sequence that might constitute a transcription termination signal. This is shown in figure 4.7.3(A). Upstream of the *fliP* homologue (*mopC*) of *E. carotovora* subspecies *atroseptica* is an open reading frame that also encodes a protein with no significant similarity to any other proteins. It is a similar size to the ORFC gene product, 140 amino acids long and has a predicted M_r of 15,024Da [204]. Although there is no sequence similarity between the two.

The non-motile (but flagellated) phenotype of the *mot-12* mutant is presumably a result of the insertion of Tn5 in ORFC. This phenotype is somewhat surprising as there should have been polar suppression of *fliP* due to Tn5 insertion and the absence of *fliP* has been shown, in *A. tumefaciens*, to lead to a flagella-less phenotype (this study). Further studies, as described for ORFA, need to be performed on ORFC to show if a protein is expressed and to elucidate its function. In particular the expression of *fliP* in *mot-12* should be determined.

Although there is no significant sequence identity between ORFA and ORFC or the predicted proteins encoded by them, there is some general similarity. Both the predicted proteins are of a similar size. ORFA is 162 amino acids long (predicted M_r 16,979Da) and ORFC is 166 amino acids long (predicted M_r 17,834 Da). The hydropathy profiles of the two proteins (see figures 4.5.1 and 4.7.1) show possible hydrophobic membrane spanning domains at the amino-termini. Both open reading frames contained possible stem-loop sequences within them and also downstream, which could constitute transcription termination signals. Both open reading frames also had similar sized intergenic gaps between the gene upstream, but their stop codons overlapped with the start codons of the genes downstream in the "ATGA" configuration. These similarities will be considered in the discussion section of this chapter.

```
  1 aaaacggcaacctcatcatcagcggctcgcaggaagtgcgtgtgaaccacgaaatccgca 60

 61 tcctcaacgtcggcggtatcgtccgtccgcaggatgtcgatgcccagaacatcatctcct 120

121 acgagcgcatcgccgaagcacgcatctcctacggcggtcgtggccgtctgacggaagtgc 180
                                                        Z
181 agcagcccccggtcgggcagcaggtcgttgacctgttctcgccgctcTGAcgccgcacac 240

241 gGGAacGGAactgaccc ATG GAA AAC GAA CAG GCT GAG GGC AAA AAA AAA 289
  1  rbs  rbs        M   E   N   E   Q   A   E   G   K   K   K   11

290 TCC TCC CCT CTG GTC ATG ACC ATC GCA GGC GTT GTG ATC CTC ACC  334
 12 S   S   P   L   V   M   T   I   A   G   V   V   I   L   T    26

335 CTG CTC GGT GCG GGC GGC GGC TGG CTC GTG GGC GGC ATG ATC GCC  379
 27 L   L   G   A   G   G   G   W   L   V   G   G   M   I   A    41
                -------  ---- --->      <--  ---- --------
380 CCG AAA GTT GCC GCA ACG GAG GCC CAT GCC ACC GCT GCG GCC GGC  424
 42 P   K   V   A   A   T   E   A   H   A   T   A   A   A   G    56

425 GGC CAT GGC GAG AAG AAG GGC GAA GGC CTC GAT AAA ATC GAC GCC  469
 57 G   H   G   E   K   K   G   E   G   L   D   K   I   D   A    71

470 GAA GCG AAC GGC ATC GTC CAG CTC GAT CCC ATC ACC ACC AAT CTC  514
 72 E   A   N   G   I   V   Q   L   D   P   I   T   T   N   L    86

515 GCC TAC CCT TCG ACC AAC TGG GTG CGG CTG GAA GTG GCG CTG ATG  559
 87 A   Y   P   S   T   N   W   V   R   L   E   V   A   L   M   101
                                                  EcoRV
560 TTC AAG GGG CCG GTG GAA GTC GGG CTG GCG GAG GAT ATC CAC CAG  604
102 F   K   G   P   V   E   V   G   L   A   E   D   I   H   Q   116

605 GAC ATC ATG GCC TAT GTC CGG ACC GTT TCC CTC CAG CAG CTG GAA  649
117 D   I   M   A   Y   V   R   T   V   S   L   Q   Q   L   E   131

650 GGC CCG CGC GGC TTT CAA TAT CTT AAG GAT GAC ATT CAG GAA CGA  694
132 G   P   R   G   F   Q   Y   L   K   D   D   I   Q   E   R   146

695 GTT GAC CTG CGC TCT CAA GGG CGC GTA TCC AAG GTC ATG TTC CGG  739
147 V   D   L   R   S   Q   G   R   V   S   K   V   M   F   R   161
                      M    (fliP)                              -
740 ACC TTT GTC ATC GAA TGA ttcgatttcttgtaaccatcgccgtcctcctcgctt 793
162 T   F   V   I   E   *                                       167
    --- ---- --> <-- ---- ----
794 tgccgggccttgccaatgcccagcaatttccgtccgatctgttcaacacgcagatcgacg 853

854 gctccgtcgcggcatggatcatccgcaccttcggcctgttgacggttctctccgtcgcgc 913

914 ccggcatcctgatcatggtcacgagctttccgcgcttcgtgatcgccttttcgatcctgc 973

974 gctccggcatgggacttgcgtcgacacc                                 1001
```

Figure 4.7.3 The putative protein encoded by the ORFC homologue of *R. meliloti*, the downstream DNA sequence and a gap alignment with ORFC of *A. tumefaciens*.

Diagram (A) shows the partial ORFC homologue from *R. meliloti* and the protein encoded by it. The start codon of *fliP* is overlined and labelled with an "M" at nucleotide positions 224-226. The possible stem-loop sequence downstream of ORFC that might constitute a transcription termination signal is also shown between nucleotides 248-263 with staggered arrows.

Diagram (B) shows a gap alignment between the proteins encoded by the ORFC's from *R. meliloti* (*Rm*) and *A. tumefaciens* (*At*). This was performed using the UWGCG program GAP.

(A)

```
   1      CTC GAG GTC GCG CTC CAG TTC GAT CGA GCT CCC GAC GTT GCA  42
   1      L   E   V   A   L   Q   F   D   R   A   P   D   V   A   14

  43 CTG GCG GAG ACG ATA CAC CAG GAC ATT GCA GCC TAT CTG AAG ACC  87
  15 L   A   E   T   I   H   Q   D   I   A   A   Y   L   K   T   29

  88 GTG TCG CTG CAG CAG ATT CAA GGA CCG CGC GGC TTC CAA TAT CTC  132
  30 V   S   L   Q   Q   I   Q   G   P   R   G   F   Q   Y   L   44

 133 CGG GAT GAC ATC CAG GAG CGG GTT GAC CTG CGC TCC GAG GGC CGC  177
  45 R   D   D   I   Q   E   R   V   D   L   R   S   E   G   R   59

 178 GTA ACG AAT GTG ATG TTC CGC ACC TTC GTC ATC CAA TGA ttcgctgA 224
  60 V   T   N   V   M   F   R   T   F   V   I   Q   *           72
     M   (of fliP)          -------> <-------
 225 TGctccggtttgccaccttcataatcgccatgatggcgatgtcgggaattgccggggctc 284

 285 agagcttccccgccgatatcctgaacacgccggtcgacggctccgtcgcctcgtggatca 344
```

(B)

```
       Percent Similarity: 88.889   Percent Identity: 75.000

 At   51 ATAAAGGHGEKKGEGLDKIDAEANGIVQLDPITTNLAYPSTNWVRLEVAL 100
                                                      |||||
 Rm    1 .....................................................LEVAL 5

 At  101 MFKGPVEVGLAEDIHQDIMAYVRTVSLQQLEGPRGFQYLKDDIQERVDLR 150
          |. :.:|:|||.||||| ||::|||||::||||||||:||||||||||
 Rm    6 QFDRAPDVALAETIHQDIAAYLKTVSLQQIQGPRGFQYLRDDIQERVDLR 55

 At  151 SQGRVSKVMFRTFVIE 166
          |:|||..|||||||:
 Rm   56 SEGRVTNVMFRTFVIQ 71
```

4.8 Similarity between the flagellar gene homologues of *A. tumefaciens* and homologous DNA sequences in *R. meliloti*

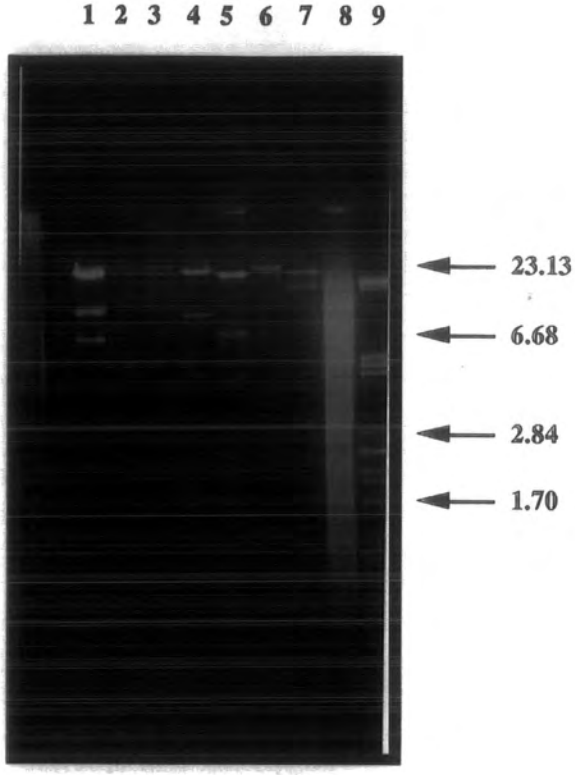
A. tumefaciens and *R. meliloti* share a similar type of cell motility [172]. Sequence comparisons have shown extensive similarities between genes from the two species that are involved in motility, see sections 4.5, 4.7 and chapter 5. Previous work using the flanking sequences from three of the *A. tumefaciens* motility mutants (*che-2*, *fla-7* and *fla-11*) as radiolabelled probes for Southern blots has revealed homologous DNA sequences within *R. meliloti* genomic DNA and also within the pRZ cosmids [47, 267]. This section describes the continuation of this work. However, rather than radiolabelling flanking sequence DNA from various mutants as probes, defined regions of the various flagellar genes and unknown ORFs were used.

The cosmids pRZ1, 2 and 4 were isolated and (approximately) 5µg samples of each were digested with *Eco*RI, and *Bam*HI. *R. meliloti* genomic DNA was also isolated and aliquots digested to completion with *Eco*RI. The digested DNA was electrophoresed on large (18x15cm) 0.7% agarose gels. Photographs of both gels are shown in figure 4.8.1. The DNA, from each gel, was transferred to duplicate nylon filters by 2-way Southern blotting.

The three pRZ cosmids complement a number of chemotactic and motility mutants in *R. meliloti*, which are clustered in a 25kb region on the chromosome. The three cosmids overlap giving a ~35kb stretch of sequence that contains this region. The sequence has been mapped by restriction enzyme digestion (*Eco*RI and *Bam*HI) and the position of the mutations complemented determined [341]. Two copies of the gene encoding the flagellin protein have also been found within this region of *R. meliloti* [27]. An approximate diagram of the above region is shown in figure 4.8.2.

The restriction enzymes used to generate the probe DNA fragments from the flagellar genes and unidentified ORFs have been shown in the last six sections. The sixty base pairs thought to encode the partial carboxy-terminus of *flgG* were excised with ORFA. After hybridisation to a particular flagellar gene probe, washing and exposure to film, the blots were stripped as described in section 2.13.6. Film sheets were then exposed to the blots for 168 hours at -80°C to ensure the blot was clean before it was re-probed. Hybridisations were carried out at 65°C and high stringency washes were performed as described in sections 2.13.3 and 2.13.4.

A.



B.

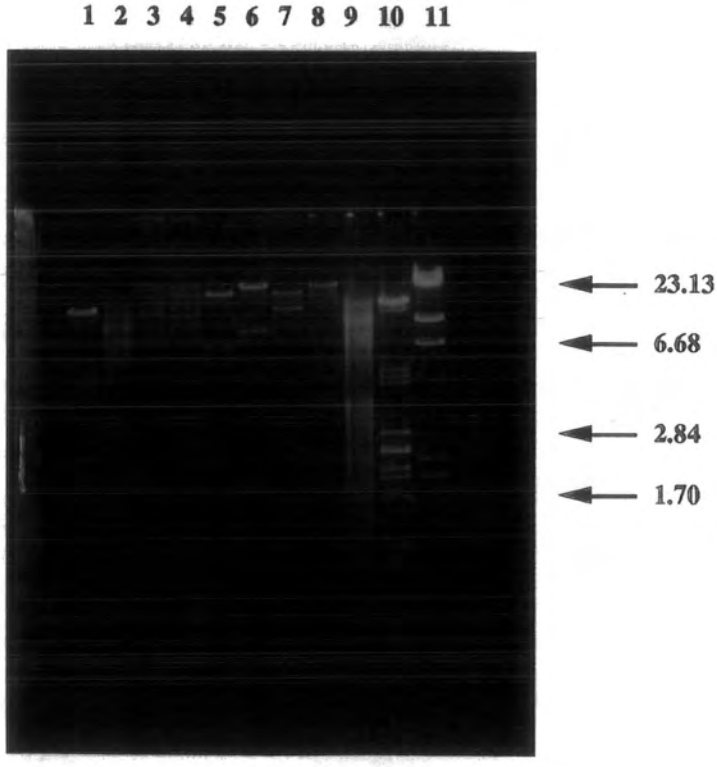


Figure 4.8.2 A diagram mapping three of the pRZ cosmids that complement behavioural mutants within the *fla-che* region of *R. meliloti*.

This diagram was modified and redrawn (with permission) from that of Ziegler *et al.* [341]. The top line shows the relative positions of a number of the *R. meliloti* behavioural mutants. The pRZ cosmids (labelled 1,2 and 4) are shown below the mutants that they complement. The thick line below these shows an approximate *Eco*RI (R) and *Bam*HI (B) restriction map of the region. Three *Eco*RI fragments, (labelled RA, RB and RC) and one *Bam*HI fragment (labelled BA) are highlighted below and will be discussed in the text. The positions of the two cloned and sequenced flagellin genes (*flaA* and *flaB*) are shown at the left hand end of pRZ4 with arrows labelled A and B. The pRZ cosmids were created by cloning *Eco*RI partially digested fragments of *R. meliloti* chromosomal DNA into the cosmid pLAFR1. Accordingly the cloned chromosomal DNA of the cosmids all have *Eco*RI endpoints that align with specific sites on the restriction enzyme map. In the original diagram, pRZ2 and pRZ4 share an *Eco*RI site, this was found not to be so and the new positions are shown. Experimental evidence for this will be presented in this section.


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

All six of the radiolabelled *A. tumefaciens* open reading frames ORFA, *flgI*, ORFB, *flgH*, ORFC and *fliP* hybridised to fragments within the pRZ cosmids and the *R. meliloti* chromosomal DNA.

The radiolabelled fragments of ORFA, *flgI* and ORFB all hybridised to the same bands of pRZ1, pRZ2 and the genomic digests, see figure 4.8.3. A band of ~2.5kb was hybridised to in the *Eco*RI digests of pRZ1 and pRZ2. The only *Eco*RI DNA fragment of this size common to both of the cosmids is that labelled "RA" on figure 4.8.2. A band of ~2.5kb was also hybridised to in the *Eco*RI digest of the *R. meliloti* chromosomal DNA, presumably this is the same DNA fragment. In the *Bam*HI digest of pRZ1, a band of ~11.5kb was hybridised to. The only band of a similar size is the approximately 10kb *Bam*HI fragment "BA" (on figure 4.8.2). The *Bam*HI fragment of pRZ2 to which the (above) probes hybridised to was >25kb. This was because the DNA (to which the probes hybridised) was not excised from the pLAFR1 vector (~23kb) by the *Bam*HI digest.

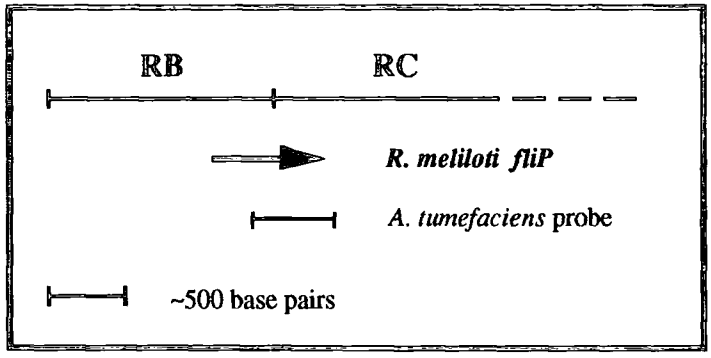
The *flgI* and ORFB (but not ORFA) probes also hybridised to a (faint) ~3.8kb *Eco*RI fragment of pRZ2. This fragment was however thought to be a result of a partial digestion of pRZ2, since no band of a corresponding size is seen on the original agarose gel photograph (figure 4.8.1). Furthermore no such band was hybridised to in the *Eco*RI genomic digests of *R. meliloti.*

Figure 4.8.4(A) shows that the radiolabelled fragment of *flgH* hybridised to a ~2.5kb *Eco*RI fragment in the *Eco*RI digests of pRZ1, pRZ2 and also the genomic DNA - presumably "RA". In the *Bam*HI digest of pRZ1 a ~11.5kb band was hybridised to, and in the *Bam*HI digest of pRZ2 the band hybridised to was too big to be sized accurately (>25kb). The ~11.5kb band was presumably "BA" and the large band probably the pRZ2 part of "BA" attached to pLAFR1. The ~3.8kb partial band in the pRZ2/*Eco*RI digest was also hybridised to by the *flgH* probe. In this lane another large (>25kb) faint band was hybridised to, this is possibly undigested DNA. A ~1.5kb band was also hybridised to by the *flgH* probe in the pRZ1/*Eco*RI and genomic DNA/*Eco*RI lanes. This was thought unlikely to be a partial as it appeared in two distinct digests and there was a band approximately this size visible on the agarose gel photo, see figure 4.8.1(A). Presumably the sequence homologous to the *flgH* homologue of *A. tumefaciens* overlapped one of the *Eco*RI sites defining "RA". However on the original map of the pRZ cosmids [341] no fragment of this size bordered "RA". Since the fragment was not found in the pRZ2/*Eco*RI or pRZ4/*Eco*RI digests but was in the pRZ1/*Eco*RI digest, it was postulated that this ~1.5kb band lay between pRZ2 and pRZ4. A personal communication from Dr. K. Bergman confirmed there is an unmapped 1.5kb band within pRZ1. The communication of the *R. meliloti fliP* sequence [86] also confirmed this. Part of the *R. meliloti fliP* gene resides on

this 1.5kb fragment and the rest on pRZ4. There is an internal *Eco*RI site within *fliP*. The fragment was not completely sequenced, but Dr. T. M. Finan also communicated sequence from the other *Eco*RI site, which putatively formed the border for pRZ2. This sequence has similarity to the *flgH* homologue of *A. tumefaciens*, data not shown. This fragment is shown, and labelled "RB", on figure 4.8.2 between pRZ2 and pRZ4.

The only cosmid the radiolabelled fragment of ORFC was found to hybridise to was pRZ1, see figure 4.8.4(B). A ~11.5kb *Bam*HI fragment and an ~1.5kb *Eco*RI fragment were hybridised to, presumably "BA" and "RB" respectively. In the *R. meliloti* genomic DNA/*Eco*RI digest only an ~1.5kb fragment ("RB") was hybridised to. The presence of DNA sequence homologous to *A. tumefaciens* ORFC on "RB" was confirmed by analysis of the upstream sequence of *R. meliloti fliP* [86] see section 4.7.

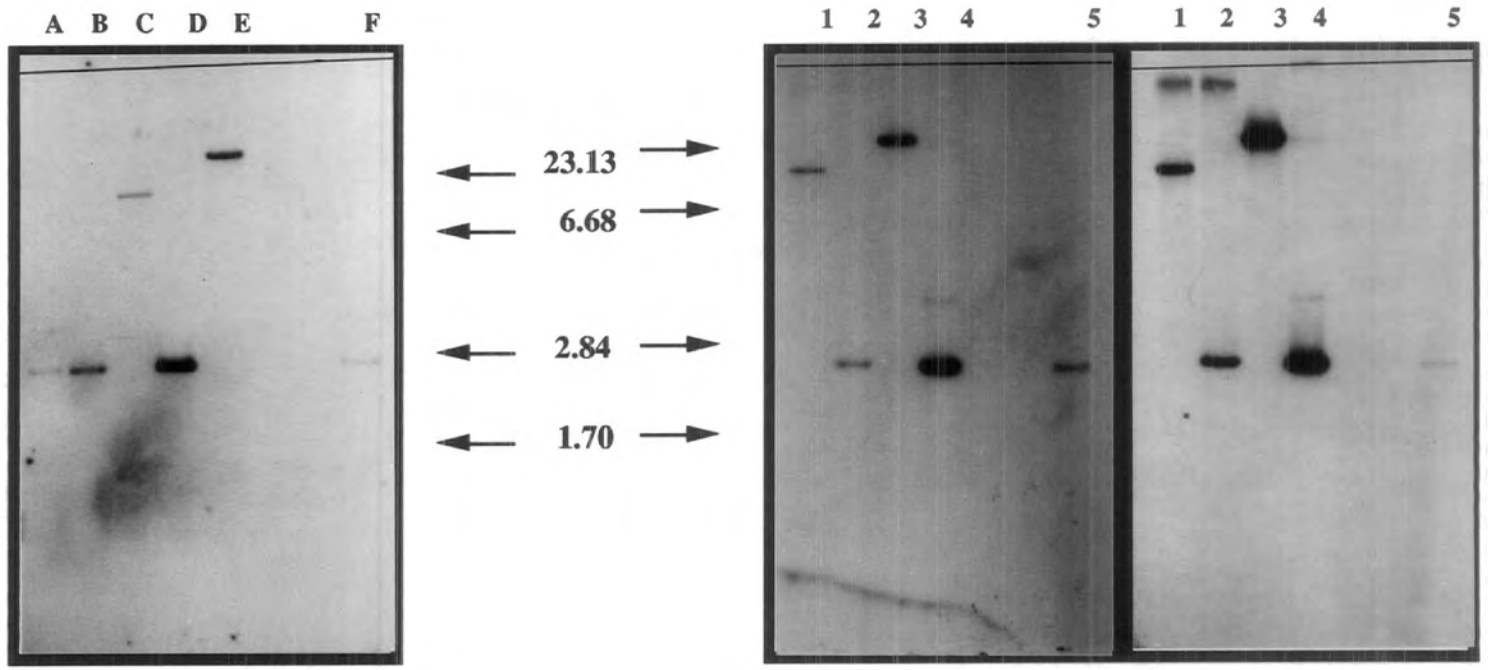
The radiolabelled fragment of *fliP* hybridised to homologous DNA sequences within the cosmids pRZ1 and pRZ4, see figure 4.8.4(C). In each of the *Eco*RI digests (and the genomic DNA) a band of ~11.5kb was hybridised to, labelled "RC" on figure 4.8.2. A band of ~11.5kb was hybridised to in the *Bam*HI digest of pRZ1, presumably "BA". In the *Bam*HI digest of pRZ4 a band too large to be sized accurately (>25kb) was hybridised to. This was probably a result of hybridisation to "BA", which is not detached from pLAFR1 by a *Bam*HI digest of pRZ4. Surprisingly the *Eco*RI fragment "RB" of pRZ1 was not hybridised to. The nucleotide sequence of *R. meliloti fliP* predicts that approximately 400 base pairs of it lie on "RB" and only approximately 360 base pairs on "RC". However the fragment of *A. tumefaciens fliP* that was used to create the probe does lack some of the sequence that would be homologous to that on "RB", see below.



There should still be ~150 base pairs of the *A. tumefaciens fliP* probe that is homologous to "RB". Evidently however this was not sufficient to hybridise to "RB", perhaps a band would have been visible if the film had been exposed for longer (>168 hours). Alternatively the complete nucleotide sequence encoding the amino-terminus could be radiolabelled and used as a probe to identify the homologous sequence on "RB".

All six of the *A. tumefaciens* flagellar gene and unidentified open reading frame fragments were found to have homologous DNA sequences within *R. meliloti*. Three of these were found in the same order as in *A. tumefaciens*; *flgH*, ORFC and *fliP* (as well as the two flagellin genes). The other three homologues were found clustered close to each other and adjacent to *flgH*/ORFC/*fliP*. However no evidence regarding their order was found. Figure 4.8.5 shows the relative positions of these homologues.

These results show that the similarity between the types of cell motility of *A. tumefaciens* and *R. meliloti* extends to both the sequence of the genes involved and the location (and to a degree the order) of these genes. However an attempt to show functional homology was unsuccessful. The cosmids pRZ1 and pRZ2 were introduced into the *A. tumefaciens* behavioural mutants *fla-15* and *mot-12* by tri-parental mating. Their presence was confirmed by isolation from the *A. tumefaciens* cells. Analysis of these strains by light microscopy and swarm plate analysis showed them still to be non-motile. An attempt to complement the *R. meliloti* mutant *mot-1* with pDUB1900 was also unsuccessful. This does not rule out the possibility that other mutants of the two species might be complemented by heterologous cloned DNA. However it is clear that there are some differences, perhaps at a regulatory level, between these related motility genes *A. tumefaciens* and *R. meliloti*.


A B C D E F
23.13
6.68
2.84
1.70
1 2 3 4 5
1 2 3 4 5
A.
B.
C.

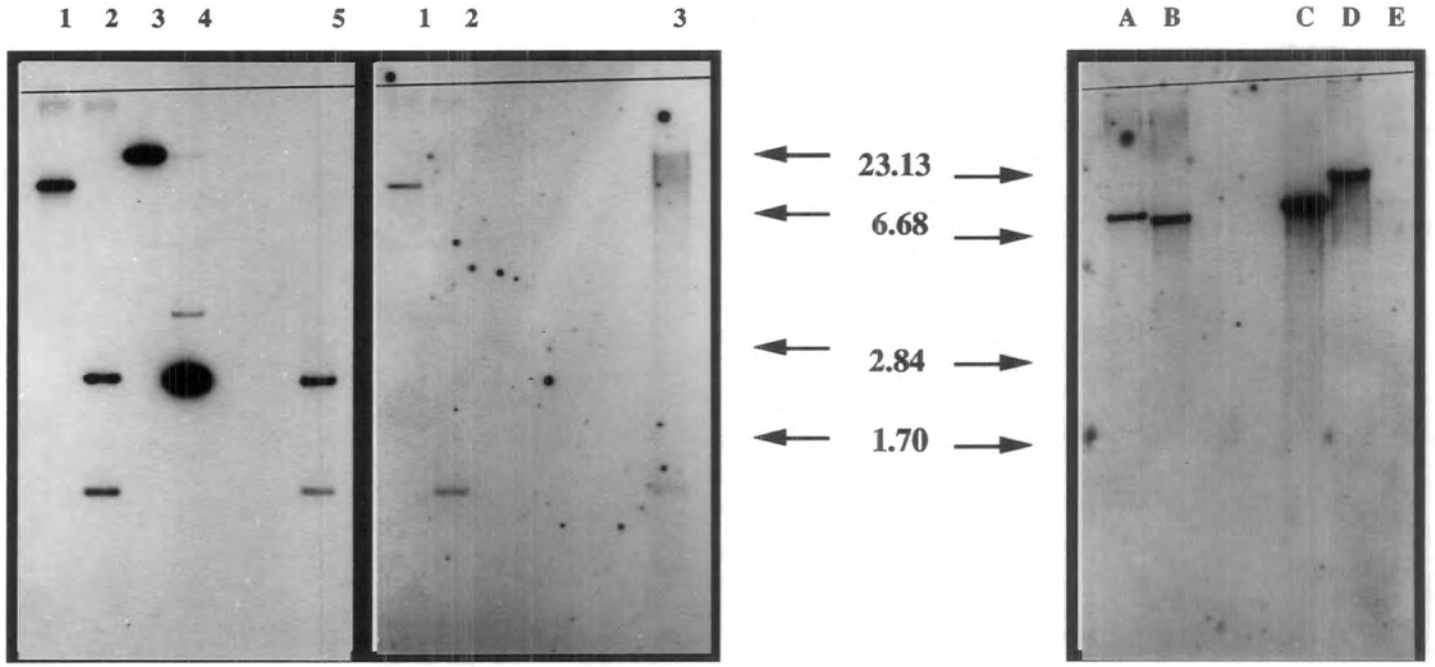
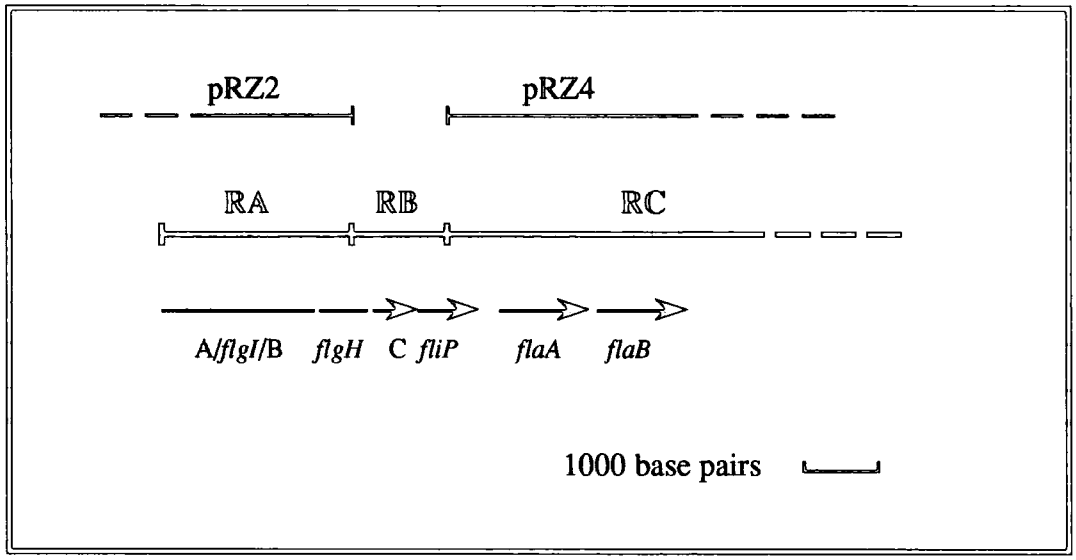

1 2 3 4 5
1 2 3
A B C D E
23.13
6.68
2.84
1.70
A.
B.
C.

Figure 4.8.5 Diagrammatic representation of the positions of the flagellar gene homologues in *R. meliloti*.

The *Eco*RI fragments RA, RB and RC, as discussed in the text are shown. The positions of the cosmids pRZ2 and 4 are marked above the *Eco*RI fragments. The flagellar gene homologues are labelled according to the gene names, the unidentified open reading frames (ORFs A, B and C) are labelled A, B and C.



In the diagram the sizes of the ORF ABC, *flgI* and *flgH* homologues are assumed to be the same as their *A. tumefaciens* counterparts. This is known to be so for the *fliP* homologues. The actual size of the *Eco*RI fragment "RB" is ~1.3kb [86]. Providing there is no large intergenic gap between the *R. meliloti flgH* and ORFC (there isn't in *A. tumefaciens*), this will mean there is ~400bp of *flgH* on RB and ~350bp on RA. The sequence data, discussed in sections 4.4 and 4.7, predicts the direction of transcription of the *R. meliloti* ORFC and *fliP* homologues to be the same (relative to the flagellin genes) as in *A. tumefaciens*. The partial sequence of the *R. meliloti flgH* homologue, see this section, would also predict the direction of its transcription to be the same as in *A. tumefaciens*. The order and direction of transcription of the ORFA/*flgI*/ORFB homologues could not be predicted from the Southern blot data. All three hybridised only to the *Eco*RI fragment RA. Assuming their size to be the same as for the *A. tumefaciens* homologues; 489, 1122 and 501 base pairs respectively (and that there are no large intergenic regions) all three can fit upon RA. The ORFA radiolabelled fragment also contained ~60bp of a putative *flgG* homologue. It is assumed that the fragment hybridised predominantly to an *R. meliloti* ORFA homologue because only a relatively small part of *flgG* was included.

4.9 Similarity between the *A. tumefaciens* flagellar gene homologues and homologous DNA from other Gram-negative bacteria

The strong homologies seen between the *A. tumefaciens* flagellar gene homologues and DNA sequences within *R. meliloti* suggested that the similarity might extend to other Gram-negative bacteria. The radiolabelled DNA fragments of *flgI*, *flgH* and *fliP* were also used as probes against zoo blots containing digested chromosomal DNA of four *Rhizobium* strains, three strains of *Pseudomonas* and an *E. coli* strain. The previous work using the flanking sequences of three of the mutants [47, 267] revealed only homologous DNA sequences in *R. meliloti*. The use of flanking sequence DNA as probes would undoubtedly involve having a larger amount of non-specific DNA being radiolabelled and thus less chance of hybridisation to any homologous DNA sequences. Thus by using defined gene fragments there should be more chance of accurate hybridisation.

Genomic DNA was isolated using the methods described in sections 2.10.5 and 2.10.6. Approximately 5µg samples of each were digested with either *Eco*RI or *Bam*HI. The digested DNA was electrophoresed on a large (18x15cm) 0.7% agarose gel and then transferred to duplicate nylon filters by two-way Southern blotting. A photograph of the gel is shown in figure 4.9.1. Despite attempts to limit the amount of chromosomal DNA digested to 5µg, figure 4.9.1 clearly shows relatively larger amounts of *R. leguminosarum* biovar *phaseoli* and *P. talassii* DNA were used. Only 2µg of the positive control (*A. tumefaciens*) DNA was digested and electrophoresed. Hybridisations were carried out at 65°C, but only one wash with 2x SSC at 65°C for ten minutes was performed. Films were exposed to the blots for 24, 72 and 168 hours.

All three of the radiolabelled DNA fragments (of *flgI*, *flgH* and *fliP*) hybridised to a ~12kb band in the *A. tumefaciens* genomic DNA/*Eco*RI lane, see figure 4.9.2. This ~12kb band is presumably *Eco*RI fragment A of pDUB1900, and thus hybridisation to this was expected.

The radiolabelled fragment of *flgI* also hybridised to a ~2.6kb band in the sym^+ and sym^- strains of strains of *R. leguminosarum* biovar *viciae*. The radiolabelled fragment of *flgH* hybridised to the same ~2.6kb band and an ~3.6kb band of *R. leguminosarum* biovar *viciae* (sym^+ and $^-$). The fragment also hybridised to a ~3.65kb band of *R. leguminosarum* biovar *phaseoli*. The radiolabelled fragment of *fliP* hybridised to the ~3.6kb band in *R. leguminosarum* biovar *viciae* (sym^+ and sym^-) and the ~3.65kb band of *R. leguminosarum* biovar *phaseoli*. The three Southern blots are all shown in figure 4.9.2.

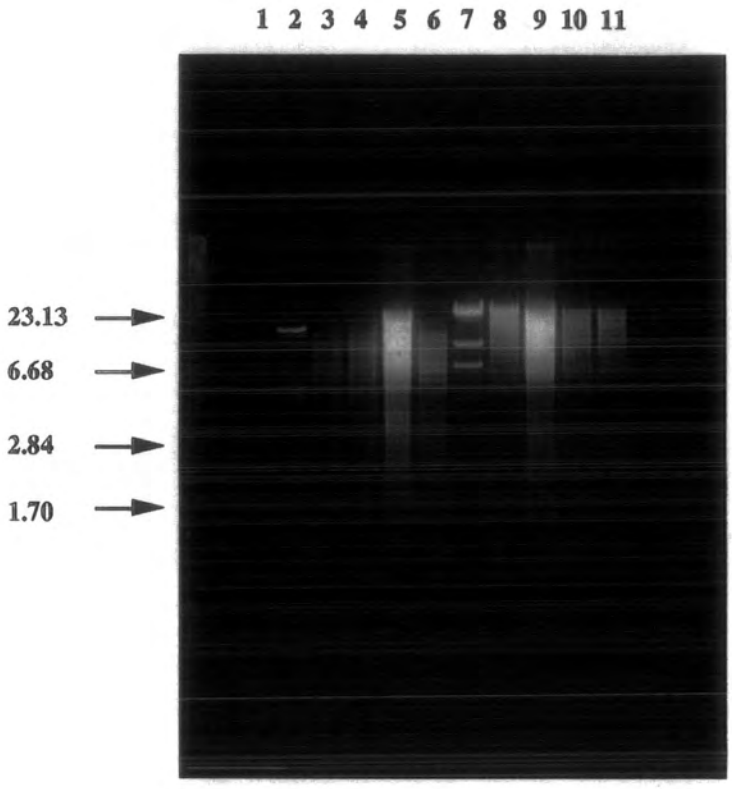
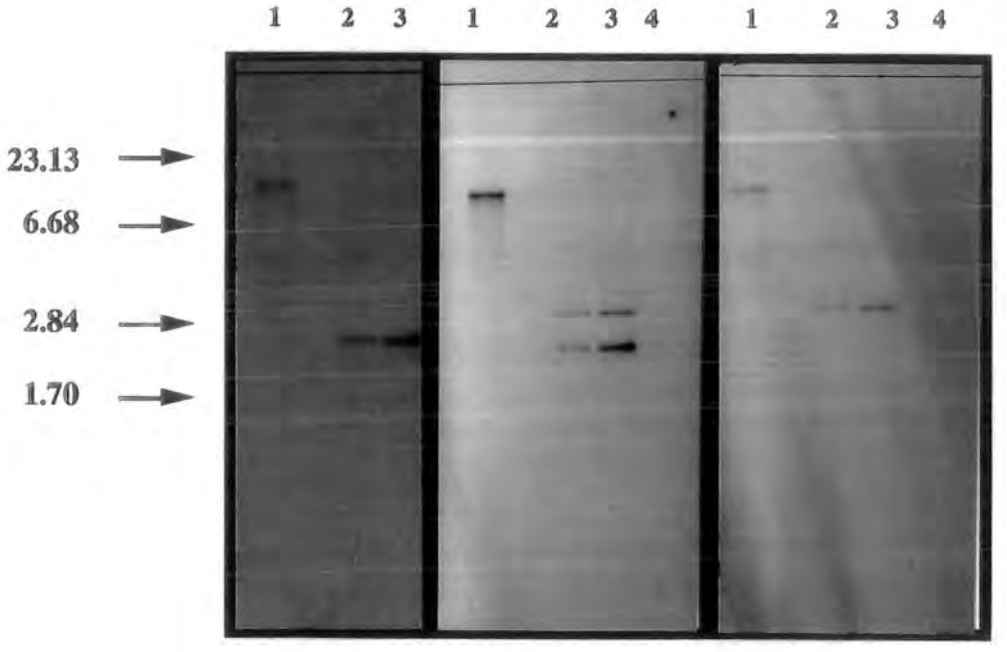

1 2 3 4 5 6 7 8 9 10 11
23.13
6.68
2.84
1.70

FIGURE 4.9.2 Southern blots of a variety of genomic DNA digests with radiolabelled probes from *flgI*, *flgH* and *fliP* of *A. tumefaciens*.

The agarose gel from which the Southern blots were made is shown in figure 4.9.1. Hybridising bands were only seen in the digests listed below.

LANE 1	*A. tumefaciens* C58C1	*Eco*RI.
LANE 2	*R. leguminosarum* biovar *viciae* (sym+)	*Eco*RI.
LANE 3	*R. leguminosarum* biovar *viciae* (sym-)	*Eco*RI.
LANE 4	*R. leguminosarum* biovar *phaseoli*	*Bam*HI.

All three films were exposed for 168 hours to the blots.



A. *flgI* **B. *flgH*** **C. *fliP***

These results demonstrate the presence of DNA sequences homologous to the *A. tumefaciens* flagellar gene homologues in another *Rhizobium* species, *R. leguminosarum*. The hybridisation of the radiolabelled gene fragments to the same bands in the sym^+ and sym^- strains of *R. leguminosarum* biovar *viciae* suggests that these genes are found upon the chromosome and not upon the sym plasmid. Similarly in *A. tumefaciens* the flagellar genes found are known to be upon the chromosome rather than the Ti plasmid [267]. A clustering of the motility genes is also suggested by the same chromosomal bands being hybridised to, for each of the *A. tumefaciens* radiolabelled gene fragments. In *R. leguminosarum* biovar *viciae* the two hybridising bands are likely to be adjacent since both are hybridised to by the same *A. tumefaciens* probe (*flgH*). The ~2.6kb band also contains DNA sequence homologous to *flgI* and the ~3.6kb band has DNA sequence homologous to *fliP*. The ~3.65kb *Bam*HI fragment of *R. leguminosarum* biovar *phaseoli* was shown to have DNA sequence homologous to *flgH* and *fliP* of *A. tumefaciens*, suggesting that these too are found close to one another in this species.

Hybridisation was not detected to any of the other genomic DNA. Further attempts were made to obtain hybridisation by lowering the stringency of the hybridisation conditions. This was done by lowering the temperature at which hybridisation and washes were carried out to 50°C. However no further hybridising bands were detected (data not shown).

4.10 Creation of MAN1, a *fliP*- strain of *A. tumefaciens* due to the insertion of a neomycin resistance cassette within the gene

This mutation was created in an attempt to clarify the phenotype of the *mot-12* mutant. The Tn5 insertion site of this mutant is within ORFC, the open reading frame immediately upstream of *fliP*. Its phenotype is of non-motile cells possessing flagella. If the flagellar gene homologues and unidentified open reading frames are found in a putative operon, it would be expected that polar effects, due to the insertion on Tn5, would also prevent transcription of *fliP*. In *E. coli*, *B. subtilis* and most significantly *R. meliloti* (in which the FliP protein is highly similar to that of *A. tumefaciens*) a mutation in *fliP* resulted in a non-flagellated strain [33, 87, 184]. The putative difference in phenotypes was unexpected and suggested a possible difference in the motility system of *A. tumefaciens*.

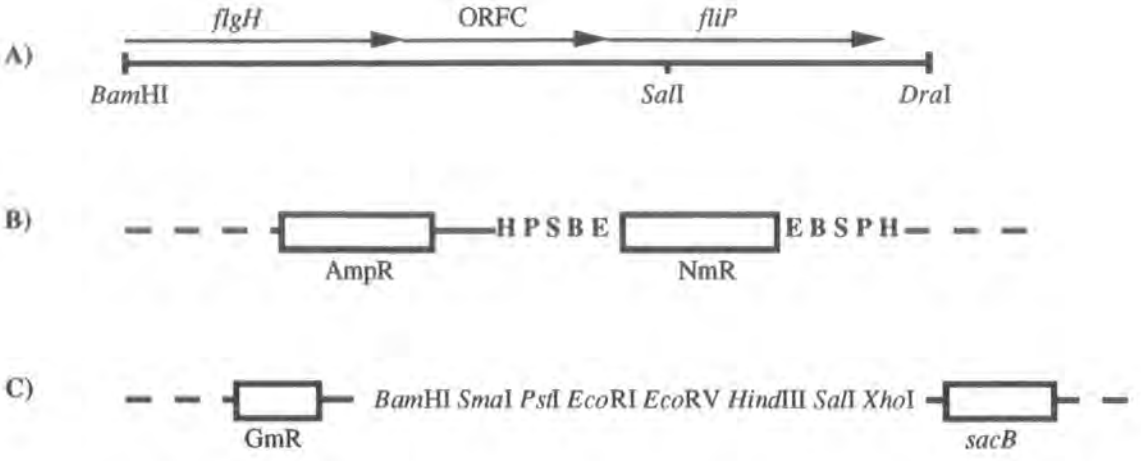
The gene replacement mutagenesis was performed according to the protocol, and using the plasmids, of J. Quandt and M. F. Hynes [240]. The experimental procedures and basic theory are explained in section 2.14. Diagrams of the relevant portions of the plasmids used and created are shown in figure 4.10.1. The subcloning steps performed are described below.

The first step in the subcloning procedure was to remove the *Sal*I site from the multiple cloning site of pJQ200SK. To do this pJQ200SK was double digested with *Pst*I and *Xho*I. The digestion products were electrophoresed, the ~5kb DNA fragment of pJQ200SK (minus the *Pst*I-*Xho*I fragment) was excised and the DNA isolated. The fragment was made "blunt ended" using the Klenow fragment (see section 2.11.9) and the DNA religated to create the plasmid pUTD26, as shown in figure 4.10.1(D).

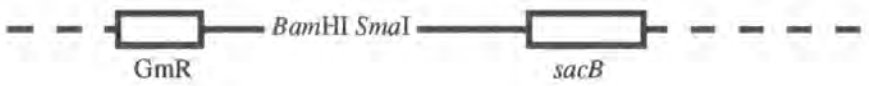
The second step was to clone the *Bam*HI-*Dra*I fragment of pWJD4 into pUTD26. Digestion by *Dra*I creates a blunt end and hence the fragment could be cloned into pUTD26 using the *Bam*HI and *Sma*I (which also generates a blunt end) sites shown. The presence of β–galactosidase within pJQ200SK (and hence pUTD26) allowed blue/white selection for the insertion. The resultant plasmid, pUTD27, is shown in figure 4.10.1(E). The sites used for subcloning are labelled; B *Bam*HI and D/S *Dra*I/*Sma*I composite. The positions of *flgH* (H), ORFC (C) and *fliP* (P) are also shown.

The third step was to disrupt the *fliP* gene by inserting the neomycin resistance cassette within it. The cassette was excised from pDUB1100 with *Sal*I and isolated. It was subcloned into pUTD27 at the *Sal*I site within *fliP* (the only *Sal*I site within this plasmid) to

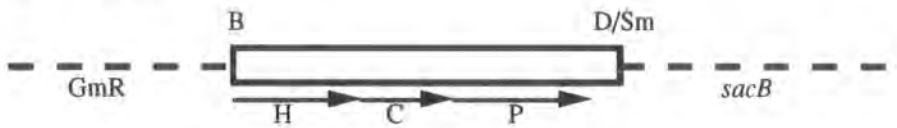
INITIAL PLASMIDS



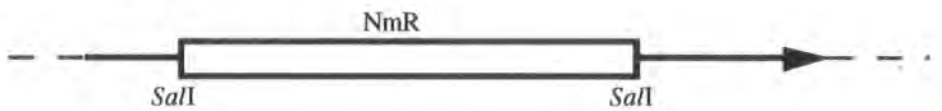
D) STEP ONE - pUTD26



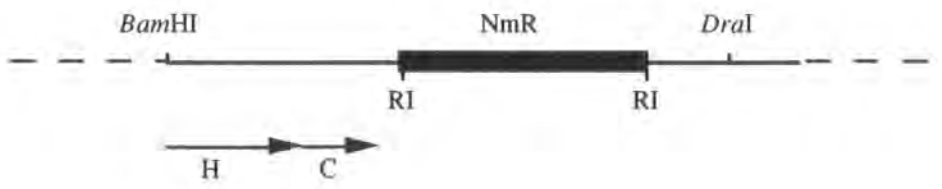
E) STEP TWO - pUTD27



F) STEP THREE - pUTD29



G) END RESULT - MAN1



create pUTD29, which could be selected for due to its double resistance to neomycin and gentamycin. An expanded diagram of the *fliP* gene with the neomycin cassette inserted is shown in figure 4.10.1(F).

Finally pUTD29 was conjugated into *A. tumefaciens*. Strains which had undergone double recombination events could be selected for due to their ability to grow on sucrose and neomycin plates, see section 2.14. These strains should have exchanged the *Bam*HI-*Dra*I fragment containing *flgH*, ORFC and *fliP* for the mutated copy containing the neomycin resistance cassette on the chromosome. Figure 4.10.1(G) shows the altered chromosome of MAN1. By excising the neomycin resistance cassette from pDUB1100 with *Sal*I, the two *Eco*RI sites flanking the neomycin cassette were retained, thus within the chromosome of the mutated strain should be two new *Eco*RI sites, which are also shown on figure 4.10.1(G).

Eight colonies which grew on the neomycin, rifampicin and sucrose plates were grown up in culture and chromosomal DNA isolated from them. Rifampicin was also included as a selectable marker for C58C1, the *A. tumefaciens* strain used for mutagenesis. The chromosomal DNA was digested with *Eco*RI and electrophoresed alongside a sample of wild-type chromosomal DNA, also *Eco*RI digested. The DNA was transferred to a nylon filter by Southern blotting, and the *fliP* specific probe (as described in section 4.4) used in a hybridisation reaction. Figures 4.10.2(A) and (B) show photographs of the initial agarose gel and the resultant radiolabelled bands on the blot, respectively. The probe hybridised to a band of ~13kb in the wild-type/*Eco*RI digest. This band is presumably *Eco*RI fragment A of pDUB1900 to which the *fliP* probe is expected, and has been shown, to hybridise to. The *Eco*RI fragments hybridised to in the putative mutant genomic DNA digests were between ~8.5-9.5kb. The slight variation between the sizes of the hybridised bands (for the mutant genomic DNA digests) is probably because of the variation in the amount of DNA loaded upon the gel, as the photograph in figure 4.10.2(A) shows. The truncation of *Eco*RI fragment A in the mutant digests was expected because of the creation of the extra *Eco*RI sites by the insertion of the neomycin resistance cassette, see figure 4.10.1(G). Figure 4.10.3 shows the relevant *Eco*RI fragments generated and hybridised to. The bands hybridised to in the mutant genomic digests are approximately the same sizes as that predicted to have been created, and hence it appeared likely that the eight colonies picked all contained the neomycin resistance cassette within the *fliP* gene, *i.e.* all were *fliP* mutants.

Figure 4.10.2(A) Agarose gel electrophoresis of the genomic DNA of eight putative *fliP* Agrobacteria mutants digested with *Eco*RI.

LANES A *Pst*I digested λ-DNA.
LANE 1 MAN1 *Eco*RI.
LANE 2 MAN2 *Eco*RI.
LANE 3 MAN3 *Eco*RI.
LANE 4 MAN4 *Eco*RI.
LANE 5 MAN5 *Eco*RI.
LANE 6 MAN6 *Eco*RI.
LANE 7 MAN7 *Eco*RI.
LANE 8 MAN8 *Eco*RI.
LANE 9 *A. tumefaciens* C58C1 genomic DNA *Eco*RI.
LANE B *Hind*III digested λ-DNA.

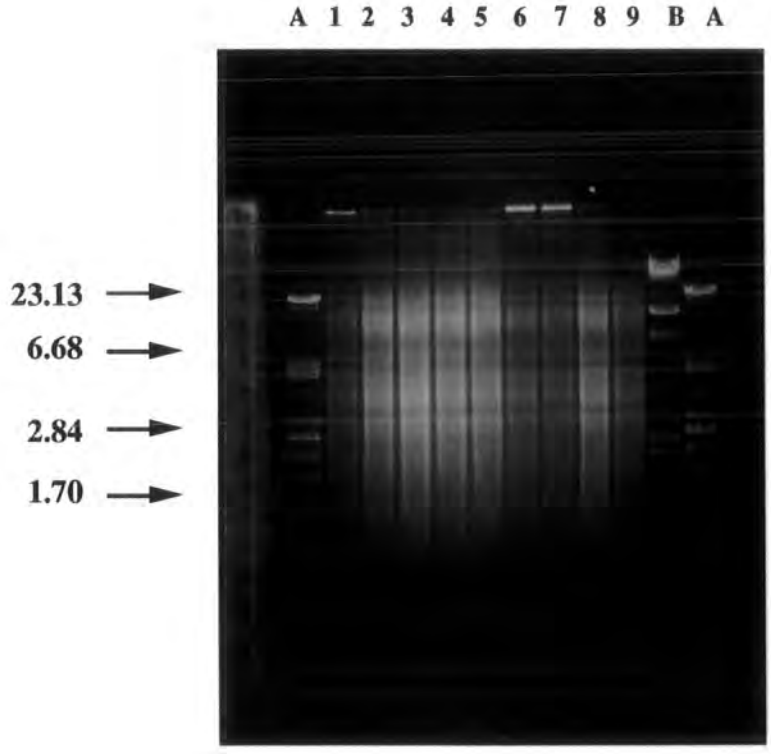


Figure 4.10.2(B) Southern blot of the putative *fliP* mutants probed with a *fliP* specific probe.

The lane labelling is as figure 4.10.2(A). The film was exposed to the blot for 168 hours at -80°C. As discussed in the text the slight variation in the sizes of the hybridising bands in lanes 1-8 is probably explained by variations in the amount of genomic DNA loaded in the agarose gel of figure 4.10.2(A).

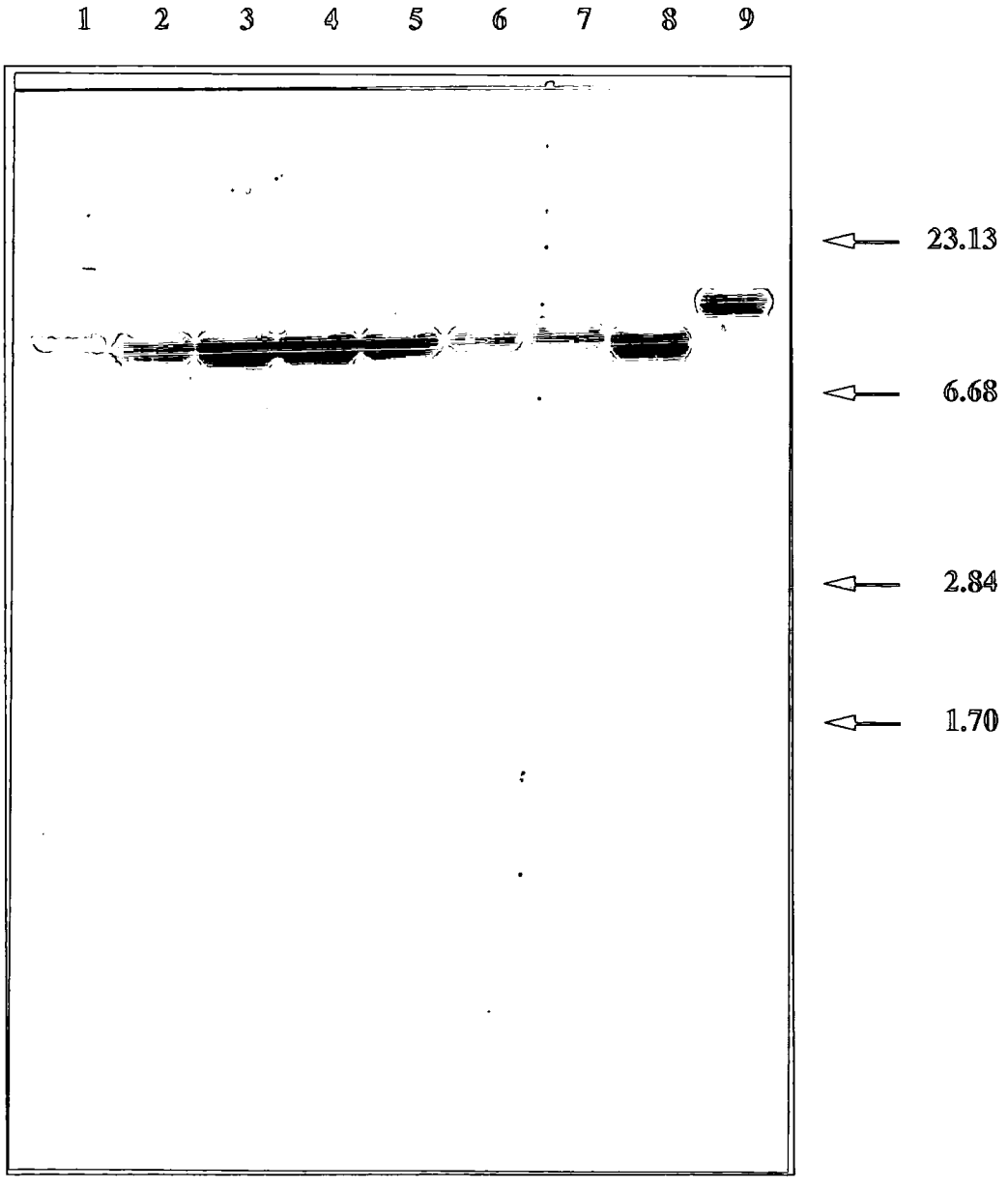
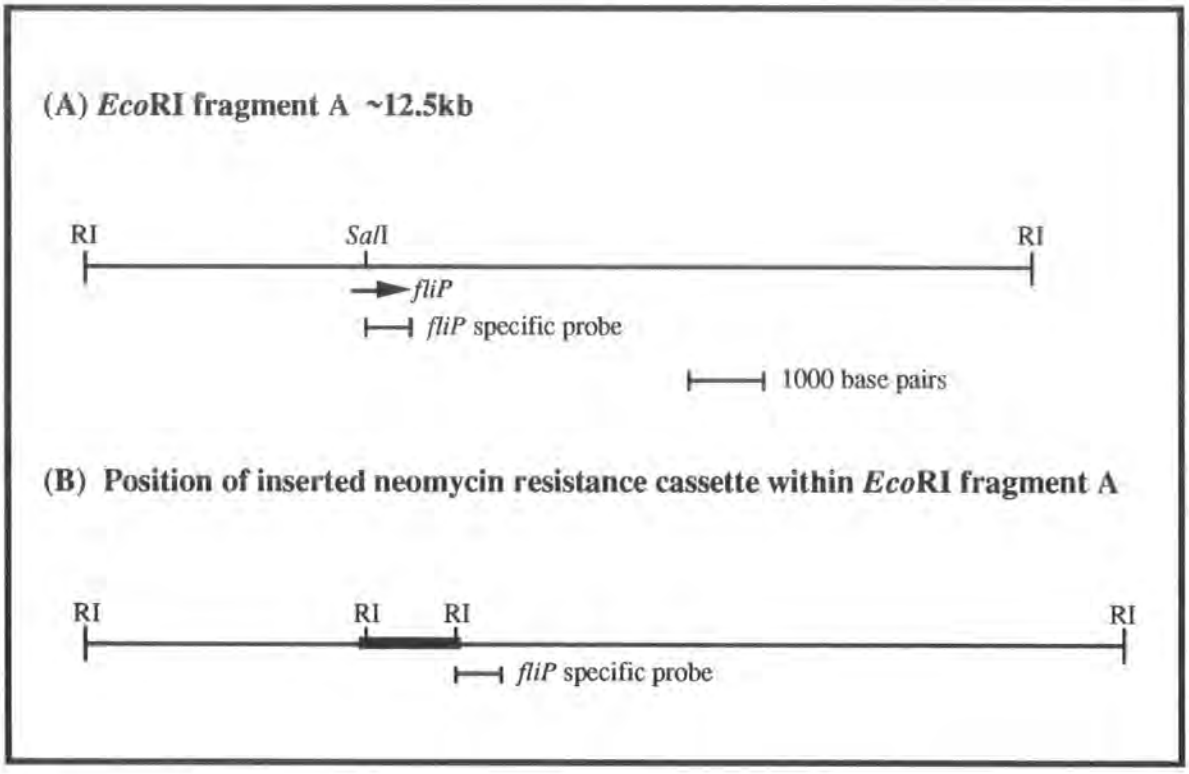


Figure 4.10.3 The different *Eco*RI fragments hybridised to by the radiolabelled DNA fragment of *fliP* in the wild-type and mutant genomic DNA digests.

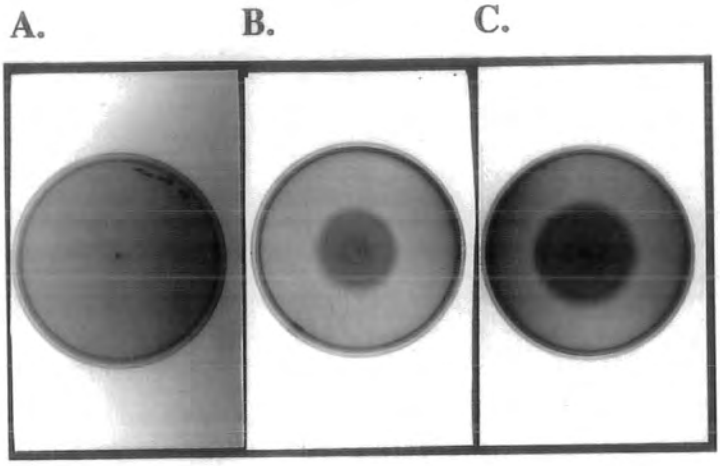
Diagram (A) shows *Eco*RI fragment A, the position of the *fliP* gene and the *Sal*I site within the gene used to mutate it. The radiolabelled DNA fragment specific for *fliP* was also created using the *Sal*I site (see section 4.4). The *Sal*I site is 3703bp from the 5' *Eco*RI site. In each diagram *Eco*RI sites are labelled "RI".

Diagram (B) shows *Eco*RI fragment A with the inserted neomycin resistance cassette and the extra *Eco*RI sites created by the insertion. The diagram shows that the *fliP* specific probe only hybridises to the largest *Eco*RI fragment of ~8.8kb.



The phenotypes of the eight potential mutants (called MAN1-8 according to their original lane positions on the agarose gel) were now analysed. Visualisation by light microscopy showed all eight to be non-motile, as did swarm plate analysis. Figure 4.10.4(A) shows a photograph of MAN1 on a swarm agar plate. The eight mutants were now observed by electron microscopy to determine if the cells had flagella. No flagella were observed. Figure 4.10.4(D) shows an electron micrograph of MAN1 and the absence of any flagella. Finally pDUB1900 was conjugated into MAN1 to try and complement the motility defect. Motility was restored as observed by light microscopy and by swarm plate analysis, see figure 4.10.4(B).

Thus a mutation in the *A. tumefaciens fliP* gene did result in a non-flagellated cell as has been observed in many other bacteria. This phenotype (in *A. tumefaciens*) was unexpected, since a putatively polar mutation in the open reading frame upstream of *fliP* results in flagellated but non-motile cells. This phenotypic anomaly will be discussed in the next section. Time limitations prevented further analyses of the MAN1 strain from being performed. An attempt to complement MAN1 by the *R. meliloti fliP* gene would show whether the two proteins have functional homology as well as high sequence similarity. The complementation analysis could be extended to use the *spa24* and ORF2 genes of *S. flexneri* and *X. campestris* pv. glycines, respectively, the gene products of which are potentially also functionally homologous to FliP. More detailed electron microscopic analyses of the structures (if any) present in the flagellar basal body of MAN1 could also be carried out.


A.
B.
C.


D.
1μm

4.11 Discussion

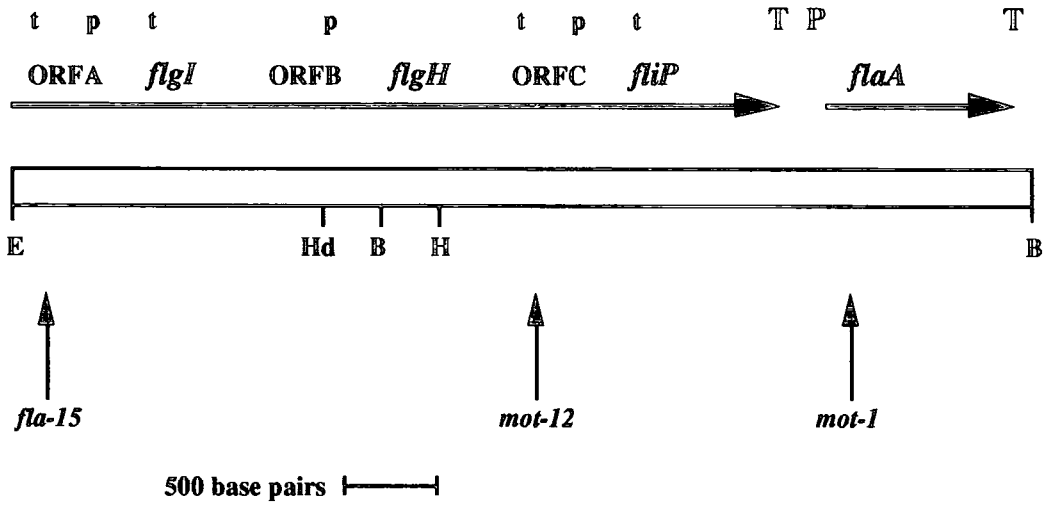
The data presented describes the identification and sequencing of several flagellar gene homologues and three open reading frames encoding potential proteins without significant sequence similarity to any other proteins in the Owl database. These flagellar gene homologues and unidentified open reading frames were initially postulated to be transcribed together as an operon, although as discussed earlier in this chapter, several possible transcription signals were identified within the "operon". A diagram showing the approximate positions of the possible transcription signals found is shown in figure 4.11.1, and their possible functions are discussed later in this section.

The gene products of the *A. tumefaciens flgI, flgH* and *fliP* homologues had significant sequence identity to the corresponding proteins of several other bacteria in the Owl database. Southern blotting using the three flagellar gene homologues as probes revealed homologous counterparts in three *Rhizobium* strains, *R. meliloti, R. leguminosarum* biovar *viciae* and *R. leguminosarum* biovar *phaseoli.* Furthermore radiolabelled DNA fragments of ORFs A, B and C hybridised to the same or adjacent DNA fragments of *R. meliloti* suggesting a clustering of the genes/ORFs in *R. meliloti* as in *A. tumefaciens*. This cluster was found in a similar position, upstream of the flagellin genes in *R. meliloti* as it is in *A. tumefaciens*. DNA sequence from *R. meliloti* revealed that three of the genes/ORFs are also found in the same order in *R. meliloti* (*flgH*, ORFC and *fliP*). However, functional homology was not demonstrated between the two species, since cosmids containing the *R. meliloti* counterparts did not complement the *A. tumefaciens* behavioural mutants *mot-12* and *fla-15*.

Unlike the flagellar gene homologues, functional roles for the gene products of ORFs A, B and C cannot be inferred as a result of homologies to previously identified proteins. Their position within a putative flagellar gene operon suggests they too may be involved in flagellar assembly. The vast majority of the previously identified flagellar proteins are all structural components of the flagellum. Given the conservation between the basic flagellum structure in a large number of bacteria it is thus unlikely that the gene products of ORFs A, B and C are structural components, unless *A. tumefaciens* has extra structural components, such as the E ring of *C. crescentus* [67], or the large basal disk of *Wolinella succinogenes* [81]. It is hence more likely that the gene products of ORFs A, B and C have roles in regulating flagellar gene expression or in the export and assembly of flagellar components. There is some experimental evidence that four putative flagellar genes in *S. typhimurium,* whose gene products have no significant sequence identity to any other previously sequenced proteins, may have such roles. *fliS* and *fliT* are found

Figure 4.11.1 Diagrammatic representation of the putative flagellar gene operon and the potential transcription signals within it.

The DNA region sequenced containing the putative flagellar gene operon and *flaA* is shown as a box. Below this are the relative positions of the Tn5 insertion sites of the three mutants, as well as selected restriction enzyme sites, B=*Bam*HI, E=*Eco*RI, H=*Hpa*I and Hd=*Hind*III.



The diagram shows the approximate positions of the flagellar gene homologues and the unidentified open reading frames within the putative operon, and the downstream flagellin *flaA*. Also shown (in lower case) above the putative operon are the potential transcription signals identified within this chapter (where "p" is a possible promoter and "t" a possible transcription termination signal). Their positions are approximate, but within the open reading frame they are labelled above. Shown in upper case downstream of the *fliP* homologue, as well as upstream and downstream of *flaA* are transcription signals which are thought to function as labelled. The termination signal downstream of *fliP*, although not shown experimentally to function as such, presumably does so as the flagellin *flaA* is known to be transcribed individually (see section 5.4).

downstream of (and are probably transcribed with) *fliD*, which encodes the filament capping protein. *fliS* and *fliT* (like ORFs A, B and C in *A. tumefaciens*) have a high guanine and cytosine content, are predicted by TESTCODE to be coding regions, are both necessary for flagellar assembly and the proteins they encode have no homologies to previously sequenced proteins. The postulated functions of FliS and FliT are regulatory, since no proteins with their molecular masses are found in the external portion of the flagellum and the *fliD* operon to which they belong is known to have a regulatory effect on flagellar synthesis. The regulatory effect is a repressor-like activity (RflA) on the late flagellar operons. Since *fliD* is a structural gene (at the very external tip of the flagellum) it was unclear how it could have such a role, but if FliS and FliT were to have this regulatory role, both it and their functions could be explained [139, 241]. The gene products of *fliU* and *fliV* in *S. typhimurium* also have no homology to any previously sequenced proteins in the databases. Their functions are also unknown, but it is thought they could be involved in the secretion, or regulation of the secretion, of the external components of the flagellum in *S. typhimurium*. If either *fliU* or *fliV* are inactivated there is a build up of the flagellin proteins in the cytoplasm and overexpression of FliU and FliV results in exceptionally long filaments [71, 72]. Open reading frames encoding proteins with unknown functions and no homologies to previously sequenced proteins have also been found within a flagellar operon of *B. subtilis* [7]. The *flaA* locus of *B. subtilis* contains two such open reading frames, Orf6 and Orf8. As discussed in section 4.6 the gene product of Orf6 has some sequence identity to the putative protein encoded by ORFB of *A. tumefaciens*. *E. carotovora* subspecies *atroseptica* also has an open reading frame (*mopB*) encoding a gene product of unknown function, with no significant sequence identity to any other proteins amongst a number of flagellar gene homologues. *mopB* is found upstream of the *fliP* homologue (*mopC*) in a position analogous to that of ORFC. Although there is no significant sequence identity between the gene products of these unidentified open reading frames, they are all of a predicted similar size; FliS 14,670Da, FliT 13,689Da, FliU ~19kDa, FliV ~20kDa, MopB 15,024Da, Orf6 22,946Da, Orf8 15,643Da, ORFA 16,979Da, ORFB 19,862Da and ORFC 17,834Da. However for ORFs A, B and C of *A. tumefaciens*, it should first be shown that they are expressed before further analyses are performed to determine their functions.

The position of the flagellar gene homologues and ORFs A, B and C within *A. tumefaciens* initially suggested that they might be transcribed together as an operon. Attempts to show this by Northern blotting using radiolabelled DNA fragments of various parts of the putative operon were unsuccessful. In every case no hybridisation was observed. The most probable reason for this was that the transcripts were below the detection limit. In the enteric bacteria it is known that there are only c. 26 copies of each of the FlgI and FlgH proteins per flagellum [277]. Given the conservation shown between the FlgI and FlgH proteins of *A. tumefaciens* and their enteric counterparts, and that presumably

similar structural constraints are placed upon the *A. tumefaciens* homologues within the flagellum. It is probable that a similar subunit stoichiometry exists in *A. tumefaciens*, *i.e.* c. 26 copies of the FlgI and FlgH homologues are used in each flagellum. The maximum number of flagella per *A. tumefaciens* cell is six [267], and therefore only (relatively) small amounts of these proteins will be required so the level of transcription will be very low and hence the probability of hybridisation low. If *fliP* and ORFs A, B and C are transcribed with *flgI* and *flgH*, the level of their transcription will also be low. No small bands were identified on Northern blots by probes specific for *fliP* and ORFs A, B and C, which would be consistent with separate transcription. The identification of possible promoter sequences for *flgI*, *flgH* and *fliP* and putative stem-loop sequences which may form transcription termination signals, indicated that these genes may not be in an operon and ORFs A, B and C are merely intergenic regions. Arguments for and against the presence of the operon are discussed below. However the possible function of the promoter sequences could be investigated by primer extension assays or S1 nuclease protection which would also show whether the flagellar gene homologues are transcribed individually.

Despite the presence of the possible transcription signals within the putative operon, there is still good evidence that ORFs A, B and C are coding regions. The UWGCG program TESTCODE predicts them (at the 95% confidence level) to be coding regions and their guanine and cytosine content is significantly higher than DNA regions predicted to be non-coding. Sequence obtained upstream of *fliP* in *R. meliloti* revealed at least the carboxy-terminus of an ORFC homologue with high sequence identity to the *A. tumefaciens* copy. Radiolabelled DNA fragments of *A. tumefaciens* ORFs A, B and C hybridised to homologous counterparts within *R. meliloti*, indicating all three are conserved in *R. meliloti*. It is unlikely that non-coding DNA regions would be so highly conserved between the two species. Assuming ORFs A, B and C are transcribed, the question of how this transcription is brought about still remains. No promoter-like sequences were found for the three open reading frames, which suggests their transcription may be coupled to the gene or genes upstream. The close proximity of the adjacent genes/open reading frames in the putative operon (most overlap) gives some indication that this may be so.

If the transcription signals do function as such, it is possible that *flgI*, *flgH* and *fliP* could all be transcribed individually and ORFs A, B and C are merely regulatory regions. Transcription signals have been found within flagellar gene operons of *S. typhimurium*, *E. coli* and *C. crescentus*. The *flgB* operon of *S. typhimurium* has a 54 base pair intergenic region between the *flgG* and *flgH* genes which contains a *rho*-independent transcription termination-like sequence, but there is no obvious promoter within this region [135, 179]. Its presence could be necessary because of the different structural properties of the genes either side of it. Upstream are genes encoding the hook and rod proteins which require the

earlier components of the flagellum to be assembled before they can be added. Downstream are the genes encoding the L- and P rings, exported by the *sec*-dependent secretory pathway, which can be synthesised and exported to the outer membrane and periplasm (respectively) regardless of the state of the partially assembled flagellum. The differences in export pathways and structural locations of the gene products of this operon probably require variations in regulation, which may involve the transcription termination signal [135, 179]. Between the *fliN* and *fliO* genes, within the *fliL* operon of *E. coli*, is a 62 base pair intergenic region containing a potential stem-loop sequence. Apart from this the position of the genes within the operon is compact [184]. The authors postulate that differential expression may occur within the operon partly as a result of this stem-loop sequence. Examples of this differential expression are not unusual within operons containing one or more genes encoding integral membrane proteins, such as this operon and the putative *A. tumefaciens* flagellar operon. The *flgH* gene of *C. crescentus* can be transcribed individually or as part of the *flgF, flgG, flaD* and *flgH* operon. Transcription is initiated from an internal promoter within the upstream gene (*flaD*) of the operon as well as the promoter upstream of *flgF*. There are no transcription termination signals apparent between the *flaD* and *flgH* genes. This overlapping transcription has been identified in other *C. crescentus* flagellar genes [67].

Since the various flagellar gene products must interact to form a complex structure it is perhaps unsurprising to find evidence of higher levels of gene regulation. Any individual protein involved in flagellar structure and assembly may be subject to temporal and spatial regulation, as well as regulation to ensure the relative stoichiometries of the various proteins are correct. The transcription signals described for *E. coli, S. typhimurium* and *C. crescentus* (as well as *A. tumefaciens*) may be present to bring about this regulation. The fact that these transcription signals are found between the same genes in other species as in *A. tumefaciens* (for example *flgH*) is suggestive of the need for more complex regulation of their expression. Putative transcription signals which may be involved in more complex levels of regulation have also been found amongst operons in *E. coli*. The *rplKAJLrpoBC* gene cluster of *E. coli* which encodes ribosomal proteins and subunits of RNA polymerase contains at least four different promoters and a transcriptional attenuator [242]. There are also overlapping transcriptional units, such as that seen for the *flgH* gene in *C. crescentus*, amongst a cluster of *E. coli* cell division genes. The *ddl, ftsQ, ftsA, ftsZ* and *envA* genes all have at least one promoter upstream, and the promoter sequences for the *ftsQ, A* and *Z* genes are all within the upstream coding region. Transcription could often be initiated from more than one promoter, for example *ftsZ* transcription could either be initiated from a promoter within the *ftsA* gene immediately upstream or a promoter within the *ddl* gene, two genes further upstream [246, 247].

The observed phenotypes of the mutants caused by insertions within ORFC and *fliP* do give some experimental evidence that *fliP* can probably be transcribed individually. The Tn5 insertion within ORFC in the non-motile (but flagellated) mutant *mot-12* would be expected to exert polar effects and prevent the transcription of the downstream gene, *fliP*, if both are found in a putative operon. However as the *fliP* mutant has a non-flagellated phenotype, the Tn5 insertion in *mot-12* is presumably not preventing transcription of *fliP* or its phenotype would be non-flagellated too. Thus in *mot-12 fliP* is presumably transcribed from its proximal promoter (as described in section 4.4) and the non-motile phenotype is probably a result of the disruption of ORFC or an interference in the transcription termination signal of *flgH*. That *fliP* can be transcribed individually is not surprising because of the location of *fliP* within the putative operon containing structural genes required fairly late in the process of flagellar assembly. In *E. coli* and *S. typhimurium* the gene product of *fliP* is known to be required at a very early stage of flagellar assembly, and *fliP* is found within an operon containing genes required at a similar stage. It is unlikely that *fliP* in *A. tumefaciens* (if FliP has a similar function) would thus be solely under the transcriptional control of an operon containing genes whose products are required at a much later stage of flagellar assembly. However the possibility that ORFC is also transcribed cannot be discounted, for the reasons described earlier. It may be that *fliP* can be transcribed under the control of an upstream promoter (that also brings about transcription of ORFC) as well as a proximal promoter within ORFC. This arrangement would be similar to that found for *flgH* of *C. crescentus* described earlier.

At this stage in the project it is not possible to state categorically how the flagellar gene homologues are transcribed and regulated, and even whether the gene products of ORFs A, B and C are expressed. The three flagellar gene homologues completely sequenced may all be transcribed individually or may be part of an operon subject to complex regulation. Further experimental work will be necessary to resolve how and if the flagellar gene homologues and ORFs A, B and C are transcribed. Initially protein expression studies should be performed to see whether ORFs A, B and C are transcribed. If it is found that they are not, it would be far more likely that the flagellar gene homologues are transcribed individually and this could be shown by primer extension assays. If expression of ORFs A, B and C is shown, the existence of the putative operon would be far more likely albeit possibly under extra levels of regulation. The DNA sequence upstream of the partial *flgG* homologue could now be sequenced to determine the 5' boundary of the operon as well as any other putative transcription signals within it. To resolve which of the putative promoter sequences do function, primer extension assays should be performed. Site-directed mutagenesis could also be performed on possible transcription signals (without altering the coding regions) to try and elucidate their functions in the regulation. Alternatively the reporter transposons Tn5*lacZ* or Tn5*luc* [273], could be used to mutate the

operon. These would also allow an analysis of gene regulation as well as demonstrating possible roles in flagellar assembly and structure for the mutated gene. Polar effects from the inserted transposons could be partially alleviated by using Tn5-*P*-out derivatives [273], which contain a promoter reading out from the transposon, thus allowing constitutive expression of downstream genes. These would be particularly useful within ORFs A, B and C to determine whether their gene products are essential for flagellar structure and/or assembly.

5. The *A. tumefaciens* flagellin genes

5.1 The *mot-1* mutant

The phenotype of the *mot-1* mutant was non-motile upon swarm agar plates and visualisation by electron microscopy revealed truncated flagellar filaments (see Introduction). These truncated filaments were generally straighter than the sinusoidally curved filaments of wild-type *A. tumefaciens* flagella. In *E. coli,* mutants with truncated, straight filaments were found to be a result of single base mutations within the gene *fliC*, which encodes the major structural protein (flagellin) of the filament itself [36]. A null mutant of the major flagellin, *flaA*, of *R. meliloti* also has a phenotype of truncated filaments. The mutant was motile, although it tumbled more frequently than the wild-type. However, this mutant was non-motile when grown in media containing elevated concentrations of salts. Under these conditions the truncated filaments were found to be straight [27]. The filaments of the *A. tumefaciens* mutant *mot-1* were also presumed to rotate, since the *mot-1* cells were observed by light microscopy to rotate when attached to glass coverslips by their flagella. Selection for motile *mot-1* cells by repetitive subculturing from the edges of a culture on a swarm agar plate, see section 2.5, resulted in the isolation of cells capable of forming a partial swarm. Figure 5.1.1 shows a comparison between the initial *mot-1* cells and the motile cells selected for after five rounds of subculturing. There is a clear size difference between the swarms. When observed by light microscopy these motile *mot-1* cells were more tumbly than the wild-type control. This behavioural similarity between the *A. tumefaciens mot-1* cells and the *R. meliloti flaA* null mutant was partially explained upon sequencing the DNA adjacent to the *mot-1* Tn5 insertion. The *mot-1* insertion was found to be close to the 5' end of the *A. tumefaciens flaA* gene.

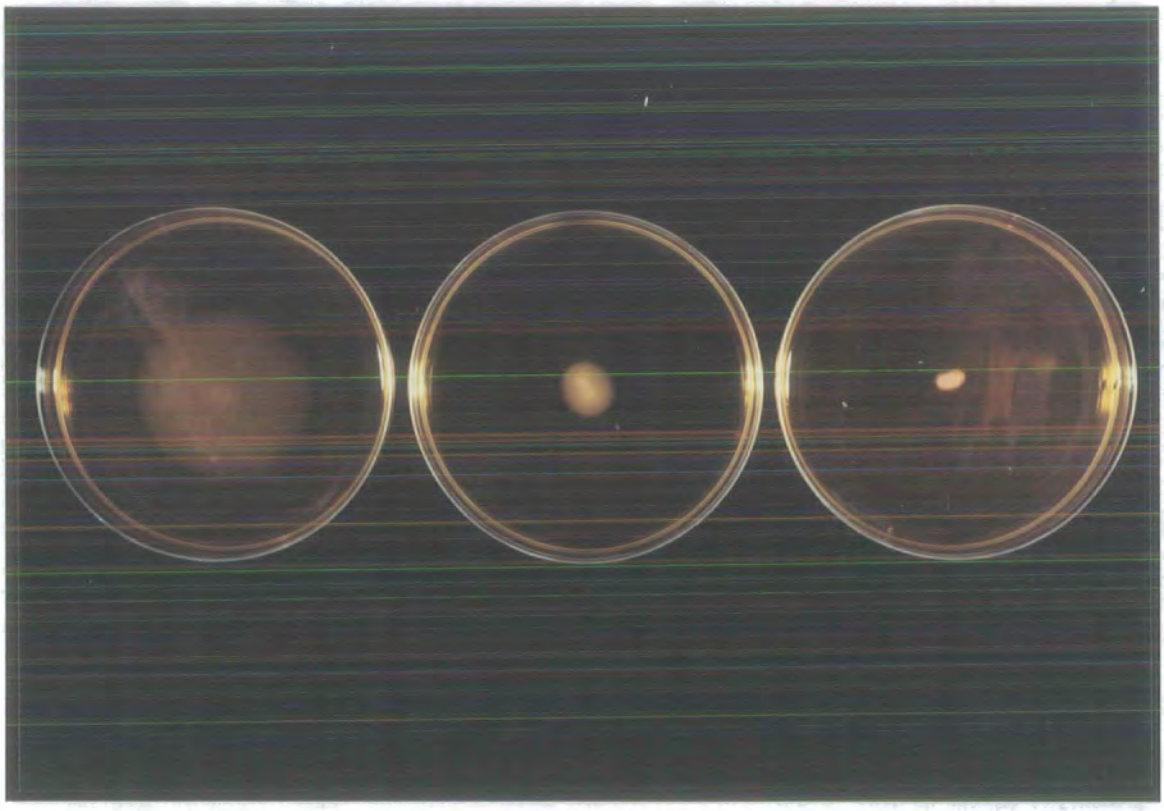
5.2 Sequencing of the *A. tumefaciens* flagellin genes

The Tn5 insertion site of the *mot-1* mutant was found to be within the *flaA* gene of *A. tumefaciens*. Two further copies of the flagellin genes (*flaB* and *flaC*) were found downstream of *flaA*. Their relative positions are shown in figure 3.5. As discussed in chapter 4, their location downstream of *fliP* is similar to that found for the *R. meliloti* flagellins *flaA* and *flaB*, to which the *A. tumefaciens* flagellin genes are similarly arranged [27, 87, 235]. A third copy of the flagellin gene has now been found in *R. meliloti*, *flaC* [258], this is postulated to be a redundant copy as a result of functional analysis experiments performed on the *flaA* and *flaB* genes of *R. meliloti* [236]. A fragment of the *A. tumefaciens flaA* coding region was excised using the restriction enzymes *Dra*I and *Bam*HI (see figure 5.2.3) radiolabelled and used as a probe for a Southern blot containing *A. tumefaciens* genomic DNA and pDUB1900 cut with *Bam*HI, *Eco*RI and *Hind*III. The *flaA* probe cross-

Figure 5.1.1 Phenotypic comparisons by swarm plate analysis of the *mot-1* mutant after selection for "motile" cells.

A. C58C1 wild-type control after 48 hours growth at 28°C.
B. *mot-1* swarm after five rounds of subculturing on swarm plates and 48 hours growth at 28°C.
C. *mot-1* swarm without any subculturing rounds after 48 hours growth at 28°C.

The subculturing protocol for the selection of "motile" *mot-1* cells shown on plate B is described in section 2.5.



A. **B.** **C.**

hybridised with the *flaB* and *flaC* DNA sequences. Only bands of predicted sizes from the restriction map of pDUB1900 were hybridised to, and no further copies of the flagellin genes were identified downstream of *flaC* (data not shown). The program TESTCODE predicts all three of the *A. tumefaciens* flagellins to be coding regions (at the 95% confidence level), see figure 3.4.

The sizes of the *A. tumefaciens* flagellin gene open reading frames are; *flaA* 921bp, *flaB* 963bp and *flaC* 942bp. These encode predicted proteins of 306 amino acids (predicted M_r 31,637Da), 320 amino acids (predicted M_r 32,966Da) and 313 amino acids (predicted M_r 32,843Da), respectively. The proteins were highly similar to one another. A multiple alignment of all three, produced using the program CLUSTAL V [119], is shown in figure 5.2.1. Gap alignments between the putative protein sequences revealed FlaA and FlaB to have 64% identity (78% similarity), FlaA and FlaC to have 60% identity (73% similarity) and FlaB and FlaC to have 66% identity (78% similarity). The regions of greatest divergence between the *A. tumefaciens* flagellins were within (approximately) the third quarter of the proteins. The flagellins of *R. meliloti* are larger than those of *A. tumefaciens* and are more highly conserved. In *R. meliloti* 10406 *flaA* encodes a protein of 395 amino acids (predicted M_r 40,459Da) and *flaB* one of 396 amino acids (predicted M_r 41,044Da), which have 87% sequence identity [235]. In *R. meliloti* 1021 *flaA* encodes a protein of 394 amino acids (predicted M_r 40,693Da) and *flaB*, 394 amino acids also (predicted M_r 40,717Da) which have 93% sequence identity [27].

The three *A. tumefaciens* flagellins, when compared against the Owl database, had significant sequence identity to a number of other flagellin proteins from a wide range of bacteria. The greatest sequence identities were to the flagellins of *R. meliloti*, with the *A. tumefaciens* FlaA protein having 67% sequence identity and 80% sequence similarity to FlaA of *R. meliloti*. The values for the other flagellins of *A. tumefaciens* and *R. meliloti* were approximately the same. For comparison the *A. tumefaciens* FlaA protein has 27% sequence identity and 48% sequence similarity to the FliC protein of *E. coli*. The flagellins of the following bacteria were identified by the FlaA protein of *A. tumefaciens*, the figure in brackets after the name is the sequence identity to FlaA: *B. subtilis* (23%), *C. crescentus* 28.5kD flagellin (39%), *Campylobacter coli* (22%), *Roseburia cecicola* (20%), *S. typhimurium* (20%), *Borrelia burgdorferi* (25%), *Pseudomonas aeruginosa* (17%). A multiple alignment of the *A. tumefaciens* FlaA protein to some of these flagellins is shown in figure 5.2.2. As figure 5.2.2 shows all the flagellin proteins have conserved amino- and carboxy-termini. For the FlaA proteins of *R. meliloti* and *A. tumefaciens* the sequence identity in these regions is between 70 and 90% and between the *A. tumefaciens* FlaA and FliC of *E. coli* the sequence identity is between 40 and 50%. The region between these highly conserved termini exhibits the greatest variability in amino acid sequence and length.

Figure 5.2.1 Multiple alignment of the *A. tumefaciens* flagellins, FlaA, FlaB and FlaC.

Identities are indicated by asterisks, and conservative substitutions are denoted by dots. The alignment was performed using the SEQNET program CLUSTAL V [119].

```
FlaA  MASILTNNNAMAALSTLRSIASDLSTTQDRISSGLKVGSASDNAAYWSIATTMRSDNKAL
FlaB  MTSIITNVAAMSALQTLRSIGQNMESTQARVSSGLRVGDASDNAAYWSIATTMRSDNMAL
FlaC  MTSILTNTAAMSALQTLRAISGQLEDTQSRVSSGLRVKSASDNAAYWSIATTMRSDNMAL
      *.**.**  **.** ***.*. ..  ** *.****.*  ****************** **

FlaA  GAVSDALGMGAAKVDTASAGMDAAIKVVTDIKAKVVAAKEQGVDKTKVQEEVSQLLDQLK
FlaB  SSVSDALGLGAAKVDTASAGMSSAIDVVKEIKAKLVTATEEGVDRTKVQEEIGQLQKQLA
FlaC  SAVQDALGLGAAKVDVAYSAMESTVEVVKEIKSKIVAATEEGVDKTKIQEEIDQLKKQLE
      ..* ****.****** * ..* ... ** .**.*.*.* *.***.**.***. **  **

FlaA  SIGTSASFNGENWLVS---------SANATKTVVSGFVRDAGGTVSVKTTDYALD-----
FlaB  SISQGASFYGENWLVGVSTLGAATPGTDPDKSVVAGFVRASGGAVSVTTTKYALDNTATG
FlaC  SIAQGASFSGENWLLGTGA----------KTVVSGFVRDGGGTVSVTKTDYTLIDTTAG
      **. .*** *****..              *.**.**** .**.***  * *.*

FlaA  --ANSMLYTE-GTPGTIDANSGILNATGATTTVGAKTYTQISVLDMNVGTD---DLDNAL
FlaB  ----NVLFGSVDGTGTPDAS-GILGTVGTFTTVAAQSVYTLDITQYAAADRAT-NMAEAL
FlaC  TTTANVLFGLTGSPATLDTTKGIIGQPGTASGI---SVWDIDLKLFNSATPPTYTIGNLL
          ..*.   .  .* *.  **.   *. . .   .   . .     .      . . *

FlaA  YSVETALTKMTSAGAKLGSLSARIDLQSGFADKLSDTIEKGVGRLVDADMNEESTKLKAL
FlaB  TLVENSLKAMTSAAAKLGSLSMRIGLQEDFASKLSDSVEKGIGRLVDADMNEESTRLKAL
FlaC  TDVETAFQSISSASAALGSIKMRIGLQEDFVSKLTDSIDKGIGRLVDADMNEESTKLKAL
        ** ..  ..**.* ***.  **.** .*  **.*...**.*************.****

FlaA  QTQQQLAIQALSIANSDSQNILSLFRX-
FlaB  QTQQQLGVQALSIANSNSESILSLFRX-
FlaC  QTQQQLGIQSLSIANTSSENILSLFRQX
      ******..*.*****. *..******
```

Figure 5.2.2 Multiple alignment of the flagellins of *A. tumefaciens*, *R. meliloti*, *B. subtilis* and the 28.5kDa flagellin of *C. crescentus*.

Identities are indicated by asterisks, and conservative substitutions are denoted by dots. The abbreviations are, *At* = *A. tumefaciens*, Fla3 (*Cc*) = the 28.5kDa flagellin of *C. crescentus*, *Rm* = *R. meliloti* and FliC (*Bs*) = the flagellin of *B. subtilis*. The alignment was performed using the SEQNET program CLUSTAL V [119].

```
Fla3 Cc  ALSVNTNQPALIALQNLNRTNDDMQAVQTRINTGEAISTAKDTAAVWSH----RPGAGDM
FlaA At  MASILTNNNAMAALSTLRSIASDLSTTQDRISSGLKVGSASDNAAYWSIATTMRSDNKAL
FlaA Rm  MTSILTNNSAMAALSTLRSISSSMEDTQSRISSGLRVGSASDNAAYWSIATTMRSDNQAL
FliC Bs  M-RINHNIAALNTLNRLSSNNSASQKNMEKLSSGLRINRAGDDAAGLAISEKMRGQIRGL
            .  *  *. .*  *             ....*  .  * * **  .     *     .

Fla3 Cc  SGLAREDEPGSGDIDRGRGPRAGESVSDLLKLMREKVVAAKDTSLTTTSRQ--ALNADFQ
FlaA At  GAVSDALGMGAAKVDTASAGM--DAAIKVVTDIKAKVVAAKEQGVD---KT--KVQEEVS
FlaA Rm  SAVQDALGLGAAKVDTAYSGM--ESAIEVVKEIKAKLVAATEDGVD---KA--KIQEEIT
FliC Bs  EMASKNSQDGISLIQTAEGALTETHA--ILQRVRELVVQAGNTGTQDKATDLQSIQDEIS
                 * . .. . .         ..  ..  .* * . .          .. .

Fla3 Cc  GLIKNLNQVLRSATFDGANLLDGSQAADM-------SFLADADAGQAITLTLQNLSLGGT
FlaA At  QLLDQLKSIGTSASFNGENWL---VSSANATKTVVSGFVRDAGGTVSVKTTDYALDANSM
FlaA Rm  QLKDQLTSIAEAASFSGENWLQADLSGGPVTKSVVGGFVRDSSGAVSVKKVDYSLNTDTV
FliC Bs  ALTDEIDGISNRTEFNGKKLLDGT-------------YKVDTATPANQKNLVFQIGANAT
           *  ..  .   . * * . *                .  *.           .

Fla3 Cc  INTLTATDDILD------------PVNAAG--------------------------------
FlaA At  LY-TEGTPGTIDANSGI--------LNATGATTTVGAKTY--------------------
FlaA Rm  LFDTTGNTGILDKVYNVSQASVTLPVNVNGTTSEYTVGAYNVDDLIDASATFDGDYANVG
FliC Bs  -----------------QQISV-----------------NIEDM--------------G

Fla3 Cc  ------------------------------------------------------------
FlaA At  ----------------------------------------------------TQISVL
FlaA Rm  AGALAGDYVKVQGSWVKAVDVAATGQEVVYDDGTTKWGVDTTVTGAPATNVAAPASIATI
FliC Bs  ADALG----------IKEADGSIAALHSVNDLDVTKFA----------------------

Fla3 Cc  -------------VLTRLDATLSAVNQAVGNIGTQAKQIDAHNTFVAKLNDVLETGVGNL
FlaA At  DMNVGTD--DLDNALYSVETALTKMTSAGAKLGSLSARIDLQSGFADKLSDTIEKGVGRL
FlaA Rm  DITIAAQAGNLDALIAGVDEALTDMTSAAASLGSISSRIDLQSDFVNKLSDSIDSGVGRL
FliC Bs  DNAADTADIGFDAQLKVVDEAINQVSSQRAKLGAVQNRLEHTINNLSASGENLTAAESRI
                       .  .. .. .    . .*.   ...          . .  . . .

Fla3 Cc  VDADLAKESARLQALQVKQPLGAQALSIANGAPQIILSLFKGG
FlaA At  VDADMNEESTKLKALQTQQQLAIQALSIANSDSQNILSLFRX-
FlaA Rm  VDADMNEESTRLKALQTQQQLAIQALSIANSDSQNVLSLF-R-
FliC Bs  RDVDMAKEMSEFTKNNILSQASQAMLAQANQQPQNVLQLLR--
          * *.  * . .   .     .   *. **  .* .* *.
```

X-ray diffraction analyses of the flagella of straight-filament producing mutants of *S. typhimurium* and biochemical studies using proteolytic digestions of flagellin monomers have allowed the determination of the structure of the bacterial flagellum [64, 206]. Within the flagellum, individual flagellin monomers are folded back upon themselves, so that the amino- and carboxy-termini are found close to each other and define the boundaries of the central channel. These domains are thought to be important for filament formation and interaction with the neighbouring subunits. The highly variable central domains of the flagellin monomers are postulated to be on the filament surface and hence responsible for the antigenic properties of the filament. The flagellar filaments of *S. typhimurium* are composed of only one type of flagellin monomer and are referred to as "plain" filaments. A second type of filament, the "complex" filament, has also been observed in a number of other bacterial species. Such filaments are thought to consist of at least two different flagellin monomers, although the underlying structural interactions are similar, in which the amino- and carboxy-termini are important for assembly and found adjacent to each other within the filament. The monomers of a complex filament are capable of additional bonding interactions which are thought to make it more rigid than the plain filament. These extra bonding interactions also result in a distinct set of ridges along the complex filament which can be observed by electron microscopy [306]. The flagella of *R. meliloti* have been shown to have this ridge-like pattern [101, 155], and are known to be composed of two different types of flagellin monomers [236]. Given the sequence similarity between the flagellins of *A. tumefaciens* and *R. meliloti*, and the phenotypic similarity between mutations in the *flaA* gene of each species, it would seem possible that *A. tumefaciens* also possesses complex flagellar filaments. This will be discussed further in sections 5.5 and 5.6 of this chapter.

a) The *flaA* sequence

The 921 base pair *flaA* open reading frame is shown in figure 5.2.3, with the translated protein sequence below. The guanine and cytosine content is 59%, significantly higher than that of DNA regions thought to be non-coding. The *flaA* open reading frame of *A. tumefaciens* is found 371 base pairs downstream of *fliP*. In *R. meliloti* the intergenic region is ~300 base pairs [87]. Approximately 150 base pairs upstream of the open reading frame is a possible class III flagellar gene promoter.

Consensus class III promoter	TAAA	N_{15}	GCCGATAA.
Putative *A. tumefaciens* class III promoter	TAAT	N_{16}	GCTGCGAA.
R. meliloti class III promoter of *flaA*	TAAA	N_{16}	GCTGCGAA.

(*fliP*)

```
   1 TGAcggcccttataccccgctgatacagattaaccctgtttttttaaccccgcacgcccaa  60

  61 ggccgcggggttttcgcattttagtgctttggttaccattttcactagcgactgggtaac 120
                DraI
 121 tatattTTTAAAtaaaaaagttactattaagctaaatcaaaaagtaaccataagtatttt 180

 181 acagtatttcttaacgctcaaatTAATgattgcgcaattattcGCTGCGAAattcttccc 240

 241 acacgcagcgggtttggtttctggaactaagcgtttctgaaccgagtggcatgaagccga 300

 301 aacgttgcgaccggtaatatccaacccggtatgtccccactccgtatctcgtaaaaacaA 360

 361 GGGAcacatttatt ATG GCA AGC ATT CTG ACC AAC AAC AAC GCA ATG    407
   1 rbs            M   A   S   I   L   T   N   N   N   A   M     11

 408 GCC GCT CTC TCG ACC CTG CGT TCC ATC GCT TCC GAC CTG TCG ACC  452
  12 A   A   L   S   T   L   R   S   I   A   S   D   L   S   T    26
                                          Tn5insertion
 453 ACG CAG GAC CGT ATT TCT TCC GGC CTG AAG GTT GGC TCG GCT TCG  497
  27 T   Q   D   R   I   S   S   G   L   K   V   G   S   A   S    41

 498 GAC AAC GCT GCT TAC TGG TCG ATC GCG ACC ACC ATG CGC TCC GAC  542
  42 D   N   A   A   Y   W   S   I   A   T   T   M   R   S   D    56

 543 AAC AAG GCT CTC GGC GCA GTT TCT GAC GCG CTG GGC ATG GGC GCT  587
  57 N   K   A   L   G   A   V   S   D   A   L   G   M   G   A    71

 588 GCC AAG GTT GAC ACC GCT TCC GCC GGT ATG GAC GCC GCC ATC AAG  632
  72 A   K   V   D   T   A   S   A   G   M   D   A   A   I   K    86

 633 GTC GTG ACC GAC ATC AAG GCA AAG GTC GTT GCC GCC AAG GAA CAG  677
  87 V   V   T   D   I   K   A   K   V   V   A   A   K   E   Q   101

 678 GGC GTT GAC AAG ACC AAG GTT CAG GAA GAA GTT TCT CAG CTT CTC  722
 102 G   V   D   K   T   K   V   Q   E   E   V   S   Q   L   L   116

 723 GAT CAG CTG AAG TCG ATC GGC ACG AGC GCG TCT TTC AAC GGT GAA  767
 117 D   Q   L   K   S   I   G   T   S   A   S   F   N   G   E   131

 768 AAC TGG CTC GTT TCC AGC GCC AAT GCT ACC AAG ACC GTT GTA TCC  812
 132 N   W   L   V   S   S   A   N   A   T   K   T   V   V   S   146

 813 GGC TTT GTT CGT GAC GCT GGC GGC ACC GTT AGC GTC AAG ACG ACG  857
 147 G   F   V   R   D   A   G   G   T   V   S   V   K   T   T   161

 858 GAC TAC GCG CTG GAC GCC AAC TCC ATG CTT TAC ACG GAA GGT ACA  902
 162 D   Y   A   L   D   A   N   S   M   L   Y   T   E   G   T   176

 903 CCG GGC ACG ATC GAC GCA AAC TCC GGT ATC CTG AAC GCG ACC GGT  947
 177 P   G   T   I   D   A   N   S   G   I   L   N   A   T   G   191

 948 GCA ACG ACC ACC GTC GGC GCC AAG ACC TAC ACA CAG ATC TCC GTG  992
 192 A   T   T   T   V   G   A   K   T   Y   T   Q   I   S   V   206

 993 CTC GAC ATG AAC GTC GGC ACC GAC GAC CTC GAC AAC GCT CTC TAC 1037
 207 L   D   M   N   V   G   T   D   D   L   D   N   A   L   Y   221

1038 TCC GTT GAA ACC GCT CTG ACG AAG ATG ACC AGC GCT GGC GCC AAG 1082
 222 S   V   E   T   A   L   T   K   M   T   S   A   G   A   K   236
```

```
1083 CTC GGT TCG CTG TCT GCA CGT ATC GAC CTG CAG AGC GGC TTC GCC  1127
 237 L   G   S   L   S   A   R   I   D   L   Q   S   G   F   A    251

1128 GAC AAG CTG TCC GAC ACC ATC GAA AAG GGT GTA GGC CGT CTC GTC  1172
 252 D   K   L   S   D   T   I   E   K   G   V   G   R   L   V    266

1173 GAC GCT GAC ATG AAC GAA GAG TCC ACC AAG CTG AAG GCT CTG CAG  1217
 267 D   A   D   M   N   E   E   S   T   K   L   K   A   L   Q    281

1218 ACG CAG CAG CAG CTG GCA ATC CAG GCT CTG TCG ATC GCC AAC AGC  1262
 282 T   Q   Q   Q   L   A   I   Q   A   L   S   I   A   N   S    296

1263 GAC TCG CAG AAC ATC CTG TCG CTC TTC CGC TAA gagccgagcttctcat 1311
 297 D   S   Q   N   I   L   S   L   F   R   *                   307
                     BamHI
1312 agcccttcggggctgggGGATCCatttcttcccacaagcaatcgggggctgaccgaacgc 1371

1372 agattgggatagaaaGCATAAAGGTCCGAAAGGCTGaatcggactgaatgcatccacgag 1431
                                                                -
1432 gccgcatccggaaacggatgcggccttatctttttttaaacttttgcagatacgcgcgc 1491
     ---------->    <----------     ........
1492 gtgcgcgtgagggttgcaattttaattttgccgggaatgaatttctttccgcgcttatca 1551

1552 attcattaaccaaacccgcgttacctaatgctcatcgaaacggcgagctaacgttaaaaa 1611

1612 tcagccagttagcaggcatgatgctgtgcgccgttcccggtgaaaccggaatgtcccttc 1671

1672 tttaatcagccattcaaggggcacactactATG                           1704
                                   (flaB)
```

As described in the introduction, the flagellin genes of the enteric bacteria are known to be under the control of class III flagellar gene promoters. The *R. meliloti* promoter described has similarity to the enteric bacterial consensus class III promoter and has been shown to be transcriptionally active in *R. meliloti* 10406 [235]. Ten base pairs upstream of the start codon of *flaA* is a possible ribosome binding site, AGGGA [269]. The sequence downstream of *flaA* was searched for possible transcription termination signals using the UWGCG program TERMINATOR [46]. Approximately one hundred and fifty base pairs downstream is a putative rho-independent terminator, shown in figure 5.2.3, consisting of a guanine/cytosine-rich stem-loop motif and eight thymidine residues. This potential stem-loop motif had given problems, in the form of band compressions, during sequencing when using the Universal primers. An oligonucleotide ("F") was synthesised upstream (see figure 5.2.3) to resolve the sequencing of this region. The position of the Tn5 insertion site within the *flaA* gene of the mutant *mot-1* is also shown in figure 5.2.3. Although the Tn5 insertion causes a disruption in the *flaA* transcript it is not known if transcription is terminated close to the insertion site or whether the transcript degrades within the cell. Translation of the partial *flaA* transcript could result in a polypeptide that might interfere with the assembly of the flagellin filament and may help produce the *mot-1* phenotype.

b) The *flaB* sequence

The 963 base pair *flaB* open reading frame is found 406 base pairs downstream of *flaA*. By comparison the *flaB* gene in *R. meliloti* 10406 is found 345 base pairs downstream, and *flaB* of *R. meliloti* 1021 340 base pairs downstream [27, 235]. The guanine and cytosine content of the *A. tumefaciens flaB* open reading frame is 61%, significantly higher than the average (47%) for DNA regions thought to be non-coding. A putative class III flagellar gene promoter is found approximately 140 base pairs upstream of *flaB*. The *A. tumefaciens* putative promoter also has sequence similarity to a class III flagellar gene promoter found upstream of *flaB* in *R. meliloti* 10406 that has been shown to be transcriptionally active [235], see below.

Consensus class III promoter	TAAA	N_{15}	GCCGATAA.
Putative *A. tumefaciens flaA* promoter	TAAT	N_{16}	GCTGCGAA.
Putative *A. tumefaciens flaB* promoter	TAAC	N_{19}	GCTCATCG.
R. meliloti 10406 *flaB* promoter	TAAC	N_{18}	ACCCATCG.
Putative *R. meliloti* 1021 *flaB* promoter	TAAC	N_{18}	GACCATCG.

The "-10" sequence of the *flaB* promoter of *A. tumefaciens* and *R. meliloti* have poor sequence identity to that of the consensus of the enteric bacteria, although they are similar to each other. There is also a larger gap between the "-35" and "-10" sequences of the *R.*

```
   (flaA)                                          BamHI
   1 TAAgagccgagcttctcatagcccttcggggctgggGGATCCatttcttcccacaagcaa 60

  61 tcgggggctgaccgaacgcagattgggatagaaagcataaaggtccgaaaggctgaatcg 120

 121 gactgaatgcatccacgaggccgcatccggaaacggatgcggccttatctttttttaaa 180
                          ---------->    <----------     ........

 181 cttttgcagatacgcgcgcgtgcgcgtgagggttgcaattttaatttgccgggaatgaa 240

 241 tttctttccgcgcttatcaattcatTAACcaaacccgcgttacctaatGCTCATCGaaac 300

 301 ggcgagctaacgttaaaaatcagccagttagcaggcatgatgctgtgcgccgttcccggt 360

 361 gaaaccggaatgtcccttctttaatcagccattcaAGGGGcacactact ATG ACG    415
   1                                        rbs          M   T      2

 416 AGC ATT ATC ACG AAT GTC GCA GCA ATG TCT GCG CTC CAG ACC CTG  460
   3 S   I   I   T   N   V   A   A   M   S   A   L   Q   T   L    17

 461 CGC TCT ATC GGC CAG AAC ATG GAA TCC ACC CAG GCA CGC GTC TCC  505
  18 R   S   I   G   Q   N   M   E   S   T   Q   A   R   V   S    32

 506 TCC GGC CTT CGC GTC GGC GAT GCT TCC GAC AAC GCT GCC TAC TGG  550
  33 S   G   L   R   V   G   D   A   S   D   N   A   A   Y   W    47

 551 TCG ATC GCA ACC ACC ATG CGC TCC GAC AAC ATG GCT CTC TCT TCC  595
  48 S   I   A   T   T   M   R   S   D   N   M   A   L   S   S    62

 596 GTT TCC GAC GCT CTC GGC CTC GGC GCC GCA AAG GTG GAC ACT GCT  640
  63 V   S   D   A   L   G   L   G   A   A   K   V   D   T   A    77

 641 TCC GCC GGT ATG AGC TCG GCA ATT GAC GTC GTT AAG GAA ATC AAG  685
  78 S   A   G   M   S   S   A   I   D   V   V   K   E   I   K    92

 686 GCA AAG CTG GTC ACC GCG ACT GAA GAA GGC GTC GAC CGC ACC AAG  730
  93 A   K   L   V   T   A   T   E   E   G   V   D   R   T   K   107

 731 GTT CAG GAA GAA ATC GGC CAG CTG CAG AAG CAG CTC GCA TCG ATC  775
 108 V   Q   E   E   I   G   Q   L   Q   K   Q   L   A   S   I   122

 776 TCG CAG GGC GCT TCC TTC TAC GGC GAA AAC TGG CTC GTC GGC GTC  820
 123 S   Q   G   A   S   F   Y   G   E   N   W   L   V   G   V   137

 821 TCG ACG CTC GGC GCA GCA ACG CCC GGC ACT GAC CCG GAC AAG TCT  865
 138 S   T   L   G   A   A   T   P   G   T   D   P   D   K   S   152

 866 GTC GTC GCC GGC TTC GTC CGC GCT TCC GGC GGT GCA GTT AGC GTT  910
 153 V   V   A   G   F   V   R   A   S   G   G   A   V   S   V   167

 911 ACG ACC ACG AAA TAC GCC CTC GAC AAC ACC GCT ACC GGC AAC GTT  955
 168 T   T   T   K   Y   A   L   D   N   T   A   T   G   N   V   182

 956 CTG TTC GGT TCG GTC GAC GGT ACC GGC ACT CCC GAC GCA TCC GGC 1000
 183 L   F   G   S   V   D   G   T   G   T   P   D   A   S   G   197

1001 ATT CTC GGC ACT GTT GGC ACG TTC ACG ACG GTC GCT GCC CAG TCT 1045
 198 I   L   G   T   V   G   T   F   T   T   V   A   A   Q   S   212

1046 GTC TAC ACA CTG GAT ATC ACC CAG TAC GCC GCC GCA GAC CGC GCC 1090
 213 V   Y   T   L   D   I   T   Q   Y   A   A   A   D   R   A   227
```

```
1091 ACC AAC ATG GCA GAA GCC CTG ACG CTG GTT GAA AAC TCG CTT AAG  1135
 228 T   N   M   A   E   A   L   T   L   V   E   N   S   L   K    242

1136 GCA ATG ACC AGC GCC GCT GCA AAG CTC GGC TCG CTC TCC ATG CGT  1180
 243 A   M   T   S   A   A   A   K   L   G   S   L   S   M   R    257

1181 ATC GGC CTG CAG GAA GAC TTC GCT TCC AAG CTG TCC GAC TCC GTC  1225
 258 I   G   L   Q   E   D   F   A   S   K   L   S   D   S   V    272

1226 GAA AAG GGC ATC GGC CGC CTC GTG GAC GCT GAC ATG AAC GAA GAG  1270
 273 E   K   G   I   G   R   L   V   D   A   D   M   N   E   E    287

1271 TCC ACC CGC CTC AAG GCT CTG CAG ACA CAG CAG CAG CTC GGC GTT  1315
 288 S   T   R   L   K   A   L   Q   T   Q   Q   Q   L   G   V    302

1316 CAG GCT CTC TCG ATC GCC AAC AGC AAC TCC GAA AGC ATC CTG TCG  1360
 303 Q   A   L   S   I   A   N   S   N   S   E   S   I   L   S    317

1361 CTC TTC CGT TAA tcgaaagctgcagccgaaaacatgtaaccgcgccagaaaggcgc  1416
 318 L   F   R   *                                   ---->  <----  321

1417 cggtttttgtttgtgctcttacctgcgaaattcacgagatttttttgattaagagttcatt  1476
        .....                                HpaI                 _

1477 aaccaaatcacggctaaatcagcgcatcgaaacgactcGTTAACcaagacgaaaaagagG  1536
     HpaI
1537 TTAACaggcatcaagccgctcagtcgtttccggcgatcccggaatgtcccttcccattta  1596

1597 gccagctcagaggggcaattATG                                       1619
                          (flaC)
```

meliloti and *A. tumefaciens* putative class III promoters.

Nine base pairs upstream of the start codon of *flaB* is a possible ribosome binding site, AGGGG [269]. The sequence downstream of *flaB* was searched for possible transcription termination signals using the UWGCG program TERMINATOR [46]. Approximately thirty base pairs downstream is a putative rho-independent terminator, consisting of a short guanine/cytosine-rich stem-loop motif and five thymidine residues. The *flaB* open reading frame, the transcription signals described and the translated FlaB product are shown in figure 5.2.4.

c) The *flaC* sequence

The 942 base pair *flaC* open reading frame is found 244 base pairs downstream of *flaB*, this intergenic gap is 218 base pairs in *R. meliloti* 1021. Although no sequence data is available for *R. meliloti* 10406, it is known to possess a third copy of the flagellin genes [258]. The guanine and cytosine content of the *A. tumefaciens flaC* open reading frame is 56%, significantly higher than DNA regions thought to be non-coding. There is a possible ribosome binding site, AGGGG, five base pairs upstream of the start codon [269]. Approximately one hundred and fifty base pairs upstream there is a putative class III flagellar gene promoter, for comparison see below.

Consensus class III promoter	TAAA	N_{15}	GCCGATAA.
Putative *A. tumefaciens flaA* promoter	TAAT	N_{16}	GCTGCGAA.
Putative *A. tumefaciens flaB* promoter	TAAC	N_{19}	GCTCATCG.
Putative *A. tumefaciens flaC* promoter	TAAC	N_{18}	GCGCATCG.
Putative *R. meliloti* 1021*flaC* promoter	TAAC	N_{18}	GCTCAACG.

The sequence downstream of the *A. tumefaciens flaC* open reading frame was searched for potential transcription termination signal sequences using the UWGCG program TERMINATOR [46]. A possible rho-independent terminator was found approximately eighty base pairs downstream. There were also a number of other potential stem-loop sequences within the downstream sequence. Figure 5.2.5 shows the *flaC* open reading frame and the translated protein it might encode, as well as the transcription signals discussed above. The remainder of DNA sequence downstream of the *flaC* ORF (upto the *Hind*III site) is also shown. The results of the program TESTCODE on this sequence (see figure 3.4) predict that, approximately, the last one hundred base pairs are a coding region at the 95% confidence level. Database searches with this DNA sequence using the FASTA

```
   (flaB)                                    ---->     <----    .....
  1 TAAtcgaaagctgcagccgaaaacatgtaaccgcgccagaaaggcgccggtttttgtttg  60

 61 tgctcttacctgcgaaattcacgagatttttttgattaagagttcatTAACcaaatcacgg 120
                                  HpaI                  HpaI
121 ctaaatcaGCGCATCGaaacgactcGTTAACcaagacgaaaaagagGTTAACaggcatca 180

181 agccgctcagtcgtttccggcgatcccggaatgtcccttcccatttagccagctcagAGG 240
                                                              rbs
241 GGcaatt ATG ACA AGT ATT CTG ACG AAC ACC GCC GCC ATG TCG GCG  286
  1         M   T   S   I   L   T   N   T   A   A   M   S   A   13

287 CTC CAG ACC TTG CGC GCA ATC AGC GGC CAG CTC GAA GAC ACA CAA  331
 14 L   Q   T   L   R   A   I   S   G   Q   L   E   D   T   Q   28

332 TCC AGA GTT TCT TCT GGA TTA CGC GTA AAA TCG GCC TCC GAT AAC  376
 29 S   R   V   S   S   G   L   R   V   K   S   A   S   D   N   43

377 GCC GCC TAT TGG TCG ATC GCG ACG ACC ATG CGT TCC GAC AAC ATG  421
 44 A   A   Y   W   S   I   A   T   T   M   R   S   D   N   M   58

422 GCT CTC TCC GCC GTT CAG GAT GCA TTG GGT CTC GGC GCG GCC AAG  466
 59 A   L   S   A   V   Q   D   A   L   G   L   G   A   A   K   73

467 GTC GAC GTC GCT TAT TCC GCG ATG GAA AGT ACC GTC GAA GTC GTC  511
 74 V   D   V   A   Y   S   A   M   E   S   T   V   E   V   V   88

512 AAG GAA ATC AAG TCG AAA ATC GTC GCG GCA ACC GAA GAG GGT GTC  556
 89 K   E   I   K   S   K   I   V   A   A   T   E   E   G   V   103

557 GAC AAG ACC AAA ATC CAG GAA GAA ATC GAC CAG CTG AAA AAA CAG  601
104 D   K   T   K   I   Q   E   E   I   D   Q   L   K   K   Q   118

602 CTT GAA TCG ATT GCA CAA GGT GCA TCT TTC AGT GGC GAG AAC TGG  646
119 L   E   S   I   A   Q   G   A   S   F   S   G   E   N   W   133

647 CTG CTG GGC ACG GGC GCC AAG ACC GTC GTT TCC GGT TTT GTG CGC  691
134 L   L   G   T   G   A   K   T   V   V   S   G   F   V   R   148

692 GAT GGC GGC GGC ACC GTC AGC GTT ACG AAA ACG GAC TAT ACG CTT  736
149 D   G   G   G   T   V   S   V   T   K   T   D   Y   T   L   163

737 ATC GAC ACC ACC GCC GGC ACA ACA ACC GCC AAC GTG CTT TTC GGC  781
164 I   D   T   T   A   G   T   T   T   A   N   V   L   F   G   178

782 CTG ACC GGC TCG CCC GCA ACG CTT GAC ACA ACA AAG GGT ATC ATC  826
179 L   T   G   S   P   A   T   L   D   T   T   K   G   I   I   193
                                                  EcoRV
827 GGC CAG CCC GGC ACG GCT TCC GGC ATT TCC GTA TGG GAT ATC GAT  871
194 G   Q   P   G   T   A   S   G   I   S   V   W   D   I   D   208

872 CTC AAG CTG TTC AAT TCT GCA ACG CCG CCG ACC TAC ACC ATC GGG  916
209 L   K   L   F   N   S   A   T   P   P   T   Y   T   I   G   223

917 AAC CTT CTC ACC GAT GTC GAA ACG GCA TTT CAG TCG ATC AGC AGT  961
224 N   L   L   T   D   V   E   T   A   F   Q   S   I   S   S   238

962 GCA TCC GCC GCT CTC GGT TCG ATC AAG ATG CGC ATT GGC CTG CAG 1006
239 A   S   A   A   L   G   S   I   K   M   R   I   G   L   Q   253
```

```
1007 GAA GAC TTC GTC TCG AAA CTC ACC GAC TCC ATC GAC AAG GGC ATC  1051
 254 E   D   F   V   S   K   L   T   D   S   I   D   K   G   I    268

1052 GGC CGT CTC GTC GAT GCG GAC ATG AAC GAA GAA TCC ACC AAG CTC  1096
 269 G   R   L   V   D   A   D   M   N   E   E   S   T   K   L    283

1097 AAG GCT CTT CAG ACA CAG CAG CAG CTG GGT ATC CAG TCG CTC TCC  1141
 284 K   A   L   Q   T   Q   Q   Q   L   G   I   Q   S   L   S    298

1142 ATC GCC AAT ACC AGT TCC GAA AAT ATC CTG TCG CTA TTC CGC CAG  1186
 299 I   A   N   T   S   S   E   N   I   L   S   L   F   R   Q    313
                                                      SspI
1187 TAA gcggcgcccgttggctacgggcggtttcgtcccgctctggcagcaacagAATATTc 1245
 314 *       ------>     <------                                   314

1246 ggatatggagaccgcgcttcgcaagaggcgcggttttccatttgacgcggcctgatgcaa 1305
               -------->    <--------....
1306 cgcgtgcgaaccgaagcgcgacgacgccttcggcgtttgcgattatctaagtatctgttt 1365

1366 tatgggagaaaatggtgctgctagagagatttgaactctcggcctctcccttaccaaggg 1425

1426 agtgctctacccctgagctatagcagcatccggtgccgaagcgtctgcttcagcatcaag 1485

1486 cgtggcggcctattgccataggtttttgacgagcgcaagtcgcaaaacgatattcttcat 1545

1546 ccagtggggcaaaaaagcgtgccccctgttgaaaaacgcatttttcagatattgctgaat 1605

1606 ctatgaacgaacaacatgacaaacaggcgaaggcggcggtttccgtggacatttcgaagg 1665

1666 aaccgtcggccgtagtggcaaacggcaacacgcaggccggccccggtgaaaaggcgcgcc 1725

1726 agcgtgaggcggaggcaaggcgcgagcggcgaagaagctt                     1765
```

program, revealed no significant similarities to any other previously sequenced DNA. However when one of the open reading frames in this region was translated to its putative gene product and the Owl database searched with this predicted polypeptide, it was shown to have slight sequence similarity to part of the chemotaxis protein CheA. The functions of this protein are described in the Introduction. The open reading frame had no obvious ribosome binding site preceding it, although the guanine/cytosine content is 62%, significantly higher than that of DNA regions predicted to be non-coding. Further investigation of this region of pDUB1900 will be required before any other possible genes can be identified within it. Any chemotactic or flagellar genes would probably be detected by mutagenesis of this region, and their initial identification could then be obtained by further DNA sequencing.

All three of the flagellin open reading frames were predicted to be coding regions by the program TESTCODE and all had guanine and cytosine contents greater than those predicted for non-coding regions of DNA. Each open reading frame had putative promoter and transcription termination sequences, thus they are probably individually transcribed. However only *flaA* was shown to be definitely transcribed, since a mutation within it led to truncated flagellar filaments. In *R. meliloti* only *flaA* and *flaB* are thought to be expressed, with *flaC* being a redundant copy of the flagellin genes, as shown from mutagenesis work [236, 258]. In *A. tumefaciens,* however, there is no reason from the sequence data, why *flaC* should not be expressed.

The flagellin proteins are known not to possess conventional amino-terminal signal sequences to direct their export and instead are exported *via* the flagellum-specific export pathway [179]. However it is not known which part of the protein is recognised as the signal for its export by this pathway. In *E. coli* the axial flagellar proteins, (four rod proteins, the hook protein, three HAPs and flagellin) which are thought to be exported by this pathway, do have localised consensus sequences at their amino-termini. The *A. tumefaciens* flagellins also possess some similarity to this identified consensus sequence within FliC of *E. coli*, although the similarity is more distant to the consensus of the other axial proteins, see below.

E. coli FlgG	TGL	X_7	DVIANNLAN.
E. coli flagellin (FliC)	SGL	X_7	DAAGQAIAN.
A. tumefaciens flagellins	SGL	X_7	NAAYWSIAT.

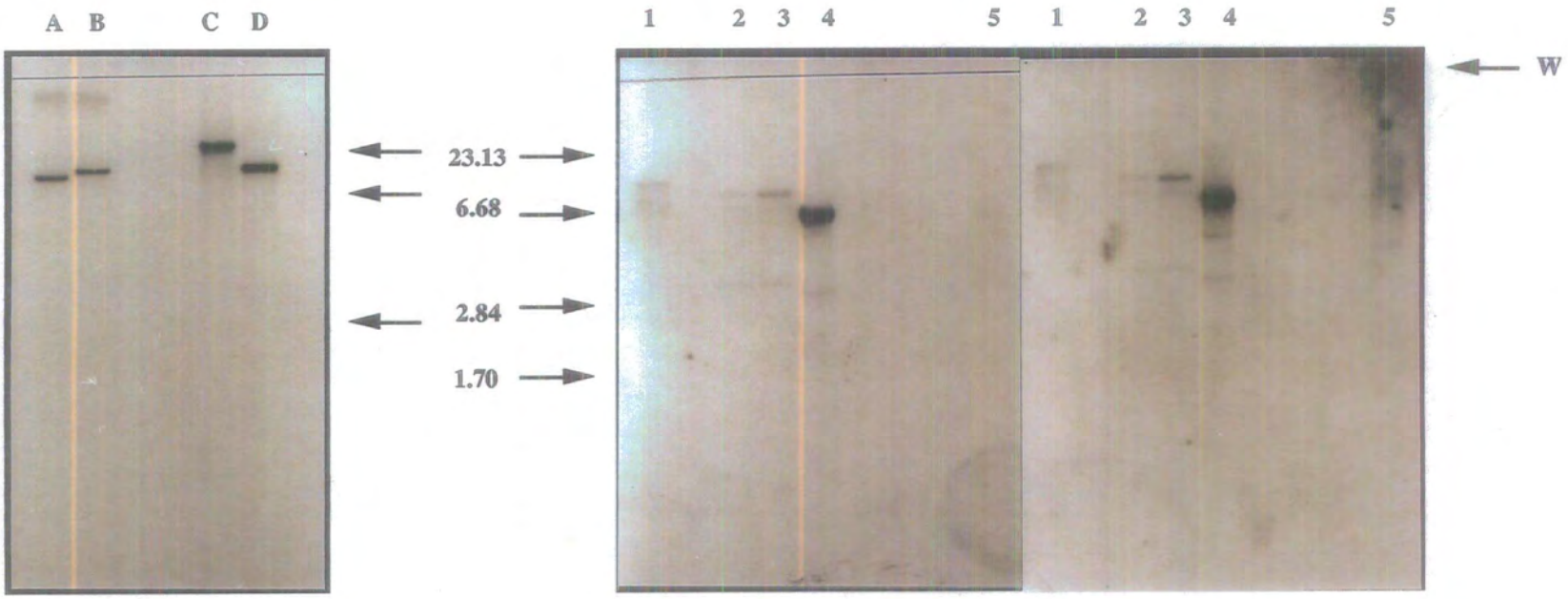
However, the axial proteins are also known to possess structural similarities, and the consensus sequences may be necessary for structural interactions within *E. coli*. The

consensus sequence directing flagellin export in *A. tumefaciens* by the flagellum-specific export pathway remains to be determined.

5.3 Detection of DNA sequences homologous to the *A. tumefaciens* flagellin open reading frames in *R. meliloti* and a variety of other Gram-negative bacteria

The data presented in this section are the results of Southern blotting experiments using the blots described in sections 4.8 and 4.9, under high stringency conditions. The blot produced from the agarose gel shown in figure 4.8.1(A) was probed with a radiolabelled *Dra*I-*Bam*HI DNA fragment containing the *flaA* gene of *A. tumefaciens*. This blot contained *Eco*RI and *Bam*HI digests of the pRZ cosmids and *R. meliloti* genomic DNA digested with *Eco*RI, as described previously. The *flaA* probe hybridised to bands of ~12kb in the *Eco*RI digests of pRZ1 and pRZ4, ~11kb in the *Bam*HI digest of pRZ1 and a band too large to be sized accurately in the *Bam*HI digest of pRZ4. A photograph of the Southern blot is shown in figure 5.3.1(A). The position of the *R. meliloti flaA* and *flaB* genes upon the cosmids pRZ1 and pRZ4 is known, see figure 4.8.2. *flaC* can also be assumed to be on these cosmids since it is known to be present just downstream of *flaB* [258]. Thus using the same nomenclature as in section 4.8, the *Eco*RI fragment hybridised to is probably RC, the *Bam*HI fragment of pRZ1 hybridised to is BA and the large band hybridised to in the *Bam*HI digest of pRZ4, is vector sequence and BA, see section 4.8. The *flaA* probe of *A. tumefaciens* hybridised strongly to the *R. meliloti* flagellin genes on the cosmids pRZ1 and pRZ4. Presumably the *flaA* probe cross-hybridised to all three copies of the *R. meliloti* flagellin genes, because of the high sequence similarity between them. The prolonged washing necessary after hybridisation probably removed too much of the hybridised probe in the genomic digest lanes for any bands to be visible on the X-ray film.

The chromosomal zoo blot described in section 4.9 was probed with a *Hpa*I-*Ssp*I radiolabelled DNA fragment containing the *flaC* open reading frame. The *flaC* open reading frame was used as the probe because the enzymes used to excise it almost flanked the coding region, and thus less non-specific DNA upstream and downstream would be liberated in comparison with the *flaA* and *flaB* probes. The genomic DNA of *P. rhodos* and *R. lupini* were included on this gel because they were known to possess complex flagella [101, 259] and hence may possess flagellin genes similar to *A. tumefaciens*. However antibodies raised against *R. meliloti* complex flagella had no immunochemical relationship to those of *P. rhodos* and *R. lupini*, suggesting differences between these complex flagella [155]. The *flaC* probe hybridised to a number of bands on the chromosomal DNA blot. A band of ~12kb was revealed in the *A. tumefaciens* genomic DNA cut with *Eco*RI which

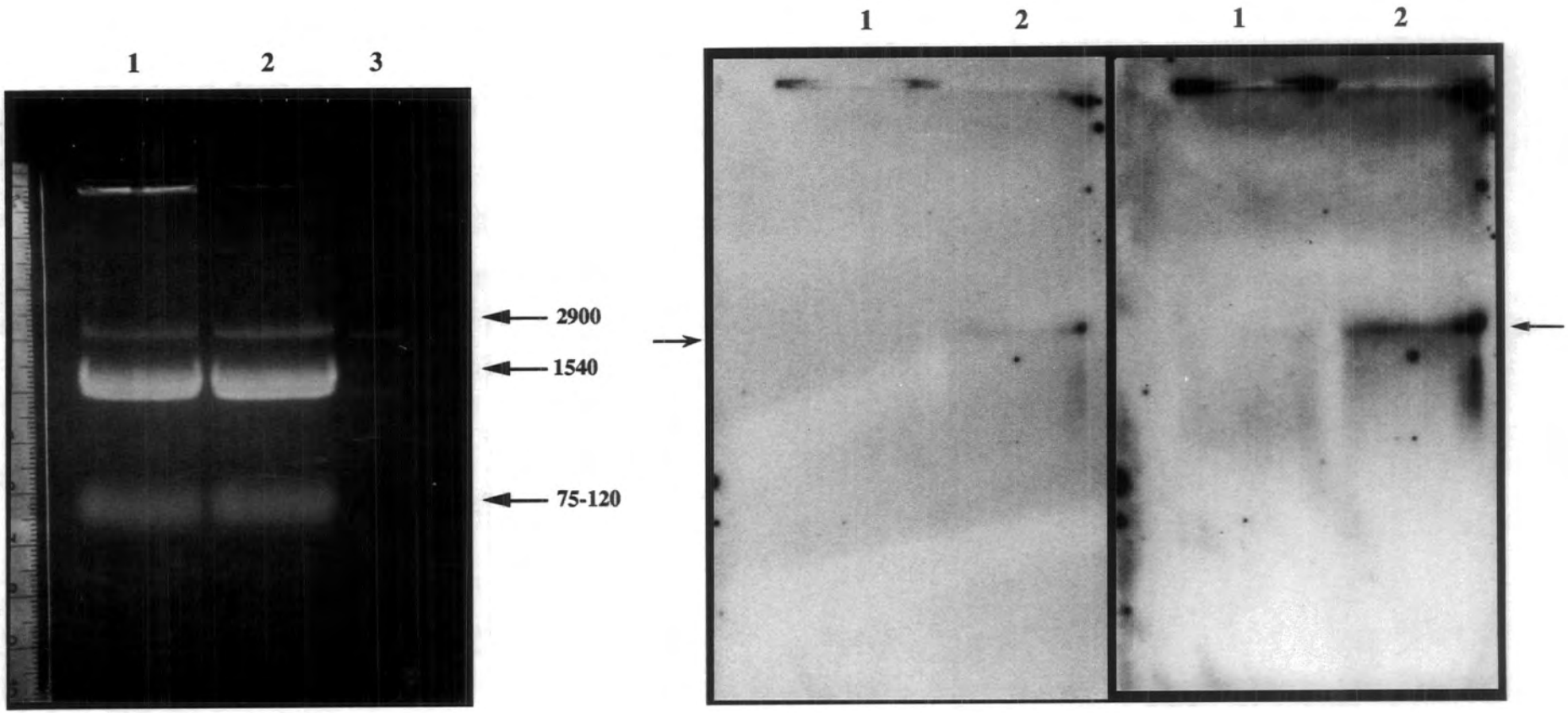

A B C D
23.13
6.68
2.84
1.70
1 2 3 4 5
1 2 3 4 5
W
A.
B.
C.

presumably is *Eco*RI fragment A of pDUB1900. Three bands of the same sizes (~10kb, ~8kb and ~3.5kb) were hybridised to in both the *R. leguminosarum* biovar *viciae* (sym^+ and sym^-) lanes. Two bands were hybridised to in the *R. leguminosarum* biovar *phaseoli* lane, a strong band of ~7kb and a weaker one of ~3kb. Finally a very faint band of ~7kb was hybridised to in the *P. talassii* chromosomal DNA cut with *Bam*HI lane. A picture of this blot is shown in figure 5.3.1(B). It was probable that bands were obtained in the *R. leguminosarum* biovar *phaseoli* and *P. talassii* lanes because more chromosomal DNA was loaded in each case, relative to the other lanes. To try and obtain further hybridising bands a smaller portion of *flaC* was radiolabelled, that contained only the highly conserved carboxy-terminal encoding DNA, see figure 5.2.5. The Southern blot obtained with this probe is shown in figure 5.3.1(C). The same bands were hybridised to in the *A. tumefaciens* and *R. leguminosarum* biovar *viciae* lanes. An extra band of ~4.5kb was hybridised to in the *R. leguminosarum* biovar *phaseoli* lane and only one band of ~4.3kb was hybridised to in the *P. talassii* lane. Perhaps by loading more genomic DNA on a subsequent gel or by lowering the stringency of the hybridisation conditions, hybridising bands could have been detected in other species.

The *R. meliloti* flagellins had been shown to have high sequence identity to those of *A. tumefaciens*. This was demonstrated experimentally by the strong hybridisations to the *A. tumefaciens flaA* probe. However functional homology could not be shown, as the pRZ1 and pRZ4 cosmids did not complement the *A. tumefaciens mot-1* mutant. Similarly pDUB1900 was unable to complement the *R. meliloti* mutants *che-1* and *che-3*, which are known to be within the *R. meliloti* flagellin genes [27].

5.4 Transcriptional analysis of the flagellin genes

The DNA sequence of the three *A. tumefaciens* flagellin open reading frames suggested that all three, if transcribed, would be transcribed individually, and that the primary transcripts of *flaA, B* and *C* would be approximately the same size, ~1.3kb. Transcription was investigated using Northern blotting experiments with total RNA extracted from the *mot-1* mutant and C58C1. The radiolabelled *flaA Dra*I-*Bam*HI DNA fragment was used as the probe since *flaA* was already thought to be transcribed. Approximately 20μg of RNA from *mot-1* and C58C1 was electrophoresed on a 1.2% formaldehyde-agarose gel, see figure 5.4.1(A). An unusual pattern of ribosomal RNA (rRNA) bands was always produced from *A. tumefaciens* RNA, as an "extra" band of ~1100 base pairs was also visible. This was suspected of being a stable degradation product of the larger rRNA fragment (usually ~2900 base pairs), and had been observed by other workers



A. B. C.

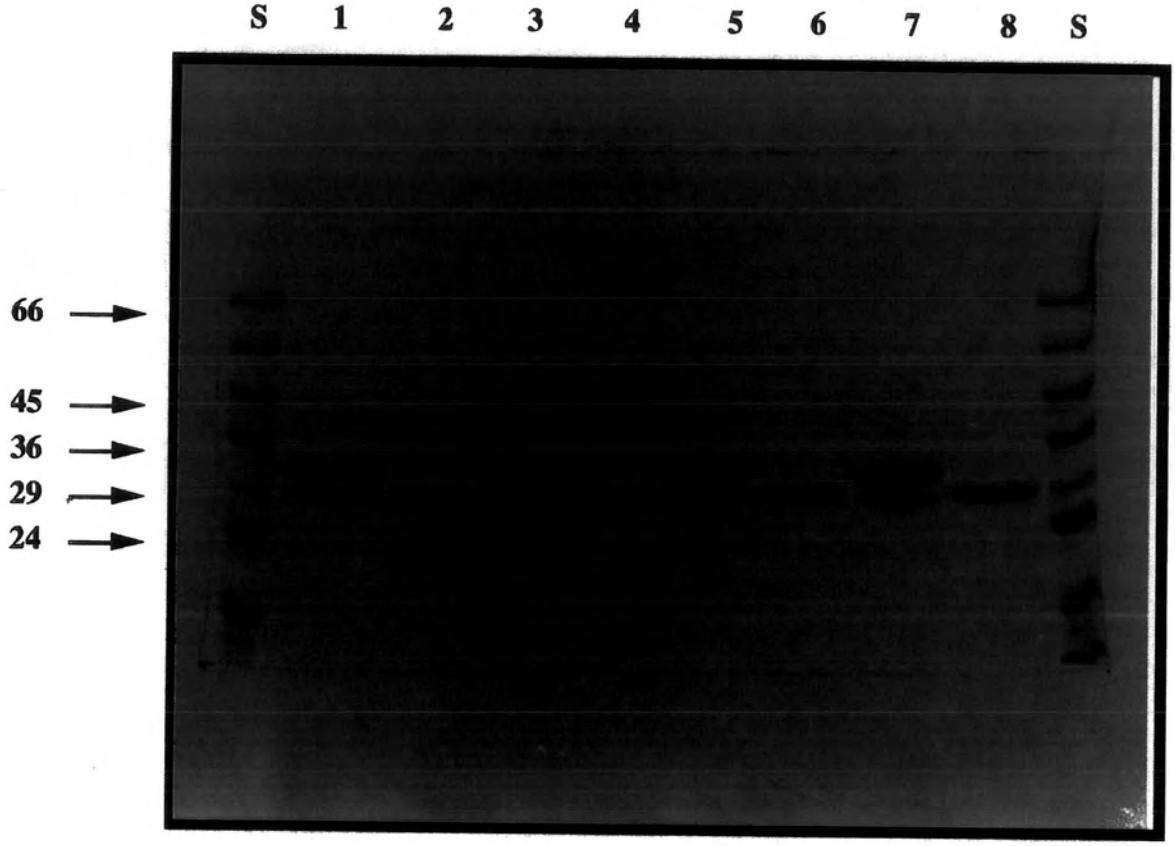
using *A. tumefaciens* RNA [330]. The large amount of RNA loaded, and the presence of this extra band, meant that the exact distance migrated by the ~1540 base pair rRNA fragment was difficult to judge. Size markers were also electrophoresed using single stranded DNA of known sizes, ~2.7kb and ~1.2kb.

Photographs of the resulting Northern blot are shown in figures 5.4.1(B and C). The only band visible was of ~1.2kb in the wild-type (C58C1) lane. No bands were seen using RNA from the *mot-1* mutant. Although the *flaA* probe would probably have cross-hybridised to any *flaB* and *flaC* transcripts produced, presumably it is the *flaA* transcript which makes up the majority of the RNA hybridised to in the C58C1 lane. (Isolation of the flagella from *A. tumefaciens*, described in the next section, shows FlaA to be the major flagellin component of the flagellum.) Further work, using primer extension analyses, will be necessary to show expression of the individual flagellins.

5.5 Isolation and investigation of the flagellin proteins of *A. tumefaciens*

Flagella were isolated from *A. tumefaciens* C58C1 and *mot-1* cells using the protocol described in section 2.17.1. Aliquots of the isolated filaments, resuspended in HEPES buffer, were analysed by SDS-PAGE, see section 2.17.4. The resulting gel was stained with Coomassie Brilliant Blue and is shown in figure 5.5.1. Two protein bands of ~29kDa and ~32kDa were observed from the samples isolated from C58C1 and one band of ~29kDa from the *mot-1* cells. Since the *flaA* gene is known to be mutated in *mot-1* cells, it seems probable that the ~29kDa bands are due to either the FlaB or FlaC flagellins (or both). The ~32kDa band isolated only from the wild-type cells is presumably FlaA. However the assignment of the above flagellins to the bands is not in accordance with the predicted protein sizes from the DNA sequence. The predicted sizes of FlaB and FlaC are 32.9kDa and 32.8kDa respectively, although the predicted size of FlaA is 31.6kDa. The SDS-PAGE analysis was repeated twice, but in each case a similar gel resulted. Thus it appears that the FlaB and/or FlaC proteins have a higher mobility and/or the FlaA proteins a lower mobility, relative to each other, during SDS-PAGE.

Examples of anomalous behaviour of polypeptides in SDS-PAGE (relative to their predicted sizes) are common. Ainouz *et al.* [5], describe the β- and γ-subunits of the lectin from *Dioclea grandiflora* which despite being shown to have similar molecular sizes by a variety of biochemical techniques, ran as 13-14kDa and 8-9kDa polypeptides (respectively) on SDS-PAGE. The flagellar protein FliP of *E. coli*, has also been shown to have abnormal migration within SDS-PAGE [184].


S
1
2
3
4
5
6
7
8
S
66
45
36
29
24

There are a number of reasons that might cause abnormal migration of the FlaA, FlaB or FlaC proteins. The FlaB and/or FlaC proteins could be processed, which would result in smaller polypeptides and thus greater mobility (relative to FlaA) on SDS-PAGE. This is unlikely, however, as the flagellins are thought to be exported by the flagellum-specific export pathway which does not involve the removal of a signal peptide. The only processing of the flagellins that occurs in *R. meliloti*, is the removal of the amino-terminal methionine residues [235].

A more plausible reason for the increased mobility of the FlaB and/or FlaC flagellins within SDS-PAGE, would be an increased binding of SDS to the proteins. Electrophoretic mobility is increased if more than 1.4g of SDS binds per gram of protein [5]. The differences in mobility of FlaB and/or FlaC relative to FlaA would be further accentuated if less than 1.4g of SDS bound per gram of FlaA - which would lead to it having decreased mobility within SDS-PAGE. The amount of SDS binding can be influenced by unusual aromatic amino acid distributions or localised highly charged domains [5]. The substitution of a single amino acid residue in a histidine-binding protein of a *S. typhimurium* mutant causes the protein to migrate in SDS-PAGE as if its molecular weight has been increased by 2kDa [213].

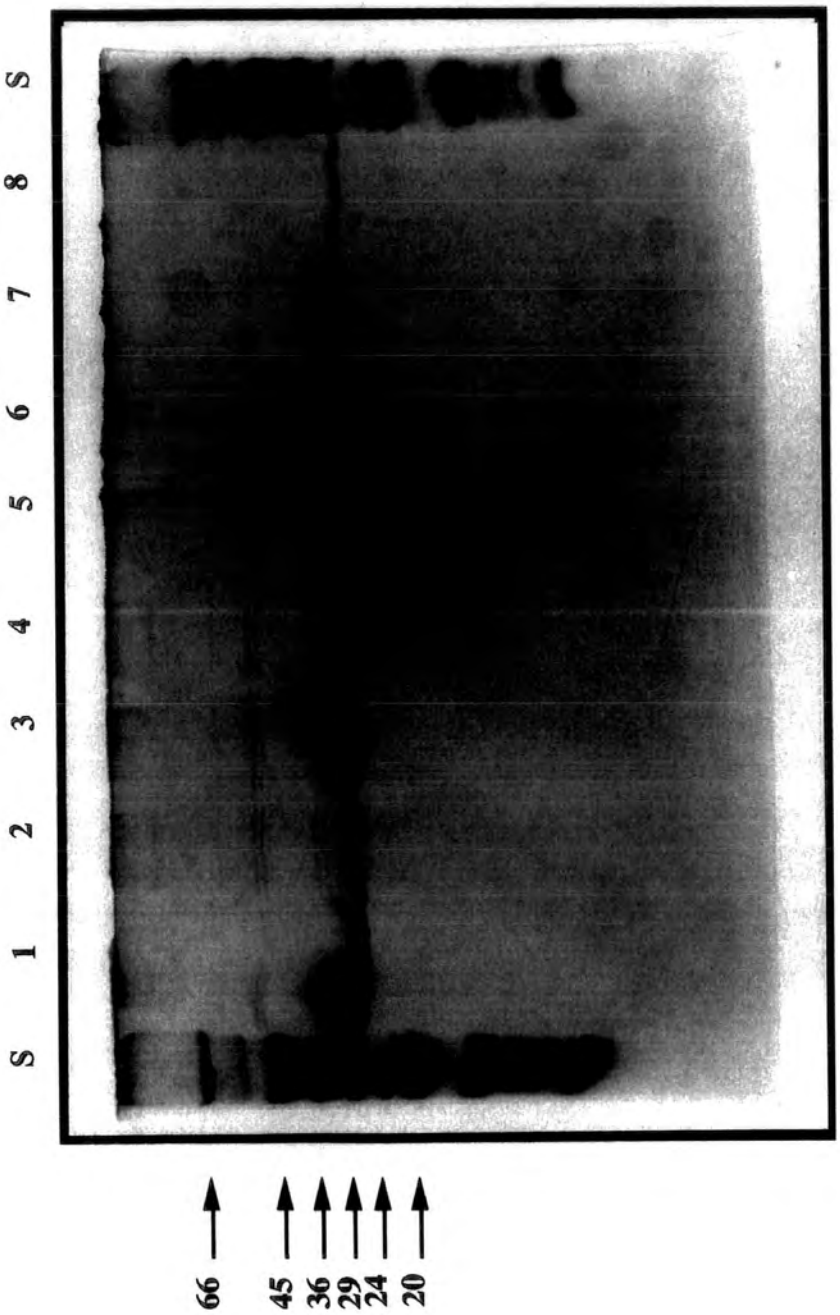
The mobility of FlaA would be decreased during SDS-PAGE if its molecular weight were increased by post-translational modification. This has been shown to occur in a number of flagellins from different bacteria. The flagellins of *S. typhimurium* are methylated at certain lysine residues by the gene product of *fliB* [179]. The single polar flagellum of the plant pathogen *P. aeruginosa* consists of flagellin subunits which are phosphorylated at tyrosine residues [140]. The flagellins of the complex flagella of *Campylobacter coli* are also thought to be post-translationally modified (possibly by phosphorylation) at serine residues [106, 173]. More extensive modifications occur to the flagellins of the archaebacterium *Halobacterium halobium*, which are glycosylated [98].

Figure 5.5.1 also shows that the relative amounts of each protein present in the wild-type are not equivalent. There is more of the ~32kDa polypeptide, thought to be FlaA. Densitometry was performed (according to section 2.17.7) on each of the wild-type lanes on the gel shown in figure 5.5.1. The ratio of the relative amounts of the ~32kDa polypeptide band (FlaA) to the ~29kDa polypeptide bands (FlaB and/or FlaC) was approximately 2:1, *i.e.* twice as much of the putative FlaA proteins were present. This is not unexpected since the phenotype of the *mot-1* mutant also suggests that the FlaA protein is the major component of the flagellum, especially the distal portion. However it is possible that FlaB and/or FlaC also form part of the distal portion of the flagellum, but are unable to assemble properly because of the absence of FlaA in *mot-1*. This could lead to a build up of FlaB

and/or FlaC in a cellular compartment. Cell protein fractions were extracted from both *mot-1* mutant and C58C1 cells, according to the protocol described in section 2.17.3, and each fraction analysed (side by side) by SDS-PAGE. There were no extra polypeptides (of ~29kDa) in any of the *mot-1* cellular fractions, relative to the wild-type control (results not shown). There was also no sign of any extra ~29kDa polypeptides in the extracellular fraction of *mot-1* cells, suggesting the "redundant" flagellins (if synthesised) were not being exported out of the cells. The results do not rule out the possibility that the absence of FlaA from *mot-1* cells has some sort of inhibitory effect on the transcription of the *flaB* and/or *flaC* genes. However given the amount of FlaA isolated from wild-type flagella and also the increased abundance of flagellin RNA transcripts in wild-type cells (relative to *mot-1* cells) it is probable that FlaA in *A. tumefaciens* forms the major component of the flagellum (especially the distal portion) as the corresponding protein does in *R. meliloti* [236].

The complex flagella of *R. meliloti* have been shown to require divalent cations (especially calcium ions) in their media to prevent dissociation [249]. As a consequence of this the isolated flagellins of *R. meliloti*, when analysed by SDS-PAGE, could not be visualised well with silver staining methods - supposedly a characteristic of calcium-binding proteins [249, 257]. To continue the comparison between the flagellins of *A. tumefaciens* and *R. meliloti*, identical aliquots of the isolated flagellins of *A. tumefaciens* were analysed by SDS-PAGE. The gel was silver stained according to section 2.17.6, and is shown in figure 5.5.2. Similar sized polypeptides to those in figure 5.5.1 can be seen, with the wild-type lanes having two polypeptide bands of ~32kDa and ~29.5kDa, and the *mot-1* lanes only one of ~29.5kDa. The greater sensitivity of the silver staining method, allowed the detection of two further bands (in all of the lanes) of ~46kDa and ~48kDa. These could possibly be HAPs (see Introduction) which would also presumably be isolated with the flagellar filaments.

The visualisation of the *A. tumefaciens* flagellins by silver staining was unexpected, given all the similarities between the *R. meliloti* and *A. tumefaciens* flagellins previously described. Considering this result on its own, it would suggest a potential difference in the *A. tumefaciens* flagella composition (relative to those of *R. meliloti*), in that it would appear not to have the characteristics of a calcium-binding protein. However the result could have occurred because of differences between the silver staining techniques used. *R. meliloti* flagellins can be detected, if the gel is prefixed with glutaraldehyde before performing the silver stain [248, 257]. The possible role of divalent cations in the subunit associations of *A. tumefaciens* flagella was further investigated, to see if this potential difference (to the


S 1 2 3 4 5 6 7 8 S
66
45
36
29
24
20

flagella of *R. meliloti*) was merely an artefact of the silver staining techniques, or if there were differences in the flagella structure.

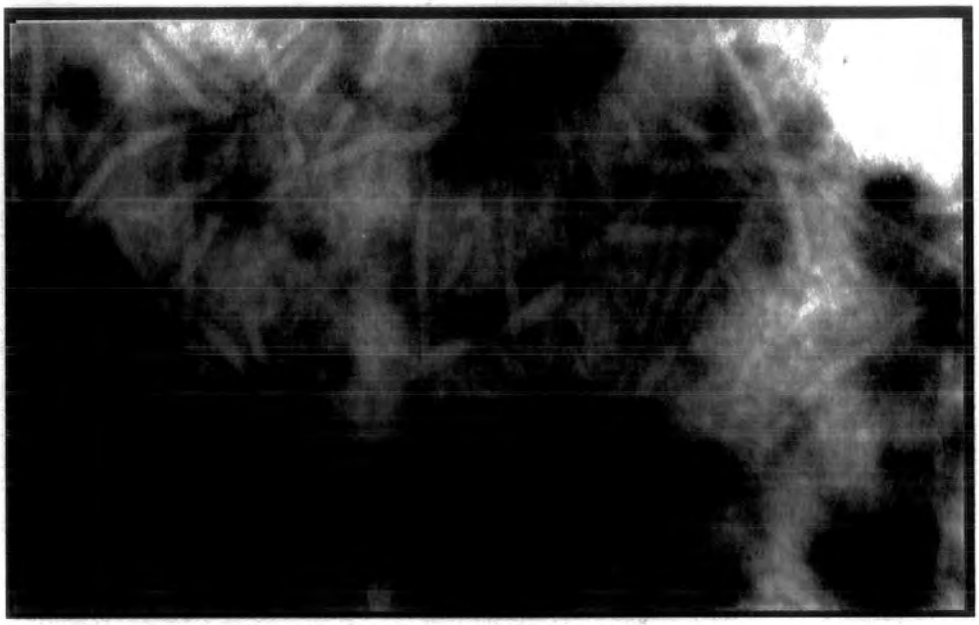
Previously, the effects of divalent cations (in particular calcium and magnesium ions) were investigated by Robinson *et al.* [249]. *R. meliloti* cells were isolated and resuspended in buffer lacking calcium and magnesium ions. The cells became non-motile and microscopic observations of the cells showed them to have lost their flagella. Motility could be restored upon the addition of calcium or magnesium ions, but only after ninety minutes - presumably to allow for the synthesis of new flagella. Isolated flagellar filaments were also treated with EDTA (which chelates divalent cations). When viewed by electron microscopy, no filaments were visible suggesting extensive dissociation. Furthermore the EDTA-treated flagellins could be analysed by native PAGE (*i.e.* with no denaturing agents present) as the dissociation resulted in polypeptides small enough to enter the gel matrix. Wild-type flagella are large polymers of flagellin subunits, and too big to enter the gel [249]. For comparison, the flagella isolated from *A. tumefaciens* C58C1 and *mot-1* cells were treated with EDTA. At first the flagella were exposed for sixteen hours to 2mM EDTA, the same conditions as used on the *R. meliloti* flagella, after which the samples were analysed by native PAGE, as described in section 2.17.4. After staining with Coomassie Brilliant Blue, no bands were visible in the resolving gel, although protein was stained in the stacking gel (at the boundary with the resolving gel). This suggests that this protein (presumably the flagellins) was too large to enter the resolving gel and hence that the EDTA treatment had not caused substantial dissociation of the *A. tumefaciens* flagella. The experiment was repeated with a range of EDTA concentrations 10, 50, 125, 250mM. Again the flagella were exposed for sixteen hours to the EDTA. EGTA (a calcium ion specific chelator) was also used, initially at 2mM and the same range of concentrations for sixteen hours. Finally a combination of EDTA and EGTA was used, both at 10mM for sixteen hours. In each case, analysis by native PAGE, revealed protein only in the stacking gel - implying large scale dissociation of the flagellar subunits had not occurred.

A culture of C58C1 cells were spun down and resuspended in 10mM HEPES and 10mM EDTA. When viewed by light microscopy a comparable number of cells (relative to the control, whose resuspension buffer also contained 200μM $CaCl_2$) retained their motility. There was some reduction in motility compared to cells viewed directly from the culture, probably because of mechanical damage to the flagella filaments during centrifugation and resuspension. The cells were viewed hourly, over six hours, and there was no difference in motility between the cells resuspended in buffer lacking calcium ions and those in buffer containing calcium ions.

The isolated flagella filaments were viewed under the electron microscope. The resulting electron micrographs are shown in figure 5.5.3. Figure 5.5.3(A) shows the filaments isolated from C58C1 cells, and demonstrates that the polypeptides seen by SDS-PAGE were the flagellins. The *A. tumefaciens* flagellar filaments do not appear to possess the "characteristic" cross-hatching seen on other complex flagella, including those of *R. meliloti*. Previous electron micrographs, with greater resolution, of *A. tumefaciens* flagellar filaments also fail to show any cross-hatching along their lengths (C. H. Shaw, unpublished data). Figure 5.5.3(B) shows the C58C1 isolated filaments after exposure to 10mM EDTA for sixteen hours. If any dissociation of the subunits has occurred, it is not visible. The filaments isolated from the *mot-1* cells were also analysed, but are not shown, as the individual filaments were not clear due to their small size and tendency to "clump" together under the electron microscope.

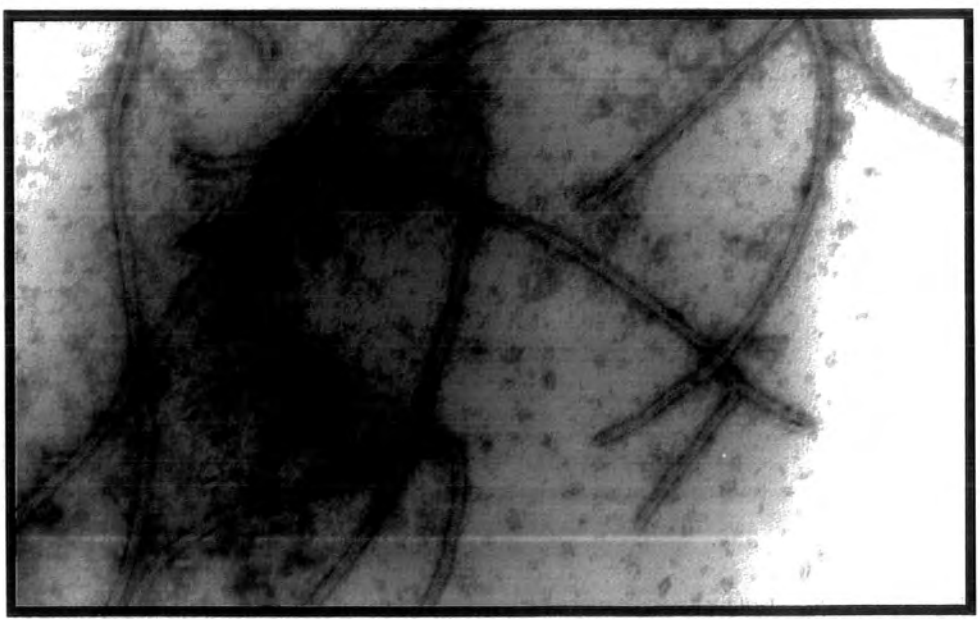
These results show a structural difference between the flagella of *A. tumefaciens* and *R. meliloti*. The flagella of *A. tumefaciens* do, however, possess similarities to the flagella of *R. leguminosarum* 8401, whose flagella also did not possess cross-hatching when viewed by electron microscopy and were unaffected by overnight exposure to EDTA [249].

A.



100 nm

B.



5.6 Discussion

The amino acid sequences of the *A. tumefaciens* flagellins showed significant sequence identity to a number of flagellins from different bacteria, especially at the conserved amino- and carboxy-termini. Hybridisation to other possible flagellin gene homologues was also observed when DNA fragments containing the *A. tumefaciens* genes were used as probes on chromosomal DNA zoo Southern blots. The greatest sequence identity was to the flagellins of the closely related bacteria *R. meliloti*. This bacterium is known to possess complex flagella. The phenotype of the *mot-1* mutants of *A. tumefaciens* suggests that the wild-type flagellar filaments possess (at least) two types of flagellin subunits. Since, in the absence of FlaA subunits (as in the *mot-1* mutant) a partial filament (presumably formed from FlaB and/or FlaC subunits) is still made. This partial filament phenotype of the *mot-1* mutant is the same as that observed for FlaA null mutants of *R. meliloti*. Thus it is probable that *A. tumefaciens* also possesses complex flagella. The complex flagella of *A. tumefaciens* do appear, however, to have subtle structural differences from those of *R. meliloti*. The flagella of *A. tumefaciens* do not dissociate in media lacking divalent cations or in the presence of chelating agents such as EDTA, as those of *R. meliloti* do. Furthermore the flagellar filaments of *A. tumefaciens* do not have the "characteristic" cross-hatching patterns of complex flagella. These properties, although different to *R. meliloti*, were the same as those observed for *R. leguminosarum* biovar *phaseoli*. The radiolabelled DNA fragments containing the *A. tumefaciens* flagellin genes hybridised to DNA sequences (presumably homologous flagellin genes) in chromosomal DNA blots of *R. leguminosarum* biovar *phaseoli*. The *A. tumefaciens* strain used (GMI9050) by Robinson *et al.* in their study of the effects of divalent cations upon flagella, did lose some motility when resuspended in media lacking divalent cations [249] implying some dissociation of the flagellin subunits from the flagella. The authors also report another member of the Rhizobiaceae, *Bradyrhizobium japonicum* USDA 110, known to have complex flagella filaments, that remains motile in buffers lacking divalent cations. The authors give two possible explanations for their observations: that divalent cations are not required in all complex flagella for subunit associations; or that the strains retaining motility in media lacking divalent cations do so because of differences in strength of the "cation-protein binding constants". Partial evidence is demonstrated for this by the difference in rates of subunit dissociation of the complex flagella of various *R. meliloti* strains [249]. Given the strong similarities previously described between the flagella of *A. tumefaciens* and *R. meliloti*, the latter reason is more probable, rather than two types of flagella with different forms of subunit associations, as is suggested by the previous explanation.

As discussed in the Introduction recent papers [304, 305], have demonstrated a role for calcium ions in the chemotaxis of *E. coli*. Calcium channel blockers and "caged"

calcium ion compounds were used to alter the levels of calcium ions intracellularly. It should be noted that changes in external calcium ion concentrations did not result in the same effects as observed in *R. meliloti* and that motility was not lost, only chemotactic ability. The role of calcium ions for the Rhizobiaceae is probably to increase the rigidity of the filaments. The natural environments of these bacteria normally have high concentrations of divalent cations, in which the bacteria may have adapted to take advantage of for flagellar assembly [249].

Another unusual feature of the *A. tumefaciens* flagella was the abnormal migration of the individual flagellin subunits in SDS-PAGE. The flagellin subunits of *R. meliloti* run together at approximately the size predicted from the amino acid sequence [249]. Possible reasons for the abnormal migration of the *A. tumefaciens* flagellins are discussed in the text.

It may be possible to determine which of the flagellins are present in the flagella of the *mot-1* mutant (and hence the faster migrating flagellin band on SDS-PAGE) by determining the protein sequence of the amino-terminal regions of the isolated flagellins. This is providing the amino-terminus is not blocked, in *R. meliloti* this does not occur although the amino-terminal methionine is removed [235]. The flagella could be isolated from *mot-1* cells as described, purified by gel filtration chromatography to remove the HAPs, and then sequenced. The first ten residues of FlaB and FlaC have two differences in amino acid sequence, shown below.

FlaB	M T S I **I** T N **V** A A
FlaC	M T S I **L** T N **T** A A

These changes (especially the valine/threonine change) would be sufficient to distinguish between the flagellins if only one type was present. If the isolated flagellins were a mix of both FlaB and FlaC subunits, this too should be shown by two elution profiles at these positions. Although this would depend on the relative amounts of each present.

A better way of determining the roles of the various flagellins, and if all three flagellins are required for filament assembly, would be by the genetic analyses described for *R. meliloti* [27, 236].

Pleier and Schmitt deleted the majority of the *flaA* and *flaB* genes with a kanamycin resistance cassette [236]. This resulted in a non-motile, filament-less mutant, to which wild-type copies of the *flaA*, *flaB* and the *flaA/flaB* genes together, were added on a plasmid. The authors found that addition of the *flaA* gene alone could not restore motility or filament production to the mutant. Addition of *flaB* alone resulted in some motility, as shortened

flagellar filaments were produced. Complementation with both the *flaA* and *flaB* genes resulted in cells with reduced motility (relative to wild-type), probably because fewer (full-sized) flagella were apparent on these cells. It was also shown, using anti-flagellin antibodies, that *flaA* expression was higher than that of *flaB*. This is in agreement with the *A. tumefaciens* data, as the protein suspected of being FlaA on the SDS-PAGE gels of the isolated flagellins is present in larger amounts than that of the FlaB and/or FlaC flagellins. The authors conclude from their results that FlaB forms the proximal part of the complex filament and FlaA the distal portion. They postulate that *flaB* is not expressed at sufficient levels to produce a full length filament in the absence of *flaA*, resulting in the truncated filaments observed. There is an obvious similarity to the phenotype of the *A. tumefaciens mot-1* mutant, which also does not possess a functional *flaA* gene. There were, however, problems with this strategy. It would have been expected that when wild-type copies of the *flaA* and *flaB* genes were both added, restoration to wild-type behaviour (or close to it) would occur. However only 20-30% of wild-type motility was observed. A possible reason for this is that the wild-type copies of the genes are upon a plasmid which is present in 5-8 copies per cell. The increased number of promoters of the highly expressed *flaA* gene are suggested to reduce the availability of the σ^{28}-like factors responsible for the expression of the class 3 motility genes (see Introduction). Thus lower amounts of other flagellar components might be synthesised resulting in (the observed) fewer flagella per cells [236].

Bergman *et al.* investigated the roles of the *flaA* and *flaB* genes in the formation of the *R. meliloti* flagella in a slightly different way [27]. The two flagellin genes were removed individually and the phenotypic effects observed. Removal of both flagellins again resulted in a filament-less, non-motile, mutant. Deletion of the *flaA* gene produced a mutant with shorter flagellar filaments, as had been observed by Pleier and Schmitt [236], and similar to that of the *mot-1* mutant of *A. tumefaciens*. However, removal of the *flaB* gene resulted in a strain with wild-type behaviour - Pleier and Schmitt had previously observed that strains expressing only *flaA* were non-motile and did not produce flagellar filaments. Possible reasons for this discrepancy are unclear, although the different methods employed and the exclusion of *flaC* from the study may have had some role.

Genetic analyses, similar to those described, could be carried out on all three of the *A. tumefaciens* flagellin genes. Gene replacement mutagenesis, as described in section 2.14, could be performed on the flagellins individually, collectively and on adjacent pairs, *i.e.* *flaA* and *flaB* and *flaB* and *flaC*. If possible, the flagellin genes could be reintroduced back onto the chromosome by recombination. This would allow the analysis of the roles (and possible interactions) of the flagellin genes in filament formation, whilst avoiding problems resulting from reintroduction *via* multicopy plasmids, as seen in *R. meliloti* [236]. If it is not practical for recombination to be carried out, the flagellin genes could be reintroduced on a

low copy number plasmid such as pGV1106. This is present at 1-3 copies per cell [90] and thus should reduce detrimental effects of the presence of multiple copies of the complementary flagellin genes (and promoters). It would also be interesting to try and complement an *A. tumefaciens* mutant with no flagellin genes with the flagellin genes of *R. meliloti*. Although complementation had not previously been observed (this study), in the absence of any *A. tumefaciens* flagellin genes and given the similarity of the *R. meliloti* flagellin genes (and promoters), complementation may now occur.

In *C. crescentus* a combination of gene replacement mutagenesis and specific antisera was employed to determine the organisation of the three different flagellin monomers in the flagellar filaments [76, 199]. This was very effective at clarifying the positions of the individual flagellins. In *A. tumefaciens* however, given the similarity between the flagellin subunits, it is probable that any antibodies raised would be non-specific and thus would cross-react with the other flagellins. Thus the genetic approach described would be the method of choice to determine the roles of the flagellin subunits in the *A. tumefaciens* flagellar filament.

SUMMARY

This work describes experiments carried out to characterise three of the *Agrobacterium* C58C1 motility mutants (*mot-1*, *mot-12* and *fla-15*), shown to map close to each other on pDUB1900 [171]. Initially approximately 9kb of DNA sequence, to which the Tn5 insertion sites of the mutants mapped, was determined. Comparing this sequence against databases (containing previously sequenced genes), revealed the presence of several putative genes involved in flagellar structure and assembly, as well as three open reading frames with no significant homology to any previously sequenced genes. The analysis of this sequence, and any subsequent work was divided according to the putative transcriptional regulation of the possible genes.

The DNA sequence revealed the presence of at least part of a putative flagellar gene operon, and genetic studies were carried out to attempt to clarify the regulation of the final two open reading frames in this putative operon. However the work described in this study represents only the beginning of a large amount of work required to determine the extent of this putative operon and the regulation (or levels of regulation) of it. This work is discussed in Chapter 4. Investigation of ORFs A, B and C by the biochemical and genetic methods described could prove particularly interesting, since these open reading frames have not been sequenced in the highly investigated motility systems of the enteric bacteria. If shown to be transcribed, their functions may clarify some of the differences between the motility systems of *A. tumefaciens* (and possibly the Rhizobiaceae) and the enteric bacteria. Further open reading frames may be found when the full extent of the putative operon is determined, which may also have no significant homology to previously sequenced genes. If so, these open reading frames should be analysed and investigated as ORFs A, B and C, as they may provide information about the different motility systems of the members of the Enterobacteriaceae and Rhizobiaceae so far studied.

The function of FliP is also unclear, although it is known to be involved in the flagellum-specific export pathway in *S. typhimurium*, as is the gene product of *fliI*. Partial DNA sequence has been found on pDUB1900, between the switch protein gene homologues and the putative operon, with homology to *fliI* (results not shown). The determination of the 5' boundary of the putative operon may even show it to contain the *fliI* homologue. The mechanism of flagellum-specific export is poorly understood, although from DNA sequence comparisons it is thought to be similar to some protein export pathways of other bacteria. Thus investigation of the *A. tumefaciens* FliP homologue may clarify not only its role in

flagellum-specific export, but also the roles of its homologues in other protein export systems.

The primary transcriptional regulation of the three flagellin gene homologues sequenced, appeared more straight forward, with each gene having its own promoter and terminator. Although possible environmental effects on regulation could be studied by creating reporter genes from the flagellins. Methods to elucidate the roles of the individual flagellins in the flagellar filament are discussed in Chapter 5. Finally, the possibility that the abnormal migration of the *flaA* gene product could be caused by post-translational modification should be investigated, since such a modification may be an important functional difference between three very similar proteins.

REFERENCES

1. Abouhamad, W. N., M. Manson, M. M. Gibson and C. F. Higgins. 1991. Peptide transport and chemotaxis in *E. coli* and *S. typhimurium* - characterisation of the dipeptide permease (DPP) and the dipeptide binding protein. Mol. Microbiol. 5: 1035-1047.

2. Adler, J. 1973. A method for measuring chemotaxis and use of the method to determine optimum conditions for chemotaxis by *Escherichia coli*. J. Gen. Microbiol. 74: 77-91.

3. Adler, J. 1975. Chemotaxis in bacteria. Annu. Rev. Biochem. 44: 341-356.

4. 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.

5. Ainouz, I. L., R. A. Moreira, F. D. A. P. Campos, M. Richardson, R. Begbie, J. C. Stewart, W. B. Watt and A. Pusztai. 1987. The isolation and amino acid sequence of the β- and γ-subunits of the lectin from the seeds of *Dioclea grandiflora*. Phytochem. **26**: 1435-1440.

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

7. Albertini, 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.

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

9. Ambler, R. P. and M. W. Rees. 1959. ε-N-Methyl lysine in bacterial flagella protein. Nature. **184**: 56-57.

10. Ames, P., S. A. Schluederberg and K. Bergman. 1980. Behavioral mutants of *Rhizobium meliloti*. J. Bacteriol. **141**: 722-727.

11. Ankenbauer, R. G. and E. W. Nester. 1990. Sugar-mediated induction of *Agrobacterium tumefaciens* virulence genes: structural specificity and activities of monosaccharides. J. Bacteriol. 172: 6442-6446.

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

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

14. 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.

15. Armitage, J. P. and R. M. Macnab. 1987. Unidirectional intermittent rotation of the flagellum of *Rhodobacter sphaeroides*. J. Bacteriol. **169**: 514-518.

16. Arnosti, D. N. and M. J. Chamberlin. 1989. Secondary σ factor controls transcription of flagellar and chemotaxis genes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. **86**: 830-834.

17. Asakura, S., G. Eguchi and T. Iino. 1966. *Salmonella* flagella: *in vitro* reconstruction and over-all shapes of flagellar filaments. J. Mol. Biol. **166**: 302-316.

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

19. 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 monocotyledonous and dicotyledonous plants. J. Bacteriol. **170**: 4181-4187.

20. 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.

21. Atsumi, T., L. McCarter and Y. Imae. 1992. Polar and lateral flagellar motors of marine *Vibrio* are driven by different ion-motive forces. Nature. **355**: 182-184.

22. Barry, T., S. Geary, S. Hannify, C. MacGearailt, M. Shalloo, D. Heery, F. Gannon and R. Powell. 1992. Rapid mini-preparations of total RNA from bacteria. Nucl. Acids Res. 20: 4940.

23. Beauchamp, C. J., W. S. Chilton, P. Dion and H. Antoun. 1990. Fungal catabolism of crown gall opines. Appl. Env. Microbiol. 56: 150-155.

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

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

26. Berg, H. C. and P. M. Tedesco. 1975. Transient response to chemotactic stimuli in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 72: 3235-3239.

27. Bergman, K., E. Nulty and L. Su. 1991. Mutations in the two flagellin genes of *Rhizobium meliloti*. J. Bacteriol. 173: 3716-3723.

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

29. 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.

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

31. Bischoff, D. S. and G. W. Ordal. 1992. *Bacillus subtillis* chemotaxis: a deviation from the *Escherichia coli* paradigm. Mol. Microbiol. 6: 23-28.

32. 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.

33. Bischoff, D. S., M. D. Weinreich and G. W. Ordal. 1992. Nucleotide sequences of *Bacillus subtilis* flagellar biosynthetic genes *fliP* and *fliQ* and identification of a novel flagellar gene, *fliZ*. J. Bacteriol. 174: 4017-4025.

34. Blair, D. F. and H. C. Berg. 1990. The MotA protein of *E.coli* is a proton-conducting component of the flagellar motor. Cell. 60: 439-449.

35. Blair, D. F. and H. C. Berg. 1991. Mutations in the MotA protein of *Escherichia coli* reveal domains critical for proton conduction. J. Mol. Biol. 221: 1433-1442.

36. Block, S. M. 1990. Personal Communication.

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

38. 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.

39. Bomhoff, G., P. M. Klapwijk, H. C. M. Kester, R. A. Schilperoort, J. P. Hernalsteens and J. Schell. 1976. Octopine and nopaline synthesis and breakdown genetically controlled by a plasmid of *Agrobacterium tumefaciens*. Mol. Gen. Genet. 145: 177-181.

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

41. Bourret, R. B., K. A. Borkovich and M. I. Simon. 1991. Signal transduction pathways involving protein phosphorylation in prokaryotes. Annu. Rev. Biochem. 60: 401-444.

42. Bourret, R. B., J. Davagnino and M. I. Simon. 1993. The carboxy-terminal portion of the CheA kinase mediates regulation of autophosphorylation by transducer and CheW. J. Bacteriol. 175: 2097-2101.

43. Bouzar, H., D. Ouadah, Z. Krimi, J. B. Jones, M. Trovato, A. Petit and Y. Desaux. 1993. Correlative association between resident plasmids and the host chromosome in a diverse *Agrobacterium* soil population. Appl. Environ. Microbiol. 59: 1310-1317.

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

45. Braun, V. 1975. Covalent lipoprotein from the outer membrane of *Escherichia coli*. Biochim. Biophys. Acta. **415**: 335-377.

46. Brendel, V., G. H. Hamm and E. N. Trifonov. 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomolec. Struct. Dyn. **3**: 705-723.

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

48. Buchman, I., F. J. Marner, G. Schroeder, S. Waffenschmidt and J. Schroeder. 1985. Tumour genes in plants: T-DNA encoded cytokinin biosynthesis. EMBO J. **4**: 853-859.

49. 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.

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

51. Cangelosi, G. A., G. Martinetti, J. A. Leigh, C. C. Lee, C. Theines and E. W. Nester. 1989. Role of *Agrobacterium tumefaciens* ChvA protein in export of β-1,2-glucan. J. Bacteriol. **171**: 1609-1615.

52. Chang, C. and S. C. Winans. 1992. Functional roles assigned to the periplasmic, linker and receiver domains of the *Agrobacterium tumefaciens* VirA protein. J. Bacteriol. **174**: 7033-7039.

53. Chilton, M.-D., M. H. Drummond, D. J. Merlo, D. Sciaky, A. L. Montoya, M. P. Gordo and E. W. Nester. 1977. Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumourigenesis. Cell. **11**: 263-271.

54. Chilton, M.-D., R. K. Saiki, N. Yadav, M. P. Gordon and F. Quetier. 1980. T-DNA from *Agrobacterium* Ti plasmid is in the nuclear DNA fraction of crown gall tumour cells. Proc. Natl. Acad. Sci. USA. **77**: 4060-4064.

55. Chou, P. Y. and G. D. Fasman. 1978. Empirical predictions of protein conformation. Ann. Rev. Biochem. **47**: 251-276.

56. Chun, S. Y. and J. S. Parkinson. 1988. Bacterial motility: membrane topology of the *Escherichia coli* MotB protein. Science. 239: 276-278.

57. Citovsky, V., M. Lie Wong and P. C. Zambryski. 1989. Co-operative interaction of *Agrobacterium* VirE2 protein with single-stranded DNA: implications for the T-DNA transfer process. Proc. Natl. Acad. Sci. USA. 86: 1193-1197.

58. Close, T. J., R. C. Tait and C. I. Kado. 1985. Regulation of Ti plasmid virulence genes by a chromosomal locus of *Agrobacterium tumefaciens*. J. Bacteriol. 164: 774-781.

59. Conner, A. J. and E. M. Dommisse. 1992. Monocotyledonous plants as hosts for *Agrobacterium*. Int. J. Plant Sci. 153: 550-555.

60. Cooley, M. B., M. R. D'Souza and C. I. Kado. 1991. The *virC* and *virD* operons of the *Agrobacterium* Ti-plasmid are regulated by the *ros* chromosomal gene: analysis of the cloned *ros* gene. J. Bacteriol. 173: 2608-2616.

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

62. Dailey, F. E. and H. C. Berg. 1993. Mutants in disulfide bond formation that disrupt flagellar assembly in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. **90**: 1043-1047.

63. 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.

64. DeRosier, D. J. 1992. Whipping flagellin into shape. Curr. Op. Struct. Biol. **2**: 280-285.

65. Devereux, J., P. Haeberli and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12**: 387-395.

66. 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.

67. Dingwall, A., J. D. Garman and L. Shapiro. 1992. Organization and ordered expression of *Caulobacter* genes encoding flagellar basal body rod and ring proteins. J. Mol. Biol. 228: 1147-1162.

68. Dingwall, A., J. W. Gober and L. Shapiro. 1990. Identification of a *Caulobacter* basal body structural gene and a *cis*-acting site required for activation of transcription. J. Bacteriol. 172: 6066-6076.

69. Dingwall, A., L. Shapiro and B. Ely. 1990. Flagella biogenesis in *Caulobacter*, pp. 155-177. *In* Armitage and Lackie (ed.), Molecular Biology of the Chemotactic Response. Cambridge University Press, Cambridge.

70. 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.

71. Doll, L. and G. Frankel. 1993. Cloning and sequencing of two new *fli* genes, the products of which are essential for *Salmonella* flagellar biosynthesis. Gene. 126: 119-121.

72. Doll, L. and G. Frankel. 1993. *fliU* and *fliV*: two flagellar genes essential for biosynthesis of *Salmonella* and *Escherichia coli* flagella. J. Gen. Microbiol. 139: 2415-2422.

73. Dow, J. M., G. Scofield, K. Trafford, P. C. Turner and M. J. Daniels. 1987. A gene cluster in *Xanthomonas campestris* pv. *campestris* required for pathogenicity controls the excretion of polygalacturonate lyase and other enzymes. Physiol. Mol. Plant Pathol. 31: 261-271.

74. Dowson, W. J. 1957. Plant diseases due to bacteria. Cambridge University Press, Cambridge.

75. Dreyfus, G., A. W. Williams, I. Kawagishi and R. M. Macnab. 1993. Genetic and biochemical analysis of *Salmonella typhimurium* FliI, a flagellar protein related to the catalytic subunit of the F_0F_1 ATPase and to virulence proteins of mammalian and plant pathogens. J. Bacteriol. 175: 3131-3138.

76. 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.

77. Dylan, T., L. Ielpi, S. Stanfield, L. Kashyap, C. Douglas, M. Yanofsky, E. W. Nester, D. R. Helinski and G. Ditta. 1986. *Rhizobium meliloti* genes required for nodule development are related to chromosomal virulence genes of *Agrobacterium tumefaciens*. Proc. Natl. Acad. Sci. USA. 83: 4403-4407.

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

79. Eisenbach, M., A. Wolf, M. Welch, S. R. Caplan, I. R. Lapidus, R. M. Macnab, H. Aloni and O. Asher. 1990. Pausing and speed fluctuation of the bacterial flagellar motor and their relation to motility and chemotaxis. J. Mol. Biol. 211: 551-563.

80. Ellis, J. G., A. Kerr, A. Petit and J. Tempe. 1982. Conjugal transfer of nopaline and agropine Ti-plamids: the role of agrocinopines. Mol. Gen. Genet. 186: 269-274.

81. Engelhardt, H., S. C. Schuster and E. Baeuerlein. 1993. An archimedian spiral: the basal disk of the *Wolinella* flagellar motor. Science. 262: 1046-1048.

82. Farrand, S. K. 1993. Personal Communication.

83. Fenselau, S., I. Balbo and U. Bonas. 1992. Determinants of pathogenicity in *Xanthomonas campestris* pv. *vesicatoria* are related to proteins involved in secretion in bacterial pathogens of animals. Mol. Plant-Microbe Interact. 5: 390-396.

84. Fickett, J. W. 1982. Recognition of protein coding regions in DNA-sequences. Nucl. Acids Res. 10: 5303-5318.

85. Figurski, D. H. and D. R. Helsinki. 1979. Replication of an origin containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. Proc. Natl. Acad. Sci. USA. 76: 1648-1652.

86. Finan, T. M. 1993. The nucleotide sequence of the *R. meliloti fliP* gene and upstream region. Personal Communication.

87. Finan, T. M., C. Gough and G. Truchet. 1993. Similarity between the *Rhizobium meliloti fliP* gene and pathogenicity-associated genes from animal and plant pathogens. J. Bacteriol. Submitted.

88. Francis, N. R., V. M. Irikura, S. Yamaguchi, D. J. DeRosier and R. M. Macnab. 1992. Localization of the *Salmonella typhimurium* flagellar switch protein FliG to the cytoplasmic M-ring face of the basal body. Proc. Natl. Acad. Sci. USA. 89: 6304-5308.

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

90. Franklin, F. C. H. and R. Spooner. 1989. Broad-host-range cloning vectors., pp. 247-267. *In* Thomas (ed.), Promiscuous plasmids of Gram-negative bacteria. Academic Press Ltd., London.

91. Galán, 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.

92. Gardina, P., C. Conway, M. Kossman and M. Manson. 1992. Aspartate and maltose-binding protein interact with adjacent sites in the Tar chemotactic signal transducer of *Escherichia coli*. J. Bacteriol. 174: 1528-1536.

93. Garnier, J., D. J. Osguthorpe and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97-120.

94. 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.

95. Gegner, J. A. and F. W. Dahlquist. 1991. Signal transduction in bacteria: CheW forms a reversible complex with the protein kinase CheA. Proc. Natl. Acad. Sci. USA. 88: 750-754.

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

97. Genetello, C., N. Van Larebeke, M. Holsters, A. Depicker, M. Van Montagu and J. Schell. 1977. Ti-plasmids of *Agrobacterium* as conjugative plasmids. Nature. 265: 561-563.

98. Gerl, L. and M. Sumper. 1988. Halobacterial flagellins are encoded by a multigene family. J. Biol. Chem. 263: 13246-13251.

99. 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.

100. Gillen, K. L. and K. T. Hughes. 1993. Transcription from two promoters and autoregulation contribute to the control of expression of the *Salmonella typhimurium* flagellar regulatory gene *flgM*. J. Bacteriol. 175: 7006-7015.

101. Götz, R., N. Limmer, K. Ober and R. Schmitt. 1982. Motility and chemotaxis in two strains of *Rhizobium* with complex flagella. J. Gen. Microbiol. 128: 789-798.

102. Götz, R. and R. Schmitt. 1987. *Rhizobium meliloti* swims by unidirectional, intermittent rotation of right-handed flagellar helices. J. Bacteriol. 169: 3146-3150.

103. 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.

104. Grisebach, H. 1981. Lignins, pp. 457-478. *In* Conn and Stumpf (ed.), The Biochemistry of Plants, a Comprehensive Treatise. Academic Press, London.

105. Grübl, G., A. P. Vogler and J. W. Lengeler. 1990. Involvement of the histidine protein (HPr) of the phosphotransferase system in chemotactic signaling of *Escherichia coli* K-12. J. Bacteriol. 172: 5871-5876.

106. Guerry, P., R. A. Alm, M. E. Power, S. M. Logan and T. J. Trust. 1991. Role of two flagellin genes in *Campylobacter* motility. J. Bacteriol. 173: 4757-4764.

107. 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.

108. Hahlbrock, K. 1981. Flavonoids, pp. 425-456. *In* Conn and Stumpf (ed.), The Biochemistry of Plants, a Comprehensive Treatise. Academic Press, London.

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

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

111. 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.

112. 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.

113. Hawley, D. K. and W. R. McClure. 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucl. Acids Res. 11: 2237-2255.

114. 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.

115. Helmann, J. D. 1991. Alternative sigma factors and the regulation of flagellar gene expression. Mol. Microbiol. 5: 2875-2882.

116. 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 β-galactosidase fusion protein into tobacco nuclei. Proc. Natl. Acad. Sci. USA. 87: 9534-9537.

117. Hess, J. F., R. B. Bourret and M. I. Simon. 1988. Histidine phosphorylation and phosphoryl group transfer in bacterial chemotaxis. Nature. **336**: 139-143.

118. Hess, J. F., K. Oosawa, N. Kaplan and M. I. Simon. 1988. Phosphorylation of three proteins in the signalling pathway of bacterial chemotaxis. Cell. 53: 79-87.

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

120. Holsters, M., R. Villaroel, J. Gielen, J. Seurinck, H. de Greve, M. Van Montagu and J. Schell. 1983. An analysis of the boundaries of the octopine TL-DNA in tumours induced by *Agrobacterium tumefaciens*. Mol. Gen. Genet. 190: 35-41.

121. Homma, M., D. J. DeRosier and R. M. Macnab. 1990. Flagellar hook and hook-associated proteins of *Salmonella typhimurium* and their relationship to other axial components of the flagellum. J. Mol. Biol. 213: 819-832.

122. Homma, M. and T. Iino. 1985. Excretion of unassembled hook-associated proteins by *Salmonella typhimurium*. J. Bacteriol. 164: 1370-1372.

123. 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.

124. Hooykaas, P. J. J., P. M. Klapwijk, M. P. Nuti, R. A. Schilperoort and A. Rörsch. 1977. Transfer of the *Agrobacterium tumefaciens* Ti-plasmid to avirulent *Agrobacteria* and to *Rhizobium ex planta*. J. Gen. Microbiol. 98: 477-484.

125. Howard, E. and V. Citovsky. 1990. The emerging structure of the T-DNA complex. Bioessays. 12: 103-108.

126. Howie, W. J., R. J. Cook and D. M. Weller. 1987. Effects of soil matrix potential and cell motility on wheat root colonization by fluorescent pseuodonomads suppressive to take-all. Phytopathol. 77: 286-292.

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

128. Huang, Y., P. Morel, B. Powell and C. I. Kado. 1990. VirA, a coregulator of Ti-specified virulence genes, is phosphorylated *in vitro*. J. Bacteriol. 172: 1142-1144.

129. 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.

130. Hwang, I., S. M. Lim and P. D. Shaw. 1992. Cloning and characterization of pathogenicity genes from *Xanthomonas campestris* pv. glycines. J. Bacteriol. 174: 1923-1931.

131. Iino, T., Y. Komeda, K. Kutsukake, R. M. Macnab, P. Matsumura, J. S. Parkinson, M. I. Simon and S. Yamaguchi. 1988. New unified nomenclature for the flagellar genes of *Escherichia coli* and *Salmonella typhimurium*. Microbiol. Rev. 52: 533-535.

132. 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 switching of the flagellar motor. J. Bacteriol. 175: 802-810.

133. Jin, S., R. K. Prusti, T. Poitsch, R. G. Ankenbauer and E. W. Nester. 1990. Phosphorylation of the VirG protein of *Agrobacterium tumefaciens* by the autophosphorylated VirA protein: essential role in biological activity of VirG. J. Bacteriol. 172: 4945-4950.

134. Jin, S., T. Roitsch, R. G. Ankenbauer, M. P. Gordon and E. W. Nester. 1990. The VirA protein of *Agrobacterium tumefaciens* is autophosphorylated and is essential for *vir* gene induction. J. Bacteriol. 172: 525-530.

135. 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.

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

137. Judicial Commission. 1970. Conservation of the generic name *Agrobacterium* Conn 1942. Int. J. Syst. Bacteriol. 20: 10.

138. Kahl, G. 1982. Molecular biology of wound healing: the conditioning phenonenom, pp. 211-265. *In* Kahl and Schell (ed.), Molecular Biology of Plant Tumours. Academic Press Inc., New York.

139. Kawagishi, I., V. Müller, 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.

140. Kelly-Wintenberg, K., S. L. South and T. C. Montie. 1993. Tyrosine phosphate in a- and b-type flagellins of *Pseudomonas aeruginosa*. J. Bacteriol. 175: 2458-2461.

141. Kennedy, B. W. and S. M. Alcorn. 1980. Estimates of USA crop losses to prokaryotic plant pathogens. Plant Dis. 64: 674-676.

142. Kerr, A. 1969. Crown gall of stone fruit 1: isolation of *Agrobacterium tumefaciens* and related species. Aust. J. Biol. Sci. 22: 111-116.

143. Kerr, A. 1974. Soil microbiological studies on *Agrobacterium radiobacter* and biological control of crown gall. Soil Sci. 118: 168-172.

144. Kerr, A. 1992. The genus *Agrobacterium*, pp. 2214-2235. *In* Balows, Trüper, Dworkin, Harder and Schleifer (ed.), The Prokaryotes. Springer-Verlag, New york.

145. Kerr, A., P. Manigault and J. Tempe. 1977. Transfer of virulence *in vivo* and *in vitro* in *Agrobacterium*. Nature. 265: 560-561.

146. Kerr, A. and C. G. Panagopoulos. 1982. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. Phytopathol. 90: 172-179.

147. Kersters, K. and J. DeLey. 1984. Section 4, Family III, Genus III: *Agrobacterium* Conn 1942, pp. 244-254. *In* Krieg (ed.), Bergey's Manual of Systematic Bacteriology. Williams and Wilkins, Baltimore.

148. Khambaty, F. M. and B. Ely. 1992. Molecular genetics of the *flgI* region and its role in flagellum biosynthesis in *Caulobacter crescentus*. J. Bacteriol. 174: 4101-4109.

149. 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.

150. Khan, S. 1993. Gene to ultrastructure: the case of the flagellar basal body. J. Bacteriol. 175: 2169-2174.

151. Khan, S., M. Dapice and T. S. Reese. 1988. Effects of *mot* gene expression on the structure of the flagellar motor. J. Mol. Biol. 202: 575-584.

152. Kleffel, B., R. M. Garavito, W. Baumeister and J. P. Rosenbusch. 1985. Secondary structure of a channel-forming protein: porin from *E. coli* outer membranes. EMBO J. 4: 1589-1592.

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

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

155. Krupski, G., R. Götz, K. Ober, E. Pleier and R. Schmitt. 1985. Structure of complex flagellar filaments in *Rhizobium meliloti*. J. Bacteriol. 162: 361-366.

156. 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.

157. Kuo, S. C. and D. E. Koshland. 1987. Roles of *cheY* and *cheZ* gene products in controlling flagellar rotation in bacterial chemotaxis of *Escherichia coli*. J. Bacteriol. 169: 1307-1314.

158. Kutsukake, K. 1994. Excretion of the anti-sigma factor through a flagellar substructure couples flagellar gene expression with flagellar assembly in *Salmonella typhimurium*. Mol. Gen. Genet. 243: 605-612.

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

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

161. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227: 680-685.

162. 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.

163. Larsen, S. H., J. Adler, J. J. Gargus and R. W. Hogg. 1974. Chemomechanical coupling without ATP: the source of energy for motility and chemotaxis in bacteria. Proc. Natl. Acad. Sci. USA. 71: 1239-1243.

164. 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.

165. Lee, K., M. W. Dudley, K. M. Hess, D. G. Lynn, R. D. Joerger and A. N. Binns. 1992. Mechanism of activation of *Agrobacterium* virulence genes: identification of phenol-binding proteins. Proc. Natl. Acad. Sci. USA. 89: 8666-8670.

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

167. Levesley, M. H. 1994. Potential applications of *Agrobacterium* virulence gene promoters in plant-protecting microbial inoculants. Ph.D. Thesis, University of Durham.

168. 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.

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

170. Liu, J. and J. S. Parkinson. 1991. Genetic evidence for interaction between the CheW and Tsr proteins during chemoreceptor signaling by *Escherichia coli*. J. Bacteriol. 173: 4941-4951.

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

172. 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.

173. Logan, S. M., T. J. Trust and P. Guerry. 1989. Evidence for posttranslational modification and gene duplication of *Campylobacter* flagellin. J. Bacteriol. 171: 3031-3038.

174. Losick, R. and L. Shapiro. 1993. Checkpoints that couple gene expression to morphogenesis. Science. 262: 1227-1228.

175. 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.

176. Lynch, B. A. and D. E. Koshland. 1991. Disulfide cross-linking studies of the transmembrane regions of the aspartate sensory receptor of *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 88: 10402-10406.

177. 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.

178. 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.

179. 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.

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

181. Macnab, R. M. and D. E. Koshland. 1972. The gradient sensing mechanism in bacterial chemotaxis. Proc. Natl. Acad. Sci. USA. 69: 2509-2512.

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

183. Maddock, J. R. and L. Shapiro. 1993. Polar location of the chemoreceptor complex in the *Escherichia coli* cell. Science. 259: 1717-1723.

184. Malakooti, J., B. Ely and P. Matsumura. 1994. Molecular characterization, nucleotide sequence, and expression of the *fliO*, *fliP*, *fliQ*, and *fliR* genes of *Escherichia coli*. J. Bacteriol. 176: 189-197.

185. 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. Natl. Acad. Sci. USA. 74: 3060-3064.

186. Mantis, N. J. and S. C. Winans. 1992. The *Agrobacterium tumefaciens vir* gene transcriptional activator *virG* is transcriptionally induced by acid pH and other stress stimuli. J. Bacteriol. 174: 1189-1196.

187. Marck, C. 1986. Fast analysis of DNA and protein sequence on Apple IIe: restriction sites search, alignment of short sequence and dot matrix analysis. Nucl. Acids Res. 14: 583-590.

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

189. Matthysse, A. G., K. V. Holmes and R. H. Gurlitz. 1981. Elaboration of cellulose fibrils by *Agrobacterium tumefaciens* during attachment to carrot cells. J. Bacteriol. 145: 583-595.

190. Mayerhofer, R., Z. Koncz-Kalman, C. Nawrath, G. Bakkeren, A. Crameri, K. Angelis, G. P. Redei, J. Schell, B. Hohn and C. Koncz. 1991. T-DNA integration: a mode of illegitimate recombination in plants. EMBO J. 10: 697-704.

191. McCarter, L. and M. Silverman. 1990. Surface-induced swarmer cell differentiation of *Vibrio parahaemolyticus*. Mol. Microbiol. 4: 1057-1062.

192. 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 parahaemolyticus*. J. Bacteriol. 175: 3361-3371.

193. McCleary, W. M. and J. B. Stock. 1993. Phosphorylation in bacterial chemotaxis, pp. 17-41. *In* Kurjan and Taylor (ed.), Signal transduction. Academic Press, Inc., San Diego.

194. 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.

195. Messens, E., A. Lenaerts, M. Van Montagu and R. W. Hedges. 1985. Genetic basis for octopine secretion from crown gall tumours. Mol. Gen. Genet. 199: 344-348.

196. Metts, J., J. West, S. H. Doares and A. G. Matthysse. 1991. Characterization of three *Agrobacterium tumefaciens* avirulent mutants with chromosomal mutations that affect induction of *vir* genes. J. Bacteriol. 173: 1080-1087.

197. Miller, J. H. 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, New York.

198. Milligan, D. L. and D. E. Koshland. 1988. Site directed cross linking - establishing the dimeric nature of the aspartate recptor of bacterial chemotaxis. J. Biol. Chem. 263: 6268-6275.

199. Minnich, S. A., N. Ohta, N. Taylor and A. Newton. 1988. Role of the 25-, 27-, and 29-kilodalton flagellins in *Caulobacter crescentus* cell motility: method for construction of deletion and Tn5 insertion mutants by gene replacement. J. Bacteriol. 170: 3953-3960.

200. Moore, L., G. Warren and G. Strobel. 1979. Involvement of a plasmid in the hairy root disease of plants caused by *Agrobacterium rhizogenes*. Plasmid. 2: 617-626.

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

202. 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.

203. Mozo, T. and P. J. J. Hooykass. 1992. Factors affecting the rate of T-DNA transfer from *Agrobacterium tumefaciens* to *Nicotiana glauca* plants. Pl. Mol. Biol. 19: 1019-1030.

204. Mulholland, V., J. C. D. Hinton, J. Sidebotham, I. K. Toth, L. J. Hyman, M. C. M. Pérombelon, P. J. Reeves and G. P. C. Salmond. 1993. A pleiotropic reduced virulence (Rvi-) mutant of *Erwinia carotovora* subspecies *atroseptica* is defective in flagella assembly proteins that are conserved in plant and animal bacterial pathogens. Mol. Microbiol. 9: 343-356.

205. Müller, 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.

206. Namba, K., I. Yamashita and F. Vonderviszt. 1989. Structure of the core and central channel of bacterial flagella. Nature. 342: 648-654.

207. Needleman, S. B. and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48: 443-453.

208. Nester, E. W., M. P. Gordon, R. M. Amasino and M. F. Yanofsky. 1984. Crown gall - a molecular and physiological analysis. Annu. Rev. Plant Physiol. 35: 387-413.

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

210. Neu, H. C. and L. A. Heppel. 1965. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. J. Biol. Chem. 240: 3685-3692.

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

212. Ninfa, E. G., A. Stock, S. Mowbray and J. Stock. 1991. Reconstitution of the bacterial chemotaxis transduction pathway from purified components. J. Biol. Chem. 266: 9764-9770.

213. Noel, D., K. Nikaido and G. F. Ames. 1979. A single amino acid substitution in a histidine-transport protein drastically alters its mobility in sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Biochemistry. 18: 4159-4165.

214. Nowlin, D. M., D. O. Nettleton, G. W. Ordal and G. L. Hazelbauer. 1985. Chemotactic transducer proteins of *E. coli* exhibit homology with methyl accepting proteins from distantly related bacteria. J. Bacteriol. 163: 262-266.

215. Ohnishi, K., K. Kutsukake, H. Suzuki and T. Iino. 1990. Gene *fliA* encodes an alternate sigma factor specific for flagellar operons in *Salmonella typhimurium*. Mol. Gen. Genet. 221: 139-147.

216. Ohnishi, K., K. Kutsukake, H. Suzuki and T. Iino. 1992. A novel transcriptional regulation mechanism in the flagellar regulon of *Salmonella typhimurium*: an anti-sigma

factor inhibits the activity of the flagellum-specific sigma factor, σ^F. Mol. Microbiol. 6: 3149-3157.

217. 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.

218. Oliver, D. 1985. Protein secretion in *Escherichia coli*. Ann. Rev. Microbiol. 39: 615-648.

219. Oosawa, K., J. F. Hess and M. I. Simon. 1988. Mutants defective in bacterial chemotaxis show modified protein phosphorylation. Cell. 53: 89-96.

220. Ordal, G. W., L. Màrquez-Magaña 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.

221. 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.

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

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

224. Parke, D., L. N. Ornston and E. W. Nester. 1987. Chemotaxis to plant phenolic inducers of virulence genes is constitutively expressed in the absence of the Ti-plasmid in *Agrobacterium tumefaciens*. J. Bacteriol. 169: 5336-5338.

225. Parkinson, J. S. 1988. Protein phosphorylation in bacterial chemotaxis. Cell. 53: 1-2.

226. Parkinson, J. S. 1993. Signal transduction schemes of bacteria. Cell. 73: 857-871.

227. Parkinson, J. S. and E. C. Kofoid. 1992. Communication modules in bacterial signaling proteins. Annu. Rev. Genet. **26**: 71-112.

228. Pazour, G. J. and A. Das. 1990. Characterisation of the VirG binding site of *Agrobacterium tumefaciens*. Nucl. Acids Res. **18**: 6909-6913.

229. Pazour, G. J. and A. Das. 1990. *virG*, an *Agrobacterium tumefaciens* transcriptional activator, initiates at a UUG codon and is a sequence-specific DNA-binding protein. J. Bacteriol. **172**: 1241-1249.

230. Pearson, W. R. and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA. **85**: 2444-2448.

231. Peralta, E. G., R. Hellmiss and L. W. Ream. 1986. Overdrive, a T-DNA transmission enhancer on the *A. tumefaciens* tumour inducing plasmid. EMBO J. **5**: 1137-1142.

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

233. Piper, K. R., S. B. Von Bodman and S. K. Farrand. 1993. Conjugation factor of *Agrobacterium tumefaciens* Ti plasmid regulates Ti plasmid transfer by autoinduction. Nature. **362**: 448-450.

234. Plano, G. V., S. S. Barve and S. C. Straley. 1991. LcrD, a membrane-bound regulator of the *Yersinia pestis* low-calcium response. J. Bacteriol. **173**: 7293-7303.

235. Pleier, E. and R. Schmitt. 1989. Identification and sequence analysis of two related flagellin genes in *Rhizobium meliloti*. J. Bacteriol. **171**: 1467-1475.

236. Pleier, E. and R. Schmitt. 1991. Expression of two *Rhizobium meliloti* flagellin genes and their contribution to the complex filament structure. J. Bacteriol. **173**: 2077-2085.

237. Poole, P. S., M. J. Smith and J. P. Armitage. 1993. Chemotactic signalling in *Rhodobacter sphaeroides* requires metabolism of attractants. J. Bacteriol. **175**: 291-294.

238. Postma, P. W., J. W. Lengeler and G. R. Jacobson. 1993. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Micobiol. Rev. **57**: 543-594.

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

240. 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.

241. Raha, M., M. Kihara, I. Kawagishi and R. M. Macnab. 1993. Organization of the *Escherichia coli* and *Salmonella typhimurium* chromosomes between flagellar regions IIIa and IIIb, including a large non-coding region. J. Gen. Microbiol. 139: 1401-1407.

242. Ralling, G. and T. Linn. 1984. Relative activities of the transcriptional regulatory sites in the *rplKAJLrpoBC* gene cluster of *Escherichia coli*. J. Bacteriol. 158: 279-285.

243. Ream, W. 1989. *Agrobacterium tumefaciens* and interkingdom genetic exchange. Annu. Rev. Phytopathol. 27: 583-618.

244. Rice, M. S. and F. W. Dahlquist. 1991. Sites of deamidation and methylation in Tsr, a bacterial chemotaxis sensory transducer. J. Biol. Chem. 266: 9746-9753.

245. Robertson, J. L., T. Holliday and A. G. Matthysse. 1988. Mapping of Agrobacterium tumefaceins genes affecting cellulose synthesis and bacterial attachment to host cells. J. Bacteriol. 170: 1408-1411.

246. Robinson, A. C., D. J. Kenan, G. F. Hatfull, N. F. Sullivan, R. Spiegelberg and W. D. Donachie. 1984. DNA sequence and transcriptional organization of essential cell division genes *ftsQ* and *ftsA* of *Escherichia coli*: evidence for overlapping transcriptional units. J. Bacteriol. 160: 546-555.

247. Robinson, A. C., D. J. Kenan, J. Sweeney and W. D. Donachie. 1986. Further evidence for overlapping transcriptional units in an *Escherichia coli* cell envelope-cell division gene cluster: DNA sequence and transcriptional organization of the *ddl ftsQ* region. J. Bacteriol. 167: 809-817.

248. Robinson, J. B. 1993. The silver staining of *R. meliloti* flagellins. Personal Communication.

249. Robinson, J. B., O. H. Tuovinen and W. D. Bauer. 1992. Role of divalent cations in the subunit associations of complex flagella from *Rhizobium meliltoi*. J. Bacteriol. 174: 3896-3902.

250. Ronson, C. W., B. T. Nixon and F. M. Ausubel. 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell. 49: 579-581.

251. Rosenberg, C. and T. Huguet. 1984. The pAtC58 plasmid of *Agrobacterium tumefaciens* is not essential for tumour induction. Mol. Gen. Genet. 196: 533-536.

252. Rossignol, G. and P. Dion. 1985. Octopine, nopaline and octopinic acid utilization in *Pseudomonas*. Can. J. Microbiol. 31: 68-74.

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

254. 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.

255. Schell, J., M. Van Montagu, M. de Brueckeleer, M. de Block, A. Depicker, M. de Wilde, G. Engler, C. Genetello, J. P. Hernalsteens, M. Holsters, J. Seurinck, B. Silva, F. Van Vliet and R. Villarroel. 1979. Interactions and DNA transfer between *Agrobacterium tumefaciens*, the Ti-plasmid and the plant host. Proc. R. Soc. Lond. B. 204: 251-266.

256. Scher, F. M., J. W. Kloepper, C. Singleton, J. Zaleska and M. Laliberte. 1988. Colonization of soybean roots by *Pseudomonas* and *Serratia* species: relationship to bacterial motility, chemotaxis and generation time. Phytopathol. 78: 1055-1059.

257. Schleicher, M. and D. M. Watterson. 1983. Analysis of differences between Coomassie Blue stain and silver stain procedures in polyacrylamide gels: conditions for the detection of calmodulin and troponin C. Anal. Biochem. 131: 312-317.

258. Schmitt, R. 1993. Identification of *R. meliloti flaC*. Personal Communication.

259. Schmitt, R., I. Raska and F. Mayer. 1974. Plain and complex flagella of *Pseudomonas rhodos*: analysis of fine structure and composition. J. Bacteriol. 117: 844-857.

260. Schroeder, G., S. Waffenschmidt, E. W. Weiler and J. Schroeder. 1984. The T-region of the Ti-plasmid codes for an enzyme synthesizing indole-3-acetic acid. Eur. J. Biochem. 138: 387-391.

261. 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.

262. Segall, J. E., S. M. Block and H. C. Berg. 1986. Temporal comparisons in bacterial chemotaxis. Proc. Natl. Acad. Sci. USA. 83: 8987-8991.

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

264. Shaw, C. H. 1993. New insights on T-DNA transfer. Trends Microbiol. 1: 325-327.

265. Shaw, C. H., A. M. Ashby, A. P. Brown, C. Royal, G. J. Loake and C. H. Shaw. 1988. VirA and VirG are the Ti-plasmid encoded functions required for chemotaxis of *Agrobacterium tumefaciens* towards acetosyringone. Mol. Microbiol. 2: 413-417.

266. Shaw, C. H., G. J. Loake, A. P. Brown and C. S. Garrett. 1989. Molecular biology of chemotaxis in *Agrobacterium*, pp. 117-123. *In* Lugtenberg (ed.), Signal Molecules in Plants and Plant-Microbe Interactions. Springer-Verlag KG, Berlin.

267. 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.

268. Shaw, C. H., M. D. Watson, G. H. Carter and C. H. Shaw. 1984. The right hand copy of the nopaline Ti-plasmid is required for tumour formation. Nucl. Acids Res. 12: 6031-6041.

269. Shine, J. and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16s ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA. 71: 1342-1346.

270. Shioi, J., R. C. Tribhuwan, S. T. Berg and B. L. Taylor. 1988. Signal transduction in chemotaxis to oxygen in *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. 170: 5507-5511.

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

272. Silverman, M. and M. Simon. 1974. Flagellar rotation and the mechanism of bacterial motility. Nature. 249: 73-74.

273. Simon, R., J. Quandt and W. Klipp. 1989. New derivatives of transposon Tn5 suitable for mobilization of replicons, generation of operon fusions and induction of genes in Gram-negative bacteria. Gene. 80: 161-169.

274. 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.

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

276. 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.

277. Sosinsky, G. E., N. R. Francis, D. J. DeRosier, J. S. Wall, M. N. Simon and J. Hainfield. 1992. Mass determination and estimation of subunit stoichiometry of the bacterial hook-basal body flagellar complex of *Salmonella typhimurium* by scanning transmission electron microscopy. Proc. Natl. Acad. Sci. USA. 89: 4801-4805.

278. 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.

279. Spencer, P. A. and G. H. N. Towers. 1988. Specificity of signal compounds detected by *Agrobacterium tumefaciens*. Phytochem. 27: 2781-2785.

280. Springer, W. R. and D. E. Koshland. 1977. Identification of a protein methyltransferase as the *cheR* gene product in the bacterial sensing system. Proc. Natl. Acad. Sci. USA. 74: 533-537.

281. Stachel, S. E., E. Messens, M. Van Montagu and P. Zambryski. 1985. Identification of the signal molecules produced that activate T-DNA transfer in *Agrobacterium tumefaciens*. Nature. 318: 624-629.

282. Stachel, S. E., B. Timmerman and P. C. Zambryski. 1986. Generation of single-stranded T-DNA molecules during the initial stages of T-DNA transfer from *Agrobacterium tumefaciens*. Nature. 322: 706-712.

283. Stachel, S. E. and P. C. Zambryski. 1986. *Agrobacterium tumefaciens* and the susceptible plant cell: a novel adaptation of extracellular recognition and DNA conjugation. Cell. 47: 155-157.

284. Stallmeyer, M. J. B., S.-I. Aizawa, R. M. Macnab and D. J. DeRosier. 1989. Image reconstruction of the flagellar basal body of *Salmonella typhimurium*. J. Mol. Biol. 205: 519-528.

285. 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.

286. Stock, A. M. and J. B. Stock. 1987. Purification and characterization of the CheZ protein of bacterial chemotaxis. J. Bacteriol. 169: 3301-3311.

287. Stock, J. 1990. Role of protein carboxyl methylation in bacterial chemotaxis., pp. 275-284. *In* Paik and Kim (ed.), Protein Methylation. CRC Press Inc., Boca Raton Fl.

288. Stock, J. B., G. S. Lukat and A. M. Stock. 1991. Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks. Annu. Rev. Biophys. Biophys. Chem. 20: 109-136.

289. Stock, J. B., A. J. Ninfa and A. M. Stock. 1989. Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. 53: 450-490.

290. Stock, J. B., A. M. Stock and J. M. Mottonen. 1990. Signal transduction in bacteria. Nature. 344: 395-400.

291. Stock, J. B., M. G. Surette, W. R. McCleary and A. M. Stock. 1992. Signal transduction in bacterial chemotaxis. J. Biol. Chem. 267: 19753-19756.

292. 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.

293. Summers, W. C. 1970. A simple method for extraction of RNA from *E.coli* utilizing diethyl pyrocarbonate. Anal. Biochem. 33: 459-463.

294. Sutherland, I. W. 1985. Biosynthesis and composition of Gram-negative bacterial extracellular and wall polysaccharides. Annu. Rev. Microbiol. 39: 243-270.

295. Suzuki, T. and T. Iino. 1981. Role of the *flaR* gene in hook formation in *Salmonella* spp. J. Bacteriol. 148: 973-979.

296. Swart, S., G. Smit, J. J. Lugtenberg and J. W. Kijne. 1993. Restoration of attachment, virulence and nodulation of *Agrobacterium tumefaciens chvB* mutants by rhicadhesin. Mol. Microbiol. 10: 597-605.

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

298. 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.

299. 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.

300. 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.

301. Thomashow, M. F., J. E. Karlinsey, J. R. Marks and R. E. Hurlbert. 1987. Identification of a new virulence locus in *Agrobacterium tumefaciens* that affects polysaccharide composition and plant cell attachment. J. Bacteriol. 169: 3209-3216.

302. Thomashow, M. F., C. G. Panagopoulos, M. P. Gordon and E. W. Nester. 1980. Host range of *Agrobacterium tumefaciens* is determined by the Ti-plasmid. Nature. 283: 794-796.

303. Thorstenson, Y. R., G. A. Kuldau and P. C. Zambryski. 1993. Subcellular localization of seven VirB proteins of *Agrobacterium tumefaciens*: implications for the formation of a T-DNA transport structure. J. Bacteriol. 175: 5233-5241.

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

305. 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.

306. Trachtenberg, S. and D. J. DeRosier. 1987. Three-dimensional structure of the frozen-hydrated flagellar filament. The left-handed filament of *Salmonella typhimurium*. J. Mol. Biol. 195: 581-601.

307. Ueno, T., K. Oosawa and S.-I. Aizawa. 1992. M ring, S ring and proximal rod of the flagellar basal body of *Salmonella typhimurium* are composed of subunits of a single protein, FliF. J. Mol. Biol. 227: 672-677.

308. Van der Werf, P. and D. E. Koshland. 1977. Identification of a γ-glutamyl ester in bacterial membrane protein involved in chemotaxis. J. Biol. Chem. 252: 2793-2795.

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

310. Van Larebeke, N., C. Genetello, J. Schell, R. A. Schilperoort, A. K. Hermans, J. P. Hernalsteens and M. Van Montagu. 1975. Acquisition of tumour inducing ability by non-oncogenic *Agrobacterium* as a result of plasmid transfer. Nature. 255: 742-743.

311. Van Montagu, M. and J. Schell. 1979. The plasmids of *Agrobacterium tumefaciens*, pp. 71-95. *In* Timmis and Pülher (ed.), Plasmids of Medical, Environmental and Commercial Importance. Elsevier/North Holland Biomedical Press, Amsterdam.

312. Van Onckelen, H., E. Prinsen, D. Inze, P. Rudelsheim, M. Van Lijsebettens, A. Folgin, J. Schell, M. Van Montagu and J. de Gref. 1986. *Agrobacterium* T-DNA gene-1 codes for tryptophan 2-monooxygenase activity in tobacco crown galls. Febs. Letts. 198: 357-360.

313. Velumbathi, K., M. Krishnan, J. H. Gould, R. H. Smith and S. B. Gelvin. 1989. Opines stimulate induction of the *vir* genes of the *Agrobacterium tumefaciens* Ti plasmid. J. Bacteriol. 171: 3696-3703.

314. 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.

315. Vogel, H. and F. Jähnig. 1986. Models for the structure of outer-membrane proteins of *Escherichia coli* derived from Raman spectroscopy and prediction methods. J. Mol. Biol. 190: 191-199.

316. 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.

317. Vogler, A. P. and J. W. Lengeler. 1987. Indirect role of adenylate cyclase and cyclic AMP in chemotaxis to phosphotransferase system carbohydrates in *Escherichia coli* K-12. J. Bacteriol. 169: 593-599.

318. Vonderviszt, F., S.-I. Aizawa and K. Namba. 1991. Role of the disordered terminal regions of flagellin in filament formation and stability. J. Mol. Biol. 221: 1461-1474.

319. Vonderviszt, F., R. Ishima, K. Akasaka and S.-I. Aizawa. 1992. Terminal disorder: A common structural feature of the axial proteins of the bacterial flagellum? J. Mol. Biol. 226: 575-579.

320. Wada, K., Y. Wada, F. Ishibashi, T. Gojobori and T. Ikemura. 1992. Codon usage tabulated from the GenBank genetic sequence data. Nucl. Acids Res. 20: 2111-2118.

321. Wagenknecht, T., D. J. DeRosier, S.-I. Aizawa and R. M. Macnab. 1982. Flagellar hook structures of *Caulobacter* and *Salmonella* and their relationship to filament structure. J. Mol. Biol. 162: 69-87.

322. Watson, B., T. C. Currier, M. P. Gordon, M. Chilton and E. W. Nester. 1975. Plasmid required for virulence of *Agrobacterium tumefaciens*. J. Bacteriol. 123: 255-264.

323. Welch, M., K. Oosawa, S.-I. Aizawa and M. Eisenbach. 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. Proc. Natl. Acad. Sci. USA. **90**: 8787-8791.

324. White, F. F. and E. W. Nester. 1980. Hairy root plasmid encodes virulence traits in *Agrobacterium rhizogenes*. J. Bacteriol. **141**: 1134-1141.

325. Willmitzer, L., M. de Breuckeleer, M. Lemmers, M. Van Montagu and J. Schell. 1980. DNA from the Ti-plasmid is present in the nucleus and absent from the plastids of plant crown gall cells. Nature. **287**: 359-361.

326. 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.

327. Winans, S. C. 1990. Transcriptional induction of an *Agrobacterium* regulatory gene at tandem promoters by plant-released phenolic compounds, phosphate starvation and acidic growth media. J. Bacteriol. **172**: 2433-2438.

328. Winans, S. C. 1991. An *Agrobacterium* two-component regulatory system for the detection of chemicals released from plant wounds. Mol. Microbiol. **5**: 2345-2350.

329. Winans, S. C. 1992. Two-way chemical signaling in *Agrobacterium*-plant interactions. Microbiol. Rev. **56**: 12-31.

330. Winans, S. C. 1993. *A. tumefaciens* RNA isolation. Personal Communication.

331. Winans, S. C., P. R. Ebert, S. E. Stachel, M. P. Gordon and E. W. Nester. 1986. A gene essential for *Agrobacterium* virulence is homologous to a family of positive regulatory loci. Proc. Natl. Acad. Sci. USA. **83**: 8278-8282.

332. Winans, S. C., R. A. Kerstetter and E. W. Nester. 1988. Transcriptional regulation of the *virA* and *virG* genes of *Agrobacterium tumefaciens*. J. Bacteriol. **170**: 4047-4054.

333. Yadav, N. S., J. Vanderleyden, D. R. Bennet, W. M. Barnes and M. Chilton. 1982. Short direct repeats flank the T-DNA on a nopaline Ti-plasmid. Proc. Natl. Acad. Sci. USA. **79**: 6332-6326.

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

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

336. Yanisch-Perron, C., J. Vierra and J. Messing. 1985. Improved M13 cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene. **33**: 103-119.

337. Zambryski, P., J. Tempe and J. Schell. 1989. Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. Cell. **56**: 193-200.

338. Zambryski, P. C. 1992. Chronicles from the *Agrobacterium*-plant cell DNA transfer story. Annu. Rev. Pl. Physiol. **43**: 465-490.

339. Zanker, H., J. Von Lintig and J. Schröder. 1992. Opine transport genes in the octopine (*occ*) and nopaline (*noc*) catabolic regions in the Ti plasmids of *Agrobacterium tumefaciens*. J. Bacteriol. **174**: 841-849.

340. Zhulin, I. B. and J. P. Armitage. 1993. Motility, chemotaxis, and methylation-independent chemotaxis in *Azospirillum brasilense*. J. Bacteriol. **175**: 952-958.

341. 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.

342. Zuberi, A. R., C. Ying, D. S. Bischoff and G. W. Ordal. 1991. Gene-protein relationships in the flagellar hook-basal body complex of *Bacillus subtilis*: sequences of the *flgB*, *flgC*, *flgG*, *fliE* and *fliF* genes. Gene. **101**: 23-31.

UNIV. DUMEA