### Title (Include Security Classification)
Genomic organization and expression of Campylobacter flagellin genes

#### Performing Organization Report Number(s)
NYRl 90-28

#### Name of Performing Organization
Naval Medical Research

#### Office Symbol (If applicable)

#### Address (City, State, and ZIP Code)
Bethesda, Maryland 20814-5055

#### Name of Monitoring Organization
Naval Medical Command

#### Office Symbol (If applicable)

#### Address (City, State, and ZIP Code)
Department of the Navy Washington, D.C. 20372-5120

#### Name of Funding/Spotning Organization
Naval Medical Research and Development Command

#### Office Symbol (If applicable)

#### Address (City, State, and ZIP Code)
Bethesda, Maryland 20814-5055

#### Name of Contracting Organization
Naval Medical (If applicable)

#### Address (City, State, and ZIP Code)
Bethesda, Maryland 20814-5055

#### Title
Genomic organization and expression of Campylobacter flagellin genes

#### Journal Article
Reprinted From: Journal of Bacteriology, Apr. 1990, p. 1853-1860 Vol. 172, No. 4

#### COSATI Codes

<table>
<thead>
<tr>
<th>FIELD</th>
<th>GROUP</th>
<th>SUB-GROUP</th>
</tr>
</thead>
</table>

#### Subject Terms (Continue on reverse if necessary and identify by block number)
Campylobacter jejuni flaA gene
Campylobacter coli VCI67 flaB gene
Campylobacter flagellin genes antigenic variation (Over)

#### Distribution/Availability of Report
Approved for public release; distribution is unlimited

#### Distribution/Availability of Abstract
Unclassified/Unlimited

#### Abstract Security Classification
Unclassified

#### Name of Responsible Individual
Phyllis Bluin, Information Services Division

#### Telephone (Include Area Code)
202-295-2188

#### Office Symbol

#### Security Classification of This Page
Unclassified
18. Subject Terms continued:
    phase variation
    DNA rearrangement
    flagellin genes
    flagellin mutants
Genomic Organization and Expression of Campylobacter Flagellin Genes

PATRICIA GUERRY,¹ SUSAN M. LOGAN,² SCOTT THORNTON,¹ AND TREVOR J. TRUST²
Infectious Diseases Department, Naval Medical Research Institute, Bethesda, Maryland 20884,¹ and Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada²

Received 14 August 1989/Accepted 27 December 1989

Campylobacter coli VC167, which undergoes an antigenic flagellar variation, contains two full-length flagellin genes, flaA and flaB, that are located adjacent to one another in a tandem orientation and are 91.5% homologous. The gene product of flaB, which has an M₀ of 58,916, has 93% sequence homology to the gene product of flaA, which has an M₀ of 58,946, has 93% sequence homology to the gene product of flaA, which has an M₀ of 58,916 (S. M. Logan, T. J. Trust, and P. Guerry, J. Bacteriol. 171:3031-3038, 1989). Mutational analyses and primer extension experiments indicated that the two genes are transcribed under the control of distinct promoters but that they are expressed concomitantly in the same cell, regardless of the antigenic phase of flagella being produced. The flaA gene, which was expressed at higher levels than the flaB gene in both phases, was transcribed from a typical σ54-type promoter, whereas the flaB promoter was unusual. A mutant producing only the flaB gene product did not synthesize a flagellar filament and was nonmotile. Southern blot analysis indicated that flagellar antigenic variation involves a rearrangement of flagellin sequence information rather than the alternate expression of the two distinct genes.

Campylobacter jejuni and Campylobacter coli are important gastrointestinal pathogens of humans, most commonly producing an acute enteritis (6, 8, 39). An essential step in the establishment of Campylobacter enteritis is the colonization of the viscous mucous blanket covering the epithelium of the small intestine. The motility imparted by the polar flagellum of the Campylobacter cell appears to play a significant role in the colonization of this mucous niche (23, 30, 31), since nonmotile variants appear unable to colonize the gastrointestinal tract of experimental animals or human volunteers (4, 7). Flagellin is also the immunodominant protein antigen recognized during infection, so flagella are clearly important virulence determinants of the thermophilic Campylobacters.

Flagellar expression in Campylobacter spp. is subject to both phase and antigenic variation (7, 12). Phase variation refers to the ability of some Campylobacter strains to exhibit a bidirectional transition between flagellated and nonflagellated phenotypes (7). Antigenic variation refers to the ability of other strains to reversibly express flagella of different antigenic specificities (12). In the case of C. coli VC167, the reversible variation corresponds to the production of flagellar filaments containing antigenically distinguishable flagellin monomers of apparent subunit molecular weight 61,500 in antigenic phase 1 (P1) cells and 59,500 in antigenic phase 2 (P2) cells. Importantly, recent studies with C. coli VC167 have demonstrated an in vivo preference for one flagellar phase, P2, during colonization in an animal model (19), further emphasizing the functional importance of this locomotory organelle in gastrointestinal colonization by Campylobacter spp.

Phase and antigenic variation of surface antigens has been described for other pathogenic bacteria, including fimbrial-phase variation in Escherichia coli (11), Moraxella bovis (25), and Neisseria gonorrhoeae (27, 35, 40), antigenic variation of outer membrane proteins of Neisseria (40, 41) and Borrelia (26, 33) spp., and antigenic variation of flagella in Salmonella typhimurium (37, 38, 42). Presumably, the ability to undergo such surface changes confers advantages to the pathogen as it encounters multifarious environments. The molecular mechanisms by which pathogens alter their surface antigens are varied but often involve programmed DNA rearrangements (6). Antigenic variation of VC167 flagella is associated with an as yet uncharacterized DNA rearrangement at a site that is not adjacent to the structural genes (11, 19). As a first step in elucidating the molecular mechanisms involved in the antigenic variation of C. coli flagella, a flagellin gene from VC167 was recently cloned and the nucleotide sequence was determined (22). The open reading frame encoded by the gene was found to share strong homology with the amino and carboxy termini of flagellins from other bacteria and was found to be posttranslationally modified at serine residues (22). Hybridization data indicated the presence of flagellin-related information immediately downstream of the sequenced gene (22), suggesting the presence of a second copy of flagellin. In this study, we show that there are, in fact, two full-length flagellin genes that are adjacent to one another on the VC167 chromosome in a tandem orientation. The two genes are highly related to one another but are transcribed from separate promoters concomitantly in both P1 and P2 cells. Moreover, we provide evidence that antigenic variation of flagellin in C. coli results from a rearrangement internal to the flagellar structural genes.

MATERIALS AND METHODS

Bacterial strains and growth conditions. C. coli VC167 serogroup L10 8 was originally isolated from human feces and was obtained from H. Lior, National Enteric Reference Centre, Ottawa, Ontario, Canada. Derivatives of VC167 producing either P1 or P2 cells were selected as described by Harris et al. (12). E. coli DH5 (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) was used as the host for cloning experiments. Growth conditions were as previously described (22).

DNA sequencing. Double-stranded dideoxy sequencing was performed after alkaline denaturation by using [35S]dATP (DuPont NEN Research Products, Boston, Mass.) and Se-
The portion of the DNA cloned on pGK201 (23) encoding two tandemly oriented flagellin genes. Transcription of the genes is from left to right, as drawn. Restriction enzymes sites marked above the line are those shared by the two genes; sites unique to one copy are indicated below the line. (B) The original clone, pGK201, containing flaA and flaB. Plasmids pGK200, pGK204, and pGK205 were used for DNA sequencing. Plasmid pGK209 contains an EcoRI-EcoRV fragment subcloned from the flaA gene as indicated. Restriction sites: P, PstI; H, HindIII; R, EcoRI; RV, EcoRV; B, BstII; Bg, BglII; HaeIII; N, NdeI.

**FIG. 1.** Maps of VC167 flagellin gene clusters. (A) The portion of the DNA cloned on pGK201 (23) encoding two tandemly oriented flagellin genes. Transcription of the genes is from left to right, as drawn. Restriction enzymes sites marked above the line are those shared by the two genes; sites unique to one copy are indicated below the line. (B) The original clone, pGK201, containing flaA and flaB. Plasmids pGK200, pGK204, and pGK205 were used for DNA sequencing. Plasmid pGK209 contains an EcoRI-EcoRV fragment subcloned from the flaA gene as indicated. Restriction sites: P, PstI; H, HindIII; R, EcoRI; RV, EcoRV; B, BstII; Bg, BglII; HaeIII; N, NdeI.
RESULTS

DNA and protein sequences of the second VC167 flagellin. Plasmid pGK201, which contains a 10-kilobase-pair (kb) ClaI fragment from the VC167 chromosome, has been shown to contain a full-length flagellin gene (22), now termed flaA. Hybridization studies had suggested that a second copy of flagellin resides adjacent to the first on this plasmid (22). DNA sequence analysis of this region confirmed the presence of a second full-length flagellin gene, flaB (Fig. 3). These two genes are tandemly oriented and separated by 216 bp. The DNA sequence of flaB is compared with that of flaA in Fig. 2, and the deduced protein sequences of the two flagellins are compared in Fig. 3. The two DNA sequences are remarkably conserved, with an overall sequence similarity of 91.5%, suggesting that they arose as a result of gene duplication. There are regions of the two genes that are totally identical, lacking even third-base-pair changes. There is also a strong bias for codons with A or T in the third position in both genes (data not shown), as would be expected for an organism with a G+C content of 30 to 33% (3, 32). The flaA gene encodes a polypeptide of 571 amino acids (rather removal of the first methionine) with a predicted posttranslationally unmodified M, of 58,916 (22); the flaB gene encodes a polypeptide of 572 amino acids with a predicted posttranslationally unmodified M, of 58,946. The two proteins share 93% homology at the amino acid level. All regions of the flagellin protein known to be posttranslationally modified (22) are identical in both molecules. Although most of the changes occur in the amino and carboxy ends of the molecules, there are areas of major changes in the middle regions of the two proteins (Fig. 3). The first is caused by a frameshift mutation (Fig. 2) and accounts for the fact that flaB encodes one additional amino acid.

Analysis of the promoter regions. The sequences of the 5' noncoding regions of the two genes are compared in Fig. 4. In the case of flaB, this region represents the intergenic region, which contains two nearly perfect 49-bp repeats separated by 35 nucleotides. Within these repeats are two identical stem-loop structures. These potential terminator sequences for the flaA gene have free-energy values of -12.8 kcal/mol and -12.1 kcal/mol, and are separated by 60 nucleotides. The region downstream of the translational start of both genes, including classical AGGA putative ribosome-binding sites, is the most highly conserved region. Both AGGA sites are 8 nucleotides from the translational initiation codon. An upstream sequence of flaA is indicated which shares strong homology to the corresponding sequence for a B. gramicola promoter (13). The corresponding areas of flaB share homology at the -35 region but little homology in the -10 region. Primer extension experiments using mRNA purified from P1 and P2 cells were performed to determine the transcriptional initiation sites and to determine which gene(s) is expressed in each flagellar phase. The mRNA from each phase was primed with oligonucleotides specific for the flaA 5'-TAATAGCTCGTTTAC3') and flaB 5'-AACGTCTCTAGATTAA3'). These oligonucleotides bind to the coding strand of each gene at a position starting 52 bp downstream of the A1G start. Both genes appeared to be transcribed independently in both flagellar phases (Fig. 5). In addition, the flaA-specific oligonucleotide gave a higher relative signal in primer extension experiments than the flaB-specific oligonucleotide, and this relative signal strength was the same using mRNA templates from P1 and P2 cells (Fig. 5). The start point for flaA (Fig. 4) is that predicted from the position of the -35 promoter. The start point for flaB, which is closer to the structural gene by 17 bp, and the putative -35 and -10 promoter regions of flaB, determined on the basis of this start point, are also indicated in Fig. 4.

Mutational analysis of flagellin genes. Gene replacement techniques (17) were used to generate flagellin mutants. Plasmid pGK200 (22; Fig. 1) contains a 2.3-kb HindIII fragment containing the first 311 bp of the flaA gene and approximately 2 kb of nonflagellin DNA sequence 5' to the start of flaA. There is a unique BglII site located 162 bp downstream of the translational start of the flaA gene. A Campylobacter flagellin for kanamycin resistance was cloned as a BamHI-ends cassette from plasmid pILL500 (17) into this BglII site. The resulting HindIII fragment, in which the flaA flagellin sequence had been disrupted by the kanamycin resistance gene, was transferred into the Campylobacter suicide vector pILL560, which is capable of being conjugal mobilized by IncP plasmids into, but is not capable of autonomous replication in Campylobacter cells (17). The resulting plasmid, pGK200, was transformed into a strain of DH5 harboring the conjugal IncP plasmid RK212.2 (9). Conjugal transfer from E. coli to C. coli VC167 P1 was performed as described by Labigne-Roussel et al. (17). All kanamycin-resistant Campylobacter transconjugants obtained should represent events in which the kanamycin resistance gene was rescued by a crossover between the Campylobacter sequences flanking the kanamycin resistance gene and the corresponding regions of homology in the chromosome. Such a single crossover resulted in a duplication of sequences on the chromosome (Fig. 6A); excision of the vector DNA and one copy of the flagellin information should occur by the analogous single reciprocal crossover. In this case, in which there are two repeating regions of homology, represented by the two tandem flagellin genes, two distinct second recombination events are possible, and both were isolated (Fig. 6A). The first class, exemplified by mutant VC167-B2, occurs when the second recombinational event occurs within the flaB gene. This results in a simple replacement of the kanamycin-inactivated flagellin sequences for the corresponding flaA chromosomal sequences. If the recombinational event occurs within the flaB gene, all of the chromosomal copy of flaB is deleted. This second class, which was the predominant class isolated (three of five transconjugants), is represented by mutant VC167-B3. Figure 6B shows a Southern blot analysis of HindIII-digested DNA from wild-type VC167 (lane 1), VC167-B2 (lane 2), and VC167-B3 (lane 3) probed with pGK201. In VC167-B2, the HindIII fragment, H3, had increased in size, by approximately 1.4 kb, the size of the inserted kanamycin resistance gene, as compared with wild-type VC167 DNA (i.e., from 2.3 kb to 3.7 kb). In VC167-B3, the H3 fragment had also increased in size; in addition, all of the H4 fragment, which encodes the 3' end of flaA and the 5' end of flaB, was deleted.

Both VC167-B3, which is genetically flaA flaB, and VC167-B2, which is flaB flaA, are nonmotile. No flagellin can be detected on the surface of either VC167-B2 or VC167-B3 cells after glycerol extraction (20; data not shown). However, Western blot analysis of whole cell proteins using a polyclonal antibody directed against Campylobacter flagellin (21) demonstrated that the flaB gene product was being expressed in VC167-B2 (Fig. 6C, lane 3) at low levels. No
immunologically reactive material was produced in VC167-B3 (Fig. 6C, lane 4).

**Mechanism of antigenic variation.** The biochemical and genetic data presented above indicate that both flaA and flaB are expressed in VC167 P1 and VC167 P2 and that alternate expression of the two genes is not involved in flagellar antigenic variation. In other pathogens, antigenic variations often involve internal rearrangements of structural genes (4). To test this possibility, DNAs from VC167 P1 and VC167 P2 cells were digested with various restriction enzymes and analyzed by Southern blot hybridization. Using 32P-labeled pGK201 DNA as the probe. No rearrangements were observed when the DNAs were digested with either HindIII (Fig. 7, lanes 1 and 2) or EcoRV (lanes 5 and 6). However,
flagellar antigenic variation, contains two full-length copies. The relative signal strength in both primer extension and that this partial copy if involved in antigenic variation. During protein sequencing of tryptic peptides from both P1

2-kb fragment was observed in P1 DNA. This result indicates that the transition from P1 to P2 DNA used to probe HincI-digested P1 and P2 DNAs. The same cells is identical in both phases for the first 38 amino acids. The N-terminal amino acid sequence of flacellin, which is underlined at the -35 and -10 positions. The position of a putative promoter sequence for flaA determined from the position of the mRNA start site, is also indicated by double underlining. Arrows indicate two identical palindromic sequences that may function as transcriptional terminators of the flaA gene. These sequences are included within nearly perfect 49-bp repetitive sequences (underlined).

CAMPYLOBACTER FLAGELLIN GENES

FIG. 4. Comparison of the 5' noncoding regions of flaA and flaB. The 214-bp region upstream of the two flagellin genes are compared. The sequence 5' to flaA on the line 1. and the sequence 5' to flaB (and 3' to flaA) on line 2. Identical bases are indicated by columns. The final ATG in bold italic letters represents the translational start of each structural gene. The mRNA start points for each gene determined by primer extension experiments are indicated by asterisks. Putative ribosome-binding sites (rbs) for each gene are marked by dotted underlining. A -3 promoter lies in the appropriate position relative to the mRNA start site on the flaA sequence (double underlined at the -55 and -10 positions). The consensus -3 promoter contains the sequence CTAAA at the -35 position and CGGATAT at the -10 position (13). The position of a putative promoter sequence for flaB, determined from the position of the mRNA start site, is also indicated by double underlining. Arrows indicate two identical palindromic sequences that may function as transcriptional terminators of the flaA gene. These sequences are included within nearly perfect 49-bp repetitive sequences (underlined).

of flagellin. The two genes, which are tandemly oriented and 93% homologous, are separately transcribed in both antigenic phases, indicating that antigenic variation is not a result of the alternate expression of flaA and flaB. The genomic organization of VC167 flagellins resembles that described for Rhizobium meliloti (34), which also has two tandemly arranged flagellin genes that are 87% homologous. However, in this organism, the gene products of both genes are found in a complex flagellar filament in approximately stoichiometric amounts, as determined by amino acid sequencing data (34). The N-terminal amino acid sequence of flagellin purified from flagellar filaments of P1 and P2 VC167 cells is identical in both phases for the first 38 amino acids (20). This amino acid sequence is encoded by the flaA gene. The changes in the amino terminal sequence encoded by flaB have not been observed in amino acid sequencing of flagellin from either VC167 P1 or P2 cells. This would suggest that the flaB gene product is not present as a major constituent in the filament of either phase. However, internal peptides encoded by flaB have been observed during protein sequencing of tryptic peptides from both P1 and P2 cells in minor amounts (data not shown). Further suggesting that the flaB gene is expressed at low levels in both phases.

DISCUSSION

This work confirms that C. coli VC167, which undergoes a flagellar antigenic variation, contains two full-length copies
Western blot analyses suggests that the flaA gene is expressed at higher levels than flaB in both phases. In other systems, bacterial flagellins comprise up to 5% of the total cell protein and are transcribed under the control of strong promoters resembling that of the B. subtilis α78 promoter (2, 29). A sequence bearing a strong homology to the α78 promoter is present 5' to the flaA gene in the appropriate position for the promoter, as determined from the mRNA start point. The putative promoter of the flaB gene is unusual and presumably represents a weaker promoter. The presence of a strong α78-like promoter 5' to the flaA gene raises the question of why the flaA gene is not expressed at immunologically detectable levels in E. coli (22). Fusion of the flaA promoter to a promoterless chloramphenicol acetyltransferase gene indicates that the promoter can function in E. coli (data not shown). The lack of detectable Campylio-
bacter flagellin in E. coli may be due either to the fact that the gene product is not posttranslationally modified (22) or to codon utilization differences. There is a positive correlation between the degree of codon bias and the level of gene expression, with more highly expressed genes having the strongest constraints (156). The codon utilization index, which is a measure of codon usage bias (156), is 0.455 for the flaA gene of E. coli 168 and 0.232 for the flaA of C. coli. This lower value for flaA reflects the presence of suboptimum codons in the E. coli host. Since both the nay and flaA genes are transcribed by strong ςB-type promoters, this would suggest that translational barriers contribute to the lack of expression of flaA in the foreign host.

Most plain bacterial flagellins, including the VC167 flagellins, share considerable homology at the amino and carboxy regions of the molecule, which function in assembly and transport (14). The central nonfunctional regions of flagellins are antigenically diverse. It is surprising, therefore, that over half the amino acid changes occurring between the flaA and flaB gene products of VC167 occur in the amino and carboxy regions. The central regions of flaA and flaB do have two major areas of change, but throughout this putative variable region, there are only 15 total amino acids changes. The data indicate that the flaA gene represents the major component of the flagellar filament and that the gene product of flaB is not assembled into a functional flagellum in the absence of the flaA gene product. The simplest explanation for this finding would be that not enough gene product is synthesized under the flaB promoter to assemble a functional flagellum. The changes in the amino and carboxy regions of the flaB gene may also represent mutations affecting assembly or transport (14). Alternatively, it may be that the VC167 flagellum resembles that of Caulobacter crescentus, in which distinct flagellins assemble into different regions of the filament (28), and that both the flaA and flaB gene products are required for functional flagella in VC167. In an effort to address these questions, we have attempted to construct flaA− flaB mutants. However, all insertions of the kanamycin resistance cassette into the flaB gene in E. coli to date have resulted in deletions or rearrangements of the Campylobacter DNA. Efforts to construct such mutants are continuing in an effort to elucidate the role of the flaB gene in the structural organization of the VC167 flagellum.

The mechanism of antigenic variation of VC167 flagella involves a rearrangement of some internal flagellin sequence information with undefined endpoints. The current data suggest that the flaA gene may represent a high-level expression locus and that the P1-to-P2 shov involves insertion of alternate genetic information in the form of a partial cassette into flaA at some point downstream of the amino-terminal sequences. Some of this information must reside outside of the flagellin gene cluster that has been cloned and sequenced, since the DNA cloned in pGK201 does not contain the 2-kb HindIII fragment observed during the rearrangement. However, we cannot rule out the possibility that the flaB gene also contributes sequence information to flaA. To clarify the molecular mechanism of antigenic variation, we are raising antisera to synthetic peptides based on the observed changes between the flaA and flaB proteins of P1 cells, and we are currently determining the DNA sequences of the flaA and flaB genes cloned from P2 cells. The relationship of this rearrangement within the structural genes to the previously described rearrangement at a nonadjacent locus (11, 19) is also being investigated.

ACKNOWLEDGMENTS

This work was supported in part by U.S. Navy Research and Development Command Research Work Unit no. 6102AJM161 102SU13 AK. 111 and by a grant from the Medical Research Council of Canada to T.J.T.

We thank Agnes Labigne-Roussel and Donald G. Guiney, Jr., for strains and Li-Ting Liang, Gary Majam, and Steve Martin for expert technical assistance.

LITERATURE CITED