JOURNAL OF BACTERIOLOGY, June 1989, p. 3031–3038
0021-9193/89/063031-08$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 6

Evidence for Posttranslational Modification and Gene Duplication of *Campylobacter* Flagellin

SUSAN M. LOGAN,[1] TREVOR J. TRUST,[1] AND PATRICIA GUERRY[2]*

Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, V8W 2Y2, Canada,[1] and Infectious Diseases Department, Naval Medical Research Institute, Bethesda, Maryland 20814[2]

Received 10 November 1988/Accepted 23 March 1989

A gene encoding a flagellin protein of *Campylobacter coli* VC167 has been cloned and sequenced. The gene was identified in a pBR322 library by hybridization to a synthetic oligonucleotide probe corresponding to amino acids 4 to 9 of the N-terminal sequence obtained by direct chemical analysis (S. M. Logan, L. A. Harris, and T. J. Trust, J. Bacteriol. 169:5072–5077, 1987). The DNA was sequenced and shown to contain an open reading frame encoding a protein with a molecular weight of 58,945 and a length of 572 amino acids. The deduced amino acid sequence was identical to the published N-terminal amino acid sequence of VC167 flagellin and to four internal regions whose partial sequences were obtained by direct chemical analysis of two tryptic and two cyanogen bromide peptides of VC167 flagellin. The *C. coli* flagellin protein contains posttranslationally modified serine residues, most of which occur within a region containing two 9-amino-acid repeating peptides separated by 34 unique amino acids. Comparisons with the sequences of flagellins from other bacterial species revealed conserved residues at the amino- and carboxy-terminal regions. Hybridization data suggest the presence of a second flagellin copy located adjacent to the first on the VC167 chromosome.

The thermophilic campylobacters, *Campylobacter jejuni* and *Campylobacter coli*, are important gastrointestinal pathogens of humans (2, 5, 7, 43). These gram-negative spiral bacteria are motile by means of polar flagella, and this motility allows them to colonize the mucous lining of the intestinal tract (26, 36, 37). Once established in the mucus, *Campylobacter* species are ideally situated to produce gastrointestinal disease. Motility appears to be a primary determinant of *Campylobacter* pathogenicity, as nonmotile variants are apparently unable to colonize the gastrointestinal tracts of experimental animals or humans volunteers (5, 6). Flagella are dominant antigens during a *Campylobacter* infection and, since they are essential for motility, must be considered as virulence factors. Flagella also appear to be serodeterminants in several of the serogroups in the Lior heat-labile serotyping scheme for campylobacters (27, 28, 47, 48).

Flagellar expression is subject to both phase and antigenic variation in *Campylobacter* species. Caldwell et al. (6) showed that some strains exhibit a bidirectional transition between flagellated (Fla$^+$ phenotype) and non-flagellated (Fla$^-$ phenotype) cells, and Harris et al. (14) showed that other strains of *Campylobacter* can reversibly express flagella of different antigenic specificities. The flagellar antigenic variations in certain strains are accompanied by a unique and reversible DNA rearrangement (13). Because of the importance of flagella in virulence and their potential value as vaccine components, it is important to define the antigenic structure and expression of *Campylobacter* flagella at the molecular level.

Unfortunately, despite the importance of this pathogen, little is known at the molecular level about *Campylobacter* species. The absence of experimental genetic systems until recently (24) and the problems encountered by workers attempting to clone *Campylobacter* genes have impeded basic understanding of these organisms. In a first attempt to elucidate the molecular mechanisms of flagellar regulation in this important pathogen and the antigenic structure of the immunodominant *Campylobacter* antigen, we have cloned a flagellin structural gene from a strain of *C. coli* that produces two antigenically distinct phases of flagella. Here we report the nucleotide sequence of a *C. coli* flagellin gene and, importantly, provide evidence that the gene product is posttranslationally modified in a unique manner. We also present evidence that the *C. coli* chromosome contains two tandemly oriented copies of the flagellin gene.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The campylobacters used in this study, all originally isolated from human feces, were *C. jejuni* 81116 (D. Newell, Public Health Laboratory Service Centre for Applied Microbiology and Research, Salisbury, United Kingdom), VC74 (T. J. Trust), VC41 (NCTC 11351; National Collection of Type Culture Collections, London, United Kingdom), VC156, VC159, and *C. coli* VC167 serogroup LIO 8 (H. Lior, National Enteric Reference Centre, Ottawa, Ontario, Canada). Derivatives of *C. coli* VC167 producing either antigenic phase 1 (P1) or antigenic phase 2 (P2) flagella were selected as described by Harris et al. (14). Stock cultures were maintained at −70°C in 15% (vol/vol) glycerol-Trypticase soy broth (BBL Microbiology Systems, Cockeysville, Md.). Cultures were grown at 37°C in anaerobic jars on Mueller-Hinton agar (Difco Laboratories, Detroit, Mich.). An atmosphere containing 5% oxygen and 10% CO_2 was produced with a gas-generating kit for campylobacters (Oxoid Ltd., Basingstoke, United Kingdom).

Escherichia coli DH5 (Bethesda Research Laboratories, Gaithersburg, Md.) was used as the host for the cloning experiments with the vector pBR322. Some clones were transformed into *E. coli* GM2199 (*dam-13*::Tn*9 thr-1 ara-14 leuB6 tonA31 lacY1 tsx-79 supE44 galK2 galT22* λ$^-$ *hisG4 rpsL136 xyl-5 mtl-1 thi*-1) to map *Bcl*I sites. *E. coli* cultures were routinely grown in Luria medium (34) at 37°C with the appropriate antibiotics.

DNA extractions and manipulations. Total campylobacter

* Corresponding author.

DNAs were isolated by the method of Hull et al. (17). Plasmid DNAs were extracted by the method of Clewell and Helinski (8) followed by banding in cesium chloride-ethidium bromide. For sequence analysis, plasmid DNAs were banded twice in cesium chloride-ethidium bromide gradients. Restriction enzymes were purchased from Boehringer Mannheim Biochemicals, Indianapolis, Ind., or New England BioLabs, Inc., Beverly, Mass., and were used under the conditions recommended by the supplier.

For molecular cloning experiments, DNA fragments generated by complete digestion of VC167 DNA with various restriction endonucleases were fractionated on sucrose gradients as described by Maniatis et al. (34). Samples of each gradient fraction were electrophoresed on a 0.7% agarose gel, transferred to a nitrocellulose membrane, and hybridized to various probes. The fraction that gave the strongest signal was cloned into restriction enzyme-digested, dephosphorylated pBR322 by using standard procedures.

Oligonucleotide synthesis. Synthetic oligonucleotides were synthesized on a Biosearch Bio Sam One DNA Synthesizer or a Biosearch 8700 Synthesizer (Biosearch, San Rafael, Calif.) and purified by polyacrylamide gel electrophoresis.

Hybridizations. Oligonucleotides were labeled for hybridization probes with [γ-^{32}P]ATP (Dupont, NEN Research Products, Wilmington, Del.) by using T4 polynucleotide kinase (34; New England BioLabs), and labeled oligonucleotide was purified from unincorporated isotope by passage through a Sephacryl S200 column (Pharmacia, Piscataway, N.J.). The flagellin-specific oligonucleotide was hybridized to colony lifts of a VC167 library in *E. coli* DH5 prepared by the method of Maniatis et al. (34). Hybridizations were for 18 h at 37°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–20 mM sodium phosphate (pH 7.2)–2 mM EDTA–5× Denhardt solution (34)–100 μg of calf thymus DNA per ml–5 × 10^6 dpm of probe per ml. Filters were washed four times for 30 min each in 2× SSC at 37°C, air dried, and exposed to X-ray film (XAR-2; Eastman Kodak Co., Rochester, N.Y.) with intensifying screens for 18 h at −70°C.

Plasmid DNAs were nick translated with [α-^{32}P]dCTP by using a commercial kit (Dupont-NEN). Conditions of hybridization to Southern blots were as described by Guerry et al. (13).

Purification of flagellin. Flagella from strain VC167 P1 were separated from 24-h cultures by homogenization, and flagellin was partially purified by the sequential acid pH dissociation, ultracentrifugation, and neutral pH association procedure previously described (32). A Pharmacia fast-protein liquid chromatography system was used to separate flagellin from minor contaminants by using Superose 12 column in a buffer of 20 mM Tris–0.1 M NaCl (pH 7.4) at a flow rate of 30 ml/h.

Purification of peptides. For trypsin digestion, purified flagellin (500 μg) was incubated with 1 μg of trypsin in 20 mM Tris–10 mM $CaCl_2$ (pH 7.6) at 37°C for 12 h. Cyanogen bromide (CNBr) cleavage of purified flagellin was performed as previously described (33). Peptides were initially separated on a Superose 12 column (Pharmacia) by using conditions outlined above. After dialysis against distilled water, the peptides were concentrated under vacuum and stored at −20°C. Where necessary, further purification was achieved by using a ProRPC HR5/10 column (Pharmacia). Chromatography conditions were as follows. Solvent A was 0.1% trifluoroacetic acid in high-performance liquid chromatography water, and solvent B was 100% acetonitrile. The flow rate was 0.7 ml/min, and chromatography was performed at room temperature. Elution of peptides was achieved by using a gradient of 0 to 60% acetonitrile in 60 min. Peptides purified in this manner were dried under vacuum on a Speed-Vac (Savant Instruments Inc., Hicksville, N.Y.) and stored at −20°C.

N-terminal amino acid sequencing. Amino acid sequencing of purified peptides was performed on an Applied Biosystems 470A gas phase sequencer.

Immunological reactivity of peptides. The ability of purified peptides to bind a campylobacter flagellin-specific monoclonal antibody, Mab39 (33), was determined by either Western blot (immunoblot) as described previously (32) or by immunodot blot assay (10) with a 1:10,000 dilution of Mab39 and goat anti-mouse immunoglobulin G antibody peroxidase conjugate (Tago Inc., Immunodiagnostic Reagents, Burlingame, Calif.) at a dilution of 1:3,000.

Electrophoresis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed with a mini-slab gel apparatus (Hoeffer Scientific Instruments, San Francisco, Calif.) by the method of Laemmli (25). Protein or peptides solubilized in sample buffer were stacked in 4.5% acrylamide (100 V, constant voltage), separated in 12.5% acrylamide (200 V, constant voltage), and stained with Coomassie blue.

DNA sequencing. Double-stranded dideoxy sequencing was performed following alkaline denaturation by using [^{35}S]dATP (Dupont-NEN) and Sequenase (United States Biochemical Corp., Cleveland, Ohio) according to the manufacturer's protocol. The templates used were three plasmids (pGK200, pGK202, and pGK204) containing overlapping sequences of the flagellin gene. Primers were either commercially available pBR322 primers (New England BioLabs) or were synthesized as flagellin gene sequence became available. Primers were synthesized at approximately 250-base-pair (bp) intervals on both strands.

RESULTS

Molecular cloning of flagellin genes. The following mixture of synthetic oligonucleotides (termed A1) capable of encoding amino acid residues 4 to 9 of VC167 flagellin (30) was synthesized: ATTAATAC(A/T)AATGT(A/T)GC. When this oligonucleotide was labeled with ^{32}P and used to probe VC167 DNA digested with *Hin*dIII restriction endonuclease in a Southern blot, a single band of approximately 2.0 to 2.5 kilobase pairs (kb) was visible (data not shown). *Hin*dIII fragments of this size range were cloned into pBR322 and hybridized to ^{32}P-labeled oligonucleotide as described in Materials and Methods. Approximately 2,000 colonies were initially screened, and five clones were identified that hybridized to the A1 oligonucleotide probe. Plasmid DNA was purified from these positive clones, digested with *Hin*dIII, and transferred to nitrocellulose membranes. All five clones contained a 2.3-kb insert that hybridized to the oligonucleotide. One of the plasmids so obtained, pGK200, was used as a template in a dideoxy sequencing reaction with the A1 oligonucleotide as a primer. The sequence obtained confirmed that the clone contained a 300-bp portion of the gene that included coding information for the N terminus of flagellin (30). The position of the flagellin gene on this plasmid is shown in Fig. 1; approximately 1.5 kb of this cloned segment represents nonflagellin sequence.

To identify restriction fragments large enough to include the entire gene, the A1 oligonucleotide was hybridized to VC167 DNA, digested with various restriction enzymes, in a Southern blot analysis. When VC167 DNA was digested

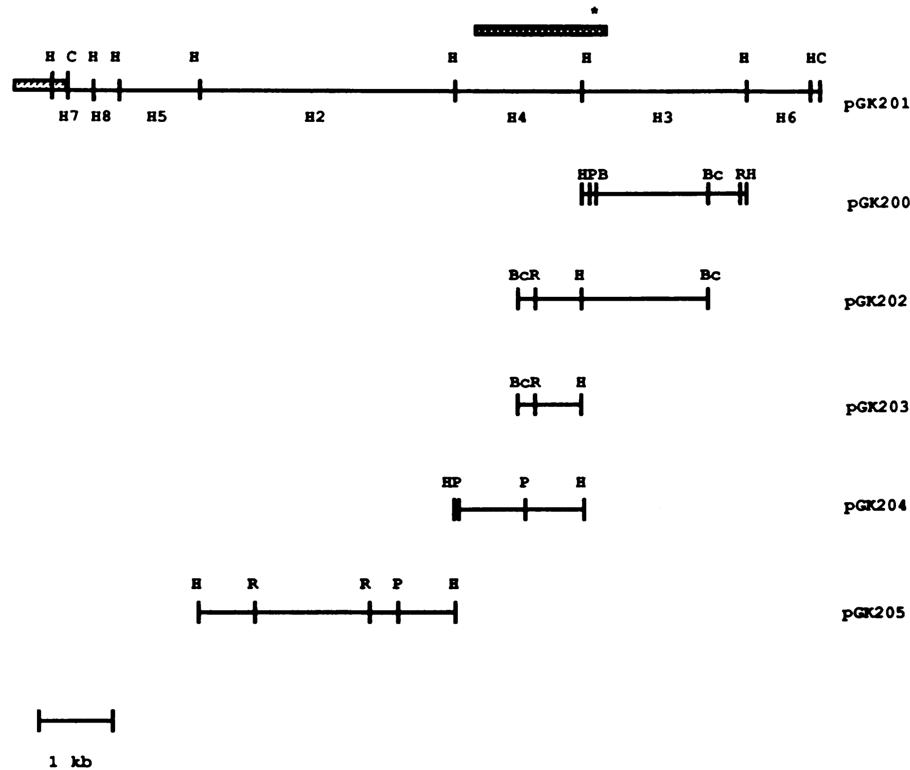

FIG. 1. Physical maps of VC167 flagellin gene clones. ——, VC167 DNA cloned into pBR322; [hatched box], adjacent pBR322 sequence; [dotted box], location of the flagellin gene; *, position at which the A1 oligonucleotide hybridizes at the 5′ end of the gene; transcription is right to left as drawn. Plasmids pGK201, pGK200, and pGK202 were independently isolated from three separate libraries. The numbers below the pGK201 map (H7, H8, etc.) refer to the fragments obtained upon digestion of pGK201 with *Hin*dIII. The largest fragment, H1, contains mostly pBR322 sequence and is not shown. Plasmid pGK203 is a deletion derivative of pGK202 obtained by deletion between the *Hin*dIII site of pGK202 and the *Hin*dIII site of pBR322 which maps to the right on the pGK202 map as drawn; plasmids pGK204 and pGK205 are subclones of the H4 and H2 fragments, respectively, of pGK201. The insert in pGK200 is equivalent to the H3 fragment of pGK201. Restriction sites are marked as follows: C, *Cla*I; H, *Hin*dIII; B, *Bgl*II; Bc, *Bcl*I; R, *Eco*RV; P, *Pst*I.

with *Cla*I restriction endonuclease, a fragment of approximately 10 kb hybridized to the A1 oligonucleotide probe. Digestion of VC167 DNA with *Bcl*I yielded a fragment of approximately 2.4 kb that hybridized to the A1 oligonucleotide. Both of these fragments were cloned into pBR322 by using procedures described above. The resulting plasmid, containing the *Cla*I fragment that hybridized to the A1 oligonucleotide, was termed pGK201. The plasmid containing the *Bcl*I fragment (cloned into the *Bam*HI site of pBR322) that hybridized to the A1 oligonucleotide was termed pGK202. Plasmid pGK201 contained within it a *Hin*dIII fragment identical in size to that originally cloned in pGK200 and a *Bcl*I fragment identical to that cloned in pGK202. The inserts in all three plasmids cross-hybridized (data not shown). Restriction maps of pGK200, pGK201, and pGK202 and some derivative plasmids are shown in Fig. 1. Attempts to detect expression of flagellin in lysates of *E. coli* containing any of the recombinant plasmids shown in Fig. 1, by using anti-*Campylobacter* flagellin Mab39 or several polyclonal antibody probes directed against *Campylobacter* flagellin, have been unsuccessful. Furthermore, preliminary attempts to subclone portions of the DNA into *E. coli* expression vectors also have not yielded immunologically detectable gene products. In many of these subcloning experiments, however, the plasmid constructions were unstable, often producing deletions (data not shown). To further confirm the identity of the clone as *Campylobacter* flagellin, we decided to obtain additional amino acid sequence information by chemical methods and compare this data with the deduced amino acid sequence from DNA sequencing.

Purification of flagellin peptides. Tryptic peptides of VC167 flagellin were isolated by chromatography on a Superose 12 column, and minor contaminants were removed from two of the major Superose peaks (Trypsin pk2 and Trypsin pk3) by using a ProRPC HR5/10 column. Peptides contained in two major Superose peaks of the CNBr cleavage (CNBr pk2 and CNBr pk3) were sequenced directly following dialysis against distilled water and concentration under vacuum. Electrophoretic analysis (data not shown) of these purified peptides revealed an apparent M_r of 29,000 for the CNBr peptide in pk2 and an apparent M_r of 25,000 for the major peptide in pk2 of the tryptic digest. When subjected to SDS-PAGE analysis, the major peptide in pk3 of the tryptic digest ran close to the dye front, indicating an apparent M_r of $<$10,000, and the peptide in pk3 of the CNBr digest was not visible on a 12.5 or 18% acrylamide gel (data not shown).

Immunological reactivity of flagellin peptides. Mab39, which was produced by immunization with *C. jejuni* VC74 flagellin, has been shown to react with a limited number of campylobacter flagellins, including VC167 (27). Two peptides selected for amino acid sequencing, CNBr pk2 and trypsin pk2, reacted with this flagellin-specific monoclonal antibody by Western blot and immunodot blot (data not shown). The smaller peptides, CNBr pk3 and trypsin pk3, failed to react with Mab39.

TABLE 1. Amino-terminal amino acid sequence of peptides from VC167 flagellin

Peptide	Residues[a]
CNBr pk2	MGFNAYNGGGAKQIIFA--IAGFM-QAGSGF-AGSGF-VGSGKNY-AIL-ASIQIV-SAA-I-ST
TRY pk2	QIIFA--IAGFM-QAGSGF-AGSGF-VGSGKNY-AIL-ASIQIV-SAA-I-STYVV-TGSGF-AGSGN
CNBr pk3	DIAETAITNLDQIRADIGSVQNQITSTINN
TRY pk3	NYNGIEDFKFDSVVI-TSVGTGLGALAEEIN

[a] Amino acid residues are designated by the single-letter nomenclature. -, Residue unidentifiable by sequencing. The underlining indicates an amino acid repeat. The sequences of CNBr pk2 and TRY pk2 are aligned to indicate the overlapping residues.

N-terminal amino acid sequence of flagellin peptides. The N-terminal amino acid sequences of these purified peptides are presented in Table 1. The two peptides which bind Mab39 have overlapping amino acid sequences, suggesting that they form a continuous region of the primary sequence of the flagellin monomer. These sequences, however, do not show homology with flagellin sequences from other organisms (9, 20). The sequences of both these peptides have a number of unidentified residues that are most likely due to posttranslational modifications. The two other peptides purified, CNBr pk3 and trypsin pk3, do not appear to be as highly modified as the N-terminal amino acid sequence of the two peptides which bind Mab39, and the sequences appear to be unique, indicating that they are from different regions of the primary sequence of the flagellin monomer. The N-terminal amino acid sequence of CNBr pk3 displays homology with the C-terminal amino acid sequence of flagellin from *S. typhimurium* (20) and *Bacillus subtilis* (9; see below and Fig. 3).

DNA sequence of campylobacter flagellin gene. The complete DNA sequence of campylobacter flagellin was determined by sequencing the double-stranded DNA contained in pGK200, pGK202, and pGK204 (Fig. 1). The results, depicted in Fig. 2, indicate an open reading frame of 1,716 bp capable of encoding 572 amino acids. Base composition analysis of the open reading frame indicates an overall G+C

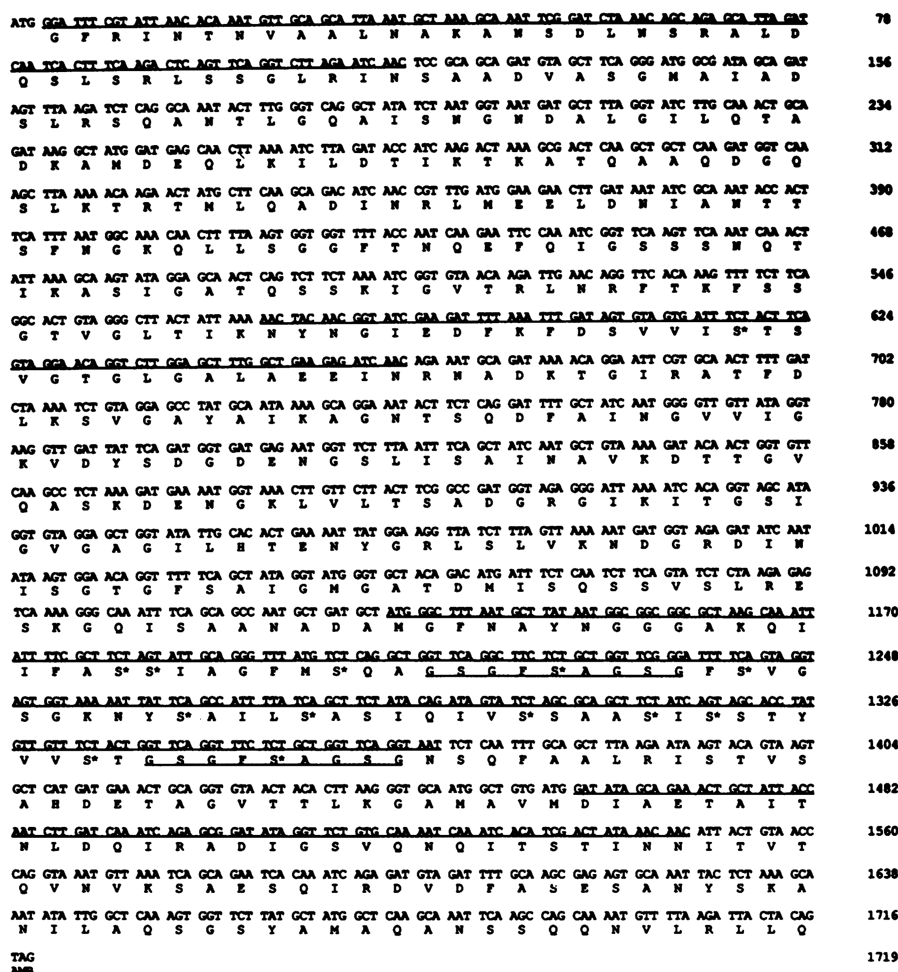

FIG. 2. DNA and protein sequence of *C. coli* VC167 flagellin. The overlined peptides are those whose amino acid sequences were also determined chemically. *, Posttranslationally modified residue whose identity was determined by DNA sequencing. The 9-amino-acid repeating peptides are underlined. The numbers at the right indicate the base pairs. The N-terminal methionine residue is known to be removed from the mature protein (28). Amino acids are designated by the single-letter nomenclature.

```
a)  MAQVINTNSLSLLTQNNLNKSQSALGTAIERLSSGLRINSAKDDAAGQAIANRFTANIKGLTQASRNANDGISIAQTTEGALNEINNNLQRVRELAVQSA
b)   GFRINTNVAALNAKANSDLNSRALDQSLSRLSSGLRINSAADVASGMAIADSLRSQANTLGQAISNGNDALGILQTADKAMDEQLKILDTIKTKATQAA
c)    MRINHNIAALNTLNRLSSNNSASQKNMEKLSSGLRINRAGDDAAGLAISEKMRGQIRGLEMASKNSQDGISLIQTAEGALTETHAILQRVRELVVQAG

a)  N-STNSQSDL-DSIQAEITQRLNEIDRVNGQTQFSGVKVLAQD-NTLTIQVGANDGETIDIDLKQINSQTLG-------(223 aa)------------
b)  Q-DGQSLKTR-TMLQADINRLMEELDNIANTTSFNGKQLLSGGFTNQEFQIGSSSNQTIKASIGATQSSKIG-------(399 aa)------------
c)  NTTGQDKATDLQSIQDEISALTDEIDGISNRTEFNGKKLLDGTYKVDTATPANQKNLVFQIGANATQQISVN-------(36 aa)-------------

a)  LAEQAAKTTENPLQKIDAALAQVDTLRSDLGAVQNRFNSAITNLGNTVNNLSSARSRIEDSDYATEVSNMSRAQILQQAGTSVLAQANQVPQNVLSLLR
b)  AGVTTLKGAMAVMDIAETAITNLDQIRADIGSVQNQITSTINNITVTQVNVKSAESQIRDVDFASESANYSKANILAQSGSYAMAQANSSQQNVLRLLQ
c)  AADTADIGFDAQLKVVDEAINQVSSQRAKLGAVQNRLEHTINNLSASGENLTAAESRIRDVDMAKEMSEFTKNNILSQASQAMLAQANQQPQNVLQLLR
```

FIG. 3. Comparisons of the amino acid (aa) sequences of the flagellins of *S. typhimurium* (16) (a), *C. coli* VC167 (b), and *B. subtilis* (8) (c). The amino- and carboxy-terminal sequences of each protein (separated by a region of variable length in the middle) are shown in the single-letter code. Identical residues are indicated by vertical dashes.

content of 37.7%, slightly higher than the overall 30 to 35% G+C content reported for the *C. coli* genome (4, 39). However, codon usage shows a strong bias for codons with A or T in the third position. The amino-terminal peptide (30) and the four internal peptides that had been sequenced (Table 1) are overlined in the amino acid sequence. The DNA sequence indicates that the blocked amino acids observed during protein sequencing in three of these peptides are serine residues. The position of these posttranslationally modified serines is designated by an asterisk in the sequence. Another interesting feature in this region is a 9-amino-acid peptide, GSGFSAGSG, that is repeated twice; these peptides are underlined in the amino acid sequence. The repeats contain a posttranslational modification at the middle serine but not at the first or third. The repeats are separated by 34 amino acids that include 7 posttranslationally modified serines and 4 unmodified serines.

The FASTP program of Lipman and Pearson (29), which aligns similar amino acid sequences while allowing for insertions and deletions, was used to compare the deduced flagellin sequence of *C. coli* VC167 with those of *B. subtilis* and *S. typhimurium* (Fig. 3). Regions of homology were evident in the amino and carboxy termini but not in the internal region of the sequence. At the N terminus, amino acids 3 to 169 of VC167 flagellin displayed 36.3% sequence identity to *B. subtilis* flagellin and 33.7% sequence identity to *S. typhimurium* flagellin. At the C terminus of VC167 flagellin, amino acids 474 to 572 displayed 36.4 and 41.9% identity with *B. subtilis* and *S. typhimurium*, respectively. No significant homology was observed between residues 169 to 474 of VC167 flagellin and either *B. subtilis* or *S. typhimurium* flagellin.

Evidence for a second copy of flagellin. The Al oligonucleotide hybridized to only a single band of *Hin*dIII-digested VC167 DNA in a Southern blot. However, when the pGK200 clone, containing approximately 300 bp of coding information at the 5′ end of the gene, is used to probe *Hin*dIII-digested VC167 DNA by Southern blot hybridization, two bands of approximately 2.3 and 1.3 kb hybridize (data not shown). Since the fragment cloned in pGK200 contains nonflagellin sequences also, the second band could be due to duplication of these nonflagellin gene sequences. To determine the number of copies of flagellin in VC167, a probe was constructed containing only DNA internal to the flagellin gene. Plasmid pGK202 was deleted, as seen in Fig. 1, from a *Hin*dIII site internal to the insert to the *Hin*dIII site of pBR322. When this plasmid, termed pGK203 (Fig. 1), was used to probe VC167 DNA digested with *Hin*dIII restriction endonuclease, two bands hybridized that were of the same size as the H2 and H4 fragments of pGK201 (Fig. 1; data not shown). When the same probe was used in a Southern blot with *Eco*RV-restricted VC167 DNA (Fig. 4, lane A) three bands were seen. Since there is an *Eco*RV site within the pGK203 insert, digestion with *Eco*RV would be expected to produce two fragments (of approximately 2.6 and 1.8 kb) hybridizing to this probe (Fig. 1). The third band, of 1.6 kb, indicates that there is a second copy of at least the internal region of flagellin present in VC167. Furthermore, this 1.6-kb *Eco*RV fragment mapped to within the H2 fragment of pGK201 (Fig. 1 and below). The hybridization blot shown in Fig. 4 also shows the reaction of the internal flagellin probe with other strains of campylobacter. Lanes B and C show that VC159 and VC156, two *C. jejuni* strains of the Lior 8 serotype, show a hybridization pattern identical to that of VC167. Hybridization to VC41 (unknown Lior serotype), VC74 (LIO 11), and 81116 (LIO 6) is seen in lanes F, E, and D, respectively. The intensity of hybridization to the DNA from these strains is greatly reduced, suggesting that this

A B C D E F

FIG. 4. Southern blot hybridization of campylobacter DNAs to pGK203. Campylobacter DNAs (2 μg) were digested with restriction enzyme *Eco*RV, run on a 0.7% agarose gel, and transferred to a nylon membrane. Hybridization to plasmid pGK203 was as described in Materials and Methods. Lanes: A, VC167 (LIO 8); B, 156 (LIO 8); C, 159 (LIO 8); D, 81116 (LIO 6); E, VC74 (LIO 11); F, VC41 (unknown LIO serotype). Hybridization of the vector sequences (pBR322) to campylobacter DNAs does not result in any bands under the conditions used (11).

internal portion of the gene differs among Lior serotypes. It is also interesting that there are three bands hybridizing in strain 81116 (Fig. 4, lane D), which is a strain that undergoes an on-off flagellar switch but which has not yet been demonstrated to undergo an antigenic variation.

When pGK200, pGK202, or pGK203 was used as a probe in Southern hybridization reactions against VC167 DNA digested with *Cla*I, only a single band hybridized, and this single band was the piece of DNA cloned in plasmid pGK201 (data not shown). These data suggest that all copies of flagellin are represented on the large *Cla*I fragment in pGK201. When pGK200, which contains only the N terminus of the flagellin gene, was used to probe a *Hin*dIII digest of pGK201, three bands hybridized: H1 (pBR322), H3 (pGK200), and H4, a 1.3-kb fragment that mapped downstream of the *Hin*dIII fragment represented in pGK200 and overlapped pGK202 (Fig. 1). When pGK202 was used to probe *Hin*dIII-digested pGK201 DNA, fragments H1 (pBR322), H3 (pGK200, which overlaps pGK202), H4 (which also overlaps pGK202), and H2, a 3.5-kb fragment that maps downstream of H4, hybridized. These data suggest that a second N terminus of flagellin exists on H4 and a second internal fragment of the gene is present on H2. In addition, hybridizations with many of the synthetic oligonucleotides used as sequencing primers also show that sequences homologous to the sequenced gene lie just downstream of it in the same orientation (data not shown).

DISCUSSION

This work represents the first report of molecular cloning of a surface antigen, flagellin, from the *C. coli-C. jejuni* group into *E. coli*, and the first report of a DNA sequence of a chromosomally encoded structural gene from this genus. The cloned gene was identified by synthetic oligonucleotide hybridization because it did not express protein at immunologically detectable levels in the foreign host. This is not surprising, however, since there are few reports of expression of *Campylobacter* genes in *E. coli* (45). Significantly, comparison of the DNA sequence with the amino acid sequence determined for several internal peptides, including immunogenic regions of the molecule, revealed that the *Campylobacter* flagellin was posttranslationally modified at numerous serine residues. Since this modification likely does not occur in *E. coli*, this may explain the inability to express detectable levels of immunologically reactive recombinant VC167 flagellin.

The most likely modification of the serine residues in the *Campylobacter* flagellin is phosphorylation. This is a common modification of this amino acid in eucaryotic systems, and radioisotopic studies have shown that inorganic phosphorus is incorporated into *Campylobacter* flagellin (31). This is an unusual finding, since the most common modification of amino acids in surface protein appendages of procaryotes is methylation. For example, the flagellins from *Salmonella typhimurium* (1, 21) and *Spirillum serpens* (12) have been shown to contain methyl-lysine residues, while the fimbrins of *Neisseria*, *Moraxella*, *Pseudomonas*, *Bacteroides*, and *Vibrio* species have methylated phenylalanine as the first residue (11, 15, 35, 40, 44).

Charge imparted to the isolated internal flagellin peptides by the modification of serine residues presumably accounted for discrepancies between the apparent M_rs of these peptides measured by SDS-PAGE and their M_rs as deduced from DNA sequence. Certainly, aberrant migration characteristics of charged peptides on SDS-PAGE are well described. However, charge effects during SDS-PAGE determination of intact flagellin M_r were less noticable. Purified VC167 P1 flagellin has an M_r of 61,500 and P2 flagellin has an M_r of 59,500, as determined by SDS-PAGE (14). The calculated M_r of 58,814 (following removal of the first methionine residue) for the predicted unmodified product of the flagellin gene sequenced in this study is clearly an underestimate. Our data indicate the presence of a minimum of 13 modified residues, which, if phosphate groups, would increase the M_r to 59,841. Therefore, the identity of this flagellin gene as P1 or P2 cannot be determined at this time by M_r. In fact, the current data do not rule out the possibility that phase 1 and phase 2 flagellins differ only in the nature or degree of their posttranslational modifications. However, since the hybridization data indicate the presence of at least a partial copy of a second flagellin gene, the most likely explanation for the ability of the strain to synthesize serologically distinct flagellins is the alternate expression of distinct genes.

Flagellins from numerous bacteria have been shown to be conserved at the amino and carboxy termini, and it would appear that the *Campylobacter* flagellin gene sequenced here conforms to this general rule. The structural conservation at the N-terminal end of the flagellin molecule may reflect an involvement of this area of the gene in regulatory functions (46), and both the amino and carboxy termini of flagellin have been shown to be involved in polymerization and excretion of *S. typhimurium* flagellin monomers (16). Indeed, Kuwajima (23) recently showed by deletion analysis that the smallest *E. coli* flagellin capable of forming flagellar filaments retained only the N-terminal 193 residues and the C-terminal 117 residues. The central portion of the *E. coli* flagellin sequence forms a surface-exposed antigenic domain (18, 22), and this is in keeping with the demonstration by Joys (19) that a determinant involved in the H-antigen serospecificity is located in the region of highest sequence diversity, between residues 182 and 299 of the 493-residue *Salmonella* flagellin molecule. It seems that this central region may also be variable in *Campylobacter* flagellins. Indeed, the DNA cloned in plasmid pGK203, an 1,176-bp fragment from the central region of the gene, may include the region involved in Lior serospecificity. This is based on the observation that this plasmid hybridizes much more intensely to Lior 8 strains than to strains of other serospecificities. While the number of serotypes used in this study is limited, preliminary experiments with 15 additional serotypes have supported the idea that the LIO 8-specific determinants reside within the DNA cloned in pGK203 (S. A. Thornton, S. M. Logan, P. Guerry, and T. J. Trust, manuscript in preparation).

The epitope recognized by Mab39, an antibody which can distinguish a number of antigenically distinct *Campylobacter* flagellins (33), is also located on a peptide whose N terminus is located centrally within the intact protein and whose coding information is included in pGK203. This peptide includes a 9-amino-acid repeat separated by 34 nonrepeating amino acids. Most of the observed serine modifications occur in this region of the molecule. Tandemly arranged, multiply repeating epitopes in the surface antigens of other pathogens are immunodominant (38), and variation in such repeats can contribute to serovariation (3, 41). Thus, it will be interesting to determine whether this 9-amino-acid sequence, of which there are only two copies, contributes to the immunodominance of flagellin in campylobacter infections or to its antigenic variation.

The data presented here suggest that at least part of the flagellin gene is duplicated in *Campylobacter* VC167 and that

all the DNA sequences related to flagellin are located on the fragment of the chromosome cloned in pGK201. This clustering of flagellin gene information is distinct from the organization of flagellin genes found in *S. typhimurium*, in which the two genes, H1 and H2, are located on different parts of the chromosome (42). Phase variation in *Salmonella* species is regulated by the orientation of an invertible DNA sequence adjacent to H2 that controls expression of H2 and a repressor of H1 (42). A DNA rearrangement has been described that is associated with flagellar antigenic variation in *Campylobacter* species (13). The DNA involved in this rearrangement, which maps very near to a ribosomal RNA locus and is not located on the pGK201 plasmid (S. M. Logan, P. Guerry, and T. J. Trust, manuscript in preparation), presumably represents a nonadjacent regulatory locus. Work is ongoing in our laboratories to investigate the nature of the switching mechanism. Complete DNA sequencing of the flagellin gene cluster is also under way and should reveal information on the epitopes involved in antigenic differences between P1 and P2 flagellins, as well as information on the regulation of the two genes. Additionally, we are constructing flagellin mutants by using the methods described by Labigne-Roussel et al. (24) for studies on regulation, pathogenesis, and the biological significance of antigenic variation in *Campylobacter* species.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the Medical Research Council of Canada to T.J.T. and by U.S. Navy Research and Development Command, Research Work Unit No. 61102A 3M161102BS13 AK.111.

We thank Sharon Cubbage, Olegario Majam, Steve Martin, and Li-Ting Liang for excellent technical assistance and Malabi Venkatsan of the Walter Reed Army Institute of Research for synthesis of some of the oligonucleotides used in this study.

LITERATURE CITED

1. **Ambler, R. P., and M. W. Rees.** 1959. *N*-methyl-lysine in bacterial flagellar protein. Nature (London) **184**:56–57.
2. **Butzler, J. P., and M. B. Skirrow.** 1979. *Campylobacter* enteritis. Clin. Gastroenterol. **8**:737–765.
3. **Beachey, E. H., J. M. Seyer, and A. H. Kang.** 1978. Repeating covalent structure of streptococcal M-protein. Proc. Natl. Acad. Sci. USA **75**:3163–3167.
4. **Belland, R. J., and T. J. Trust.** 1982. Deoxyribonucleic acid relatedness between thermophilic members of the genus *Campylobacter*. J. Gen. Microbiol. **128**:2515–2522.
5. **Black, R. E., M. M. Levine, M. L. Clements, T. P. Hughes, and M. J. Blaser.** 1988. Experimental *Campylobacter jejuni* infection in humans. J. Infect. Dis. **157**:472–479.
6. **Caldwell, M. B., P. Guerry, E. C. Lee, J. P. Burans, and R. I. Walker.** 1985. Reversible expression of flagella in *Campylobacter jejuni*. Infect. Immun. **50**:941–943.
7. **Chowdhury, M. N. H.** 1984. *Campylobacter jejuni* enteritis: a review. Trop. Geogr. Med. **36**:215–222.
8. **Clewell, D. B., and D. R. Helinski.** 1972. Effect of growth conditions on the formation of the relaxation complex of supercoiled ColE1 deoxyribonucleic acid and protein in *Escherichia coli*. J. Bacteriol. **110**:1135–1146.
9. **Delange, R. J., J. Y. Chang, J. H. Sharper, and A. M. Glazer.** 1976. Amino acid sequence of flagellin of *Bacillus subtilis* 168. J. Biol. Chem. **254**:705–711.
10. **Feutrier, J., W. W. Kay, and T. J. Trust.** 1986. Purification and characterization of pili from *Salmonella enteritidis*. J. Bacteriol. **168**:357–364.
11. **Froholm, L. O., and K. Sletten.** 1977. Purification and N-terminal sequence of a fimbrial protein from *Moraxella nonliquefaciens*. FEBS Lett. **73**:29–32.
12. **Glazer, A. N., R. J. DeLange, and R. J. Martinez.** 1969. Identification of L-methyl-lysine in *Spirillum serpens* flagella and of *N*-dimethyl-lysine in *Salmonella typhimurium* flagella. Biochem. Biophys. Acta **188**:164–165.
13. **Guerry, P., S. M. Logan, and T. J. Trust.** 1988. Genomic rearrangements associated with antigenic variation in *Campylobacter coli*. J. Bacteriol. **170**:316–319.
14. **Harris, L. A., S. M. Logan, P. Guerry, and T. J. Trust.** 1987. Antigenic variation of *Campylobacter* flagella. J. Bacteriol. **169**:5066–5071.
15. **Hermondson, M. A., K. C. S. Chen, and T. M. Buchanan.** 1978. *Neisseria* pili protein: amino terminal amino acid sequences and identification of an unusual amino acid. Biochemistry **17**:422–445.
16. **Homma, M., H. Fujita, S. Yamaguchi, and T. Iino.** 1987. Regions of *Salmonella typhimurium* flagellin essential for its polymerization and excretion. J. Bacteriol. **169**:291–296.
17. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33**:933–938.
18. **Iino, T.** 1977. Genetics of structure and function of bacterial flagella. Annu. Rev. Genet. **11**:161–182.
19. **Joys, T. M.** 1976. Identification of an antibody binding site in the phase-1 flagella protein of *Salmonella typhimurium*. Microbios **15**:221–228.
20. **Joys, T. M.** 1985. The covalent structure of the phase-1 flagellar filament protein of *Salmonella typhimurium* and its comparisons with other flagellins. J. Biol. Chem. **247**:5180–5193.
21. **Joys, T. M., and H. Kim.** 1979. Identification of *N*-methyl-lysine residues in the phase-1 flagellar protein of *Salmonella typhimurium*. Microbios Lett. **7**:65–68.
22. **Kuwajima, G.** 1988. Flagellin domain that affects H antigenicity of *Escherichia coli* K-12. J. Bacteriol. **170**:485–488.
23. **Kuwajima, G.** 1988. Construction of a minimum-size functional flagellin of *Escherichia coli*. J. Bacteriol. **170**:3305–3309.
24. **Labigne-Roussel, A., P. Couroux, and L. S. Tompkins.** 1988. Gene disruption and replacement as a feasible approach for mutagenesis of *Campylobacter jejuni*. J. Bacteriol. **170**:1704–1708.
25. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
26. **Lee, A., J. L. O'Rourke, P. J. Barrington, and T. J. Trust.** 1986. Mucus colonization by *Campylobacter jejuni*: a mouse cecal model. Infect. Immun. **51**:536–546.
27. **Lior, H., and D. L. Woodward.** 1985. Serogrouping and biotyping of *Campylobacter*—an update, p. 217. *In* A. D. Pearson, M. B. Skirrow, H. Lior, and B. Rowe (ed.), *Campylobacter* III. Public Health Laboratory Service, London.
28. **Lior, H., D. L. Woodward, J. A. Edgar, L. J. Laroche, and P. Gill.** 1983. Serotyping of *Campylobacter jejuni* by slide agglutination based on heat-labile antigenic factors. J. Clin. Microbiol. **15**:761–768.
29. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
30. **Logan, S. M., L. A. Harris, and T. J. Trust.** 1987. Isolation and characterization of *Campylobacter* flagellins. J. Bacteriol. **169**:5072–5077.
31. **Logan, S. M., and T. J. Trust.** 1982. Outer membrane characteristics of *Campylobacter jejuni*. Infect. Immun. **38**:898–906.
32. **Logan, S. M., and T. J. Trust.** 1983. Molecular identification of surface protein antigens of *Campylobacter jejuni*. Infect. Immun. **42**:675–682.
33. **Logan, S. M., and T. J. Trust.** 1986. Location of epitopes on *Campylobacter jejuni* flagella. J. Bacteriol. **168**:739–745.
34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **McKern, N. M., I. J. O'Donnell, A. S. Inglis, D. J. Stewart, and B. L. Clark.** 1983. Amino acid sequence of pilin from *Bacteroides nodosus* (strain 198), the causative organism of ovine footrot. FEBS Lett. **164**:149–153.
36. **Morooka, T., A. Umeda, and K. Amako.** 1985. Motility as an

intestinal colonization factor for *Campylobacter jejuni*. J. Gen. Microbiol. **131**:1973–1980.

37. **Newell, D. G.** 1986. Monoclonal antibodies directed against the flagella of *Campylobacter jejuni*: production, characterization, and lack of effect on the colonization of infant mice. J. Hyg. **96**:131–141.
38. **Nussenzweig, V., and R. S. Nussenzweig.** 1985. Circumsporozoite proteins of malaria parasites. Cell **42**:401–403.
39. **Owen, R. J.** 1983. Nucleic acids in the classification of campylobacters. Eur. J. Microbiol. **2**:367–377.
40. **Sastry, P. A., B. B. Finlay, B. L. Pasloske, W. Paranchych, J. R. Pearlstone, and L. B. Smillie.** 1985. Comparative studies of the amino acid and nucleotide sequences of pilin derived from *Pseudomonas aeruginosa* PAK and PAO. J. Bacteriol. **164**: 571–577.
41. **Scott, J. R., S. K. Hollingshead, and V. A. Fishetti.** 1986. Homologous regions within M protein genes in group A streptococci of different serotypes. Infect. Immun. **52**:609–612.
42. **Silverman, M., and M. Simon.** 1980. Phase variation: genetic analysis of switching mutants. Cell **19**:845–854.
43. **Skirrow, M. B.** 1977. *Campylobacter* enteritis: a "new" disease. Br. Med. J. **2**:9–11.
44. **Taylor, R. K., V. L. Miller, D. B. Furlong, and J. J. Mekalanos.** 1987. Use of *pho*A fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc. Natl. Acad. Sci. USA **84**:2833–2837.
45. **Walker, R. I., M. B. Caldwell, E. C. Lee, P. Guerry, T. J. Trust, and G. M. Ruiz-Palacios.** 1986. Pathophysiology of *Campylobacter* enteritis. Microbiol. Rev. **50**:81–94.
46. **Wei, L.-N., and T. M. Joys.** 1985. Covalent structure of three phase-1 flagellar filament proteins of *Salmonella*. J. Mol. Biol. **186**:783–791.
47. **Wenman, W. M., J. Chai, T. J. Louie, C. Goudreau, H. Lior, D. G. Newell, A. Pearson, and D. E. Taylor.** 1985. Antigenic analysis of *Campylobacter* flagellar protein and other proteins. J. Clin. Microbiol. **21**:108–112.
48. **Wenman, W. M., D. E. Taylor, and H. Lior.** 1985. The flagella protein determines *Campylobacter* heat-labile serotype, p. 212. *In* A. D. Pearson, M. B. Skirrow, H. Lior, and B. Rowe (ed.), *Campylobacter* III. Public Health Laboratory Service, London.