JOURNAL OF BACTERIOLOGY, July 1989, p. 3890–3900
0021-9193/89/073890-11$02.00/0

Vol. 171, No. 7

L-, P-, and M-Ring Proteins of the Flagellar Basal Body of *Salmonella typhimurium*: Gene Sequences and Deduced Protein Sequences

CHRISTOPHER J. JONES,† MICHIO HOMMA,‡ AND ROBERT M. MACNAB*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Received 22 December 1988/Accepted 5 April 1989

The *flgH*, *flgI*, and *fliF* genes of *Salmonella typhimurium* encode the major proteins for the L, P, and M rings of the flagellar basal body. We have determined the sequences of these genes and the *flgJ* gene and examined the deduced amino acid sequences of their products. FlgH and FlgI, which are exported across the cell membrane to their destinations in the outer membrane and periplasmic space, respectively, both had typical N-terminal cleaved signal-peptide sequences. FlgH is predicted to have a considerable amount of β-sheet structure, as has been noted for other outer membrane proteins. FlgI is predicted to have an even greater amount of β-structure. FliF, as is usual for a cytoplasmic membrane protein of a procaryote, lacked a signal peptide; it is predicted to have considerable α-helical structure, including an N-terminal sequence that is likely to be membrane-spanning. However, it had overall a quite hydrophilic sequence with a high charge density, especially towards its C terminus. The *flgJ* gene, immediately adjacent to *flgI* and the last gene of the *flgB* operon, encodes a flagellar protein of unknown function whose deduced sequence was hydrophilic and may correspond to a cytoplasmic protein. Several aspects of the DNA sequence of these genes and their surrounds suggest complex regulation of the flagellar gene system. A notable example occurs within the *flgB* operon, where between the end of *flgG* (encoding the distal rod protein of the basal body) and the start of *flgH* (encoding the L-ring protein) there was an unusually long noncoding region containing a potential stem-loop sequence, which could attenuate termination of transcription or stabilize part of the transcript against degradation. Another example is the interface between the *flgB* and *flgK* operons, where transcription termination of the former may occur within the coding region of the latter.

Cells of *Salmonella typhimurium* swim by rotating helical flagellar filaments, which are attached to motors embedded in the cell wall. They modulate their swimming by switching the direction of rotation of the motors. Under control of environmental sensory information, this results in selective migration, or taxis (for review, see references 44, 45, and 62).

The basal body constitutes a major portion of the flagellar organelle and consists of four rings (L, P, S, and M) mounted on a central rod (1, 13) (Fig. 1). The M ring is integral to the inner membrane of the cell (14) and may be connected to the flagellar rod via the S ring (60). It is not known whether the M ring is actively involved in energy transduction, but its location in the cytoplasmic membrane suggests that it plays some role in the process. The S, or supramembrane, ring lies just distal to the M ring; image reconstructions of the basal body suggest that the two rings may be features of a single structure (60). The L and P rings lie in the outer membrane and the periplasmic space (possibly in interaction with the peptidoglycan layer), respectively (14). They form the outer cylinder, which is believed to participate only passively in the motor mechanism, acting as a bushing for the central rod. The P and L rings form a large pore, through which the rod passes (1, 60).

The P-ring protein possesses an N-terminal signal peptide which is cleaved during export of the protein to the periplasmic space (24). The L-ring protein is also processed during its export to the outer membrane (26). By analogy with other cytoplasmic membrane proteins of procaryotes (52), the M-ring protein is not expected to be processed by cleavage of a signal peptide.

The flagellar gene system is complex, involving about 40 genes organized into a regulon of 13 operons, with a master level of control of gene expression and several secondary levels of control (35, 36). Almost all of the genes are clustered in just three regions of the chromosome (45), one of which (region I) codes for all but one of the known components of the flagellar basal body. The genes encoding the L-, P-, and M-ring proteins have been identified as *flgH*, *flgI*, and *fliF*, respectively (1, 23, 24, 26, 30, 40). The first two of these genes are located in region I, while *fliF* is in region III.

In this paper, we present the DNA sequences and deduced product amino acid sequences of these three ring protein genes. We also present the corresponding data for *flgJ*, a flagellar gene of unknown function that follows *flgI* and is the terminal gene of the *flgB* operon (41).

Flagellar gene symbols used in this paper follow the new unified nomenclature for *S. typhimurium* and *Escherichia coli* (28). The *S. typhimurium* genes whose sequences are presented here, namely, *flgH*, *flgI*, *flgJ*, and *fliF*, correspond to *flaFVIII*, *flaFIX*, *flaFX*, and *flaAII.1*, respectively, in the old nomenclature.

MATERIALS AND METHODS

Plasmids. Plasmids pMH64 (containing *flgE* through *flgG*, and part of *flgH*), pOH20 (containing *flgH* and *flgI*), pMH41

* Corresponding author.
† Present address: ERATO, 5-9-5 Tokodai, Toyosato, Tsukuba, Ibaraki 300-26, Japan.
‡ Present address: Laboratory of Medical Mycology, Research Institute for Disease Mechanism and Control, Nagoya University School of Medicine, Showa-ku, Nagoya 466, Japan.

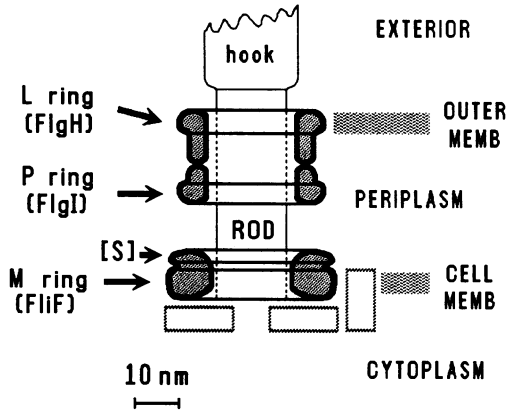

FIG. 1. Cartoon of the flagellar basal body of *S. typhimurium*, emphasizing the locations of the L, P, and M rings and the identities of their component proteins, whose deduced sequences were obtained in this study. Whether the S ring consists of a distinct protein from FliF or simply represents a different domain of FliF is not known. The proposed locations of the Mot proteins in the membrane (memb) surrounding the M ring, and of the switch proteins as a complex mounted onto the cytoplasmic face of the M ring, are indicated in hatched outline.

(containing *flgI* through *flgL*), and its *Bal* 31 deletion derivatives, and pAMH33 (containing *fliF*) have been described before (25, 26, 33).

Chemicals and enzymes. Acrylamide and bisacrylamide (Bio-Rad Laboratories, Rockville Centre, N.Y.) were electrophoresis grade; urea (International Biotechnologies, Inc., New Haven, Conn.) was molecular-biology grade; all other chemicals used were reagent grade. Restriction enzymes and T4 DNA ligase were from standard commercial sources.

DNA sequencing of the *flgH*, *flgI*, *flgJ*, and *fliF* genes. DNA sequencing of *flgH* through *flgJ* was performed on portions of pMH64, pOH20, and various *Bal* 31 deletion derivatives of pMH41. Restriction fragments were in some cases reduced in size by 5'-to-3' digestion either with exonuclease (ExoIII) or by the single-stranded sequential method (11) with the CYCLONE I Biosystem (International Biotechnologies, Inc.), according to manufacturer's instructions, except that after the tailing reaction the DNA was extracted with phenol. For ExoIII digestion of the pMH41 *Bal* 31 derivatives, the plasmid was digested with *Eco*RI, cloned into phage M13 mp18 or mp19, cut with *Bam*HI and *Pst*I at the linker region, digested first with ExoIII and then with mung bean nuclease, and ligated with T4 DNA ligase. Fragments were cloned into M13 and used to obtain the gene sequences by the strategy shown in Fig. 2. The *fliF* gene was sequenced from a *Kpn*I restriction fragment of plasmid pAMH33 and restriction subfragments and CYCLONE derivatives thereof (Fig. 2).

Manipulation of DNA was done by conventional procedures (48). Sequencing was done by the dideoxynucleotide method (55), with the modified T7 DNA polymerase (65) Sequenase (U.S. Biochemical Corp., Cleveland, Ohio), dITP in place of dGTP, and either universal primers or special primers synthesized by a facility in the Department of Chemistry, Yale University. In some instances, we used purified Klenow fragment and 7-deaza-dGTP (Boehringer Mannheim Biochemicals, Indianapolis, Ind.). Labeling was done with [^{35}S]dATP (1,000 Ci/mmol; Amersham Corp., Arlington Heights, Ill.).

Sequence analyses. Routine sequence manipulations, analysis of DNA sequences for codon usage (57) and coding probability (17), and analysis of protein sequences by the algorithm of Engelman et al. (16) were all carried out on an IBM personal computer. Sequence comparisons against the GenBank DNA database (version 56) and the National Biomedical Research Foundation protein data base (version 16) and analyses of protein sequences by the algorithms of Chou and Fasman (10), Garnier et al. (18), and Kyte and Doolittle (42) were made on a VAX computer in the Biomedical Computing Unit, Yale University.

RESULTS

Sequences of the *flgH*, *flgI*, *flgJ*, and *fliF* genes. The DNA sequences of the genes, together with the deduced amino acid sequences of their products, are shown in Fig. 3 to 6. These sequences have been deposited in GenBank under accession numbers M24466 through M24469. In the case of *fliF*, we also present a comparison with the homologous *E. coli* gene (B. Frantz, D. Bartlett, and P. Matsumura, manuscript in preparation).

The *flgH*, *flgI*, and *flgJ* genes were identified by the following criteria. (i) All three genes are in the same operon (37, 40, 41) and so must be transcribed from the same DNA strand. An examination of all six possible reading frames of the sequence revealed only three open reading frames that were adjacent, in the same orientation, and large enough to encode FlgH, FlgI, and FlgJ. (ii) The sizes of the genes were consistent with the experimentally measured apparent mo-

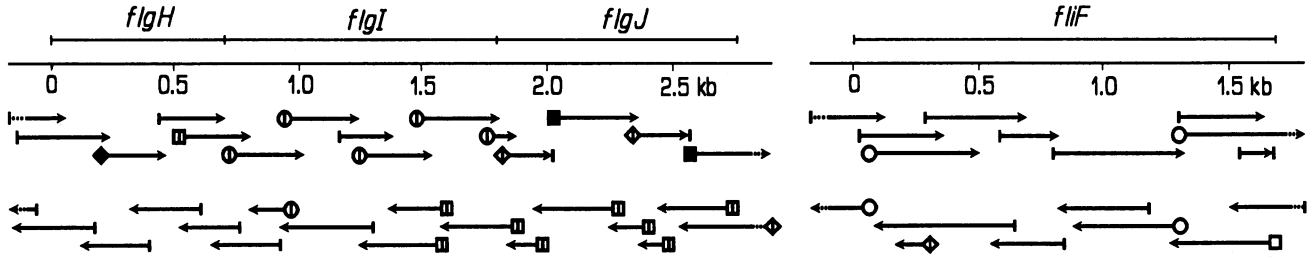

FIG. 2. Strategies used to obtain the DNA sequences of the *flgH*, *flgI*, and *flgJ* genes (left) and the *fliF* gene (right). *flgH*, *flgI*, and *flgJ* are the last three genes of the *flgB* operon in flagellar region I of the chromosome, while *fliF* is the first gene of the *fliF* operon in flagellar region III. The fragments used for sequencing the nontemplate strand are shown above, and those for the template strand are shown below. Fragments starting with a restriction site are indicated as follows: *Eco*RI (○), *Kpn*I (□), *Mlu*I (■), *Sal*I (◆). Fragments generated by deletions are shown as follows: CYCLONE (◖), ExoIII (◔), *Bal* 31 (◫). Fragments sequenced with a custom synthesized primer are shown (◈). Where the limit to the sequence available from a fragment is a restriction site at its 3' end, this is shown by a bar at the head of the arrow. Sequence beyond the region of interest is indicated by dashed lines.

TABLE 1. Apparent and deduced molecular masses of flagellar proteins

Protein	Molecular mass[a] (Da)	
	Apparent	Deduced
FlgH precursor	29,000	24,681
FlgH mature (assumed Cys-22 start)[b]	27,000	22,392
FlgI precursor	40,000	38,125
FlgI mature (Glu-20 start)	38,000	36,235
FlgJ	36,000	34,346
FliF mature (Ser-2 start)	65,000	61,030
E. coli FliF mature (assumed Asn-2 start)[c]	60,000	60,390

[a] Apparent molecular masses are based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis values reported in references 1, 4, 24, and 26, except for that for FlgJ, which is from unpublished minicell data obtained during an earlier study of the genetic region containing the hook-associated protein genes *flgK* and *flgL* (25). The deduced molecular masses are from the DNA sequences reported in the present study, except for that for *E. coli* FliF, which is from data obtained by Frantz et al. (manuscript in preparation). The N termini of the mature FlgI and FliF proteins have been determined by amino acid sequence analysis (Jones et al., manuscript in preparation).

[b] Trp-24 is an alternative start (see text), which would give a deduced molecular mass of 22,218 Da.

[c] By analogy with *S. typhimurium* FliF.

lecular masses of the gene products (Table 1). (iii) The combined molecular masses of the three proteins (101 kilodaltons [kDa]) requires a coding sequence of ca. 2.8 kilobases (kb). The length of the cloned region that includes the three genes (from the *Mlu*I site in pOH20 to the 3′ terminus of the insert of pMH41-17 [25]) was 3.2 kb, too small to include any other genes of the appropriate size. (iv) *flgH* contains an *Eco*RI site (26), seen at position 526 of the open

FIG. 3. DNA sequence of the *flgH* gene and deduced amino acid sequence of its product (the L-ring protein). Numbering of the sequence commences with the presumed initiation codon. The intergenic regions between *flgG* and *flgH* and between *flgH* and *flgI* are also shown. Two possible sites of cleavage of the N-terminal signal sequence are indicated by vertical arrows. A potential stem-loop structure in the region between *flgG* and *flgH* is indicated and discussed in the text. S-D, Potential ribosome-binding site (Shine-Dalgarno sequence).

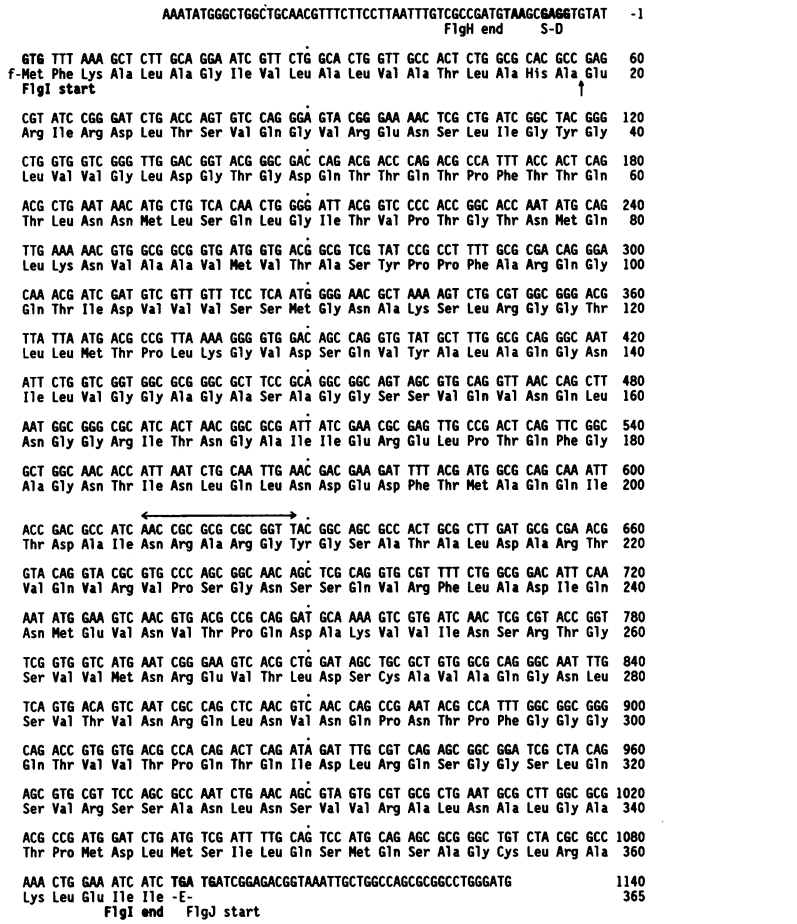

FIG. 4. DNA sequence of the *flgI* gene and deduced amino acid sequence of its product (the P-ring protein). Numbering of the sequence commences with the presumed initiation codon; bp 1 in this figure corresponds to bp 711 in Fig. 3. The intergenic region between *flgH* and *flgI* and the region of overlap between *flgI* and *flgJ* are also shown. The site of cleavage of the N-terminal signal sequence, between Ala-19 and Glu-20 (Jones et al., manuscript in preparation), is indicated by a vertical arrow. A 16-bp palindrome is indicated by a double arrow. S-D, Shine-Dalgarno sequence.

reading frame in Fig. 3; *flgI* contains a single *Mlu*I site (24), seen at position 669 in Fig. 4; and *flgJ* contains two such sites (25), seen at positions 199 and 760 in Fig. 5. (v) The 3′ end of *flgI* lay between the 3′ termini of the inserts of pMH41-21 and pMH41-41 (24). These endpoints lay about 200 bases upstream and 60 bases downstream of the 3′ end of the open reading frame in Fig. 4. (vi) In all three cases, the open reading frames were assessed at the 96% confidence level as coding regions by the program TESTCODE (17).

fliF was identified as the open reading frame shown in Fig. 6 by the following criteria. (i) Restriction mapping (data not shown) indicated that the gene contained two *Eco*RI sites, seen at positions 74 and 1299 in Fig. 6. (ii) The size of the gene is consistent with the experimentally measured apparent molecular mass of its product (Table 1). (iii) The deduced amino acid sequence of the N terminus of FliF agreed with that obtained by experimental analysis (see below). (iv) The DNA sequence closely resembled that of the homologous *E. coli* gene (4, 40; Frantz et al., manuscript in preparation) (see below). (v) The open reading frame was assessed at the 96% confidence level as a coding region by the program TESTCODE (17).

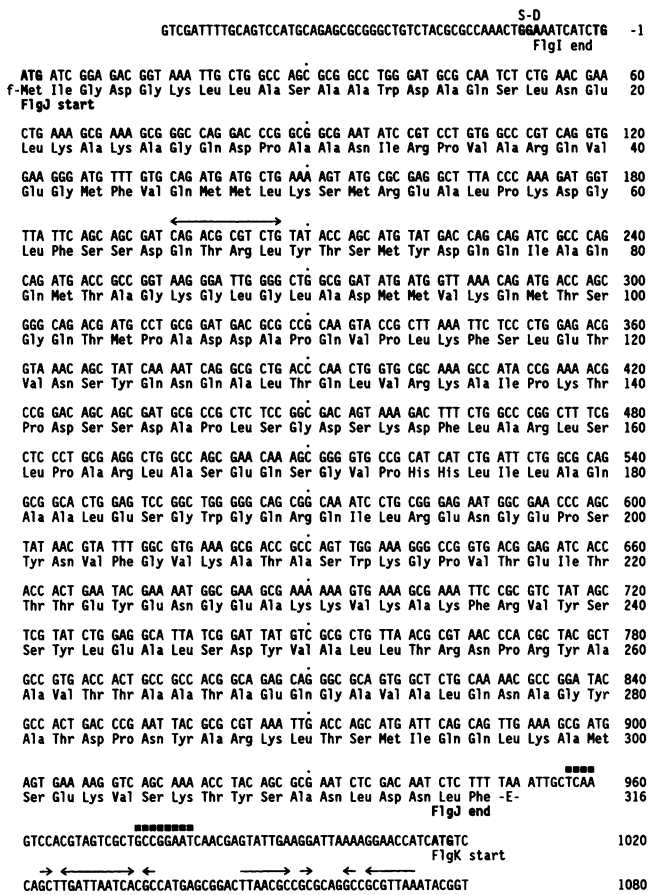

FIG. 5. DNA sequence of the *flgJ* gene and deduced amino acid sequence of its product (a flagellar protein of unknown function). Numbering of the sequence commences with the presumed initiation codon; bp 1 in this figure corresponds to bp 1098 in Fig. 4. The region of overlap between *flgI* and *flgJ*, the interoperon region between *flgJ* (the last gene of the *flgB* operon) and *flgK* (encoding one of the hook-associated proteins, and the first gene of the *flgK* operon), and a portion of the coding region of *flgK* that may constitute the transcription termination signal for the *flgB* operon are also shown. Palindromes and potential stem-loop structures are indicated by arrows. The presumed promoter of the *flgK* operon is indicated by solid squares. S-D, Shine-Dalgarno sequence.

Translational initiation sites of *flgH*, *flgI*, *flgJ*, and *fliF*. The identification of the initiation codons for the various genes is based on the absence of other reasonable candidates nearby. In the case of *flgI*, it was further supported by the knowledge that the mature protein begins at Glu-20 (see below), placing tight constraints on the start site of the precursor protein. For *flgH*, there was supporting evidence that is less direct; the protein is processed (26), and the deduced N-terminal region from the translation start shown in Fig. 3 contained a consensus signal sequence. In the case of *flgJ*, the TTG codon at position 19 could be the start in that it has an acceptable ribosome-binding sequence, but we think it unlikely since (in *E. coli* at least) TTG is used as a start codon in only about 1% of the genes for which data are available (20).

The presumed ribosome-binding site (AAGG) for *flgH* is strong (58) but unusually close to the initiation codon (Fig. 3); it may be that the somewhat weaker site upstream (GAAG) is the correct one. The *flgH-flgI* intergenic distance was only 11 base pairs (bp) and contained a strong consensus ribosome-binding site (GAGG) 6 bp upstream from the start codon of *flgI* (Fig. 4). *flgI* terminated in a 1-bp overlap (TGATG) with the start of *flgJ*, and so the ribosome-binding site for *flgJ*, if it exists, must lie within the coding region of *flgI*; the best candidate is weak (GGA, Fig. 5). *fliF* contained a strong, well-placed ribosome-binding site (GAGG, Fig. 6).

Codon usage. We examined the codon usage of *flgH*, *flgI*, *flgJ*, and *fliF* by using the codon adaptation index of Sharp and Li (57). The calculated values (0.28, 0.28, 0.32, and 0.33, respectively) predict fairly low levels of gene expression for these proteins, a reasonable prediction since there are typically only about 6 flagella per cell, and the number of subunits of each ring component in the basal body is ca. 26 (C. J. Jones, H. Okino, S.-I. Aizawa, and R. M. Macnab, manuscript in preparation). Analyses using the optimal codon frequency parameter of Ikemura (29), or the level of rare and infrequent codon usage as described by Konigsberg and Godson (38), yielded qualitatively similar results.

Transcriptional features. Upstream of *flgH* lies *flgG* (26, 37, 40; M. Homma and R. M. Macnab, unpublished data) (Fig. 3). For two genes within the same operon (the *flgB* operon), the intergenic region was unusually large, 54 bp, and contained an inverted repeat sequence that would permit the formation of a stem-loop structure with 14 paired bases in the stem and 8 bases in the loop, with an estimated stability of ca. −33 kcal/mol (66). The possible significance of this will be discussed later.

Following *flgJ* (Fig. 5) was a noncoding region of 64 bp before an extended open reading frame that corresponded to *flgK*, the first gene of a two-gene operon (25, 41; M. Homma and R. M. Macnab, unpublished data). For two genes belonging to different operons, this separation is quite short. A moderately strong flagellum-specific promoter sequence (3) for the *flgK* operon (TCAA and GCCGGAAT, separated by 15 bp) was evident within the intergenic region between *flgJ* and *flgK*. There was no stem-loop sequence in the intergenic region, such as might constitute a transcription terminator for the *flgB* operon. However, just 8 bp into the *flgK* coding sequence there was a palindromic sequence immediately followed by a potential stem-loop sequence (with an estimated stability of −12.6 kcal/mol), which may represent the termination signal for the *flgB* operon.

fliF is the first gene of a six-gene operon. The gene immediately upstream is *fliE*, which is an independent transcriptional unit; its orientation (parallel or divergent) with respect to the *fliF* operon is unknown (41). Because *fliE* is partially deleted in the parent plasmid (33), we made no attempt here to identify features associated with that gene. Centered about 34 bp upstream of the start codon of *fliF* (Fig. 6) was the sequence TTTGCAAAAAACG, whose central portion represents a five of eight match to the flagellum-specific −10 consensus promoter sequence (GC CGATAA) described by Helmann and Chamberlin (22), and whose complete sequence represents a 10 of 13 match to the extended consensus described by Bartlett et al. (3). This extended consensus, which is also present in the *fliF* operon of *E. coli*, may represent a further level of refinement in the control of flagellar gene expression, as has been discussed (3). The sequence upstream of this (AGTG) bears no resemblance to the flagellum-specific −35 consensus (TAAA) seen in a number of flagellar operons. However, there are several other flagellar operons whose −35 regions do not conform to the TAAA consensus (3, 33); indeed, evidence for the consensus stems primarily from operons encoding products, such as flagellin, the hook-associated proteins, and the Mot and Che proteins, that are required at a late stage of the

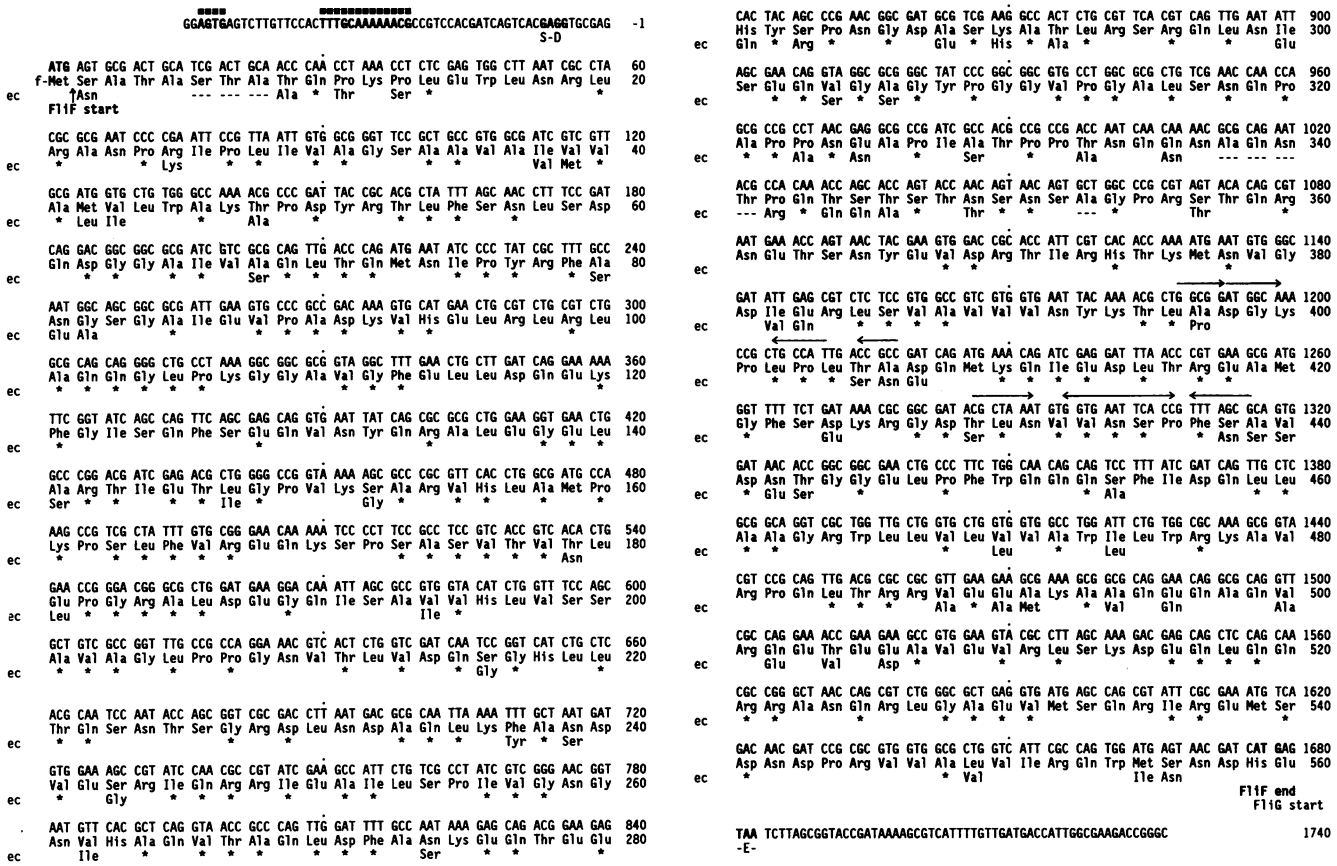

FIG. 6. DNA sequence of the *fliF* gene and deduced amino acid sequence of its product (the M-ring protein). Numbering of the sequence commences with the presumed initiation codon. The N-terminal methionine is cleaved (vertical arrow) to yield Ser-2 as the first residue of the mature protein (Jones et al., manuscript in preparation). A portion of the interoperon region between *fliE* and *fliF* and the region of overlap between *fliF* and *fliG* (one of the flagellar switch genes) are also shown (cf. Fig. 6 in reference 33). The presumed promoter of the *fliF* operon is indicated by solid squares. Symmetry elements (palindromes and inverted repeats) are indicated by arrows. Below the *S. typhimurium* amino acid sequence is shown the amino acid sequence of the homologous *E. coli* protein (ec; Frantz et al., manuscript in preparation). Only differing residues are shown; where differences in DNA sequence exist but do not alter the amino acid sequence, this is indicated by an asterisk. The *S. typhimurium* sequence contains 8 additional residues, indicated by dashes in the *E. coli* sequence. Alignment of the two sequences between *S. typhimurium* Ala-338 and Thr-346 has been chosen arbitrarily. S-D, Shine-Dalgarno sequence.

assembly process (3), whereas FliF is required at an early stage (63; C. J. Jones and R. M. Macnab, unpublished data).

There were several extensive symmetry elements in these genes. For example, there was a 16-bp palindrome in *flgI* (Fig. 4) and 12-bp ones in *flgJ* and *fliF* (Fig. 5 and 6), as well as a number of shorter ones. Two sequences existed in the *fliF* transcript (Fig. 6) that could generate stem-loops, with predicted stabilities of −26.6 and −17.0 kcal/mol (66).

N termini of the FlgH, FlgI, FlgJ, and FliF proteins. Both *flgH* and *flgI* code for proteins that are known to be processed during their export to the outer membrane and periplasmic space, respectively (24, 26). That the processing consists of cleavage of an N-terminal peptide was confirmed by the N-terminal sequences (Fig. 3 and 4), which conformed well to the consensus for signal peptides (52): each sequence had a single lysine residue near the N terminus, followed by a predominantly hydrophobic sequence of 11 to 14 residues, a threonine residue, and one (FlgI) or two (FlgH) possible cleavage sites. The predicted cleavage site of the FlgI precursor protein before Glu-20 (24) (Fig. 4) was verified by N-terminal amino acid analysis of the mature form (Jones et al., manuscript in preparation). It has not proved possible thus far to determine the N-terminal sequence of the mature form of FlgH, as it appears to be blocked; we are therefore unable to state which of the potential cleavage sites (before Cys-22 or Trp-24 of the precursor form, Fig. 3) is correct.

The N terminus of FliF did not resemble a consensus signal sequence (Fig. 6); this is the expected result for a procaryotic cytoplasmic membrane protein (52). N-terminal amino acid analysis verified that the mature protein had had the initial methionine residue cleaved (as commonly occurs in bacterial proteins) to yield Ser-2 as the first residue (Jones et al., manuscript in preparation). FlgJ, whose cellular location has not been determined, also lacked an N-terminal signal sequence (Fig. 5).

The predicted molecular masses of FlgH, FlgI, FlgJ, and FliF (Table 1) agree reasonably well with the values calculated from sodium dodecyl sulfate-polyacrylamide gels.

Primary structure and predictions of secondary structure. The ring proteins are likely to conform partially to descriptions of both soluble and membrane proteins (see Discussion), and so we elected to examine them by several predictive schemes, those of Chou and Fasman (10) and Garnier et al. (18) for secondary structure of cytoplasmic proteins, Kyte and Doolittle (42) for hydropathy, and Engelman et al. (16) for transmembrane α-helical secondary structure.

The FlgH protein was fairly hydrophilic (hydrophilicity

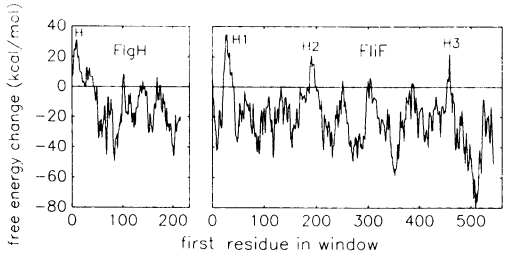

FIG. 7. Analyses of the primary sequences of the membrane-associated L- and M-ring proteins (FlgH and FliF) by the algorithm of Engelman et al. (16), which predicts the portions of a membrane protein that are likely to be membrane-spanning as an α-helix. For the FlgH protein, which is in the outer membrane, there is a predicted span (H) that represents the N-terminal signal peptide, but none in the mature protein. For the FliF protein, there is one strongly predicted transmembrane helix (H1) near the N terminus and two weakly predicted ones (H2 and H3), centered at about residues 200 and 470, respectively.

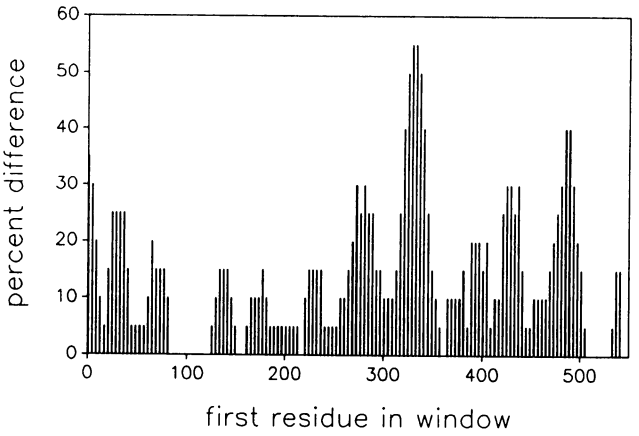

FIG. 8. Comparison of the amino acid sequences of the FliF proteins of *S. typhimurium* (this study) and *E. coli* (Frantz et al., manuscript in preparation). The histogram shows the percent difference in the sequence of a window of 20 residues as a function of position of the window within the sequence. (For clarity, only every fourth data point is shown.) The two longest stretches of identity contain 58 and 42 residues, beginning at Ser-83 and Ala-507, respectively. The maximum divergence of the two sequences occurs from about the middle until close to the C terminus, reaching as high as 55%.

index for the mature protein of 0.16 on the Kyte-Doolittle scale). The Chou-Fasman and Garnier et al. algorithms both indicated a fairly high β-sheet content for the mature protein (26 and 30%, respectively), although with extensive disagreement as to where it occurs in the primary sequence (data not shown). The α-helical content is predicted to be either comparable to (27% by Chou-Fasman) or much lower than (4% by Garnier et al.) the β-sheet content. The only transmembrane α-helix strongly predicted by the scheme of Engelman et al. (Fig. 7) corresponded to the signal peptide. The N terminus of FlgH was notably deficient in charged residues (Fig. 3); there was only one (Lys-28) in the first 45 residues of the mature protein. This uncharged region was immediately followed by a short region of very high charge density (Glu-Asp-Arg-Arg-Pro-Arg, positions 67 to 72). Within the uncharged N-terminal region of FlgH was a segment in which prolines alternated with other amino acids (Pro-Ile-Pro-Gly-Pro-Val-Pro, positions 39 to 45). The FlgH sequence also had several segments where serines or threonines alternated with other amino acids (Ser-Ala-Ser-Lys-Ser-Ser-Ser, positions 87 to 93; Ser-Asn-Thr-Phe-Ser-Gly-Thr-Leu-Thr-Val-Thr, positions 141 to 151; and Thr-Ile-Ser-Gly-Ser-Asn-Ser, positions 188 to 194).

The FlgI sequence was weakly hydrophilic (hydrophilicity index of 0.08 for the mature protein). As expected from its location in the aqueous periplasmic space, it had no predicted membrane-spanning α-helices other than that of its signal peptide (data not shown). The mature protein is predicted to have a large amount of β-structure (42% [Chou-Fasman] or 36% [Garnier et al.]). Four of the five N-terminal amino acids of the mature FlgI protein were charged (Glu-Arg-Ile-Arg-Asp, positions 20 to 24, Fig. 4); the C terminus was also charged (three of the seven terminal residues). Near the center of the sequence were two sets of three sequential charged residues (positions 172 to 174 and 191 to 193). There were three large regions containing at most one charged residue (positions 51 to 97, 131 to 163, and 272 to 310) which together accounted for 34% of the mature protein. The protein contained just two cysteine residues, in the C-terminal region of the protein (positions 273 and 357).

FlgJ was hydrophilic (hydrophilicity index of 0.31) and was not predicted to have any transmembrane α-helices, but was predicted to have a high α-helical content (58% [Chou-Fasman] or 36% [Garnier et al.]). In one predicted helix-turn-helix region (Leu-178 to Gly-197, Fig. 5), the sequence conformed to the consensus for a DNA-binding motif (53). The sequence also had a highly charged region (Glu-223 to Arg-237) that contained three acidic residues and then five basic ones.

FliF was hydrophilic (hydrophilicity index of 0.43), especially in its C-terminal half; from about residue 274 on, there was a marked increase in hydrophilicity, giving rise to a hydrophilicity index of 0.77, whereas the N-terminal region (residues 1 to 273) had an index of only 0.08. The protein had one strongly predicted transmembrane α-helix (H1, Fig. 7) at the N terminus, and two other possible ones, centered around positions 200 (H2) and 470 (H3). There were also shorter stretches of hydrophobic sequence (such as Leu-Ser-Val-Ala-Val-Val-Val, positions 385 to 391, or Val-Val-Ala-Leu-Val-Ile, just before the C terminus at positions 546 to 551; Fig. 6). The secondary structure predictions used agree on assignments for 40% of the protein outside these putative transmembrane helices and indicate a protein with a much higher α-helical content (46% [Chou-Fasman] or 35% [Garnier et al.]) than the L-ring protein, FlgH. The charge distribution throughout the sequence was quite uneven: there was a stretch of 53 amino acids from Gln-303 to Pro-355 that—though quite hydrophilic—contained just one charged residue, Glu-325, which was flanked on both sides by sequence rich in proline residues (6 residues out of 13). Following this large uncharged region was a highly charged portion of the molecule (24 charged residues out of 73 between Arg-356 and Asp-428), a region of low charge that included the possible transmembrane helix H3, and a second highly charged region (28 charged residues out of 69 between Arg-477 and Arg-545). Together, the two charged regions near the C terminus constituted 25% of the protein but contained 44% of its charged amino acids. There were two regions where serine or threonine residues alternated with other residues, one (Ser-Pro-Ser-Ala-Ser-Val-Thr-Val-Thr) at positions 171 to 177, and the other (Thr-Ser-Thr-Ser-Thr-Asn-Ser-Asn-Ser) at positions 344 to 352, as was noted above for FlgH.

Sequence similarity comparisons. None of the gene sequences or deduced protein sequences described here

showed any extended similarities to each other, such as have been found among the rod and hook proteins of the hook-basal body complex (D. J. DeRosier, M. Homma, and R. M. Macnab, unpublished data). There was weak but possibly significant similarity between FlgH and FlgI toward their C termini (Val-Thr-[Leu/Val]-Asp at positions 150 and 268, respectively, Gly-Asn-Leu-X-Val at positions 159 and 278, respectively, and Val-Val-X-Pro-X-Thr at positions 183 and 303, respectively); between FlgH and FliF ([Val/Phe]-Ala-Asn-Gly-Ser at positions 46 and 79, respectively); and between FlgI and FliF (Arg-Gln-Leu-Asn-[Val/Ile], at positions 286 and 296, respectively). Although FlgH, FlgI, and FliF are all basal-body ring proteins, they are in quite different cellular locations (outer membrane, periplasm, and cytoplasmic membrane, respectively), and so the lack of strong similarity is not surprising. Hahnenberger and Shapiro (21) have published the deduced sequence of a *Caulobacter crescentus* flagellar protein, FlaD, which is thought to be the P-ring protein of that species; aside from the presence of a consensus signal sequence, we detected no similarity to the *S. typhimurium* P-ring protein (FlgI) or to the other proteins described here. In comparisons with sequence data bases and unpublished sequence data for several other flagellar genes, the only strong similarity we encountered was the anticipated one between *fliF* and the homologous *E. coli* gene (Frantz et al., manuscript in preparation), which we discuss next.

Comparison between the FliF proteins of *S. typhimurium* and *E. coli*. The deduced FliF amino acid sequence showed 86% identity (47% from the use of identical codons, and 39% from the use of synonymous ones) to that of the homologous *E. coli* protein (4; Frantz et al., manuscript in preparation) (Fig. 6). The homology between the two genes had been deduced on the basis of complementation (40) and gene order (4) and is now fully established by the sequence comparison. The FliF protein of *S. typhimurium* contained only 8 more residues than that of *E. coli* (559 versus 551 after cleavage of the N-terminal methionine), and so the deduced molecular masses were similar (61,030 versus 60,390 Da); for unknown reasons, the apparent molecular masses show a larger difference (65 versus 60 kDa [1, 4]). The sequence differences between the two proteins were fairly uniformly scattered throughout (Fig. 8), the most divergent region being between residues 270 and 500. There were two extensive regions of identity, one centered around residue 110 and the other close to the C terminus.

The only substantial difference in amino acid content between the two proteins was in the number of threonine residues (33 in *S. typhimurium* FliF and 23 in *E. coli* FliF). Since the additional threonine residues in the *S. typhimurium* protein are scattered throughout the protein sequence, we doubt that the difference is functionally significant.

DISCUSSION

The L, P, and M rings are major features of the flagellar basal body of *S. typhimurium*. We have obtained the sequences of the genes (*flgH*, *flgI*, and *fliF*) that encode the subunits of these rings; we have also obtained the sequence of an adjacent gene, *flgJ*. We discuss first the properties of the genes and then those of the deduced gene products.

The *flgJ* gene belongs to the *flgB* operon. Earlier studies of flagellar gene expression in *E. coli* (35, 36) had suggested that *flgJ* was transcriptionally independent of the operon upstream and played a regulatory role in the later stages of morphogenesis, a result that was difficult to reconcile with the fact that FlgJ was needed for the earliest detectable basal-body structure (64). A recent study of flagellar gene expression in *S. typhimurium* (41) concluded that the genes from *flgB* through *flgJ* constitute a single operon. Our data strongly support the latter conclusion. (i) Not only is there no appreciable intergenic distance between *flgI* and *flgJ*, they in fact overlap. (ii) There is no likely promoter sequence near the 3′ end of *flgI*. (iii) There is no evidence for transcription termination at *flgI*; although there are two examples of inverted repeats in the region, they are both only 7 bp in length and the distance between the repeats is much larger than usual for transcription termination sequences. Therefore, in *S. typhimurium* at least, the evidence indicates that *flgJ* belongs to the same operon as *flgI*. We anticipate that this will also prove to be true of *E. coli*. No structural role has been assigned to the FlgJ protein; the presence of a consensus DNA-binding sequence within the protein provides weak support for a postulated regulatory role (35, 36).

Evidence for regulation of these flagellar genes. Multicistronic operons that encode products which must interact in complex kinetic pathways or macromolecular assemblies often display complex patterns of gene regulation. We expect the flagellar regulon to be a prime example of this. One aspect that may have to be regulated is the relative stoichiometries of the various gene products. For example, it seems highly likely that the cell will synthesize more filament protein than hook protein and more hook protein than ring protein. As well as regulation of the strength of gene expression, one can imagine that in the case of a complex macromolecular structure such as the flagellum, other types of control may be needed. For example, spatial and temporal control of synthesis clearly plays a role in flagellar assembly of *C. crescentus* (8) and may play such a role in *S. typhimurium* (70). We have therefore scrutinized the *flgH*, *flgI*, *flgJ*, and *fliF* gene sequences for any suggestions of potential regulatory features.

Overlap of the coding regions of two genes have been shown in at least some cases to cause coupling of translation, with the ribosome failing to dissociate fully upon reaching the first termination codon (19). The *flgI-flgJ* gene boundary contains a 1-bp overlap (TGATG), while that of *fliF-fliG* contains an 8-bp one (ATGAGTAA), and a number of other examples are known to exist elsewhere within the flagellar regulon (12, 33, 39, 46, 47, 59).

Another possible form of regulation is modulation of transcription of different parts of the same operon. We failed to find any sequences suggestive of internal promoters (primary or flagellum specific) within the portions of the *flgB* and *fliF* operons described here. However, we found one striking example of an intergenic sequence that is likely to play a regulatory role. This is the potential stem-loop structure followed by a run of 4 T's that exists between genes *flgG* and *flgH* (Fig. 3). This sequence has the characteristics of a rho-independent transcription terminator (68), in spite of the fact that these two genes belong to the same operon. The sequence does not show any similarity to the repetitive extragenic palindromic sequences that have been described (61).

Similar structures within operons have been observed before, for example in (i) the macromolecular synthesis (*rpsU*) operon of *E. coli* (7, 43); (ii) the ribosomal protein/RNA polymerase (*rpl-rpo*) gene cluster (54); (iii) the *trp* operon, where the partially effective *trp t* rho-independent terminator is followed by a strong rho-dependent terminator (49); (iv) the bacteriophage lambda *sib* site, which is respon-

sible for retroregulation of the upstream integrase gene *int* (56); and (v) the *rxcA* operon of *Rhodopseudomonas capsulata*, which encodes the subunits of the light-harvesting complexes and the reaction center (5). In these various systems, the existence of such stem-loop structures has been shown to play a role in partial termination, protection of upstream message sequence from degradation, or both. Interestingly, RNase III, which has been shown to cleave at stem-loop structures within some of the operons mentioned above, is involved in motility in *E. coli* (2), supporting the idea that message processing may contribute to control of the flagellar regulon.

Is there anything special about the location—between *flgG* and *flgH*—of this potential regulatory feature? The three genes upstream from the site (*flgE*, *flgF*, and *flgG*) encode hook and rod proteins of the hook-basal body complex (Fig. 1) (1, 26, 31, 51), while the two genes downstream from the site (*flgH* and *flgI*) encode the L and P rings of the complex (1, 24, 26, 30, 40). These represent two rather different classes of structure, in several regards. (i) The rod and hook structures are relatively narrow (ca. 15 to 20 nm in diameter) cylindrical structures, with presumably only a narrow central core (for flagellin export [15, 27]), whereas the L and P rings together constitute the outer cylinder, with an external diameter of 30 nm and an internal diameter of ca. 15 nm. (ii) The L- and P-ring proteins are exported by the main N-terminal signal peptide-dependent pathway, while the rod and hook proteins lack conventional signal peptides and are not processed (except for the removal, in some cases, of the N-terminal methionine [Jones et al., manuscript in preparation]). (iii) The L- and P-ring proteins can be synthesized and transferred to their destinations in the outer membrane and periplasmic space, respectively, regardless of the availability of other flagellar structures, whereas the rod and hook proteins are not detected in mutants blocked in earlier components of the structure (C. J. Jones and R. M. Macnab, unpublished data). Given these differences between the products of the upstream and downstream genes surrounding this stem-loop sequence, it would not be surprising if the genes required different regulation.

As well as the stem-loop sequence just described, there are other symmetry elements that might play regulatory roles. For example, there is a 16-bp palindrome in *flgI* (Fig. 4), 12-bp palindromes in *flgJ* and *fliF* (Fig. 5 and 6), and two potential stem-loop structures in *fliF* (Fig. 6).

Recall also that the putative transcription termination sequence of the *flgB* operon lay within the coding region of the first gene, *flgK*, of the next operon. This could cause an interaction between the two operons, with translation initiation of *flgK* from the *flgB* transcript interfering with termination of transcription.

As more of the sequence of the flagellar regulon becomes available, further patterns may emerge to suggest experimental tests of its regulation.

Physical environment of the ring proteins. Before discussing possible structural implications of the sequences of these flagellar proteins, we review briefly the context in which the proteins are to be found within the organelle (Fig. 1). FlgH and FlgI together constitute the outer cylinder, which appears in electron micrographs as a doubly-flanged cylinder with the distal flange constituting the L ring (in the plane of the outer membrane), the proximal flange constituting the P ring (in the periplasmic space), and the remainder of the structure constituting the cylinder wall, which appears quite thin in electron micrographs and is at a right angle to the flanged surfaces. The rod penetrates the outer cylinder and is presumed to be free to rotate within it. In *flgH* mutants, basal bodies possessing the P ring but lacking the L ring have been observed (30, 50, 63, 64); in these, the cylinder wall is missing, indicating that FlgH contributes largely to that structure. For the FlgH protein, we therefore anticipate that (i) it will have quaternary interactions with other FlgH subunits, to form the cylindrical structure, (ii) its L-ring portion will interact with the outer membrane, (iii) the outer surface of its wall portion will be exposed to the aqueous periplasmic space, (iv) the inner surface of its wall portion will be facing the rod (perhaps with an aqueous space intervening between the two), and (v) its cell-proximal edge will be in a strong quaternary interaction with FlgI.

FlgI will have (i) quaternary interactions with other FlgI subunits, (ii) the interaction with FlgH just described, (iii) a P-ring surface exposed to the periplasmic space, and (iv) an inner surface facing the rod. It may also have (v) interactions with the peptidoglycan layer.

Nothing is known experimentally of the location of FlgJ.

Although the location of FliF as the major M-ring protein has been established (23), its detailed environment is difficult to describe for several reasons. First, it is unclear whether the S-ring feature of the M,S-ring pair (Fig. 1) is contributed by the FliF protein or by the product of another flagellar gene (60). Second, it is not known with certainty whether there is a direct juxtaposition of the M ring and the rod and, if so, whether the interface is engaged in mutual rotation. Third, although there is evidence for other flagellar components adjacent to the M ring at its circumference (the Mot proteins [32]) and its inner surface (the switch proteins [69]), the relationship is not known in any detail. It is known that the outer edge of the M ring is hydrophobic (1), that the entire M,S-ring structure can exist in the membrane as part of the basal body even when the Mot proteins are lacking, and that the structure is thick enough that it probably extends into both the periplasmic and cytoplasmic spaces. We tentatively suggest therefore that the FliF protein will have (i) quaternary interactions with other FliF subunits, (ii) surfaces exposed to water on both sides of the cell membrane, (iii) a hydrophobic surface at the outer radius that may, depending on circumstances, contact lipid or another hydrophobic protein surface, (iv) sites of quaternary interaction with the rod, and (v) sites of interaction with peripherally mounted switch proteins in the cytoplasm.

Structural features of the ring proteins. FlgH has a cleaved N-terminal sequence, but the N terminus of the mature protein appears to be blocked. Of the two possible candidates for the signal sequence cleavage site, one would generate cysteine as the N-terminal residue; cleavage at this site is favored by other evidence (Jones et al., manuscript in preparation). It is interesting that the predominant outer membrane protein of gram-negative bacteria, lipoprotein, has lipid covalently attached to its N-terminal residue, cysteine (6). It remains to be seen whether the apparent N-terminal block of FlgH (also an outer membrane protein) is a consequence of a similar modification.

The mature form of FlgH, although a membrane protein, shows no evidence of transmembrane α-helical structure, but is predicted to have extensive β-structure. In a variety of other outer membrane proteins, β-sheets appear to be the dominant transmembrane structural motif (34, 67). The region of alternating prolines in FlgH resembles a sequence in OmpA, a porin protein located in the outer membrane of *E. coli* (9). This segment in OmpA resides at the inner surface of the outer membrane, dividing the protein into two domains. While it is tempting to assign such a role to the

proline-rich segment in FlgH, separating the membrane-associated L ring from the wall of the outer cylinder, this would leave little more than a dozen amino acids to form the membrane-associated portion of the proteins. Perhaps this proline-rich segment allows the larger uncharged region to double back on itself, forming a more compact domain which might interact with the outer membrane. The three regions of alternating serines or threonines in FlgH, if arranged as β-sheets, would have the potential for extensive hydrogen bonding on one side, and indeed such a structure could be extended by the use of available asparagine and glutamine residues (Fig. 3). We tentatively suggest that FlgH is organized with its N terminus at the periplasmic side of the membrane, the largely uncharged region forming the bulk of the transmembrane structure (the distal flange of the L,P-ring cylinder), and the more polar C-terminal region of the molecule constituting the cylinder wall that faces the periplasm and the rod.

FlgI has a high predicted β-content. It is not obvious whether this would generate sheets parallel to the flagellar axis, parallel to the ring plane, or neither. The two sets of charged triplets near the center of the sequence might represent the boundaries between N-terminal and C-terminal domains. One indication that the C-terminal region constitutes a domain is the fact that it contains the only two cysteine residues in the protein, and these are presumed to form a disulfide bridge (since FlgI, uniquely among all of the known flagellar proteins, shows anomalous electrophoretic mobility in the absence of 2-mercaptoethanol [1, 24]).

On the basis of its hydrophilic primary sequence, its lack of any predicted transmembrane α-helices, and its lack of a consensus signal sequence for export across the cell membrane (Fig. 5), we suggest that FlgJ is likely to be either in the cytoplasm or peripheral to the cytoplasmic membrane. In spite of the fact that its gene is adjacent to those of the two outer rings of the basal body, there is no information to suggest it plays a related role. Weak evidence for a regulatory role is provided by the presence of a consensus DNA-binding motif.

FliF is remarkably hydrophilic for a membrane protein and has a high predicted α-helical content. A simple interpretation of this would be that the only membrane-spanning regions of FliF are the three predicted α-helices shown in Fig. 7, in which case only about 10% of the protein would be in the lipid bilayer and the remainder would constitute large aqueous domains. We think this interpretation is unlikely, since the M ring appears as a substantial structure in isolated basal bodies and is hydrophobic at its bulkiest point, as judged by association of the M ring and the hydrophobic L ring in head-to-tail dimers in electron micrographs (1). We therefore imagine that there must be a number of additional membrane spans, either as hydrophobic β-sheet (perhaps the hydrophobic sequences noted in Results) and therefore consisting of too few residues to be predicted by the method of Engelman et al., or as α-helices too polar to be predicted by the method. It is important to remember that even in the transmembrane region, the FliF protein is in quaternary interaction with other proteins (itself, and possibly rod proteins and the Mot proteins). The regions of high charge density towards the C terminus may well constitute aqueous domains. If the intervening region is, as predicted, membrane-spanning, this would place the two charged regions on opposite sides of the membrane.

Axial versus ring components of the hook-basal body complex. Evidence is accumulating that a subset of the proteins in the hook-basal body complex (the rod, hook, and the hook-associated proteins) are related structurally as an axial family (DeRosier et al., unpublished data) and are probably exported by a special flagellum-specific pathway. The amino acid sequence data presented here for the ring proteins, the processing of the L- and P-ring proteins (24, 26), and the location of the M ring all argue that they are unrelated to this axial family of proteins and that they are exported (L- and P-ring proteins) or inserted into the cytoplasmic membrane (M-ring protein) by conventional pathways.

ACKNOWLEDGMENTS

We are grateful to M. Kihara for assistance with generating the sequence data; to B. Frantz, D. Bartlett, and P. Matsumura for providing unpublished sequence information regarding the *E. coli fliF* gene; and to G. Hatfull and C. Joyce for a gift of Klenow fragment.

This work has been supported by Public Health Service grants AI12202 and GM40335 from the National Institutes of Health.

LITERATURE CITED

1. **Aizawa, S.-I., G. E. Dean, C. J. Jones, R. M. Macnab, and S. Yamaguchi.** 1985. Purification and characterization of the flagellar hook-basal body complex of *Salmonella typhimurium*. J. Bacteriol. **161**:836–849.
2. **Apirion, D., and N. Watson.** 1978. Ribonuclease III is involved in motility of *Escherichia coli*. J. Bacteriol. **133**:1543–1545.
3. **Bartlett, D. H., B. B. Frantz, and P. Matsumura.** 1988. Flagellar transcriptional activators FlbB and FlaI: gene sequences and 5′ consensus sequences of operons under FlbB and FlaI control. J. Bacteriol. **170**:1575–1581.
4. **Bartlett, D. H., and P. Matsumura.** 1984. Identification of *Escherichia coli* region III flagellar gene products and description of two new flagellar genes. J. Bacteriol. **160**:577–585.
5. **Belasco, J. G., J. T. Beatty, C. W. Adams, A. von Gabain, and S. N. Cohen.** 1985. Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic *rxcA* transcript. Cell **40**:171–181.
6. **Braun, V.** 1975. Covalent lipoprotein from the outer membrane of *Escherichia coli*. Biochim. Biophys. Acta **415**:335–377.
7. **Burton, Z. F., C. A. Gross, K. K. Watanabe, and R. R. Burgess.** 1983. The operon that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in *E. coli* K12. Cell **32**:335–349.
8. **Champer, R., A. Dingwall, and L. Shapiro.** 1987. Cascade regulation of *Caulobacter* flagellar and chemotaxis genes. J. Mol. Biol. **194**:71–80.
9. **Chen, R., W. Schmidmayr, C. Krämer, U. Chen-Schmeisser, and U. Henning.** 1980. Primary structure of major outer membrane protein II* (*ompA* protein) of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **77**:4592–4596.
10. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:45–148.
11. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**:31–40.
12. **Dean, G. E., R. M. Macnab, J. Stader, P. Matsumura, and C. Burks.** 1984. Gene sequence and predicted amino acid sequence of the *motA* protein, a membrane-associated protein required for flagellar rotation in *Escherichia coli*. J. Bacteriol. **159**: 991–999.
13. **DePamphilis, M. L., and J. Adler.** 1971. Fine structure and isolation of the hook-basal body complex of flagella from *Escherichia coli* and *Bacillus subtilis*. J. Bacteriol. **105**:384–395.
14. **DePamphilis, M. L., and J. Adler.** 1971. Attachment of flagellar basal bodies to the cell envelope: specific attachment to the outer, lipopolysaccharide membrane and the cytoplasmic membrane. J. Bacteriol. **105**:396–407.

15. **Emerson, S. U., K. Tokuyasu, and M. I. Simon.** 1970. Bacterial flagella: polarity of elongation. Science **169**:190–192.
16. **Engelman, D. M., A. Goldman, and T. A. Steitz.** 1982. The identification of helical segments in the polypeptide chain of bacteriorhodopsin. Methods Enzymol. **88**:81–88.
17. **Fickett, J. W.** 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. **10**:5303–5318.
18. **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.
19. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
20. **Gren, E. J.** 1984. Recognition of messenger RNA during translational initiation in *Escherichia coli*. Biochimie **66**:1–29.
21. **Hahnenberger, K. M., and L. Shapiro.** 1988. Organization and temporal expression of a flagellar basal body gene in *Caulobacter crescentus*. J. Bacteriol. **170**:4119–4124.
22. **Helmann, J. D., and M. J. Chamberlin.** 1987. DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative σ factor. Proc. Natl. Acad. Sci. USA **84**:6422–6424.
23. **Homma, M., S.-I. Aizawa, G. E. Dean, and R. M. Macnab.** 1987. Identification of the M-ring protein of the flagellar motor of *Salmonella typhimurium*. Proc. Natl. Acad. Sci. USA **84**: 7483–7487.
24. **Homma, M., Y. Komeda, T. Iino, and R. M. Macnab.** 1987. The *flaFIX* gene product of *Salmonella typhimurium* is a flagellar basal body component with a signal peptide for export. J. Bacteriol. **169**:1493–1498.
25. **Homma, M., K. Kutsukake, and T. Iino.** 1985. Structural genes for flagellar hook-associated proteins in *Salmonella typhimurium*. J. Bacteriol. **163**:464–471.
26. **Homma, M., K. Ohnishi, T. Iino, and R. M. Macnab.** 1987. Identification of flagellar hook and basal body gene products (FlaFV, FlaFVI, FlaFVII, and FlaFVIII) in *Salmonella typhimurium*. J. Bacteriol. **169**:3617–3624.
27. **Iino, T.** 1969. Polarity of flagellar growth in *Salmonella*. J. Gen. Microbiol. **56**:227–239.
28. **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.
29. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for synonymous codon choice that is optimal for the *E. coli* translation system. J. Mol. Biol. **151**:389–409.
30. **Jones, C. J., M. Homma, and R. M. Macnab.** 1987. Identification of proteins of the outer (L and P) rings of the flagellar basal body of *Escherichia coli*. J. Bacteriol. **169**:1489–1492.
31. **Kagawa, H., S.-I. Aizawa, S. Yamaguchi, and J.-I. Ishizu.** 1979. Isolation and characterization of bacterial flagellar hook proteins from salmonellae and *Escherichia coli*. J. Bacteriol. **138**: 235–240.
32. **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.
33. **Kihara, M., M. Homma, K. Kutsukake, and R. M. Macnab.** 1989. Flagellar switch of *Salmonella typhimurium*: gene sequences and deduced protein sequences. J. Bacteriol. **171**: 3247–3257.
34. **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.
35. **Komeda, Y.** 1982. Fusions of flagellar operons to lactose genes on a Mu *lac* bacteriophage. J. Bacteriol. **150**:16–26.
36. **Komeda, Y.** 1986. Transcriptional control of flagellar genes in *Escherichia coli* K-12. J. Bacteriol. **168**:1315–1318.
37. **Komeda, Y., K. Kutsukake, and T. Iino.** 1980. Definition of additional flagellar genes in *Escherichia coli* K12. Genetics **94**:277–290.
38. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:687–691.
39. **Kuo, S. C., and D. E. Koshland, Jr.** 1986. Sequence of the *flaA* (*cheC*) locus of *Escherichia coli* and discovery of a new gene. J. Bacteriol. **166**:1007–1012.
40. **Kutsukake, K., T. Iino, Y. Komeda, and S. Yamaguchi.** 1980. Functional homology of *fla* genes between *Salmonella typhimurium* and *Escherichia coli*. Mol. Gen. Genet. **178**:59–67.
41. **Kutsukake, K., Y. Ohya, S. Yamaguchi, and T. Iino.** 1988. Operon structure of flagellar genes in *Salmonella typhimurium*. Mol. Gen. Genet. **214**:11–15.
42. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
43. **Lupski, J. R., B. L. Smiley, and G. N. Godson.** 1983. Regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon and the initiation of DNA replication in *Escherichia coli* K-12. Mol. Gen. Genet. **189**:48–57.
44. **Macnab, R. M.** 1987. Flagella, p. 70–83. *In* F. C. Neidhardt, J. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
45. **Macnab, R. M.** 1987. Motility and chemotaxis, p. 732–759. *In* F. C. Neidhardt, J. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
46. **Macnab, R. M.** 1988. The end of the line in bacterial sensing: the flagellar motor. Cold Spring Harbor Symp. Quant. Biol. **53**: 67–75.
47. **Malakooti, J., Y. Komeda, and P. Matsumura.** 1989. DNA sequence analysis, gene product identification, and localization of flagellar motor components of *Escherichia coli*. J. Bacteriol. **171**:2728–2734.
48. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
49. **Mott, J. E., J. L. Galloway, and T. Platt.** 1985. Maturation of *Escherichia coli* tryptophan operon mRNA: evidence for 3′ exonucleolytic processing after rho-dependent termination. EMBO J. **4**:1887–1891.
50. **Ohnishi, K., M. Homma, K. Kutsukake, and T. Iino.** 1987. Formation of flagella lacking outer rings by *flaM*, *flaU*, and *flaY* mutants of *Escherichia coli*. J. Bacteriol. **169**:1485–1488.
51. **Okino, H., M. Isomura, S. Yamaguchi, Y. Magariyama, S. Kudo, and S.-I. Aizawa.** 1989. Release of flagellar filament-hook-rod complex by a *Salmonella typhimurium* mutant defective in the M ring of the basal body. J. Bacteriol. **171**:2075–2082.
52. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
53. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
54. **Ralling, G., and T. Linn.** 1984. Relative activities of the transcriptional regulatory sites in the *rplKAJL rpoBC* gene cluster of *Escherichia coli*. J. Bacteriol. **158**:279–285.
55. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
56. **Schmeissner, U., K. McKenney, M. Rosenberg, and D. Court.** 1984. Removal of a terminator structure by RNA processing regulates *int* gene expression. J. Mol. Biol. **176**:39–53.
57. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.
58. **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.

59. **Stader, J., P. Matsumura, D. Vacante, G. E. Dean, and R. M. Macnab.** 1986. Nucleotide sequence of the *Escherichia coli motB* gene and site-limited incorporation of its product into the cytoplasmic membrane. J. Bacteriol. **166**:244–252.
60. **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.
61. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015–1026.
62. **Stewart, R. C., and F. W. Dahlquist.** 1987. Molecular components of bacterial chemotaxis. Chem. Rev. **87**:997–1025.
63. **Suzuki, T., T. Iino, T. Horiguchi, and S. Yamaguchi.** 1978. Incomplete flagellar structures in nonflagellate mutants of *Salmonella typhimurium*. J. Bacteriol. **133**:904–915.
64. **Suzuki, T., and Y. Komeda.** 1981. Incomplete flagellar structures in *Escherichia coli* mutants. J. Bacteriol. **145**:1036–1041.
65. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
66. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
67. **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.
68. **Yager, T. D., and P. H. von Hippel.** 1987. Transcript elongation and termination in *Escherichia coli*, p. 1241–1275. *In* F. C. Neidhardt, J. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
69. **Yamaguchi, S., S.-I. Aizawa, M. Kihara, M. Isomura, C. J. Jones, and R. M. Macnab.** 1986. Genetic evidence for a switching and energy-transducing complex in the flagellar motor of *Salmonella typhimurium*. J. Bacteriol. **168**:1172–1179.
70. **Yamamori, T., K. Ito, T. Yura, T. Suzuki, and T. Iino.** 1977. Ribonucleic acid polymerase mutant of *Escherichia coli* defective in flagella formation. J. Bacteriol. **132**:254–261.