Characterization of the fliE Genes of Escherichia coli and Salmonella typhimurium and Identification of the FliE Protein as a Component of the Flagellar Hook-Basal Body Complex

VOLKER MÜLLER,†CHRISTOPHER J. JONES,‡IKURO KAWAGISHI,¶SHIN-ICHI AIZAWA,§ AND ROBERT M. MACNAB*†

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511,1 and ERATO, 5-9-5 Tokodai, Toyosato, Tsukuba, Ibaraki 300-26, Japan2

Received 31 October 1991/Accepted 21 January 1992

Within flagellar region III of Escherichia coli and Salmonella typhimurium, the genomic organization has been largely established. An exception is fliE, a gene whose exact location and product function are not well understood. We cloned the fliE gene, obtained its DNA sequence, and identified its product. fliE was found to be a monocistronic transcriptional unit, adjacent to and divergent from the large fliF operon. It is several kilobases distant from the nearest flagellar operon in the other direction, the fliD operon, and constitutes the first operon within the newly defined region IIIb, which contains the genes fliE through fliR. fliE encodes a small, moderately hydrophilic protein with a deduced molecular mass of 11,114 Da (E. coli) or 11,065 Da (S. typhimurium). We identified a protein within the isolated hook-basal body complex as the fliE gene product on the basis of its size and comparison of its N-terminal amino acid sequence with that deduced from the gene sequence. From gel electrophoresis and autoradiography of 35S-labeled S. typhimurium hook-basal body complexes (C. J. Jones, R. M. Macnab, H. Okino, and S.-I. Aizawa, J. Mol. Biol. 212:377–387, 1990) and the deduced number of sulfur-containing residues in FIE, we estimated the stoichiometry of the protein in the hook-basal body complex to be about nine subunits. FIE does not undergo cleavage of a signal peptide, nor does it show any sequence similarity to the axial components like the rod or hook proteins, which are believed to be exported by the flagellum-specific export pathway. On the basis of this and other evidence, we suggest that FIE may be in the vicinity of the MS ring, perhaps acting as an adaptor protein between the ring and rod substructures.

The flagellar systems of Escherichia coli and Salmonella typhimurium are encoded by at least 40 genes, which until recently were thought to be organized into three regions on the chromosome (plus, in the case of S. typhimurium, the phase 2 flagellin region) (27). Region I contains mostly flagellar structural genes, and region II contains a mixture of flagellar, motility, and chemotaxis genes; for most of the genes in those two regions, the DNA sequences and the functions of their products are known (27).

Region III was known to contain genes fliA through fliR, which were thought to be contiguous. However, we have found that it is interrupted by a large stretch of DNA unrelated to flagellar function, justifying its subdivision into regions IIIa and IIIb (17). Except for fliE, the genomic organization and DNA sequences of all of these region IIIa and IIIb genes in either E. coli, S. typhimurium, or both were known, with fliA through fliD belonging to region IIIa and fliF through fliR belonging to region IIIb (9, 14, 18, 22, 23, 28, 29, 32, 33, 44). We report here that fliE lies at the start of region IIIb. It is a monocistronic operon oriented divergently from the large fliF operon. We have obtained the DNA sequence of the gene, shown that its product is a component of the flagellar basal body, and estimated its subunit stoichiometry within that structure.

MATERIALS AND METHODS

Bacterial strains. The following E. coli strains were used: for cloning, subcloning, and sequencing, DH5α (GIBCO BRL, Gaithersburg, Md.) and XL1-Blue (Stratagene, La Jolla, Calif.); for minicell production, UH869 (minD-minB) (12); for complementation analysis, YK1101 (fliD) (41), RP4191 [fliE79(Am)] (34), MS111 (fliF) (38), and RP4501 [fliG [scyB10]] (34).

S. typhimurium SJW1103 (47) was used for isolation of hook-basal body complexes and analysis of the N-terminal amino acid sequences of their component proteins and as the source of S. typhimurium region IIIA DNA for cloning. SJW1381 (fliE) (45) was used as the host for transformation with plasmids containing S. typhimurium region IIIA DNA.

Plasmids. pIK1001 (see Fig. 1a) is a pBR322-based plasmid with an E. coli chromosomal fragment containing the fliC-FE (subscripts E and S are used to distinguish the genes or proteins of E. coli and S. typhimurium, respectively) region of JAI1 (17); the subclones of pIK1001 constructed in this study are described in Results.

pOYAl (Fig. 1b) is a pUC119-based plasmid with a 6.7-kb EcoRI insert of S. typhimurium DNA extending from within fliDS through the start of fliF. Its construction is described in Results.

Chemicals and enzymes. All of the chemicals and enzymes used were obtained from standard commercial sources.
Antibiotics were used at the following concentrations: ampicillin, 50 to 100 μg/ml; kanamycin, 25 μg/ml; tetracycline, 10 μg/ml.

**Manipulation of DNA.** Isolation of plasmid DNA, manipulation of DNA, and transformation of host cells were all carried out by standard procedures (30).

**DNA sequencing.** DNA sequencing of *E. coli* DNA was performed on the XbaI-PstI fragment of pVM5 (see Fig. 1a) and subfragments thereof. Bal 31 deletions from the Smal site were obtained by digestion first with Smal and then with Bal 31 and ligation in the presence of an EcoRI linker; the EcoRI-XbaI fragments were then subcloned into M13 for sequencing. Sequencing of *S. typhimurium* DNA was performed on a 2.5-kb EcoRI-BamHI fragment subcloned from pOYA1 (see Fig. 1b) into M13. Sequencing was done by the dideoxynucleotide method (36) with the modified T7 DNA polymerase (42) Sequence 2.0 (U.S. Biochemical Corp., Cleveland, Ohio) and commercial and custom-synthesized primers.

**Sequence analysis.** Sequence manipulations and analysis were carried out as previously described (14). Amino acid sequence comparisons were made against the National Biomedical Research Foundation protein sequence data base (version 25.0; June 1990) by using the FASTA algorithm and against the GenBank DNA data base (version 67.0; March 1991) by using the TFASTA algorithm.

**N-terminal amino acid sequencing.** Hook-based body complexes from a 7.5-liter culture of SJ.W1103 were isolated essentially as previously described (1), subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (26) in two 1-cm lanes of a 1-mm-thick 12.5% polyacrylamide gel, electroblotted for 50 min onto polyvinylidene difluoride membranes (Millipore Corp., Bedford, Mass.), and stained with Coomassie blue by following the protocol described by Matsudaira (31). Bands of interest were then cut out, and amino acid sequence analysis was performed on them as previously described (16).

**Nucleotide sequence accession numbers.** The fliE* and fliE* sequences have been deposited in the GenBank data base under accession numbers M84992 and M84993.

**RESULTS**

**Cloning of fliE*.** While characterizing flagellar region III and establishing that it is interrupted by DNA unrelated to flagellation (17), we had cloned a 14.5-kb EcoRI fragment from the chromosome of *E. coli* 1A11, a strain which was wild type for all flagellar genes except flagellin gene fliC* (which had been partially deleted and disrupted with a kanamycin resistance gene cassette). The resulting plasmid, pJK1001, contains the following genes: ΔfliC*::Km*, fliD*, fliE*, and fliF*, with a 5.4-kb region unrelated to flagellar function lying between the MluI and XbaI sites shown in Fig. 1a. Plasmids such as pVM5, which lack the left-hand part of the insert but retain the right-hand part, failed to complement fliD* mutants in a swarm assay on semisolid agar but complemented fliE* and fliF* mutants. Thus, fliE* and fliF* lie to the right of the intervening nonflagellar region, within flagellar region IIIb.

To localize fliE* in more detail, various deletion derivatives of pVM5 were made and tested for the ability to complement a fliE* mutant. The results (Fig. 1a) indicated that fliE* was located within a 0.6-kb AccI-Smal fragment (plasmid pVM511). This was confirmed by loss of complementation when a small PvuII deletion was made within this fragment (plasmid pVM512) or when a small extent of Bal 31 digestion was allowed to proceed from the Smal site (plasmid pVM3-2). The inferred location of fliE* is consistent with the facts that the Smal site lies just before the putative fliF* promoter and that the PstI site lies just within the 5’ end of the fliE* coding region (2, 32).

**DNA sequence analysis of the fliE* gene.** We sequenced both strands of the PstI-XbaI fragment of pVM5 in its entirety by using various restriction and deletion derivatives. The sequence of the PstI-AccI portion of the fragment is shown in Fig. 2a.

The sequence between the PstI and Smal sites, which contains the promoter of the fliE* operon and the first five codons of the fliE* gene (Fig. 2a), is identical to that obtained by B. Frantz and P. Matsumura, who have sequenced the fliF* gene (32).

Between the Smal and AccI sites, there were no reasonable candidates for open reading frames with the same orientation as fliF*. In the orientation opposite to that of fliF*, there was an open reading frame extending from positions 231 to 542 (Fig. 2a). This open reading frame was assessed as a coding region at the 96% confidence level by...
(a) E. coli

![](http://jb.asm.org/)

(b) S. typhimurium

FIG. 2. DNA sequences of the flIE genes of E. coli and S. typhimurium and the deduced sequences of their products. (a) Sequence of the PstI-AccI fragment of pVM5 (Fig. 1a), which contains the flIE gene and its control regions. Also shown is the flIE promoter region and the start of the flIE gene, which is oriented with respect to flIE mRNA (b) Sequence of a portion of the rightward EcoRI-BamHI fragment of pOY1 (Fig. 1b), which contains the flIE gene and its control regions, and the start of the flIE gene. The start and stop codons, putative -10 promoter sequences, putative ribosome-binding sites (S-D [Shine-Dalgarno sequence]), and restriction sites referred to in the text are underlined.

The deduced amino acid sequence of the N terminus agreed with the experimentally determined sequence of a flagellar protein (see below).

Thus, the sequences of flIE and the intergenic region between flIE and flIF have now been determined. The extent of the intergenic region is 214 bp. The presumed promoter for flIE is GCCAATAG, which represents a seven-of-eight match to the flagellum-specific -10 consensus sequence (GCCGATAA) described by Helmann and Chamberlin (7). The region upstream of this shows no sequence homologous to the flagellum-specific -35 consensus (TAAA). The presumed ribosome-binding site (GGAG) is strong and well placed (7 bp upstream) with respect to the initiation codon. The codon adaptation index (37) for flIE is 0.32, suggesting a fairly low level of synthesis. There is no
open reading frame immediately following \textit{fliE}$_S$, which is consistent with genetic evidence that it is the only gene in the transcriptional unit (20). We were unable to identify the transcription termination sequence for \textit{fliE}$_E$.

If another flagellar operon were to exist in region IIIb (17) before \textit{fliE}$_E$, it would need to fall entirely within the 700 bp between the \textit{XbaI} site and the 3‘ end of \textit{fliE}$_E$. We found no features within the sequence of that region to suggest the existence of such an operon, oriented either the same as or opposite to \textit{fliE}$_E$ (data not shown).

**Deduced properties of the \textit{fliE}$_E$ protein.** The deduced amino acid sequence of \textit{fliE}$_E$ yields a molecular mass of 11,114 Da. The sequence is moderately hydrophilic (mean hydrophobicity index determined as described by Kyte and Doolittle [25], −0.09), with no extensive stretches of hydrophobic residues; analysis by the algorithm of Engelman et al. (4) indicated no potential membrane-spanning segments. The secondary-structure algorithms of Chou and Fasman (3) and Garnier et al. (6) both predict a fairly high α-helical content, with several extensive stretches connected by turns.

The N terminus does not resemble the signal sequences of proteins exported by the primary cellular pathway, nor does the protein undergo peptide cleavage (see below).

\textit{FliE}$_E$ shows no obvious sequence similarity to other flagellar proteins. Specifically, it lacks features shared by the axial proteins of the flagellum (rod proteins, hook protein, hook-associated proteins, and flagellin), namely, hydrophobic heptad repeats and various conserved local motifs near the termini (9, 11).

Comparison of the sequence of \textit{FliE}$_E$ with the National Biomedical Research Foundation protein sequence database and the translated GenBank data base revealed no proteins related to it.

**Identification of the \textit{FliE}$_E$ protein as a hook-basal body component.** We initially attempted to detect the product of \textit{fliE}$_E$ by using transformed minicells. With some (but not all) plasmids that had been shown to complement \textit{fliE}$_E$ mutants, we were able to detect a protein with an apparent molecular mass similar to the deduced value for \textit{fliE}$_E$.

At about that time, we obtained direct identification of the \textit{FliE}$_E$ protein in situ. In a study of flagellar assembly in \textit{S. typhimurium} (15), we had identified in hook-basal body preparations several proteins whose synthesis was under control of the flagellar master operon but whose genetic origins were unknown. One of these unidentified components was a small protein with an apparent molecular mass of 9 kDa (15) (or 12 kDa in an earlier study; reference 1), comparable to the deduced value of 11 kDa obtained as described above for \textit{FliE}$_E$. N-terminal amino acid analysis of this protein showed that its first 17 residues differed by only a single conservative change from residues 2 to 18 of the deduced \textit{FliE}$_E$ sequence (Fig. 3). Also, in the course of a study of the flagellar ring genes in \textit{S. typhimurium} (14), we had noted in the vicinity of \textit{fliF}$_S$ a partial open reading frame that yielded a 30-residue deduced amino acid sequence which, by comparison with the deduced \textit{FliE}$_E$ sequence, was clearly \textit{FliE}$_E$; residues 2 to 18 of this partial \textit{FliE}$_E$ sequence were identical to the N-terminal sequence of the basal body protein. We conclude that this protein is \textit{FliE}$_E$, with the mature protein having undergone cleavage of its N-terminal methionine.

**Cloning and sequence analysis of \textit{fliE}$_E$.** To use existing \textit{S. typhimurium} data (16) to estimate the stoichiometry of the \textit{FliE} protein in the basal body (see below), we needed to know the number of sulfur-containing residues in the \textit{FliE}$_E$ sequence. Although homologous proteins in \textit{E. coli} and \textit{S. typhimurium} usually have a high degree of sequence identity, we did not feel justified in using the \textit{E. coli} data for our calculation, especially in view of the small number of sulfur-containing residues involved (seven in the deduced mature \textit{E. coli} sequence).

The fragmentary \textit{fliE}$_E$ sequence described above derived from a mutant with a deletion beginning early in the gene (14). We therefore needed to clone the corresponding region from wild-type \textit{S. typhimurium}. There is an \textit{EcoRI} site just within the \textit{fliF}$_S$ gene (at bp 74 in Fig. 2 of reference 14), and the nearest known \textit{EcoRI} site to the left lies within \textit{fliD}$_S$ (9). An \textit{EcoRI} chromosomal digest from wild-type strain SJW1105 was ligated into plasmid pUC119, and the pool of plasmids was used to transform SJW1381 (a \textit{fliE}$_E$ mutant). Transformants were spotted onto semisolid nutrient agar, and swarming cells were selected. They were found to harbor a plasmid with a 6.7-kb \textit{EcoRI} insert of \textit{S. typhimurium} DNA, which we named pOYA1. Restriction mapping (Fig. 1b) indicated that the insert of pOYA1 was in fact bounded by the sites mentioned above, within the \textit{fliD}$_S$ gene and just within the \textit{fliF}$_S$ gene; this was confirmed by DNA sequence analysis of the ends of the insert (data not shown). We subcloned a 2.5-kb \textit{EcoRI} fragment from pOYA1 and verified by sequencing that it contained the 5‘ end of \textit{fliF}$_S$. This fragment was used to obtain the sequence of the \textit{fliE}$_E$ gene (Fig. 2b).

**Comparison of \textit{FliE}$_E$, \textit{FliE}$_E$, and \textit{fliE}$_S$ and their products.** The genomic organization of \textit{fliE} and its relationship to \textit{fliF} are essentially the same in the two species (Fig. 2). The locations of the \textit{fliD}$_S$ and \textit{fliE}$_S$ genes on plasmid pOYA1 establish that, as with \textit{E. coli}, there is a large distance between regions IIIa and IIIb. The deduced \textit{FliE} amino acid sequences are also very similar (Fig. 3). They contain 104 amino acids and have sequences that are 82% identical; there is one region of significant divergence, between residues 20 and 32, where eight positions differ. The numbers of Met residues (eight before cleavage of the N-terminal residue) and Cys residues (none) in the two proteins are the same.

**Stoichiometry of \textit{FliE} in the hook-basal body complex.** In a previous study involving gel electrophoresis and quantitative autoradiography of 35S-labeled hook-basal body complexes of \textit{S. typhimurium} (16), we had estimated the stoichiometry of known proteins by measuring the total number of sulfur atoms contributed to a hook-basal body complex by a given protein and dividing by the number of sulfur-containing residues in its sequence. We had also recorded the primary radiolabeling data for proteins of unknown genetic origin. By using the previous estimate of 64.8 ± 11.2 for the total...
number of sulfur atoms contributed by the protein that we now know to be FlIE (13) and dividing by the number of sulfur-containing residues per subunit of the mature protein (seven), we estimated that there are 9.3 ± 1.6 subunits of FlIE in the hook-basal body complex.

DISCUSSION

This study had as its goal the characterization of the poorly understood flIE gene and its product. We established the genomic organization of flIE and obtained its gene sequence in both E. coli and S. typhimurium. We showed that FlIE is a basal body protein, estimated its stoichiometry in that structure, and formed some tentative conclusions as to its location within the structure.

Characterization of the flIE gene and its location in flagellar region IIIb. Until recently, all of genes fliA through fliR of E. coli and S. typhimurium had been thought to be a single contiguous flagellar region, region III. We have, however, established that a nonflagellar region of the chromosome separates the fliA, fliC, and fliD genes from the fliF and fliB operons, so that region III is in fact two regions, IIIa and IIIb (17).

The exact location of the flIE gene in relation to the nonflagellar region and to the flID and flIF genes was not known. We have now shown that it lies immediately adjacent to and divergent from the flIF operon (which contains the genes flIF through flIK). There is no indication of any gene other than flIE within the flIF operon, consistent with the genetic conclusion that it is monocistronic (20). The orientation of flIE had not been established before, whereas the orientation of the multigene flagellar operons was already known from genetic analysis of polarity. Most of the flagellar genes are located in multicistronic operons. The fact that flIE is an independent transcriptional unit may reflect a unique structural or functional role.

The pattern of expression of the flIE operon (19, 24) places it within class 2, the intermediate class in the transcriptional hierarchy of the flagellar regulon (24). Its putative promoter conforms to this classification (24), since it contains the −10 consensus that is characteristic of flagellar promoters of both classes 2 and 3, but not the −35 consensus that is specific for those of class 3. The gene appears to be expressed at quite low levels, judging from its usage of rare codons and the fact that we were unable to detect the product of the cloned gene consistently in minicell experiments.

We found no evidence for further flagellar operons between flIE and the XbaI site that defines the start of the known nonflagellar region in E. coli. It is likely, therefore, that the 3′ end of flIE represents the start of flagellar region IIIb. From published and unpublished DNA sequence data (14, 18, 22, 28, 29, 32, 44), it is known that region IIIb then continues uninterrupted through to fliR. The organization of genes in region IIIb in E. coli and S. typhimurium is summarized in Fig. 4.

The role of FlIE in flagellar structure and in the flagellar assembly process. From comparison of the deduced amino acid sequence of FlIE and the experimentally determined sequence of a flagellar protein of previously unassigned genetic origin, we have demonstrated that FlIE is an 11-kDa protein located in the hook-basal body complex. This represents a further step towards a complete definition of that structure.

We do not know where FlIE is located within the hook-basal body complex or precisely what its role is, but several lines of evidence suggest that it is a relatively early component in the assembly process. Suzuki and colleagues found that the flIE gene was necessary for the earliest structure they could detect, the so-called rivet consisting of the MS ring plus the flagellar rod (Fig. 5) (40, 41). Biochemical analysis of partial structures (21) has since revealed that the rivet contains FlIE as well as the MS-ring protein FlIF (14, 43) and the four rod proteins FlgB, FlgC, FlgF, and FlgG (11). In a study of the assembly pathway using temperature-sensitive mutants and a differential radiolabeling protocol (15), the labeling pattern of FlIE in various mutant backgrounds indicated that FlIE was assembled into the basal body after the M ring and at about the same time as the proximal rod (13). Finally, the chromosomal location of flIE is suggestive: flIE lies in region IIIb along with genes for early structures such as the MS ring and the switch, whereas the rod and later components are all encoded by genes in either region I (genes for the rod, hook, hook-filament junction proteins, and outer rings) or region IIIa (genes for the filament and filament capping protein).

We can also consider the location of FlIE in the context of the processes necessary for export of components that are external to the cell membrane. It clearly is not exported by the primary cellular pathway, since it does not have a cleaved signal peptide; among flagellar proteins, only the P- and L-ring proteins appear to employ this pathway (10, 14). Nor does its sequence have the characteristics shared by those of the axial family of proteins that are believed to be exported by the flagellum-specific export pathway (9, 11). The available evidence therefore suggests that FlIE is not exported at all. If this is the case, it must be either a peripheral membrane component or a component that actu-
ally spans the membrane. Its fairly hydrophilic sequence seems to argue against the latter possibility, but it should be remembered that the algorithms for prediction of membrane proteins are not necessarily well suited to identification of components of multisubunit complexes such as the flagellum, in which a protein might span all or part of the membrane yet not be in direct contact with lipid.

A poorly understood feature of the basal body has been the S ring, and until recently it would have been tempting to suggest that FliE is responsible for that structure. This possibility is ruled out by the discovery that the M and S rings are in fact constructed from subunits of a single protein, FliF, whose domain structure apparently gives rise to the double-ring morphology (43).

The estimated stoichiometry of about nine FliE subunits per basal body is higher than that for the proximal-rod proteins FliB, FliC, and FliD (about 6 subunits each) but lower than that for the MS-ring protein FliF (about 27 subunits each) (16). This suggests FliE is incorporated into a different type of structure than either the MS ring or the rod. The MS ring has annular symmetry, whereas it seems likely that the rod (like the hook and filament) has helical symmetry (11). A possible role for FliE might then be as an adaptor between these structures.

Although we feel that overall the data are most consistent with a location for FliE near the MS ring, further work is needed to establish positively its location and its functional role in the flagellar basal body.

ACKNOWLEDGMENTS

We thank May Kihara for obtaining the sequence of the S. typhimurium fliE gene and Kenji Oosawa and Shigeru Yamaguchi for assistance in constructing plasmid pOYA1.

This work was supported by Public Health Services grant AI2202 (to R.M.M.). V.M. was supported by the Deutsche Forschungsgemeinschaft.

REFERENCES


