Characterization of the Flagellar Hook Length Control Protein FliK of Salmonella typhimurium and Escherichia coli

IKURO KAWAGISHI,† MICHIO HOMMA,† ANDREW W. WILLIAMS,‡ AND ROBERT M. MACNAB*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114

Received 27 December 1995/Accepted 14 March 1996

During flagellar morphogenesis in Salmonella typhimurium and Escherichia coli, the fliK gene product is responsible for hook length control. A previous study (M. Homma, T. Iino, and R. M. Macnab, J. Bacteriol. 170:2221–2228, 1988) had suggested that the fliK gene may generate two products; we have confirmed that both proteins are products of the fliK gene and have eliminated several possible explanations for the two forms. We have determined the DNA sequence of the fliK gene in both bacterial species. The deduced amino acid sequences of the wild-type FliK proteins of S. typhimurium and E. coli correspond to molecular masses of 41,748 and 39,246 Da, respectively, and are fairly hydrophilic. Alignment of the sequences gives an identity level of 50%, which is low for homologous flagellar proteins from S. typhimurium and E. coli; the C-terminal sequence is the most highly conserved part (71% identity in the last 154 amino acids). The central and C-terminal regions are rich in proline and glutamine residues, respectively. Linker insertion mutagenesis of the conserved C-terminal region completely abolished motility, whereas disruption of the less conserved N-terminal and central regions had little or no effect. We suggest that the N-terminal (or N-terminal and central) and C-terminal regions may constitute domains. For several reasons, we consider it unlikely that FliK is functioning as a molecular ruler for determining hook length and conclude that it is probably employing a novel mechanism.

Bacterial flagella of Salmonella typhimurium and Escherichia coli consist of three distinct parts, the filament, the hook, and the basal body (for a review, see reference 23). The basal body contains an axial rod, as well as four ring structures, two located in the outer membrane and the periplasm (the L and P rings, respectively), one embedded in the inner membrane (the MS ring), and one in the cytoplasm mounted onto the MS ring (the C ring). The flagellar filament extends up to 10 μm from the cell and is the propulsive component of the system. The hook connects the filament to the basal body, transmitting the rotation of the rod by the flagellar motor and acting as a flexible coupling or universal joint.

The hook of the bacterial flagellum has a fairly well-defined length. A recent quantitative analysis (7) has established a mean length of 55 nm for the wild-type hook, with a standard deviation of about 10%. This is in contrast to the filament, which has no active length control (although its rate of elongation decreases with length and it is susceptible to accidental shortening by breakage). The hook probably needs to be of a particular length, because if it was too short it would not generate a sufficient bend angle and if it was too long it would not be able to transmit torque efficiently to the filament. The control of hook length does not appear to be intrinsic to its structure, since purified hook monomer can polymerize indefinitely in vitro (15).

Mutants that possess hooks of indeterminate length have been isolated. These hooks have been called polyhooks, although the term is somewhat misleading since each polyhook consists of a single abnormally long structure of arbitrary length rather than multiple units of a defined length (7). Polyhook mutants not only fail to control hook length but also fail to add filament to the tips of the polyhooks. Pseudorevertants of polyhook mutants which are still largely defective in hook length control but have regained the ability to assemble filament (polyhook-filament mutants) have been isolated (33).

The gene responsible for the polyhook mutant phenotype is not the hook structural gene, flgE, but another flagellar gene, fliK (28, 33). The FliK protein has not been detected within the flagellar structure, and the means by which it controls hook length is not known. The organization of the subunits in polyhooks appears to be identical to that in normal hooks, and indeed polyhooks have been used for biochemical characterization of the hook protein (13, 14) and for analysis of the hook structure (27, 36). Also, simple overexpression of flgE in vivo can result in abnormal hook length (12). Thus, it seems unlikely that biochemical modification of the hook protein by FliK is the means of length control.

Several biological structures are believed to have their length determined by a molecular-ruler mechanism, in which a protein in an elongated state determines the number of subunits of a globular protein that assemble via quaternary interactions into a filamentous structure; these structures include the tail of phage lambda (with the GpH protein as the ruler) (16, 17) and the thin (F-actin) filaments of muscle (with nebulin as the ruler) (21). We have determined the fliK gene sequence in both S. typhimurium (fliKs) and E. coli (fliKc) and conclude from the deduced amino acid sequence that FliK is unlikely to be a molecular ruler. The sequence instead suggests that FliK may be a two- or three-domain protein, and insertion mutagenesis experiments indicate that the C-terminal region (which is the most highly conserved region) is especially important for function. The possibility that the N-terminal (or N-terminal and central) and C-terminal domains might correspond functionally to two separable processes is explored in an accompanying paper (37), in which mutations responsible for polyhook and polyhook-filament phenotypes are analyzed.
MATERIALS AND METHODS

Chemicals and enzymes. BglII linkers (dCAAGATCTG, dGAAGATCTC, and dGGAAGATCTTC) (BglII sites are in boldface) were obtained from New England Biolabs, Inc. (Beverly, Mass.). Universal primers for DNA sequencing were obtained from standard commercial sources. All other chemicals and enzymes were of at least reagent grade and were obtained from standard commercial sources.

Recombinant DNA techniques. Restriction enzymes, T4 DNA ligase, and DNA polymerase Klenow fragment were used according to the manufacturers’ recommendations. Isolation of plasmid DNA was carried out according to standard procedures (29). Subclones of fliKS and fliKE were constructed by standard techniques (29). Deletion derivatives of these clones for DNA sequencing were obtained either by digesting replicative forms of phage DNA with suitable restriction enzymes, treating them with Klenow enzyme, and then ligating them with T4 DNA ligase or by the single-strand sequential method (3) using the CYCLONE I Biosystem (International Biotechnologies, Inc., New Haven, Conn.). The dideoxynucleotide method (30), using the modified T7 DNA polymerase (34) Sequenase (U.S. Biochemical, Cleveland, Ohio) and occasionally using dITP in the place of dGTP, was employed for all DNA sequencing.

Sequence manipulation and analysis. Routine sequence manipulations were carried out using the personal computer package DNAANALYZE (Gregory Wernke, University of Cincinnati Medical School). Sequence comparisons were run at the National Center for Biotechnology Information, Washington, D.C., by using the program TBLASTN and the combined nonredundant database (31, August 1995 version).

Minicell analysis. Minicells from strain UH869 (11) were isolated, and plasmid-encoded proteins were labeled with [35S]methionine as described previously (24). For pulse-chase labeling, a 150-μl aliquot of minicells (~2 × 10^10 cells ml^-1) was incubated with 15 μCi of [35S]methionine and chased with 5 μl of an 8-mg ml^-1 solution of unlabeled methionine and the chase was terminated by addition of 16 μl of 50% trichloroacetic acid. Samples were then subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography as described previously (38).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this paper have been deposited with GenBank under accession numbers L43492 (S. typhimurium) and L43491 (E. coli).

RESULTS

Construction of plasmids carrying fliK_S and fliK_E. Published deletion mapping data had indicated where fliK should be located in cloned DNA fragments from the S. typhimurium and E. coli chromosomes (2, 9). fliK_S was subcloned from pMH11 into pUC19 as a 1.7-kb SalI-SacI fragment, yielding plasmid pK600. fliK_E was subcloned from pDB7 into pUC19 as a 1.4-kb PstI-BamHI fragment, yielding plasmid pIK800. Both plasmids were able to complement an E. coli fliK null mutant, MS694 (32), in a swarm plate assay.

Nucleotide sequences of the DNA fragments containing fliK_S and fliK_E. The nucleotide sequences of the fliK-complementing DNA fragments in plasmids pAW11 and pAW12 (see the text) are shown in brackets in panel a; only the rightmost boundary of the pAW12 deletion is shown.

FIG. 1. Nucleotide sequences of the 5′ and 3′ portions of the hook length control gene fliK in S. typhimurium (a) and E. coli (b), together with the flanking sequences, namely, the 3′ end of fliI (18, 39). The central portions of the fliK sequences (............) are not shown. The deduced amino acid sequence of FliK is shown below the corresponding DNA sequence. Potential ribosome binding sites (Shine-Dalgarno sequences [S-D]) are indicated by underlining. Potential stem-loop sequences that could function in transcription termination of the flf/GHfIK operon are indicated by horizontal arrows. The putative flagellum-specific promoter for flfL (pflL) is also indicated by underlining. The amino acid sequences that were deleted in plasmids pAW11 and pAW12 (see the text) are shown in brackets in panel a; only the rightmost boundary of the pAW12 deletion is shown.
At the position of the first discrepancy, we find that both *S. typhimurium* and *E. coli* have glutamine at the center of the conserved sequence QQQ; at the position of the second discrepancy, we find that the *S. typhimurium* and *E. coli* sequences are highly similar (GNGAVDIFA[stop] and GNSGVDIFA[stop], respectively), whereas the frameshift in the published *E. coli* sequence gives a quite different sequence (WQQRR[stop]). We therefore consider that our sequence is correct.

**Identification of the fliK and fliK genes.** The identification of the open reading frames shown in Fig. 1 as fliK and fliK is supported by several lines of evidence. (i) The orientation of the open reading frames agrees with the known orientation of the fliFGHIJK operon (20). (ii) The open reading frames start with an ATG codon which overlaps (ATGA) with the termination codon of the upstream gene, fliI (35); such overlap is a common feature for flagellar genes within the same operon (22). (iii) The 5' portion of the following open reading frames is identical to that of the published DNA sequences of fliL (18) and fliL (19), in agreement with the known organization of flagellar genes in this region. (iv) The open reading frames are preceded by a good ribosome binding site consensus (AGGA; 7 bp before the ATG codon for both species). (v) The sizes of the open reading frames are approximately those expected from the apparent molecular masses of the gene products (see below). (vi) The deduced amino acid sequences of fliK and fliK show considerable similarity (see below). (vii) For both species, the open reading frames were assessed as coding regions (at confidence levels of 96% for *S. typhimurium* and 88% for *E. coli*) by the program TESTCODE (5); codon adaptation indices (31) calculated for fliK and fliK are 0.26 and 0.30, respectively, indicating a low to moderate level of synthesis, as is typical for other flagellar proteins.

The intergenic region between the fliK and fliL coding regions is 104 bp long in both *S. typhimurium* and *E. coli* and contains a stem-loop sequence (WQQRR[stop]) that may represent a termination signal for the fliFGHIJK operon. It also contains the putative promoter for the fliLMNOPQR operon and a potential ribosome binding site for the fliL gene (18, 19).

**The deduced amino acid sequences of FliK and FliK.** The deduced molecular masses of FliK and FliK are 41,748 Da (405 residues) and 39,246 Da (375 residues), respectively. The deduced molecular mass of FliK (41,748 Da) is similar to the apparent value of 44 to 46 kDa estimated by SDS-PAGE (9; also this study). However, the deduced molecular mass of

FIG. 2. Alignment of the amino acid sequences of FliK from *S. typhimurium* (above) and *E. coli* (below). Identities are indicated by vertical dashes, conservative changes (among I-L-V-M-F, P-Y, A-G, D-E, H-K-R, N-Q, or S-T) are indicated by dots, and gaps are indicated by hyphens. The central region of FliK contains a large number of prolines (+). Note also the high glutamine content of the C-terminal region. Vertical arrows 1 to 3 show the sites of insertion of oligopeptides into the *E. coli* sequence, as described in the text, as follows: 1, pIK812; 2, pIK808; and 3, pIK815.

FIG. 3. Phenotype of fliK mutant alleles constructed in vitro. Cells of an *E. coli* fliK amber mutant strain, MS694, were transformed with plasmids carrying wild-type (pIK800) and mutant (pIK812, pIK808, and pIK815) alleles of fliK (see the text). The mutant alleles contain short in-frame insertions at the residues indicated in this figure and in Fig. 2. Cells were inoculated in triplicate onto semisolid tryptone agar plates and incubated for 8 h at 30°C, and their motility was scored.
FIG. 4. Deletion analysis of the two forms of FliK. An autoradiograph of radiolabeled minicells containing plasmids with either the wild-type fliK allele (pIK600) or mutant alleles (pAW11 and pAW12) with in-frame deletions of the codons shown; pUC19 was used as a control. The two forms of FliK are indicated by arrows and by the labels FliK for pIK600 and FliKΔ for pAW12. With pAW11, there is a small but significant shift in the positions of both forms of FliK compared with wild-type FliK with pAW12 there is a large shift of about the expected magnitude. Molecular masses (in kilodaltons) of standards are shown to the left.

FliK_E (39,246 Da) is considerably lower than the value of 54 kDa reported as its apparent molecular mass (2). We have confirmed in minicell expression experiments with plasmid pIK800 that FliK_E does have an anomalously low mobility in SDS-PAGE gels (data not shown); the reason for this is unknown.

Comparison of the sequences with the combined nonredundant database at the National Center for Biotechnology Information yielded highly significant similarity between the S. typhimurium and E. coli sequences (see below). The only other significant similarity found was a weak one to the deduced product of Bacillus subtilis orf7 (a presumed gene within a large flagellar operon), as was noted previously by Albertini et al. (1) in a comparison with our unpublished S. typhimurium and E. coli sequences.

Hydropathy analysis (4) indicates that FliK is unlikely to be a membrane protein (data not shown), consistent with its cellular location in minicell expression experiments (9). Most of the flagellar structural proteins that are external to the cell (the so-called axial family, including the hook and filament proteins) have certain sequence similarities, including N- and C-terminal hydrophobic repeats and limited consensus sequences (8, 10). FliK does not have any of these features, and so it is not likely to be a member of the axial family.

The central one-third of the protein is proline rich. For example, the 96-residue stretch between residues 139 and 234 of FliK_E contains 17 prolines, or 18 mol%, whereas the remainder of the sequence has a proline content of only 5 mol%. The C-terminal part of the protein is glutamine rich, with one stretch of about 100 residues having a content of over 20 mol%.

Alignment of the deduced amino acid sequences from the two species (Fig. 2) indicates an identity level of around 50%. This level, though highly significant statistically, is much lower than those for most other flagellar proteins, where values of greater than 80% are common. Also, the alignment entailed introduction of substantial gaps, especially in the central region of the sequence. The N-terminal and central portions of FliK are the least conserved parts. Thus, the first 251 residues are only 36% identical and contain only a few appreciable stretches of identity (the longest being 12 consecutive residues centered around residue 140). The N-terminal sequence is somewhat better conserved than the central portion of the molecule. In contrast, the last 154 residues are 71% identical and contain only a single one-residue gap. This is also the best-conserved part of the sequence in comparisons with the putative FliK protein (Orf7) of B. subtilis (1).

From the features described above, FliK can be considered to consist of three regions. The N-terminal region, representing roughly one-third of the sequence, is not strongly conserved and contains some gaps. The central one-third of the sequence is even less conserved, contains extensive gaps, and is proline rich. The C-terminal one-third of the sequence is well conserved, has essentially no gaps, and is glutamine rich. These data suggested to us that FliK might consist of two or more domains.

The unusually high glutamine content of the C-terminal region suggests that it may have a specialized function. Glutamine-rich activation domains exist in a number of eukaryotic transcription factors (26), but we have no evidence that FliK is functioning in transcriptional control.

Construction and characterization of oligonucleotide linker-insertion mutant alleles of fliK_E. In an initial attempt to address the importance of the various regions of FliK, we constructed three in-frame linker-insertion mutant alleles of the fliK_E gene in plasmid pIK800. pIK812 has a 12-mer BglII linker inserted at the HpaI site, pIK808 has a 10-mer BgII linker at the Klenow fragment-treated BsrEI site, and pIK815 has an 8-mer BgII linker at the Klenow fragment-treated HindIII site. These linkers result in the insertion of short amino acid sequences at residues 98, 208, and 324 of the 375-residue sequence, as shown by arrows 1 to 3 in Fig. 2.

Cells of the fliK null mutant MS694 were transformed with these plasmids, and the transformants were tested for motility (Fig. 3). The insertions had increasingly severe effects as they were placed progressively closer to the C terminus: pIK815 gave essentially wild-type swimming, pIK808 gave about 50% of wild-type swimming, and pIK812 gave no swimming at all.

FliK_E occurs in two forms. Previous experiments (9) had suggested that expression of the S. typhimurium fliK gene (and possibly also that of the E. coli gene [see Fig. 8 of reference 9]) resulted in two products with similar apparent molecular masses and isoelectric points, the protein with the higher apparent molecular mass being the predominant form. To test whether both proteins do in fact derive from expression of fliK_E, we constructed two in-frame deletion derivatives of plasmid pIK600, making use of existing restriction sites within the gene. pAW11 utilized two BclI sites and resulted in deletion of codons 2 to 7 (Fig. 1a). pAW12 utilized two MluI sites, which after filling in with Klenow fragment and blunt-end ligation resulted in the deletion of codons 285 to 375.

Minicell expression experiments with pIK600 confirmed the original finding of two products with apparent molecular masses of 44 and 43 kDa (Fig. 4, second lane from the left). When the deletion plasmids were used, there were subsequent shifts in the mobilities of both bands. The shifts for pAW11 (third lane) and pAW12 (fourth lane) corresponded to apparent losses of about 1 and 9.5 kDa, respectively, close to the expected values. These results demonstrate that the 44- and 43-kDa proteins are both products of the fliK_E gene. Pulse-chase labeling indicated that the 44-kDa form was the precursor of the 43-kDa form (data not shown).

These results also establish that the two forms do not reflect (i) N-terminal peptide cleavage (since the position of the second band in pAW11 would not be expected to shift in this case), (ii) modification in the immediate vicinity of the N terminus, or (iii) modification within a substantial part of the C
terminus. Removal of the N-terminal methionine is unlikely because while such removal is common with bacterial proteins, it is not observed when—as is the case with both wild-type FliK and the mutant FliK of pAW11—the second residue is isoleucine (25).

**DISCUSSION**

The two forms of FliK. Previous studies had suggested that fliK might be giving rise to two products closely similar in size and pl (9). We have now demonstrated that both proteins are products of the fliK gene. We have eliminated several possible mechanisms for generation of the second form, such as modification or cleavage of the N terminus or modification of a 91-residue region near the C terminus. If the process involved can be identified, it may be possible to block it and determine whether it is important for hook length control or filament assembly.

**FliK is unlikely to be a molecular ruler.** Several filamentous structures, such as phage tails (16, 17) and thin filaments in muscle (21), have their lengths controlled by a molecular-ruler mechanism. It appears that the rulers in these cases are predominantly stretches of polypeptide in an α-helical state. While the molecular size of FliK (ca. 400 residues) is in the right range to be an α-helical molecular ruler for the hook (400 \(\times\) 0.15 = 60 nm), its amino acid sequence makes it highly unlikely that it functions in this way. Predictions of secondary structure show no pronounced tendency towards α-helix, and the central region, with its high proline content, is almost certainly not α-helical. Furthermore, the highly asymmetric composition and degree of conservation of the molecule are not characteristics that one would expect to be associated with a molecular ruler.

There are at least two other reasons to doubt that FliK functions as a molecular ruler. (i) It would need to be responsible for terminating elongation rather than enabling it, since its absence results in abnormally long hooks rather than zero-length hooks (33). (ii) It would have to function externally to the cell, and yet it has never been found in that location. We suspect, therefore, that the length of the bacterial flagellar hook is controlled by a novel mechanism, one distinct from those that have been described for other structures, such as phage tails, thin filaments, and (with DNA or RNA rather than protein as the ruler) filamentous phages such as M13 (6).

**Possible domain structure of FliK.** The amino acid sequence of the FliK protein suggests that it consists of three regions. The fact that the central region is poorly conserved and of substantially different sizes in the two species seems to argue against any highly specific function, and its high proline content suggests that it may not have large amounts of well-defined secondary structure. This initially led us to hypothesize that FliK is a two-domain protein with the central portion of the sequence acting as a hinge or connector between the two; however, the linker insertion mutagenesis results indicate that the functional distinction between the N-terminal and central regions may not be an absolute one, since disruption of at least part of the N-terminal region can be tolerated without loss of function. Disruption of the proline-rich region caused some loss of function. However, the insertion was towards the end of the region, at a point where sequence conservation between S. typhimurium and E. coli begins to increase rapidly (Fig. 2), suggesting that the insertion might have an effect on the stability of the C-terminal region. Disruption of the conserved C-terminal region itself caused total loss of function.

Thus, there are indications that the FliK protein may have a domain structure. Is there any experimental evidence that FliK possesses multiple functions? FliK mutants are characterized by a double failure: the inability to terminate the hook at its proper length and the inability to initiate filament assembly. The existence of intragenic suppressor mutants with double mutations in fliK which still fail in hook length control but succeed in filament assembly at the tips of the polyhooks seems to argue in favor of the hypothesis that FliK is a bifunctional protein. In the companion paper (37), we explore this hypothesis by examining the nature of mutations responsible for the polyhook phenotype and by characterizing the suppressor mutations that permit filament addition.

**ACKNOWLEDGMENTS**

We thank Stavros Stavropoulos for his contribution to the sequencing of the E. coli fliK gene, Phil Matsumura (University of Illinois, Chicago) for the gift of plasmid pDB7, and May Kihara for technical assistance.

This work was supported by USPHS grant AI12202.

**REFERENCES**