JOURNAL OF BACTERIOLOGY, June 1991, p. 3573–3579
0021-9193/91/113573-07$02.00/0

Vol. 173, No. 11

The *flaA* Locus of *Bacillus subtilis* Is Part of a Large Operon Coding for Flagellar Structures, Motility Functions, and an ATPase-Like Polypeptide

A. M. ALBERTINI,[1] T. CARAMORI,[1] W. D. CRABB,[2] F. SCOFFONE,[3] AND A. GALIZZI[1,4]*

Dipartimento di Genetica e Microbiologia "A. Buzzati-Traverso," Università degli Studi di Pavia,[1] Istituto di Genetica Biochimica ed Evoluzionistica, Consiglio Nazionale delle Ricerche,[3] and Centro Interuniversitario per lo Studio delle Macromolecole Informazionali,[4] 27100 Pavia, Italy, and Genencor, South San Francisco, California 94080[2]

Received 14 December 1990/Accepted 13 March 1991

We cloned and sequenced 8.3 kb of *Bacillus subtilis* DNA corresponding to the *flaA* locus involved in flagellar biosynthesis, motility, and chemotaxis. The DNA sequence revealed the presence of 10 complete and 2 incomplete open reading frames. Comparison of the deduced amino acid sequences to data banks showed similarities of nine of the deduced products to a number of proteins of *Escherichia coli* and *Salmonella typhimurium* for which a role in flagellar functioning has been directly demonstrated. In particular, the sequence data suggest that the *flaA* operon codes for the M-ring protein, components of the motor switch, and the distal part of the basal-body rod. The gene order is remarkably similar to that described for region III of the enterobacterial flagellar regulon. One of the open reading frames was translated into a protein with 48% amino acid identity to *S. typhimurium* FliI and 29% identity to the β subunit of *E. coli* ATP synthase.

Many bacterial species swim actively by means of flagella. Swimming tends to move the bacteria towards attractants and away from repellents. The flagellar organelle is made of three parts: the basal body, the hook, and the filament, with more than 10 different proteins involved in flagellar formation. In addition, there are the motor apparatus and the switch, necessary for rotation of the flagellar filaments and control of the sense (clockwise or counterclockwise) of rotation. A number of sensory elements and components for transduction of the signal to the motor switch are also part of the motility system. A considerable body of data is available regarding the structure and functioning of the flagella of *Escherichia coli* and *Salmonella typhimurium* (reviewed in references 21 and 22).

About the gram-positive bacterium *Bacillus subtilis* much less is known. A limited number of mutants affected in motility have been isolated, and a few have been characterized (8, 32). More advanced are the studies related to the chemotactic mechanisms that operate in *B. subtilis*. A large number of mutants defective in chemotaxis have been isolated, 21 different complementation groups have been identified, and part of a chemotactic locus has been cloned (28, 29). Better knowledge of the genes and gene products of *B. subtilis* involved in flagellar assembly, motility, and taxis seemed desirable in view of the fact that *B. subtilis* is a gram-positive bacterium and that its chemotactic response is more complex than that of *E. coli* (37). From a *B. subtilis* chromosomal DNA library in λCharon 4A (5) we have isolated one clone that contains the wild-type alleles of two known *flaA* mutations, *flaA4* and *flaA15* (9).

In this report, we describe the nucleotide sequence of 8,302 contiguous nucleotides of the *flaA* locus. Insertional inactivation with integrative plasmids showed that the region sequenced is part of a single operon involved in motility (10, 44). Nine of the deduced gene products show similarity to *E. coli* and *S. typhimurium* proteins known to be involved in flagellar structures and functioning and suggest the identification of elements of the motor switch, the rod of the basal body, and the M ring. One is highly similar to the β subunit of *E. coli* ATP synthase (and the *S. typhimurium fliI* gene product). Of the 11 gene products, 2 appear to have no known counterpart in *E. coli* and *S. typhimurium*.

In the course of this work, we adopted the term *flaA* locus for the entire complex of genes involved in flagellation and motility (it has been called the *che* locus by Zuberi et al. [44]). The cistrons identified from the DNA sequence were named *flaAorf1*, *flaAorf2*, etc. (usually abbreviated *orf1*, *orf2*, etc.). The gene products are called FlaAOrf1, FlaAOrf2, etc., or abbreviated Orf1, Orf2, etc. When the nucleotide sequence of the *flaA* locus is completed, a more rational nomenclature can be adopted. For *E. coli* and *S. typhimurium*, we used the unified nomenclature (15).

MATERIALS AND METHODS

Bacterial strains. *E. coli* DP50 *supF* [F⁻ *tonA53 glnV44 tyrT58 dapD8 lacY1 supE44 supF58* Δ(*gal-uvrB*)*47 gyrA29 thyA57 hsdS3*] and LE392 (*supE44 supF58 hsdR514 galK2 galT22 metB1 trpR55 lacY1*) were used as hosts for Charon 4A recombinant derivatives. *E. coli* cultures were grown in TY medium (tryptone [Difco Laboratories, Detroit, Mich.], 10 g; yeast extract, 5 g; NaCl, 5 g; $MgCl_2$, 2 g; water to 1 liter, pH 7.1). For strain DP50, the medium was supplemented with diaminopimelic acid (100 μg/ml) and thymidine (50 μg/ml). TY medium was solidified by adding 12 g of agar per liter or 10 g of agarose per liter as required. For bacteriophage plating, the top layer contained 4 g of agar per liter or 5 g of agarose per liter. *E. coli* JM103 [*supE44 thi rpsL20 endA hspR4* Δ(*lac-proAB*) F′ (*traD36 proAB lacI*^{q} *ZΔM15*)] and DH5α [*supE44 lacU169* (φ80*lacZ ΔM15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*] were used for transformation and plasmid propagation, and LB medium (tryptone [Difco], 10 g; yeast extract, 5 g; NaCl, 10 g; water to 1 liter; pH 7.1) was used. The *B. subtilis* strain was PB1424 (*hisH2 trpC2 metD4*).

* Corresponding author.

Scoring for motility. Colonies were transferred with toothpicks to swarm plates prepared with semisolid medium B (Difco Bacto-Peptone, 10 g; Difco gelatin, 80 g; Difco agar, 4 g; NaCl, 5 g; water to 1 liter). The medium was supplemented with the required amino acids at 20 μg/ml. The plates were incubated at 37°C overnight.

Insertional inactivation. A *Bgl*II fragment of 534 bp internal to *orf4* was cloned in pGEM-4Z for sequencing purposes. To produce a plasmid suitable for insertional inactivation, the insert was excised by *Eco*RI and *Hin*dIII digestion and recloned in pJM103 (31). The resultant plasmid (pBBfla) was used to transform competent cells of *B. subtilis* PB1424. Selection was for chloramphenicol resistance.

Sequence analysis. The plasmid used for cloning and sequencing was pGEM-4Z (Promega Biotec, Madison, Wis.). Sequence analysis of both strands was accomplished by the supercoil sequencing method of Chen and Seeburg (2). The dideoxy-chain termination and elongation reactions were performed by the method of Sanger et al. (34) with Sequenase (U.S. Biochemical Corp.) or T7 polymerase (Pharmacia-LKB) in accordance with the appropriate instructions. Custom primers were synthesized with a Cyclon Plus DNA Synthesizer (Millipore).

Sequence manipulations and comparisons. DNA sequence manipulation and translation routines, etc., were carried out with the IBI Pustell Sequence Analysis Program. Protein sequences were compared and aligned by using the CLUSTAL 4 program of Higgins and Sharp (11).

Phage and plasmid DNA preparation. Southern hybridizations and nick translation were performed by standard procedures (33).

Nucleotide sequence accession number. The nucleotide sequence in Fig. 1 has been submitted to the EMBL data library and assigned accession no. X56049.

RESULTS

Cloning and sequencing of part of the *flaA* locus. We isolated a lambda clone (λUF7) from the *B. subtilis* Charon 4A library of Ferrari et al. (5) by hybridization to RNA extracted from outgrowing spores and expected to contain an outgrowth-specific locus named *outI* (7). The λUF7 clone has an insert of about 12 kb consisting of two *Eco*RI fragments of 4.0 and 8.3 kb, respectively. The 8.3-kb *Eco*RI fragment can rescue strains with the *flaA4* or *flaA15* mutations (9). As a first step in the study of the *flaA* locus, we determined the nucleotide sequence of the 8.3-kb *Eco*RI fragment.

The DNA sequence and the deduced amino acid sequences of its products are shown in Fig. 1. The gene products were deduced purely from the nucleotide sequence, since no information was available regarding the properties of the encoded proteins. We considered valid the longest possible open reading frame (ORF) starting with ATG, GTG, or TTG, preceded by a good Shine-Dalgarno sequence or overlapping the preceding ORF. Similarity of the deduced products to known proteins supported the assignments a posteriori.

Nucleotide sequence analysis. Computer analysis of the nucleotide sequence showed the presence of nine complete ORFs plus two ORFs at the beginning and end of the sequenced fragment that were incomplete and could continue into the adjacent DNA fragments (Fig. 2). The nine ORFs were all in the same direction (from left to right in the figures), and in two instances (*orf3-orf4* and *orf7-orf8*) they appeared to overlap. The overlap of *orf3* with *orf4* is not strictly required from the sequence, since an initiation codon for *orf4* is also found downstream of the presumptive stop codon of *orf3*. We have assumed an overlap on the basis of the lack of a good Shine-Dalgarno sequence upstream of *orf4*. The start codon of *orf8* overlapped with the termination codon of the preceding gene.

The intergenic regions were often short (2 to 22 bases), but there were three examples of considerable distance between the termination of one gene and the initiation of another; the spacings between *orf2* and *orf3*, *orf6* and *orf7*, and *orf9* and *orf10* were 117, 161, and 251 bp, respectively. Within these regions there are potential stem-loop structures with a ΔG of less than −15 kcal (1 cal = 4.184 J) mol^{-1}, suggesting the presence of signals potentially involved in control of transcription or translation (Fig. 1). In all of these intercistronic regions there is at least one potential promoter sequence of the type recognized by the σ^D form of RNA polymerase, which is known to be involved in the transcription of the *B. subtilis* flagellin gene (26). If active, these promoters must be secondary sites of transcription initiation, since insertional inactivation and RNase protection experiments both indicated that the 8.3-kb *Eco*RI segment is part of a longer transcriptional unit (10, 44).

One additional ORF of 534 bp (*OrfC1*) was detected on the complementary strand, in correspondance to the middle portion of *orf4*.

Amino acid sequence similarity comparisons. The translation products of the proposed reading frames were compared to the National Biomedical Research Foundation-PIR protein data base, GenBank DNA sequences translated into proteins, and a collection of *S. typhimurium* flagellar sequences (23, 38). In a number of instances, relevant similarities to bacterial gene products involved in the assembly and functioning of flagella were obtained. The similarity scores are reported in Table 1. Table 1 also summarizes some of the properties deduced for the gene products of the *B. subtilis flaA* locus.

Orf1, Orf2, and FliG. Orf1 was incomplete but nevertheless had good similarity (data not shown) to *E. coli* FliF, the major M-ring component (16). The complete sequence of the *B. subtilis fliF* homolog has been determined by Zuberi et al. (43). Orf2 showed similarity to *Salmonella* switch protein FliG (36% identity and 58% conserved substitutions; Fig. 3). As the *Salmonella* counterpart (19), the Orf2 protein has a very high α-helical content (75%) as predicted by the algorithms of Chou and Fasman (3) and Garnier et al. (6). The moderately hydrophilic nature of the FliG protein and the very high α-helical content are consistent with location at the cytoplasmic face of the membrane, where the motor switch is expected to reside (19). The primary and deduced secondary structures of Orf2 suggest a similar location for the *B. subtilis* protein. By marker rescue experiments, we localized the *flaA15* mutation in *orf2* (9).

Orf3 and Orf5. Orf3 and Orf5 have 23% identity to each other. The strongest similarity is in a stretch of 10 amino acids corresponding to residues 135 to 144 of Orf3 and 130 to 139 of Orf5. The sequence is Glu-(Val/Met)-Lys-Glu-(Tyr/Met)-Asp-Asp-Ile-Ser-Ile. The two proteins are predicted to have a high α-helical content (81%) by the algorithms of Chou and Fasman (3) and Garnier et al. (6). In addition, Orf3 shows good similarity (23% identity) to FliH (Fig. 4) and Orf5 is similar to FliJ (40) of *S. typhimurium*. The function of these *Salmonella* proteins is unknown, but we found that the N terminus of Orf3 is somewhat similar to Trg (19% identity; data not shown), a protein required for chemotaxis and involved in transduction of the signal (1).

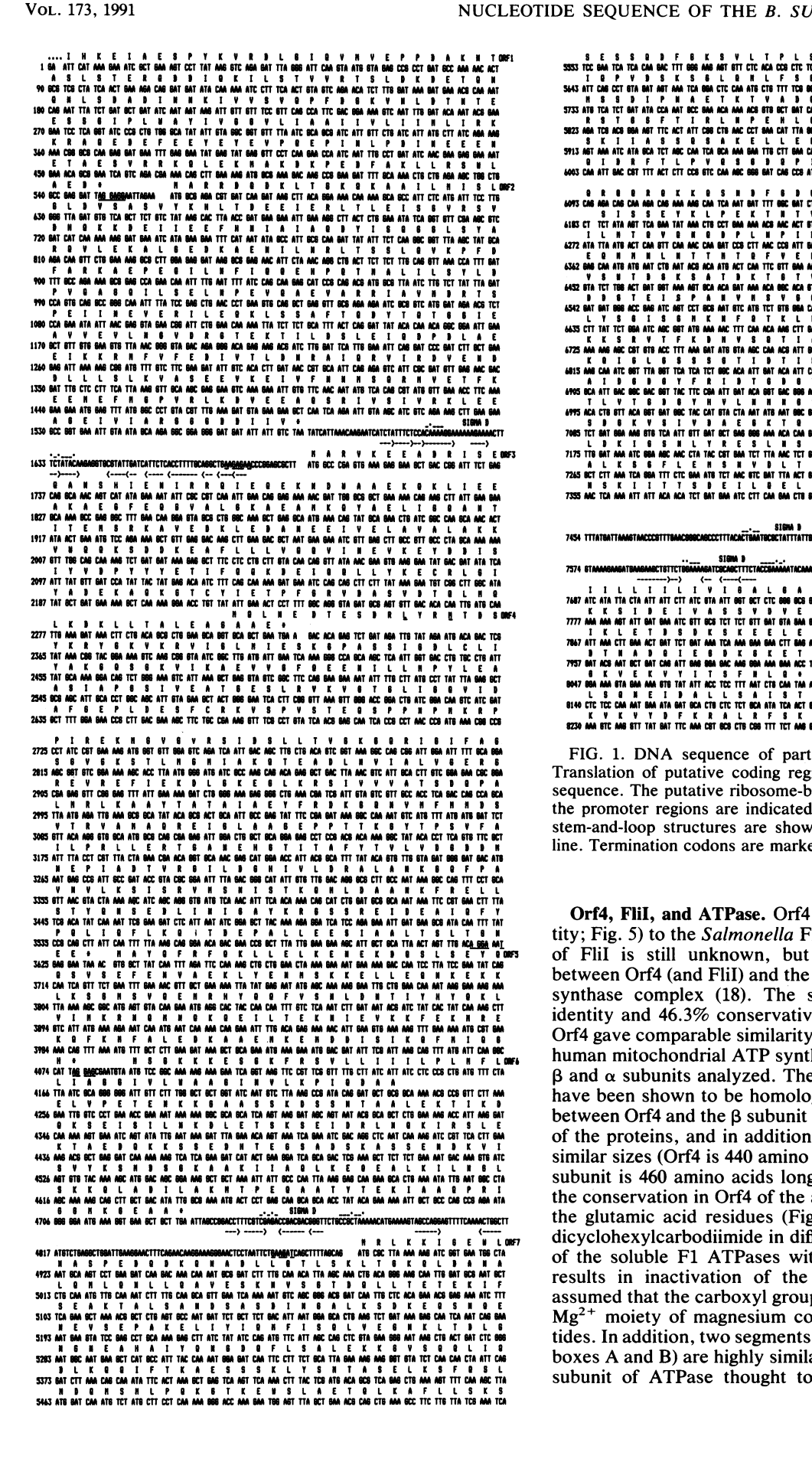

FIG. 1. DNA sequence of part of the *B. subtilis flaA* locus. Translation of putative coding regions is shown above the DNA sequence. The putative ribosome-binding sites are underlined, and the promoter regions are indicated above the sequence. Potential stem-and-loop structures are shown as dashed arrows below the line. Termination codons are marked with asterisks.

Orf4, FliI, and ATPase. Orf4 is highly similar (49% identity; Fig. 5) to the *Salmonella* FliI protein (38). The function of FliI is still unknown, but we found good similarity between Orf4 (and FliI) and the β subunit of the *E. coli* ATP synthase complex (18). The similarity scores were 29% identity and 46.3% conservative residue substitutions, and Orf4 gave comparable similarity scores with the β subunit of human mitochondrial ATP synthase (27) and all of the other β and α subunits analyzed. The α and β subunits of *E. coli* have been shown to be homologous (35, 39). The similarity between Orf4 and the β subunit is distributed over the length of the proteins, and in addition, the two polypeptides have similar sizes (Orf4 is 440 amino acids long, and the *E. coli* β subunit is 460 amino acids long). Particularly interesting is the conservation in Orf4 of the amino acid sequence around the glutamic acid residues (Fig. 5, box C) that react with dicyclohexylcarbodiimide in different species (41). Reaction of the soluble F1 ATPases with dicyclohexylcarbodiimide results in inactivation of the enzyme, and it has been assumed that the carboxyl groups might function to bind the Mg^{2+} moiety of magnesium complexes of adenine nucleotides. In addition, two segments of the Orf4 sequence (Fig. 5, boxes A and B) are highly similar to the *E. coli* sites of the β subunit of ATPase thought to be involved in nucleotide

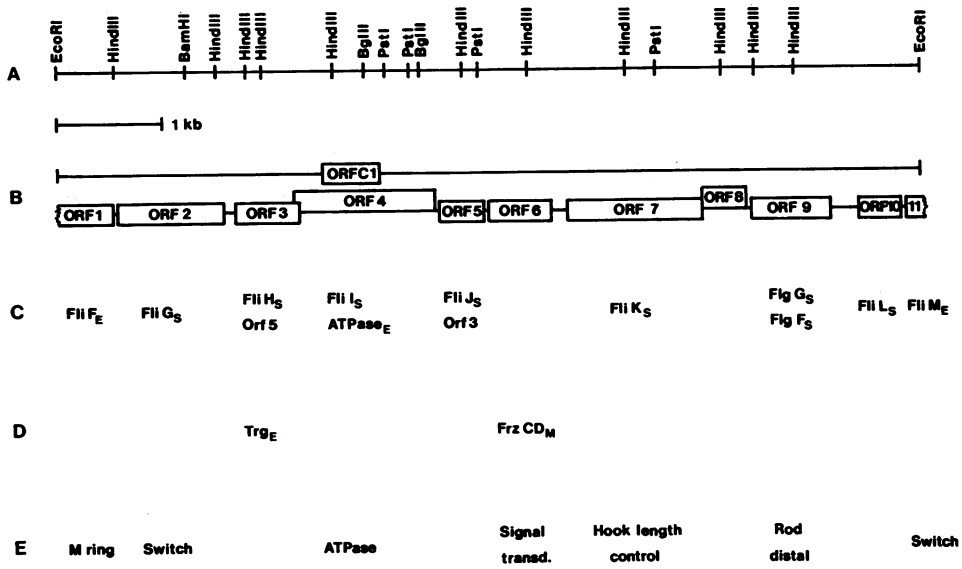

FIG. 2. Organization of the ORFs in the 8.3-kb *Eco*RI fragment. (A) Restriction map of the 8.3-kb fragment. (B) Positions and extents of the ORFs; a box above the line indicates overlapping of the ORF with the preceding one. ORFC1 represents a hypothetical reading frame on the complementary strand. The direction of transcription is from left to right. (C and D) Similarity among the deduced *B. subtilis* proteins and the gene products of other bacteria. Subscript symbols: E, *E. coli*; M, *M. xanthus*; S, *S. typhimurium*. Two levels of similarity are shown: above 20% identity (row C) and 10 to 20% identity (row D). (E) Possible structures or functions deduced from protein sequence similarity and analogy.

binding (18, 39). The two sites correlate with nucleotide-binding sites of adenylate kinase, RecA protein, ATP-ADP translocase, and phosphofructokinase (39). These observations indicate that Orf4 (and the *Salmonella* counterpart, FliI) acts as an ATPase. We performed inactivation experiments with an insertional plasmid carrying a DNA fragment internal to *orf4* (pBBfla). The only phenotype we detected among the transformants was inability to swarm on semisolid agar plates. We cannot rule out the possibility that the Mot$^-$ phenotype was due to a polar effect of the insertion, but it appears certain that the hypothetical ATPase encoded by *orf4* is not essential for *B. subtilis*.

Orf6. The Orf6 protein has a stretch of 31 hydrophobic amino acids between positions 11 and 41, with a proline residue at position 21. The hydrophobic region is preceded by a cluster of basic residues (four lysines among 10 residues) and followed by two hydrophilic domains. This organization could be interpreted as that of a protein anchored to the bacterial membrane via the hydrophobic region. Alternatively, the cluster of basic residues followed by a hydrophobic region could represent a signal sequence for translocation of the Orf6 protein through the cytoplasmic membrane.

The amino acid sequence of Orf6 had similarity (18% identity and 36.1% conserved residues; data not shown) to the C-terminal region of the FrzCD protein of *Myxococcus xanthus*, which in turn is similar to the methyl-accepting chemotaxis receptor proteins of enteric bacteria and has been shown to be modified by methylation (25).

Orf7, FliK, and Orf8. The protein encoded by *orf7* is 22% identical to the *Salmonella* FliK protein (23) and, as its counterpart, could control the length of the flagellar hook (30, 36). The Orf8 protein did not show any significant similarity to known proteins, and no particular feature could be predicted from the amino acid sequence.

Orf9, FlgG, and FlgF. The amino acid sequence of Orf9 is 40% identical to the *Salmonella* FlgG protein (13). FlgG is a structural component of the flagellar basal body, where it constitutes the distal part of the rod, and we suggest a similar role for the *B. subtilis* counterpart. Orf9 and FlgG have

```
ORF2    1   MARRDQDKLTGKQEAAILMISLGLDVSASVYEHLTDEEIERLTLEISGVRSV-DHQKKDE
FLIG    1   MS----N-LSGTDESVILLMTIGEDRAAEVFEHLSTREVQALSTAMANVRQISNEQLTD-

ORF2   61   IIEEFHNIAIAQDY-ISQGGLSYARQVLEKALGEDEA----ENILN-RLTSSLQVKPFDF
FLIG   55   VLSEFEQEAE-QFAALNINANETLRSVLVKALGEERASSLLEDILETRDTTSG-IETLNF

ORF2  114   ARKAEPEQILNFIQQEHPQTMALILSYLDPVQAGQILSELNPE-VQAEVARRIAVNDRTS
FLIG  113   M---EPQSAADLIRDEHPQIIATILVHLKRSQAADILA-LFDERLRHDVMLRIATFGGVQ

ORF2  173   PEIINEV-ERILEQKLSSAFTQDYTQTGGIE-AVVEVLNGVDRGTEKTILDSL-EIQDPD
FLIG  169   PAALAELTEV-LNGLLDGQ-NLKRSKMGGVRTAA-EIINLMKTQQEEAVITAVREF-DGE

ORF2  230   LAEEIKKRMPVFEDIVTLDNRAIQRVIRDVENDDLLLSLEVASEE-VKEIVFNNMSQRHV
FLIG  225   LAQKIIDEMPLFENLVDVDDDRSIQRLLQEVDSESLLIALEGA-EPPLREKFLRNMSQRAA

ORF2  288   ETFKEEMEFMGPVRLKDVEEAQSRIVSIVRELEEAGEIVIARGGDDIIV
FLIG  284   DILRDDLANRGPVRLSQVENEQKAILLIVRRLAETGEMVI--GSGEDTYV
```

FIG. 3. Alignment of the sequence of Orf2 with FliG of *S. typhimurium*. Identities are indicated by asterisks, and conservative changes are represented by dots. The groups of conservative substitutions are shown in the footnote to Table 1.

```
ORF3    1   MAR----------VKEEADRISEQANSHIENIRRQIEQE----------
FLIH    1   MSNELPWQVWTPDDLAPPPETFVPVEADNVTLTEDTPEPELTAEQQLEQELAQLKIGAHE

ORF3   30   --KN-DWAAE-KQKLIEEAKAEGFEQGVALGKAEAMKQYAELIGQANTI-TEMSRKAVED
FLIH   61   QGYNAGLA-EGRQKGHAQGYQEGLAQGLEQGGQAQAQTQQAPIHARMQQLVSE-FQNTLDA

ORF3   85   KLED-ANEEIVELAVALAKKVWQQKSD-DKEAFLLLVQQVINEVKETDDIS-IYVDPYYY
FLIH  119   -LDSVIASRLMQMALEAARQVIGQTPAVDNSALIKQIQQLLQQEPLFSGKPQLRVHP---

ORF3  142   ETIFQQKDEIQQLLYKECRLG---IYADEKAQKGTCYIETPF-GRVDASVDTQLMQLKDK
FLIH  175   ----DDLQRVEEMLGATLSLHGWRLRGDPTLHHGGCKV-SADEGDLDASVATRWQEL---

ORF3  199   LLTALEA-GAAE
FLIH  227   --CRLAAPG-VL
```

FIG. 4. Alignment of the sequence of Orf3 with FliH of *S. typhimurium*. The symbols are as in Fig. 3.

FIG. 5. Alignment of the sequence of Orf4 with FliI of *S. typhimurium* and the β subunit of the F_1 portion of the F_0F_1 proton-translocating ATPase of *E. coli*. Identities for all three sequences are indicated by asterisks, and conservative substitutions are represented by dots. Identities and conservative changes between only two sequences are printed in boldface. Boxes A and B represent the sites of *E. coli* ATPase thought to be involved in nucleotide binding. Box C is the sequence around the glutamic acid residues of ATPases that react with dicyclohexylcarbodiimide. The sequence of the β subunit of ATPase continues beyond the last amino acid of Orf4.

about the same size (264 amino acid residues deduced for Orf9 and 260 for FlgG), and both lack cysteines and have a fairly uniform distribution of hydrophobic, polar, basic, and acidic residues throughout their sequences (13). FlgG is structurally related to FlgF, and the similarity of Orf9 extends to the latter protein. The consensus sequence Ala-X-Asn-(Ile/Leu)-Ala-Asn (referred to as ANNLAN), found to be present in at least four basal-body proteins (13), is also largely conserved in Orf9. The sequence of the *B. subtilis* homolog to *flgG* has also been independently reported by Zuberi et al. (43).

Orf10 and FliL. The amino acid sequence of Orf10 shows similarity to the *Salmonella* FliL protein (21% identity) and that of *E. coli* (19, 20), whose function is still unknown. It has been shown that loss of the *fliL* gene product results in a nonflagellated phenotype and that the FliL protein was localized to the membrane fraction (24), thus suggesting a role in either synthesis or assembly of flagellum components.

The nucleotide sequence of *orf11* is only partial; nevertheless, good similarity to FliM was observed. This finding was supported by the comparison of the entire *orf11* sequence to FliM (42).

As for the ORF on the complementary strand, entirely contained in the *orf4* sequence, we found no similarity of the deduced product to known proteins and the significance of the ORF is purely hypothetical.

DISCUSSION

Bacterial flagella are complex organelles that require the concerted action of a relevant number of proteins for assembly and functioning. We have determined the nucleotide sequences of a number of *B. subtilis* genes involved in flagellar organization. The chromosomal region which is the subject of this presentation comprises the *flaA4* and *flaA15* mutations previously described as causing a flagellumless phenotype (8, 9, 32). In addition, an *orf7* deletion mutant examined with an electron microscope lacked flagellar filaments (10). In the latter case, the Fla^- phenotype could be due to inactivation of *orf7* or a polar effect. The deletion of 100 bp in the middle of *orf7* is out of frame and gives rise to 17 stop codons before the natural end of the protein. The start codon of *orf8* overlaps the stop codon of *orf7*, and the shift in the reading frame may be expected to have a strong polar effect. We can conclude that the *B. subtilis flaA* locus is indeed involved in the synthesis and assembly of flagella. Comparison of the deduced protein sequences of the *B. subtilis flaA* locus to proteins of *E. coli* and *S. typhimurium* known to be part of the flagellar structure further supports the above notion.

We identified with confidence at least five elements of

TABLE 1. Deduced properties of proteins encoded by *orf2* to *orf10* of the *B. subtilis flaA* locus

Protein	Length (amino acids)	Molecular mass (Da)	Predicted pI	Similarity score (%)[a]	Residue(s) missing	Hypothetical function
Orf2	338	38,188	4.46	FliG[b] (36; 58)	Trp, Cys	Motor switch
Orf3	208	23,785	4.40	FliH[b] (23; 47), Trg[c] (19; 45)		
Orf4	440	47,886	5.59	FliI[b] (49; 68), ATPase[c] (29; 46)	Trp	ATPase
Orf5	147	17,906	8.08	FliJ[b] (23; 41), Orf3 (23; 47)	Trp, Cys, Pro	
Orf6	213	22,946	9.07	FrzCD[d] (18; 36)	Cys	Signal transduction
Orf7	429	47,508	4.71	FliK[b] (22; 37)	Cys	Hook length control
Orf8	140	15,643	4.35		Cys, Arg	
Orf9	264	27,467	4.73	FlgG[b] (40; 62), FlgF[b] (21; 42)	Cys, Trp	Rod, distal
Orf10	140	15,592	5.17	FliL[b] (21; 42)	Trp, Cys, His, Pro	

[a] The first number refers to percentage of identical residues, and the second refers to the percentage of conserved substitutions. The conserved residues are grouped as follows: A-G, C, D-E-N-Q, F-Y-W, H-K-R, I-L-M-V, P, and S-T.
[b] *S. typhimurium*.
[c] *E. coli*.
[d] *M. xanthus*.

gram-positive flagella: the M-ring protein (Orf1, analogous to FliF), two components of the motor switch (Orf2 and Orf11, corresponding to FliG and FliM), the main component of the distal part of the rod (Orf9 in FlgG), and a gene involved in control of hook length (Orf7 in FliK). Other genes (*orf3*, *orf5*, and *orf6*) could be involved in the process of chemotactic signal transduction. Functions for *orf8* and *orf10* could not be deduced from the present data. They may represent genes whose products are peculiar to *B. subtilis*, or they may correspond to some of the unknown components of the flagellar structures that have been recently hypothesized on the basis of their synthetic dependence on the flagellar master operon for expression (17). It should be stressed that no data are available regarding the *B. subtilis* proteins which have been deduced from the sequenced DNA and that the present discussion is based on comparisons of the latter to known gene products of *E. coli* and *S. typhimurium*. Cloning and sequencing of the genes should greatly facilitate study of the encoded proteins.

The finding that the product of *orf4* was highly similar to the catalytic β subunit of bacterial and mitochondrial ATP synthases was totally unexpected. It is well known that the bacterial flagellum rotates and that the energy source for rotation is the transmembrane proton potential (for a review, see reference 22). Only under conditions of anaerobic glycolysis is ATP hydrolysis required, and this occurs via proton-linked membrane ATPase action. The hypothetical ATPase represented by *B. subtilis* Orf4 is highly similar to the gene product of *S. typhimurium fliI*, and this fact may help in defining the role of this gene. Both *fliI* and *orf4* are parts of a flagellar operon, and the hypothetical ATPase could be involved in some steps of the assembly of flagellar components. It may be relevant that the *S. typhimurium* FliI protein was found to be associated mostly with the membrane fraction (12).

The genetic organization of the *flaA* locus of *B. subtilis* is remarkably similar to that found for region III of enteric bacteria. With the exception of *orf8* and *orf9*, the order of the genes in *B. subtilis* is identical to the order of the *fli* genes of *E. coli* and *S. typhimurium* (12, 15).

Flagellin monomers are exported to the exterior of the bacterial cell by a specific export pathway (4, 14), and the same pathway may be used for other external components of the flagella (13). We found that a presumptive rod protein of *B. subtilis* (Orf9) has the consensus sequence ANNLAN, observed for *Salmonella* rod proteins FlgB, FlgC, FlgF, and FlgG. Whether the sequence is indeed a signal involved in export of the proteins or, rather, represents a structural constraint of a component of the axial filament remains to be demonstrated. The conservation of such a sequence in gram-positive and -negative bacteria that have diverse cell envelope organization may argue in favor of the structural constraint.

Comparison of protein sequences with similar functions from distantly related bacteria represents a valuable tool for evaluation of the structural motifs that underlie flagellar organization and for comprehension of the relationships between structure and function.

ACKNOWLEDGMENTS

We thank R. M. Macnab for communicating the sequence of *S. typhimurium* FliH, FliI, FliJ, and FliK in advance of publication and for advice about the manuscript.

This research was supported in part by Consiglio Nazionale delle Ricerche grant 89.03411.14, the Ministero dell'Università e Ricerca Scientifica e Tecnologica (MURST), and EEC grant SCI*-0211-C-EDB.

REFERENCES

1. **Bollinger, J., C. Park, S. Harayama, and G. L. Hazelbauer.** 1984. Structure of the Trg protein: homologies with and differences from other sensory transducers of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:3287–3291.
2. **Chen, E. I., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
3. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:45–147.
4. **Emerson, S. U., K. Tokuyasu, and M. I. Simon.** 1970. Bacterial flagella: polarity of elongation. Science **169**:190–192.
5. **Ferrari, E., D. J. Henner, and J. A. Hoch.** 1981. Isolation of *Bacillus subtilis* genes from a Charon 4A library. J. Bacteriol. **146**:430–432.
6. **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.
7. **Gianni, M., and A. Galizzi.** 1986. Isolation of genes preferentially expressed during *Bacillus subtilis* spore outgrowth. J. Bacteriol. **165**:123–132.
8. **Grant, G. F., and M. I. Simon.** 1969. Synthesis of bacterial flagella. II. PBS1 transduction of flagella-specific markers in *Bacillus subtilis*. J. Bacteriol. **99**:116–124.
9. **Hauser, P. M., W. D. Crabb, M. G. Fiora, F. Scoffone, and A. Galizzi.** 1991. Genetic analysis of the *flaA* locus of *Bacillus subtilis*. J. Bacteriol. **173**:3580–3583.
10. **Hauser, P. M., and A. Galizzi.** Unpublished data.
11. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.
12. **Homma, M., T. Iino, and R. M. Macnab.** 1988. Identification and characterization of the products of six region III flagellar genes (*flaAII.3* through *flaQII*) of *Salmonella typhimurium*. J. Bacteriol. **170**:2221–2228.
13. **Homma, M., K. Kutsukake, M. Hasebe, T. Iino, and R. M. Macnab.** 1990. FlgB, FlgC, FlgF and FlgG. A family of structurally related proteins in the flagellar basal body of *Salmonella typhimurium*. J. Mol. Biol. **211**:465–477.
14. **Iino, T.** 1969. Polarity of flagellar growth in *Salmonella*. J. Gen. Microbiol. **56**:227–239.
15. **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.
16. **Jones, C. J., M. Homma, and R. M. Macnab.** 1989. L-, P-, and M-ring proteins of the flagellar basal body of *Salmonella typhimurium*: gene sequences and deduced protein sequences. J. Bacteriol. **171**:3890–3900.
17. **Jones, C. J., and R. M. Macnab.** 1990. Flagellar assembly in *Salmonella typhimurium*: analysis with temperature-sensitive mutants. J. Bacteriol. **172**:1327–1339.
18. **Kanazawa, H., T. Kayano, T. Kiyasu, and M. Futai.** 1982. Nucleotide sequence of the genes for β and α subunits of proton-translocating ATPase from *Escherichia coli*. Biochem. Biophys. Res. Commun. **105**:1257–1264.
19. **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.
20. **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.
21. **Macnab, R. M.** 1987. Flagella, p. 70–83. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.

22. **Macnab, R. M.** 1987. Motility and chemotaxis, p. 732–759. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
23. **Macnab, R. M. (Yale University).** 1990. Personal communication.
24. **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.
25. **McCleary, W. R., M. J. McBride, and D. R. Zusman.** 1990. Developmental sensory transduction in *Myxococcus xanthus* involves methylation and demethylation of FrzCD. J. Bacteriol. **172**:4877–4887.
26. **Mirel, D. B., and M. J. Chamberlin.** 1989. The *Bacillus subtilis* flagellin gene (*hag*) is transcribed by the σ^{28} form of RNA polymerase. J. Bacteriol. **171**:3095–3101.
27. **Ohta, S., H. Tomura, K. Matsuda, and Y. Kagawa.** 1988. Gene structure of the human mitochondrial adenosine triphosphate synthase β subunit. J. Biol. Chem. **263**:11257–11262.
28. **Ordal, G. W., D. O. Nettleton, and J. A. Hoch.** 1983. Genetics of *Bacillus subtilis* chemotaxis: isolation and mapping of mutations and cloning of chemotaxis genes. J. Bacteriol. **154**:1088–1097.
29. **Ordal, G. W., H. M. Parker, and J. R. Kirby.** 1985. Complementation and characterization of chemotaxis mutants of *Bacillus subtilis*. J. Bacteriol. **164**:802–810.
30. **Patterson-Delafield, J., R. J. Martinez, B. A. D. Stocker, and S. Yamaguchi.** 1973. A new *fla* gene in *Salmonella typhimurium*—*flaR*—and its mutant phenotype—superhooks. Arch. Mikrobiol. **90**:107–120.
31. **Perego, M., and J. A. Hoch.** Unpublished data.
32. **Pooley, H. M., and D. Karamata.** 1984. Genetic analysis of autolysin-deficient and flagellaless mutants of *Bacillus subtilis*. J. Bacteriol. **160**:1123–1129.
33. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Saraste, M., N. J. Gay, A. Eberle, M. J. Runswick, and J. E. Walker.** 1981. The *atp* operon: nucleotide sequence of the genes for the α, β, and γ subunits of *Escherichia coli* ATP synthase. Nucleic Acids Res. **9**:5287–5296.
36. **Suzuki, T., and T. Iino.** 1981. Role of the *flaR* gene in flagellar hook formation in *Salmonella* spp. J. Bacteriol. **148**:973–979.
37. **Thoelke, M. S., J. M. Casper, and G. W. Ordal.** 1990. Methyl group turnover on methyl-accepting chemotaxis proteins during chemotaxis by *Bacillus subtilis*. J. Biol. Chem. **265**:1928–1932.
38. **Vogler, A. P., M. Homma, V. M. Irikura, and R. A. Macnab.** 1991. *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F_0F_1, vacuolar, and archaebacterial ATPase subunits. J. Bacteriol. **173**:3564–3572.
39. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
40. **Ying, C. F., F. Scoffone, A. M. Albertini, A. Galizzi, and G. W. Ordal.** 1991. Properties of the *Bacillus subtilis* chemotaxis protein CheF, a homolog of the *Salmonella typhimurium* flagellar protein FliJ. J. Bacteriol. **173**:3584–3586.
41. **Yoshida, M., W. S. Allison, F. S. Esch, and M. Futai.** 1982. The specificity of carboxyl group modification during the inactivation of the *Escherichia coli* F1-ATPase with dicyclohexyl [^{14}C]carbodiimide. J. Biol. Chem. **257**:10033–10037.
42. **Zuberi, A. R., D. S. Bischoff, and G. W. Ordal.** 1991. Nucleotide sequence and characterization of a *Bacillus subtilis* gene encoding a flagellar switch protein. J. Bacteriol. **173**:710–719.
43. **Zuberi, A. R., C. Ying, D. S. Bischoff, and G. W. Ordal.** Gene, in press.
44. **Zuberi, A. R., C. Ying, M. R. Weinreich, and G. W. Ordal.** 1990. Transcriptional organization of a cloned chemotaxis locus of *Bacillus subtilis*. J. Bacteriol. **172**:1870–1876.