JOURNAL OF BACTERIOLOGY, July 1988, p. 3305–3309
0021-9193/88/073305-05$02.00/0

Vol. 170, No. 7

Construction of a Minimum-Size Functional Flagellin of *Escherichia coli*

GORO KUWAJIMA

Shionogi Research Laboratories, Shionogi & Co., Ltd., Fukushima-ku, Osaka 553, Japan

Received 10 December 1987/Accepted 19 April 1988

Various deletions were introduced into the central region of *Escherichia coli* flagellin (497 residues) without destroying its ability to form flagellar filaments. The smallest flagellin retained only the N-terminal 193 residues and the C-terminal 117 residues, which are suggested to be the domains essential for filament formation.

The motility of flagellated bacteria stems from the rotation of the external flagellar filament. The filament is a helical, cylindrical structure composed of subunits of flagellin associated each other by noncovalent interactions (11, 24–26). The flagellin of *Escherichia coli* K-12, encoded by the *hag* gene, is composed of 497 amino acid residues (17). Genetic studies of *Salmonella typhimurium* (6, 10, 28) and analysis of the homology pattern of the amino acid sequences (2, 4, 12, 17, 27) suggest that both the N- and C-terminal regions of flagellin are important for filament formation. Previously, I reported that *E. coli* flagellins with small deletions in the central region still could assemble into filaments (16). Here, I have extended this approach and have constructed small flagellins which can form filaments despite large deletions in the central region.

I had previously constructed plasmids pFD2 and pFD3, in which the *hag* gene suffers a deletion and an insertion of the 8-mer *Hin*dIII linker (16). The deletions in the gene correspond to 16 amino acid residues from residue 239 to residue 254 (pFD2) and 20 residues from residue 259 to residue 278 (pFD3) of *E. coli* K-12 flagellin. These plasmids were linearized with *Hin*dIII, digested with exonuclease Bal31 (19), and ligated in the presence of the 18-mer linker DNA 5′-AAGCTTCCCGGGAGATCT-3′, which contains restriction sites for *Hin*dIII, *Sma*I, and *Bgl*II (neither pFD2 nor pFD3 contains *Sma*I and *Bgl*II sites). The resulting plasmids were introduced into *E. coli* C600 *hsm hsr hag*::Tn*10*, which is nonmotile because of the *hag* mutation (16). From the transformants, I selected eight strains which could swarm on lambda motility agar plates (14) and whose plasmids could be cut by *Hin*dIII, *Sma*I, and *Bgl*II. These plasmids were designated as pFD202 (derived from pFD2) and pFD301, pFD303, pFD306, pFD307, pFD311, pFD313, and pFD315 (all derived from pFD3). The ability of these plasmids to confer motility on C600 *hsm hsr hag*::Tn*10* was estimated as the swarm zone size relative to that of the same strain carrying pBR322/hag93, which has the wild-type *hag* gene (17). Four of them (pFD301, pFD303, pFD306, and pFD307) conferred ca. 10% of wild-type swarming, three (pFD311, pFD313, and pFD315) conferred ca. 50%, and one (pFD202) conferred almost 100% (Fig. 1).

*Hin*cII restriction fragments of these plasmids were compared by agarose gel electrophoresis. Only the 1.7-kilobase fragment of pBR322/hag93, which includes the entire *hag* gene (17), was shortened in each of the deletion plasmids. From the length of these fragments, it was inferred that pFD202, pFD301, pFD303, pFD306, pFD307, pFD311, pFD313, and pFD315 were about 60, 540, 340, 320, 400, 250, 160, and 220 bases, respectively, shorter than pBR322/hag93 (data not shown).

The DNA sequence of approximately 100 bases of each mutant *hag* allele was determined in both directions from the inserted 18-mer linker (Fig. 2). These alleles encode mutant flagellins which suffer deletions in the central region (Fig. 3). Artificial amino acid sequences of six residues (pFD202, pFD301, pFD306, and pFD313) or seven residues (pFD303, pFD307, pFD311, and pFD315) were present in the deleted regions of the mutant flagellins as the result of the 18-mer linker insertion. Among these, the flagellin encoded by the *hag* allele on pFD301 (pFD301 flagellin) was the smallest. It was composed of only the N-terminal 193 residues and the C-terminal 117 residues of the wild-type *E. coli* K-12 flagellin and six artificial residues.

I detected mutant flagellins which had assembled into filaments and had been released into the culture medium. Each strain was grown to the late log phase at 28°C. The culture broth was then heated at 65°C for 10 min to depolymerize filaments (13). The supernatant (400 μl) was concentrated by ultrafiltration and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (18) (Fig. 4). In each sample, one major band was found. The band from the strain carrying pBR322/hag93 had the same mobility as that of the purified wild-type flagellin, and those from the strains

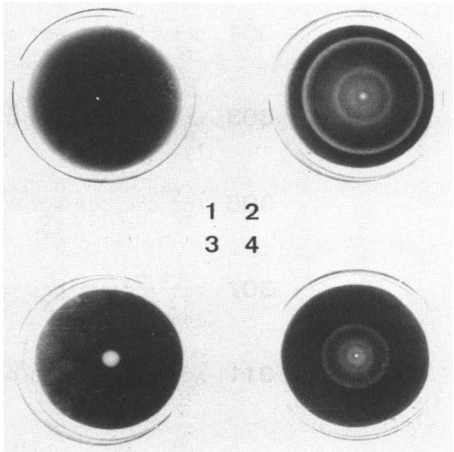
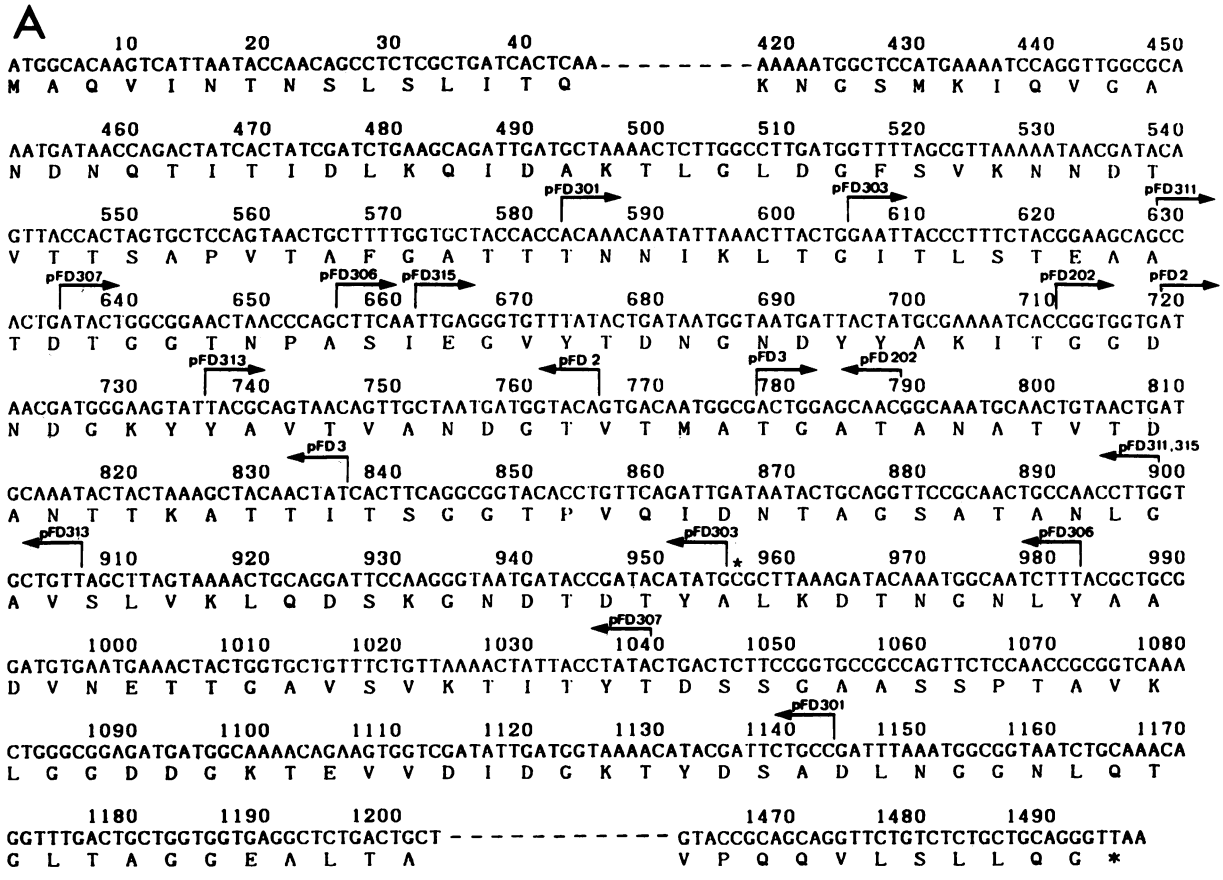


FIG. 1. Motility tests of C600 *hsm hsr hag*::Tn*10* carrying pBR322 (plate 1), pBR322/hag93 (plate 2), pFD301 (plate 3), and pFD311 (plate 4). Each strain was inoculated on a lambda motility agar plate (14) and incubated at 37°C for 7 h. The swarm zone diameter reflects the motility of each strain.



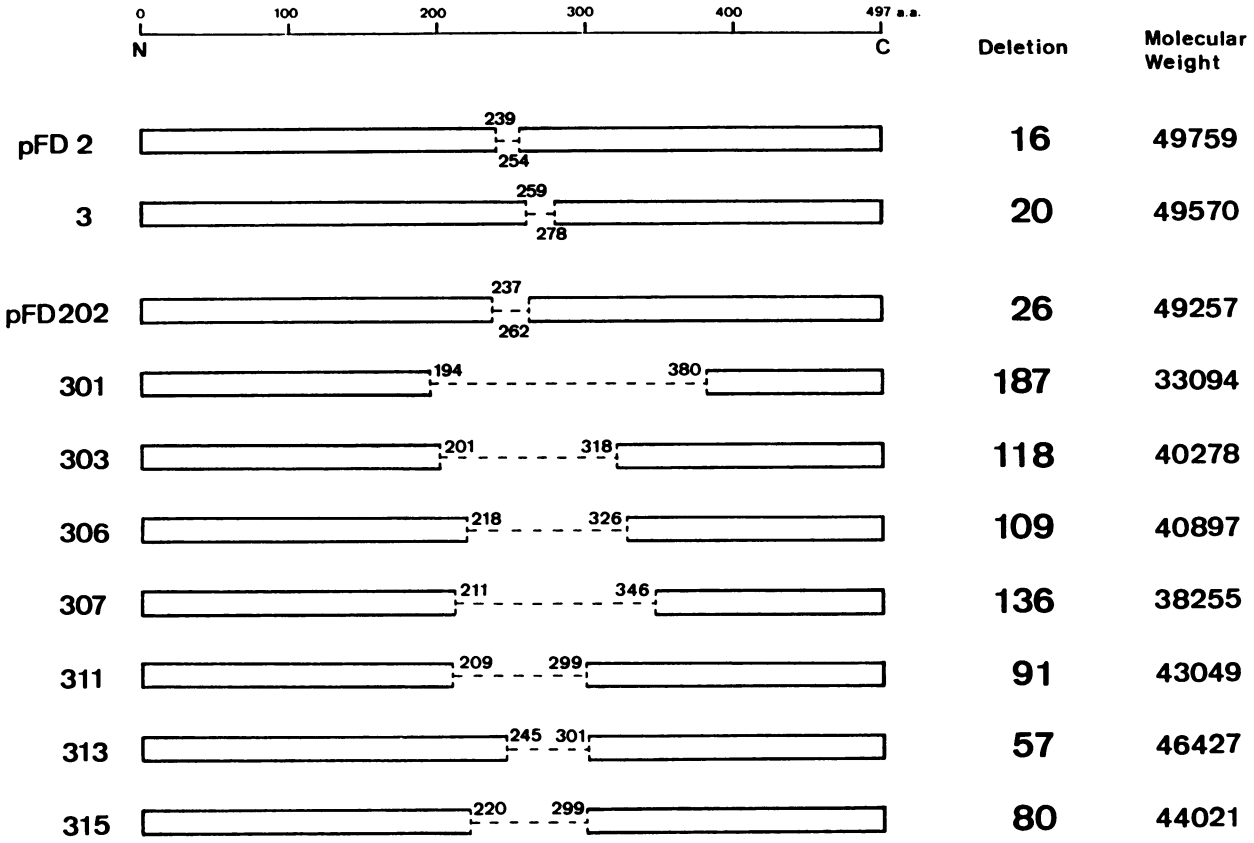


FIG. 3. Deletion maps of mutant flagellins. The amino acid (a.a.) residues of the wild-type flagellin are numbered from the N terminus. The deleted region in the flagellin encoded by each plasmid is indicated by broken lines, and the endpoints are indicated. The molecular weights calculated from the DNA sequences and the numbers of deleted amino acid residues are also shown.

carrying the other plasmids had higher mobilities, with apparent molecular weights in accordance with those predicted from the gene sequences. The amounts of flagellins in the cultures were determined by densitometry. The strain carrying pFD301 produced about half as much, on a weight basis, as that carrying pBR322/hag93 (4.4 versus 8.8 mg/ml per unit of optical density at 650 nm), corresponding to about 80% as much on a molar basis. Molar ratios for other mutant flagellins were similar (data not shown). Thus, mutant flagellins have an almost normal ability to be excreted.

The amino acid sequences of both the N- and C-terminal regions of flagellins from various bacterial strains and species are highly homologous, in contrast to the central heterologous region (2, 4, 12, 17, 27). Between *E. coli* K-12 flagellin and *S. typhimurium* phase-1 flagellin, the N-terminal 170 residues and the C-terminal 140 residues show 80 and 60% homology, respectively, whereas the central region shows only 20% homology (17). Interestingly, pFD301 flagellin is composed simply of two parts, corresponding almost exactly to these homologous domains (Fig. 5). pFD301 flagellin seems to be nearly the smallest functional form of *E. coli* flagellin. As it was excreted almost as well as the wild-type flagellin (Fig. 4) and assembled into filaments, although in small quantities (data not shown), the N-terminal 193 residues and the C-terminal 117 residues of *E. coli* K-12 flagellin appear to be the functional domains essential for filament formation, with the central region playing no such role. However, large deletions may cause some stress in the conformation of the flagellin, leading to less-stable interactions between flagellin subunits and between flagellins and other flagellar proteins (5, 7–9).

Three-dimensional images of the flagellar filament of *S. typhimurium* proposed by two groups show that the flagellin has three or four domains (24, 26). It is interesting to consider that these structural domains may correspond to the terminal and central domains in this study. The central

FIG. 2. DNA sequences of the mutant *hag* allele and the corresponding amino acid residues. Two *Hin*dIII and *Bam*HI fragments of each plasmid were cloned onto M13mp18 phage DNA (22, 29), and their DNA sequences were determined by the dideoxy method (23). (A) Deleted regions are indicated by the arrows on the wild-type *hag* gene. Into these deletions the 18-mer linker was inserted in the orientation *Hin*dIII-*Sma*I-*Bgl*II in pFD202, pFD301, pFD306, and pFD311 and in the opposite orientation in the other plasmids. The asterisk indicates the stop codon. Some parts of the sequence are abbreviated (broken lines). (B) DNA sequence and the corresponding amino acid residues of the juncture regions are shown. Nucleotide bases of the wild-type *hag* gene are indicated in uppercase letters and numbered from the initiation codon; those of artificially inserted 18-mer linkers indicated are in lowercase letters. Amino acid residues of the wild-type flagellin are indicated by one-letter abbreviations in uppercase letters and numbered from the N terminus; artificially inserted ones are indicated by three-letter abbreviations in lowercase letters. In the *hag* gene on pFD303, the C at base 956 (*) is changed to a T.

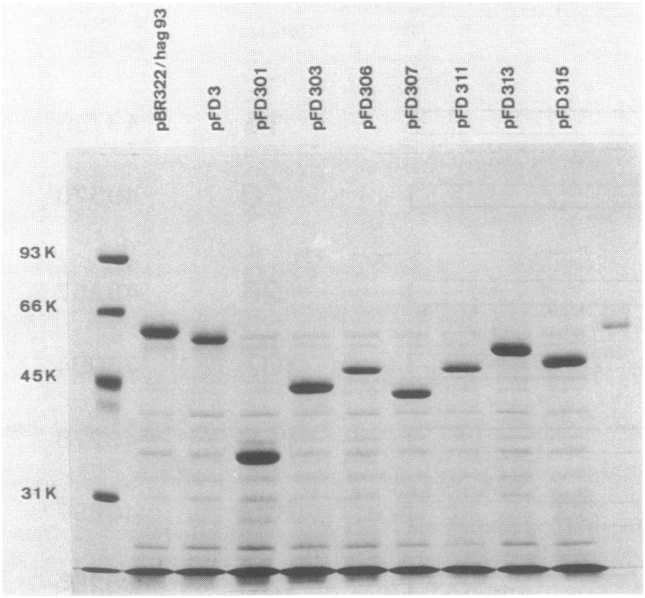


FIG. 4. Detection of mutant flagellins. Extracellular samples from C600 *hsm hsr hag*::Tn*10* carrying pBR322/hag93 (17), pFD3 (16), and pFD301, pFD303, pFD306, pFD307, pFD311, pFD313, and pFD315 (see text) were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (10% acrylamide). The gel was stained with Coomassie brilliant blue R-250. In the rightmost lane, 1 μg of the purified wild-type flagellin (17) was loaded. In the leftmost lane, markers (phosphorylase *b*, bovine serum albumin, ovalbumin, and carbonic anhydrase) were loaded. The molecular weights of these markers are indicated (in thousands [K]).

heterologous domains of the wild *E. coli* K-12 flagellin deleted from pFD301 flagellin may correspond to the "S and M domains" of Shirakihara and Wakabayashi (24) and to the "outer domain" of Trachtenberg and DeRosier (26). I suggest that such a specific organization of the functional and structural domains is common in various bacterial flagellins; such an organization may be important in explaining the large variations in flagellin molecular weight (1–4, 12, 13, 15, 17, 20, 21, 27) and H antigenicity.

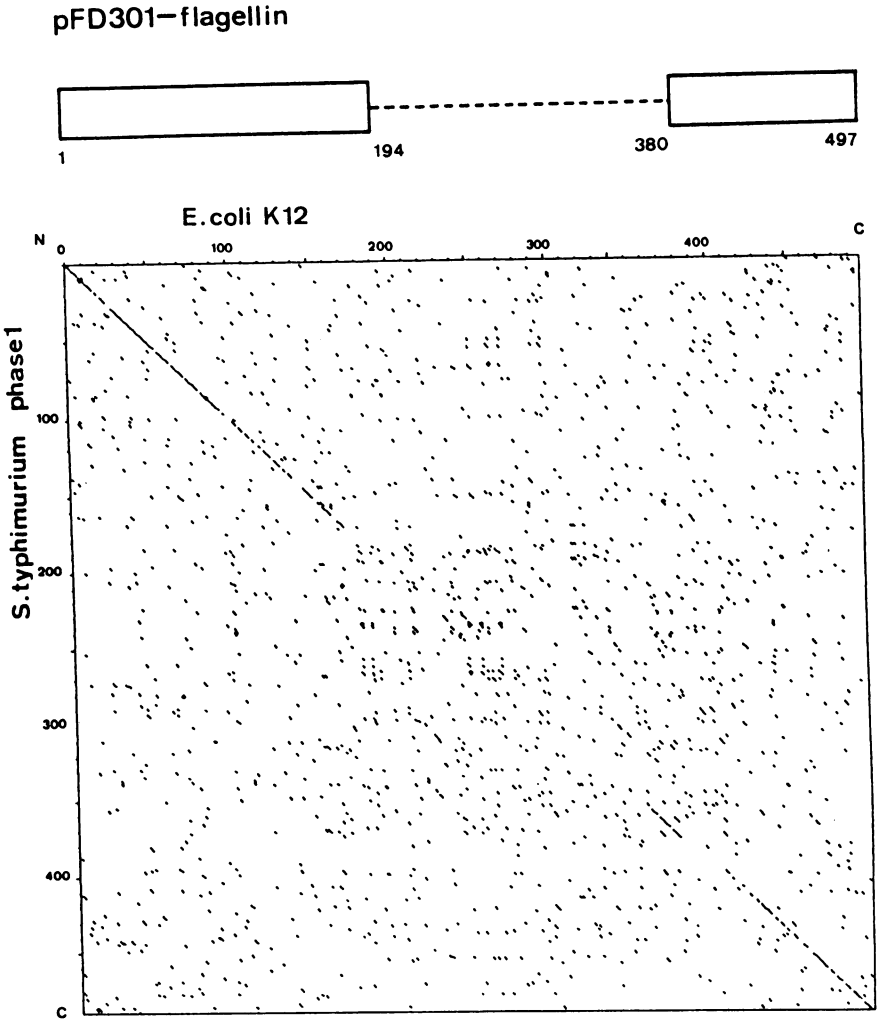


FIG. 5. Correspondence between the sequence of pFD301 flagellin and the highly homologous domains of the flagellins of *E. coli* K-12 and *S. typhimurium* (in phase 1). The numbers indicate the positions of the amino acid residues in the wild-type flagellin. pFD301 flagellin is shown in the same way as in Fig. 3. In the matrix comparison, each dot corresponds to a consecutive-residue stretch of identity between the two sequences at the respective positions shown.

We thank H. Ozeki for helpful discussions and advice in this work and R. M. Macnab and M. Homma for comments on the manuscript.

LITERATURE CITED

1. **Chang, J. Y., D. M. Brown, and A. N. Glazer.** 1969. Characterization of the subunits of the flagella of *Proteus vulgaris*. J. Biol. Chem. **244**:5196–5200.
2. **DeLange, R. J., J. Y. Chang, J. H. Shaper, and A. N. Glazer.** 1976. Amino acid sequence of flagellin of *Bacillus subtilis* 168. III. Tryptic peptides, *N*-bromosuccinimide peptides, and the complete amino acid sequence. J. Biol. Chem. **251**:705–711.
3. **DeLange, R. J., J. Y. Chang, J. H. Shaper, R. J. Martinez, S. K. Komatsu, and A. N. Glazer.** 1973. On the amino-acid sequence of flagellin from *Bacillus subtilis* 168: comparison with other bacterial flagellins. Proc. Natl. Acad. Sci. USA **70**:3428–3431.
4. **Gill, P. R., and N. Agabian.** 1983. The nucleotide sequence of the M_r=28,500 flagellin gene of *Caulobacter crescentus*. J. Biol. Chem. **258**:7395–7401.
5. **Homma, M., H. Fujita, S. Yamaguchi, and T. Iino.** 1984. Excretion of unassembled flagellin by *Salmonella typhimurium* mutants deficient in hook-associated proteins. J. Bacteriol. **159**: 1056–1059.
6. **Homma, M., H. Fujita, S. Yamaguchi, and T. Iino.** 1987. Regions of *Salmonella typhimurium* flagellin essential for its polymerization and excretion. J. Bacteriol. **169**:291–296.
7. **Homma, M., and T. Iino.** 1985. Locations of hook-associated proteins in flagellar structures of *Salmonella typhimurium*. J. Bacteriol. **162**:183–189.
8. **Homma, M., T. Iino, K. Kutsukake, and S. Yamaguchi.** 1986. *In vitro* reconstitution of flagellar filaments onto hooks of filamentless mutants of *Salmonella typhimurium* by addition of hook-associated proteins. Proc. Natl. Acad. Sci. USA **83**:6169–6173.
9. **Homma, M., K. Kutsukake, T. Iino, and S. Yamaguchi.** 1984. Hook-associated proteins essential for flagellar filament formation in *Salmonella typhimurium*. J. Bacteriol. **157**:100–108.
10. **Horiguchi, T., S. Yamaguchi, K. Yao, T. Taira, and T. Iino.** 1975. Genetic analysis of *H1*, the structural gene for phase-1 flagellin in *Salmonella*. J. Gen. Microbiol. **91**:139–149.
11. **Iino, T.** 1977. Genetics of structure and function of bacterial flagella. Annu. Rev. Genet. **11**:161–182.
12. **Joys, T. M.** 1985. The covalent structure of the phase-1 flagellar filament protein of *Salmonella typhimurium* and its comparison with other flagellins. J. Biol. Chem. **260**:15758–15761.
13. **Kondoh, H., and H. Hotani.** 1974. Flagellin from *Escherichia coli* K-12: polymerization and molecular weight in comparison with *Salmonella* flagellins. Biochim. Biophys. Acta **336**:117–139.
14. **Kondoh, H., and H. Ozeki.** 1976. Deletion and amber mutants of *fla* loci in *Escherichia coli* K-12. Genetics **84**:403–421.
15. **Krupski, G., R. Gotz, K. Ober, E. Pleier, and R. Schmitt.** 1985. Structure of complex flagellar filaments in *Rhizobium meliloti*. J. Bacteriol. **162**:361–366.
16. **Kuwajima, G.** 1988. Flagellin domain that affects H antigenicity of *Escherichia coli* K-12. J. Bacteriol. **170**:485–488.
17. **Kuwajima, G., J.-I. Asaka, T. Fujiwara, T. Fujiwara, K. Node, and E. Kondoh.** 1986. Nucleotide sequence of the *hag* gene encoding flagellin of *Escherichia coli*. J. Bacteriol. **168**:1479–1483.
18. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 94–148. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Martinez, R. J., D. M. Brown, and A. N. Glazer.** 1967. The formation of bacterial flagella. III. Characterization of the subunits of the flagella of *Bacillus subtilis* and *Spirillum serpens*. J. Mol. Biol. **28**:45–51.
21. **Maruyama, M., G. Lodderstaedt, and R. Schmitt.** 1978. Purification and biochemical properties of complex flagella isolated from *Rhizobium lupini* H13-3. Biochim. Biophys. Acta **553**:110–124.
22. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
24. **Shirakihara, Y., and T. Wakabayashi.** 1979. Three-dimensional image reconstruction of straight flagella from a mutant *Salmonella typhimurium*. J. Mol. Biol. **131**:485–507.
25. **Silverman, M., and M. I. Simon.** 1977. Bacterial flagella. Annu. Rev. Microbiol. **31**:397–419.
26. **Trachtenberg, S., and D. J. DeRosier.** 1987. Three-dimensional structure of the frozen-hydrated flagellar filament. J. Mol. Biol. **195**:581–601.
27. **Wei, L.-N., and T. M. Joys.** 1985. Covalent structure of three phase-1 flagellar filament proteins of *Salmonella*. J. Mol. Biol. **186**:791–803.
28. **Yamaguchi, S., H. Fujita, K. Sugita, T. Taira, and T. Iino.** 1984. Genetic analysis of *H2*, the structural gene for phase-2 flagellin in *Salmonella*. J. Gen. Microbiol. **130**:255–265.
29. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.