JOURNAL OF BACTERIOLOGY, June 1989, p. 3085–3094
0021-9193/89/063085-10$02.00/0

Vol. 171, No. 6

Cloning of the Flagellin Gene from *Bacillus subtilis* and Complementation Studies of an In Vitro-Derived Deletion Mutation

EDWARD R. LAVALLIE AND MARK L. STAHL*

Genetics Institute, Inc., 87 CambridgePark Drive, Cambridge, Massachusetts 02140

Received 20 October 1988/Accepted 3 March 1989

The flagellin promoter and structural gene from *Bacillus subtilis* I168 was cloned and sequenced. The amino-terminal protein sequence deduced from the coding sequence of the cloned gene was identical to that of the amino terminus of purified flagellin, indicating that the export of this protein is not directed by a posttranslationally processed N-terminal signal peptide. A sequence that was homologous to that of a consensus σ^{28} RNA polymerase recognition site lay upstream of the proposed translational start site. Amplification of this promoter region on a multicopy plasmid resulted in the formation of long, filamentous cells that accumulated flagellin intracellularly. The chromosomal locus containing the wild-type flagellin allele was replaced with a defective allele of the gene (Δ*hag-633*) that contained a 633-base-pair deletion. Transport analysis of various flagellin gene mutations expressed in the *hag* deletion strain suggest that the extreme C-terminal portion of flagellin is functionally involved in export of the protein.

Bacillus subtilis is a flagellated bacterium that assembles and maintains approximately five flagellar organelles per cell. The flagellum is a complex structure that is composed of a basal body, hook, and filament; the filament accounts for 98% of the mass of the organelle and is composed exclusively of flagellin, the product of the *hag* gene (21, 45). Genetic analysis of flagellar synthesis in *B. subtilis* has been limited by the necessity of utilizing the generalized transducing phage PBS1, which requires an intact flagellum for infectivity (22). However, in *Escherichia coli*, at least 40 genes have been identified which are apparently involved in bacterial motility; of these, 29 are involved in the assembly of the flagellum (21, 45). A pathway for the assembly of the flagellum has been proposed (21, 49). It has been suggested that the flagellar structure is assembled from the cell membrane outward and that additions are made to the nascent organelle with proteins that are transported through the core of the organelle and accumulate at the outer tip.

Our interest is to study the transport mechanism that is responsible for directing the export of flagellin. The mechanism by which flagellins are exported is unknown; however, they do not seem to be transported with the aid of a secretory signal sequence which is subsequently cleaved from the amino terminus of the protein (39, 44, 53). For example, the amino acid sequence of the amino terminus of purified flagellin from the gram-negative bacterium *Caulobacter crescentus* corresponds to the amino-terminal sequence deduced from the nucleotide sequence of the cloned structural gene (12, 13). Pulse-chase experiments confirm the absence of a larger precursor of *C. crescentus* flagellin in either intracellular or membrane fractions. Similarly, the amino terminus of purified flagellins from both *E. coli* and *Salmonella typhimurium* begin with alanine, which corresponds to the second amino acid after the translational start of the cloned structural genes (24, 28, 58). The amino-terminal sequence of the flagellin from each of these species is composed primarily of hydrophobic residues but does not resemble secretory signal sequences (37). It is therefore unlikely that a processed leader sequence mediates transport of flagellin in *Caulobacter*, *Escherichia*, or *Salmonella* species. In fact, a recent study of *S. typhimurium* (19) implicates the C terminus of both H1 and H2 flagellin as an essential region for transport.

In an effort to better understand the mechanism of flagellin export in *B. subtilis*, we decided to utilize the previously published amino acid sequence of its flagellin protein (10) to clone and sequence the *hag* gene. This approach allowed us to establish the absence of an N-terminal secretory leader sequence, to characterize the regulatory regions responsible for expression, and to construct an in vitro-derived genomic deletion mutation for complementation and transport studies.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* MM294 (F⁻ *supE44 endA1 thi-1 hsdR4*) was used as a host for screening the pUC18-based *B. subtilis* GIB1 genomic library. The *B. subtilis* strains used in this study are listed in Table 1. *B. subtilis* strains were transformed by the method of Anagnostopoulos and Spizizen (2), and auxotrophic markers were selected on minimal glucose plates supplemented with the appropriate amino acids at 50 μg/ml (47). *E. coli* was transformed by the procedure of Dagert and Ehrlich (8), with selection on L agar plates (Difco Laboratories) containing the appropriate antibiotic. The vector used in this study for subcloning manipulations was pBE3, an *E. coli-B. subtilis* shuttle vector. The plasmid was constructed by inserting the pBR322 origin of replication and the pUC18 polylinker region into pUB110. This plasmid replicates in either host and confers resistance to neomycin. The basis for the *B. subtilis* integrable plasmid pFIV-3 was pJH101 (11).

Media and reagents. Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, the Klenow fragment of DNA polymerase I, and the M13 universal sequencing primer were purchased from New England BioLabs and used according to the recommendations of the supplier. Nucleotide triphosphates labeled with ^{32}P were obtained from New England Nuclear Corp., as were ^{125}I-labeled *S. aureus* protein A and ^{35}S-labeled methionine. Dideoxy nucleotides were obtained from P-L Biochemicals, Inc. "Slow" *Bal* 31 exonuclease was purchased from International Biotechnologies, Inc. Selection for plasmid transformants in *E. coli* was on L agar

* Corresponding author.

TABLE 1. *B. subtilis* strains used in this study

Strain	Genotype	Origin
I168	*trpC2*	BGSC[a] (1A1)
W23	Prototroph	BGSC (2A2)
GIB1	Prototroph	W23 DNA × I168[b]
DB104	*his nprR2 nprE18 ΔaprA3*	R. Doi (25)
GIB16	*his nprR2 nprE18 ΔaprA3 ifm*	DB104, *ifm* selection (15)
GIB17	*his nprR2 nprE18 ΔaprA3 ifm hag*::pFIV-3	pFIV-3 DNA × GIB16[c]
GIB18	*his nprR2 nprE18 ΔaprA3 ifm Δhag-633*	Cm[s] GIB17[d]
GIB19	*his nprR2 nprE18 ΔaprA3 ifm Δhag-633 nprE*::pIEV-3	pIEV-3 DNA × GIB18

[a] Bacillus Genetic Stock Center, Columbus, Ohio.
[b] Strain I168 was transformed with W23 DNA.
[c] When plasmid pFIV-3 was introduced into strain GIB16 by transformation, a nonmotile phenotype was observed (data not shown). Since this plasmid was designed to interrupt the flagellin coding sequence, we have assigned the genotypic designation *hag*::pFIV-3 to the transformant.
[d] A spontaneous chloramphenicol-susceptible isolate of strain GIB17 as explained in the text.

containing either 50 μg of ampicillin per ml for pUC plasmids, 15 μg of neomycin per ml for pBE3 plasmids, or 15 μg of chloramphenicol per ml for pFIV-3 plasmids. Plasmid transformants in *B. subtilis* were selected on L agar containing 5 μg of neomycin per ml for pBE3 plasmids or 5 μg of chloramphenicol per ml for pFIV-3 integrants. Soft agar plates for assaying motility (15) were L plates containing 0.4% agar and 0.8% gelatin. Liquid culture of *B. subtilis* was generally performed in 1× minimal salts (2), 0.2% Difco Casamino Acids (technical grade), and 2% glucose. Metabolic labeling of *B. subtilis* proteins was accomplished by the addition of [^{35}S]methionine to liquid cultures at a final concentration of 10 μCi/ml.

Oligonucleotide probe preparation. Synthetic oligonucleotides were provided by the Genetics Institute Organic Chemistry group and were synthesized by the phosphoramidite method (4). Probes were prepared by end labeling with [γ-^{32}P]ATP and T4 polynucleotide kinase (40). The sequences of the oligonucleotides in pool 1 were 5′-AA(T/C)AT(T/C/A)GA(A/G)GATATGGG-3′, and those in pool 2 were 5′-AA(T/C)AT(T/C/A)GA(A/G)GACATGGG-3′. Each pool constituted a mixture of 12 individual 17-base oligonucleotides.

Southern blots. Southern blots were performed by previously established procedures (31, 46). *B. subtilis* chromosomal DNA was isolated by the method of Marmur (32), digested with the appropriate restriction enzymes, and electrophoretically separated on 1% agarose gels. After transfer to nitrocellulose (Schleicher & Schuell Co.), the membranes were baked in a vacuum oven at 80°C for 2 h. The blots were then prehybridized at 65°C with 5× Denhardt solution (1× Denhardt solution is 0.02% [wt/vol] each of Ficoll, polyvinylpyrrolidone, and bovine serum albumin), 6× SSC (1× SSC is 0.15 M NaCl–0.015 M sodium citrate), 0.5% sodium dodecyl sulfate (SDS), and 40 μg of soluble type XI baker's yeast RNA (Sigma Chemical Co.) per ml. Hybridizations were at 65°C in the same solution with the addition of ^{32}P-labeled probe and disodium EDTA (0.01 M final concentration). Washes were done at room temperature in 2× SSC–0.1% SDS.

Construction and screening of plasmid libraries. Plasmid DNA was prepared from both *E. coli* and *B. subtilis* by the alkaline lysis method of Birnboim and Doly (5). The separation of restriction fragments on polyacrylamide and agarose gels was performed as previously described (30). All plasmid constructions were made with DNA fragments purified by electroelution from gels. The genomic library was prepared by using pUC18 as the vector with genomic DNA from *B. subtilis* GIB1. *E. coli* MM294 was transformed with the library, and screening of bacterial colonies for plasmids with inserts containing the flagellin gene was by transfer to nitrocellulose by the method of Grunstein and Hogness (16). Hybridization conditions for the labeled oligonucleotide pools were at 37°C in a solution of 1× Denhardt solution, 0.1 mM ATP, 0.9 M NaCl, 0.5% Nonidet P-40 (Sigma), 0.2 mg of soluble type XI baker's yeast RNA per ml, 90 mM Tris hydroxide (pH 7.5), and 6 mM disodium EDTA. Washing was at 37°C in 6× SSC.

DNA sequencing. Restriction fragments were ligated into appropriate sites of M13 phage vectors mp18 and mp19 (51, 55) in preparation for sequence determination by dideoxy methods (42).

Preparation of antisera to flagellin. Polyclonal antibodies to *B. subtilis* flagellin were raised in rabbits. Intact flagella were purified by the method of Martinez (33). The resulting flagella were denatured by boiling in 1× sample buffer and subjected to SDS–10% polyacrylamide gel electrophoresis by the method of Laemmli (29). The 32,000-dalton band representing pure flagellin monomer was cut from the gel, emulsified in complete Freund adjuvant, and injected intramuscularly into rabbits. A 25-μg initial injection was followed 4 weeks later by an additional injection of 25 μg emulsified in incomplete Freund adjuvant. Serum samples were harvested 10 days later and analyzed for anti-flagellin antibodies by the ability to agglutinate Formalin-treated, flagellated *B. subtilis* cells (7).

Immunochemical analysis. Western blots were performed essentially as previously described (6), with the exception that 1% gelatin (Bio-Rad Laboratories) was substituted for bovine serum albumin where applicable. Fractionation of extracellular polymerized or unpolymerized flagellin from intracellular, nontransported material was performed before immunochemical analysis. A 0.5-ml portion of culture (typically grown to a cell density of 10^9 cells per ml) was mixed with 0.5 ml of 2× SETT buffer (1× SETT buffer is 2% Triton X-100 [Sigma], 50 mM Tris chloride [pH 8.0], 150 mM NaCl, 1 mM disodium EDTA) in a 1.5-ml microfuge tube and placed in a water bath at 100°C for 5 min to depolymerize extracellular flagellin. The cells were then centrifuged at 12,000 × *g* for 5 min, and the remaining cell pellet was sonicated to lyse the cells. The culture supernatant fraction containing total extracellular flagellin was precipitated by the addition of 100% (wt/vol) trichloroacetic acid to a final concentration of 5%, incubated on ice for 5 min, and centrifuged at 12,000 × *g* for 5 min. The remaining pellet was washed three times with cold (−20°C) acetone, with centrifugation between washes, and lyophilized to dryness. Both fractions were then electrophoresed on SDS–10% polyacrylamide gels before immunoblotting.

RESULTS

Cloning of the flagellin structural gene (*hag*). The 17-base oligonucleotide pool used for the cloning of the *hag* gene of *B. subtilis* GIB1 was designed and based on the amino acid sequence of flagellin published by Delange et al. (10). The two pools of 12 individual 17-base oligonucleotides completely covered the degeneracy of amino acids 170 through 174 and, in addition, the first two bases of the glycine codon at amino acid 175 of the sequence NIEDMG. The library was

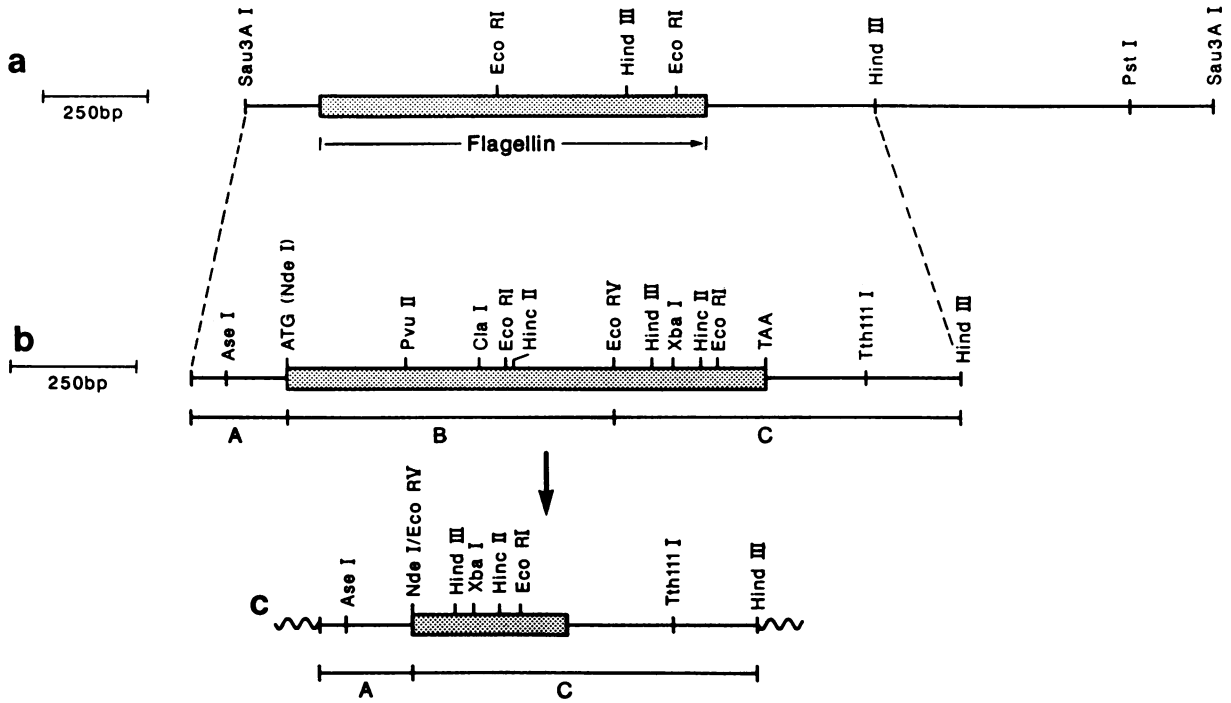

FIG. 1. Partial restriction map of the 2.3-kilobase genomic *Sau*3A I fragment containing the entire *B. subtilis* flagellin promoter and structural gene in clone p4A. (a) The shaded box depicts the extent of the flagellin-coding sequence, and the arrow indicates the direction of transcription. (b) Expanded figure showing the extent of p4A that was sequenced. The *Nde*I site at the extreme 5′ end of the flagellin gene was created by inserting a thymidine at the −1 position by oligonucleotide-directed site-specific mutagenesis (9). The sequence between this *Nde*I site and the *Eco*RV site in the 3′ end of the coding sequence was removed by fusing the two sites. Regions A and C represent the two portions of the DNA sequence that were joined to create the defective Δ*hag-633* allele (c).

constructed by inserting fragments of genomic DNA that were partially digested with *Sau*3A-I into a *Sal*I-digested pUC18 vector by the method of Hung and Wensink (20). Probe pools were mixed and labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and added to the hybridization mixture at a concentration of 0.1 pmol/ml. Two clones identified as hybridization positive, p4A and p8A, were confirmed to contain the *hag* gene by DNA sequence analysis.

Characterization of the *hag* gene. A restriction map of the *hag* gene and surrounding sequences is shown in Fig. 1, and the nucleotide sequence of the entire flagellin promoter and structural gene is shown in Fig. 2. The complete sequence of an open reading frame contained in p4A was found to encode a protein of 304 amino acids; all but two amino acids are identical to the published protein sequence of *B. subtilis* I168 flagellin (10). The exceptions are a pair of amino acids, glycine-101 and threonine-102, which are inverted in the published protein sequence.

Assignments were made for a putative ribosome-binding site (43) and translational start for the flagellin structural gene based on the proposed requirement for strong complementarity ($\Delta G < -11$ kcal) between the Shine-Dalgarno sequence of the mRNA and the 3′-OH end of the 16S rRNA for efficient translation in *Bacillus* ssp. and *Staphylococcus aureus* (Fig. 2) (34–36). An inverted and repeated sequence ($\Delta G = -17.6$ kcal) (50) located in the 3′-flanking end of the gene (Fig. 2) may function as a transcriptional terminator. Similar sequences can be found in termination regions of *E. coli* genes (41) and *Bacillus* spp. genes (48, 52, 56). A sequence that displayed strong homology to the consensus sequence for σ^{28} promoters in *B. subtilis* (14) lay approximately 100 base pairs 5′ to the proposed translational start of the flagellin gene (Fig. 2). The consensus sequence (CTAAA. . . N^{16}. . . CCGATAT) differed from the sequence found in the flagellin regulatory region (TTAAC. . . N^{16}. . . CCGATAT) by only two nucleotides.

Subcloning and amplification of *hag* sequences. The presumed flagellin promoter region and structural gene were subcloned from genomic clone p4A into pBE3 to create pFLA-1 (Fig. 3). When pFLA-1 was introduced into GIB1 by transformation, this normally motile strain displayed very weak motility on soft agar plates. Microscopically, the cells appeared as long, multiseptate filaments. In addition, immunoblot analysis (Fig. 4a, lane 4) showed that GIB1 harboring pFLA-1 overproduced flagellin intracellularly but transported little flagellin to the medium.

In an attempt to determine whether overproduction of flagellin protein was causing this aberrant phenotype, we created a translational fusion of the β-galactosidase gene of *E. coli* to the promoter and ribosome-binding site of the flagellin gene in pFLA-1. This construct, called pFZ-FLA, when introduced into GIB1 by transformation, expressed active β-galactosidase under the control of the flagellin promoter as judged by the blue appearance of colonies on plates containing X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; Sigma). The resulting filamentous phenotype in these cells was comparable to that of cells overproducing flagellin from the flagellin promoter on a plasmid. Therefore, this phenotype does not appear to be associated with overproduction of flagellin protein but may be related to the promoter region of the flagellin gene. In support of this, we replaced the flagellin promoter region in pFLA-1 with the *tac* promoter (1), a hybrid *E. coli* promoter that is identical to a *B. subtilis* σ^{43} recognition sequence (54). Wild-type cells

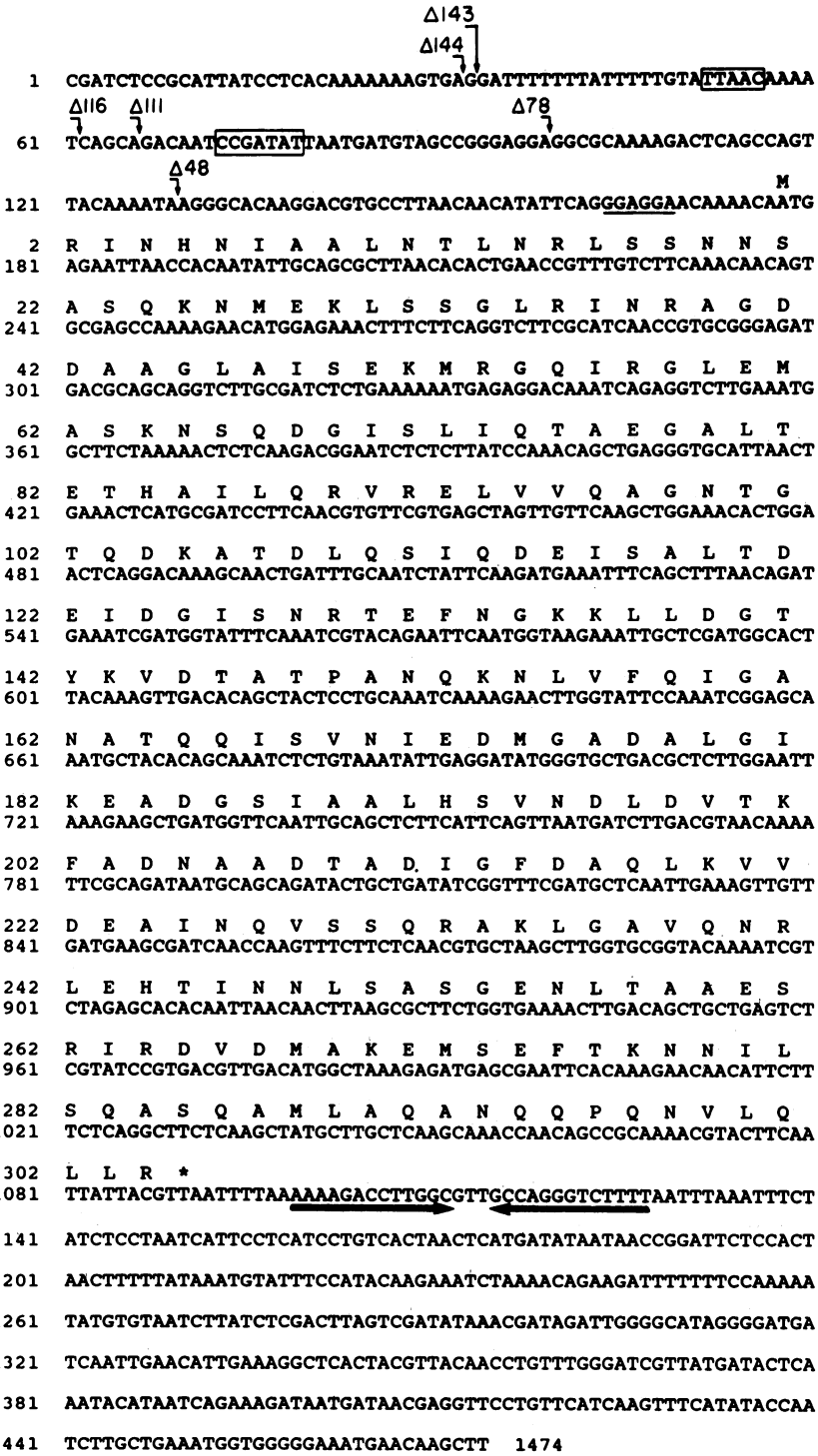

FIG. 2. Nucleotide sequence of the flagellin structural gene and adjacent regions. The inferred translated amino acid sequence is shown above the coding sequence. The two boxed sequences denote a proposed σ^{28}-dependent promoter sequence. The potential ribosome-binding site is underlined, and the two opposing arrows indicate a possible transcriptional terminator. Endpoints of deletions extending from the extreme 5′ end of the sequence into the promoter region are illustrated (e.g., Δ144).

harboring this construct exhibited normal motility and cell morphology.

To try to further implicate flagellin promoter activity with the filamentous phenotype, we constructed a series of *Bal* 31 deletions in the PFZ-FLA plasmid that spanned the promoter region (Fig. 2). All deletions began at the extreme 5′ end of the sequence shown in Fig. 2 and removed increasingly larger portions of the promoter region, ending where indicated by arrows. These deleted subclones were introduced into GIB1 by transformation, and their effect on cell

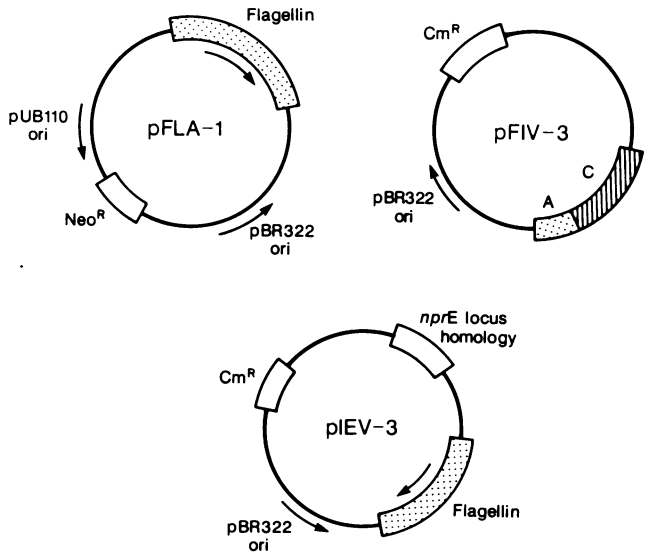

FIG. 3. Schematic representation of three plasmids containing the flagellin-coding sequence. Plasmid pFLA-1 is an *E. coli-B. subtilis* shuttle vector that confers resistance to neomycin in either host and contains the entire flagellin transcriptional unit. Plasmid pFIV-3 consists of a deleted, nonexpressable version of the flagellin gene on an integration vector that is incapable of replication in *B. subtilis* and was used to replace the wild-type flagellin allele with this altered copy. Regions A and C refer to the regions depicted in Fig. 1. Plasmid pIEV-3 is a *B. subtilis* flagellin expression vector that directs the integration of a single copy of the flagellin-coding sequence at the *nprE* locus.

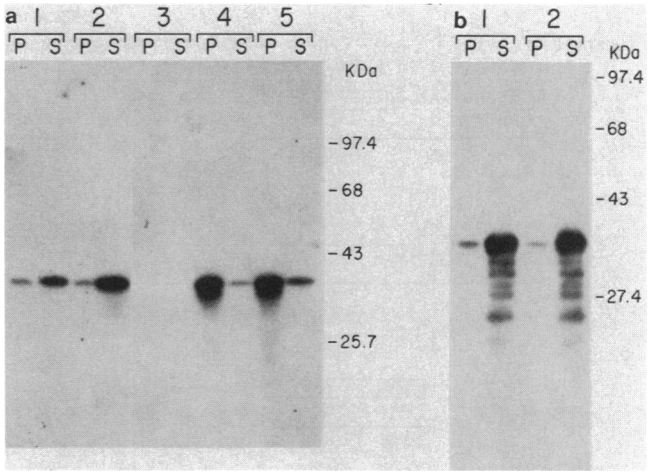

FIG. 4. Immunoblot analysis of flagellin expression in wild-type *B. subtilis* cells and in a *hag* deletion strain. Localization of expressed flagellin to the intracellular fraction (P) or the supernatant fraction (S) was performed as described in Materials and Methods. Proteins were separated on SDS–10% polyacrylamide gel electrophoresis and blotted onto nitrocellulose, and flagellin-specific bands were visualized by incubation with flagellin antisera and radioiodinated protein A before autoradiography. The strains examined (and the pertinent genotypes) are as follows (lanes): 1a, GIB1 (wild type); 2a, GIB16 (*ifm*); 3a, GIB18 (Δ*hag-633*); 4a, GIB1 transformed with pFLA-1 containing the flagellin promoter and structural gene in multicopy; 5a, GIB18 transformed with pFLA-1; 1b, GIB16 (*ifm*); 2b, GIB19 (Δ*hag-633 nprE*::pIEV-3 *ifm*), which has a single copy of the flagellin promoter and structural gene integrated at an alternate locus. Migration of protein standards in illustrated to the right with molecular masses expressed in kilodaltons.

morphology in conjunction with β-galactosidase activity was analyzed. Cells harboring deletion subclones Δ144-Lac and Δ143-Lac appeared blue on X-gal plates, indicating that the flagellin promoter was functional. Microscopically, these cells appeared identical to GIB1 transformed with intact pFLA-1 in terms of producing the characteristic filamentous morphology and impaired motility. However, deletion subclones Δ116-Lac, Δ111-Lac, Δ78-Lac, and Δ48-Lac in GIB1 did not produce blue color on X-gal plates, were highly motile, and microscopically appeared as unicellular short rods. Thus, all deletions that restored normal cell morphology and motility involved at least a portion of the sequence believed to represent the recognition sequence of σ^{28} RNA polymerase. These results strongly implicate amplification of a functional *hag* promoter region with perturbation of normal cell division, motility, and flagellin transport.

Creation of a genomic *hag* deletion mutation. To facilitate subsequent flagellin transport analysis, it was desirable to create a genomic mutation that would abolish host flagellin expression but could be complemented with a copy of the cloned flagellin structural gene. This defective allele was constructed by using a derivative of pFLA-1 that has an *Nde*I site at the beginning of the flagellin coding sequence (Fig. 1). The sequence between this *Nde*I site and the *Eco*RV site in the 3′ end of the coding sequence was removed, the two site were fused together, and the deleted gene was inserted into pJH101. The result is a 633-base-pair deletion that rendered the gene unable to be translated, since there was no longer an initiator methionine codon within the proper distance from the ribosome-binding site (26) and more than two-thirds of the flagellin coding sequence had been removed. This plasmid, called pFIV-3 (Fig. 3), does not contain a *B. subtilis* origin of replication and can be used to replace the wild-type flagellin gene with this defective copy as previously described (48). To do this, plasmid pFIV-3 was introduced into *B. subtilis* GIB16 by transformation, and integrants were selected on L plates with chloramphenicol. The resulting transformants were then picked onto motility plates, where it was noted that about 10% of the Cm^r integrants were nonmotile. This was presumably due to a double recombination event with a plasmid multimer, resulting in total replacement of the host gene with a copy of the plasmid (48). One of these Cm^r, nonmotile isolates was picked and designated GIB17. Southern hybridization indicated a genomic structure for GIB17 consistent with such a recombinational event (Fig. 5). Spontaneous Cm^s revertants of GIB17 were isolated by growing the strain overnight in broth without chloramphenicol selection, plating the cells on L plates without antibiotic, and replica plating onto L plates with and without chloramphenicol. In this way, several Cm^s isolates were obtained. One of these, called GIB18, was analyzed by Southern hybridization to verify that the wild-type allele had been replaced by the deleted copy (Fig. 5). In addition, immunoblot analysis of GIB18 cultures confirmed that the strain no longer produced flagellin (Fig. 4a).

Complementation analysis of *hag* deletion strain GIB18. Plasmid pFLA-1, containing the intact flagellin promoter and structural gene, was introduced into GIB18 by transformation to assess the ability to complement the genomic *hag* deletion with multiple copies of the gene. GIB18 containing pFLA-1 exhibited only weak motility on soft agar plates, and the cells were filamentous when viewed microscopically. Immunoblot analysis of flagellin expression and localization in these cells showed that expression levels were comparable to those of the parental strain GIB16, but almost all of the

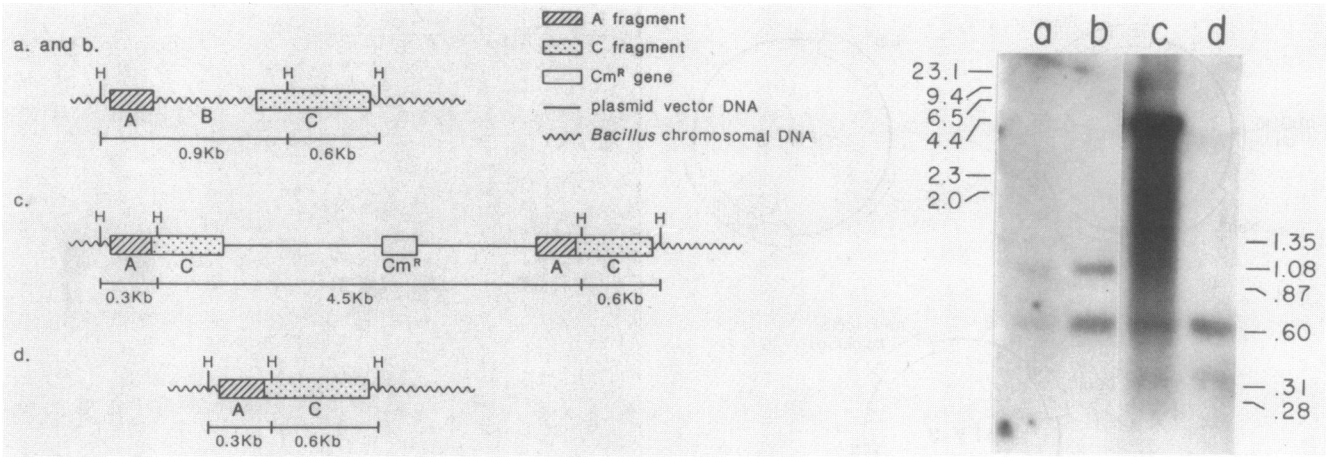

FIG. 5. Analysis of the *hag* deletion mutation and intermediates. Shown are the proposed genomic structures and Southern hybridization data of the intact flagellin locus (GIB1 and GIB16, lanes a and b, respectively), the nonmotile *hag*::pFIV-3 intermediate (lane c), and the Δ*hag-633* allele (lane d). The Southern blot shown is composed of *Hin*dIII genomic digests hybridized with labeled pFIV-3 plasmid as a probe. The *Hin*dIII sites in the drawing are designated H, and regions A, B, and C correspond to those depicted in Fig. 1. Molecular size markers (expressed in kilobase pairs) to the left of the Southern blot represent fragments 1 through 6 of *Hin*dIII-digested λ DNA, and those to the right represent fragments 1 through 6 of *Hae*III-digested φX174 RF DNA.

flagellin expressed was not transported and remained intracellular (Fig. 4a). This was not surprising in view of the fact that amplification of the flagellin promoter region invariably results in this phenotype.

Since multicopy complementation with the flagellin promoter was not possible, a single-copy vector system was prepared which would allow evaluation of complementation of the *hag* mutation with a copy of the intact *hag* gene at an alternate site. The system that was developed involved integration of a single copy of the flagellin promoter and structural gene at the *nprE* locus, which encodes the neutral protease gene in *B. subtilis*. By using the published nucleotide sequence for this gene (57), a synthetic oligonucleotide was prepared which corresponded to a region of the coding sequence of neutral protease and was used to clone a genomic *nprE* fragment from a *B. subtilis* plasmid library by colony hybridization. This fragment was then placed into a derivative of pFIV-3 that contains an intact flagellin coding sequence to create pIEV-3 (Fig. 3). This plasmid was capable of expressing flagellin, but, because it lacked an origin of replication that will function in *B. subtilis*, it could only render cells Cm^r by integrating into the genome. Since the plasmid contained homology to both the *hag* locus and the *nprE* locus and could therefore integrate at either site, Cm^r transformants of pIEV-3 in GIB18 were checked by Southern hybridization to isolate an integrant at *nprE*, which was designated GIB19 (data not shown). This strain was indistinguishable from the original parent GIB16 by all criteria; the cells were highly motile on soft agar plates, they appeared as unicellular rods microscopically, and Western analysis showed wild-type levels of exported flagellin with no intracellular accumulation (Fig. 4b). Thus, this integration system fully complemented the Δ*hag-633* allele.

Transport studies of various flagellin deletions. A number of deletions in the flagellin-coding region of pIEV-3 were created to begin to ascertain what regions of flagellin are necessary for transport. The extents and locations of these deletions are diagrammed in Fig. 6. These deletions were constructed by joining together appropriate restriction sites in the flagellin gene to eliminate the sequences between the sites but maintain the translational reading frame. There were a total of four different deletions that resulted in the removal of various internal portions of the flagellin protein. A fifth deletion prematurely terminated translation of the flagellin polypeptide after amino acid 235. Bacterial flagellins share extensive regions of homology in the amino- and carboxy-terminal portions of the polypeptides, whereas the central portions of the flagellins are divergent (Fig. 6). We designed deletions to selectively remove either highly conserved portions of the molecule (Δ76-123, Δ234-241, and fla235) or regions that do not share homology with other bacterial flagellins (Δ145-211 and Δ145-266). These deleted flagellin genes were integrated at the *nprE* locus in GIB18, and flagellin expression and transport were assayed by autoradiography of electrophoretically separated [^{35}S]methionine-labeled proteins (Fig. 7a) and by immunoblotting (Fig. 7b). Flagellin protein could not be detected in either intracellular or supernatant fractions of the Δ76-123 deletion (data not shown), and the levels of recovery of the other deleted flagellins were low in comparison with full-length flagellin (Fig. 7a), presumably because the deleted flagellin monomers were more susceptible to proteolysis. However, the four remaining mutations produced levels of flagellin that were still readily detectable. Of these, the three internal deletions (Δ234-241, Δ145-211, and Δ145-266) ranged from a small 8-amino-acid deletion of a highly conserved region to a deletion that eliminated more than one-third of the central portion of the flagellin protein. Transport analysis of these deletions (Fig. 7) showed that all three internally deleted flagellins were transported to the culture supernatant. However, the introduction of a termination codon after amino acid 235 in the fla235 construct (which removed 69 highly conserved amino acids from the C terminus of the protein) resulted in transport of only a portion of the flagellin expressed by this mutant, with at least half of the material accumulating intracellularly. These data suggest that the extreme C terminus of flagellin plays a role in transport of the protein from the cell, whereas the divergent central region does not appear to be involved in this function.

DISCUSSION

In this paper, we report the cloning and sequencing of the structural gene that encodes *B. subtilis* flagellin (*hag*). The

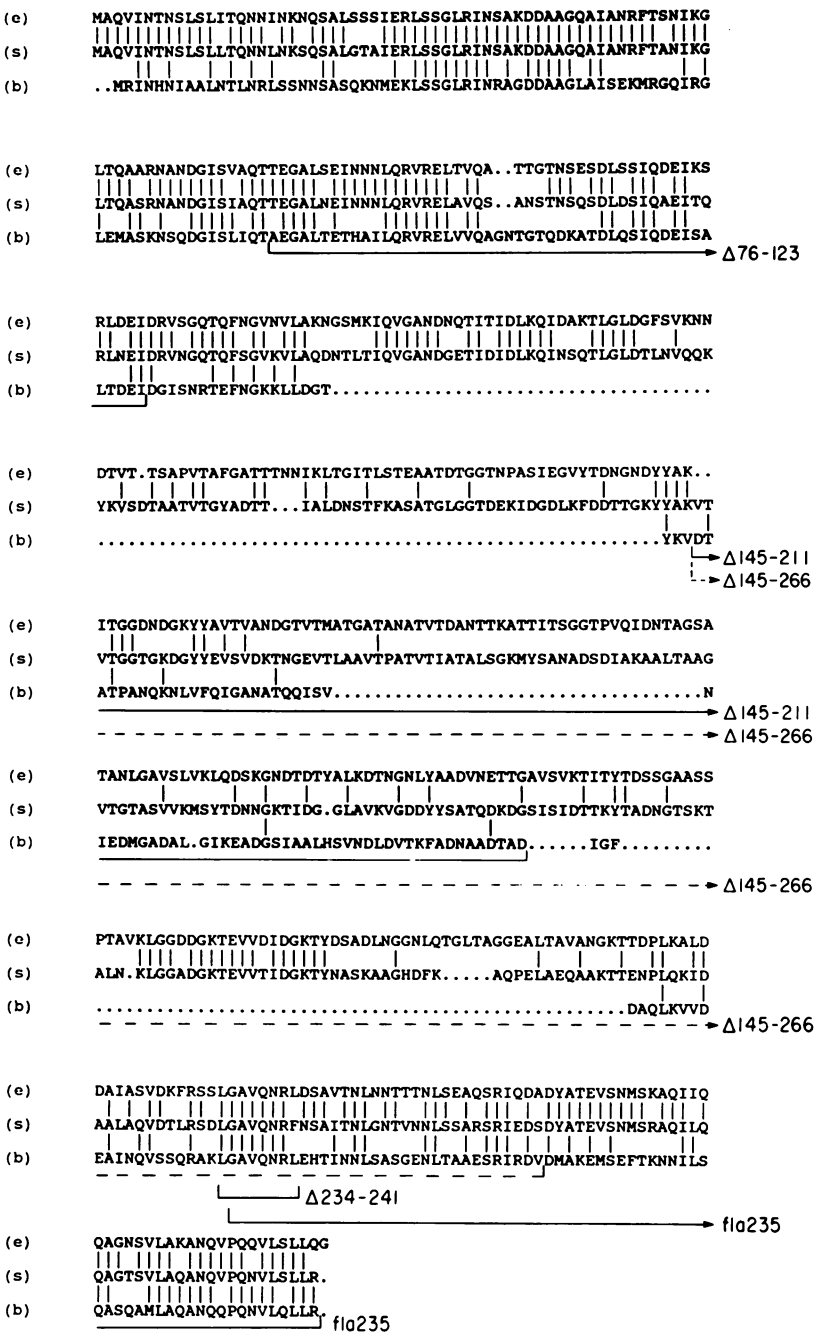

FIG. 6. Comparison of *E. coli* (e), *S. typhimurium* phase 1 (s), and *B. subtilis* (b) flagellin amino acid sequences. Identical residues are indicated by vertical lines. The entire sequence of each flagellin is shown , with gaps to aid in alignments appearing as dotted lines. Deleted regions of the *B. subtilis* protein described in the text are underlined, with arrows indicating continuation of the deleted region to the next line of the sequence. Deletion designations are listed to the right of each line of the sequence involved.

protein product of this gene (as predicted from the nucleotide sequence) is a 304-amino-acid protein, the N terminus of which corresponds to the published amino acid sequence of purified flagellin (10). Therefore, like the cloned flagellin genes of *E. coli* (28), *S. typhimurium* (23, 58), and *C. crescentus* (13), *B. subtilis* flagellin does not utilize a classic amino-terminal signal sequence to direct transport of the protein to the extracellular space.

The *hag* promoter region contains a region of strong homology to the proposed recognition sequence for *B. subtilis* σ^{28} RNA polymerase (14, 17). We have shown that amplification of this region results in intracellular accumulation of flagellin, inability of cells to properly divide, and compromised motility in otherwise wild-type cells. Helmann et al. (18) recently reported a similar phenotype in cells that are incapable of producing σ^{28} due to interruption of the *sigD* gene. This mutation results in loss of motility, failure to synthesize flagellin, and a filamentous cell morphology. This same phenotype is often seen in *lyt* mutants, which are deficient in both autolysin and flagellin production (38).

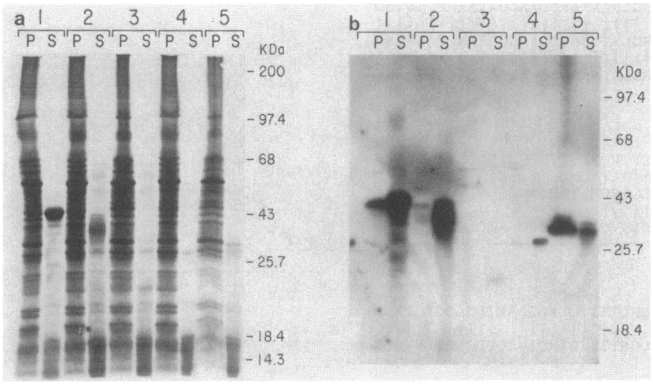

FIG. 7. Transport analysis of various flagellin deletion mutations. (a) Total [^{35}S]methionine-labeled proteins were separated by SDS–10% polyacrylamide gel electrophoresis and (b) Western blot of intracellular (P) and extracellular (S) fractions as described in the text and in the legend to Fig. 4. All constructs were expressed as integrated copies at the *nprE* locus in the Δ*hag-633* strain GIB18. Lanes: 1, pIEV-3 (expressing full-length flagellin); 2, Δ234-241; 3, Δ145-266; 4, Δ145-211; 5, fla235. Migration of standard protein markers is shown to the right of the figure, with molecular masses expressed in kilodaltons.

Helmann et al. proposed that σ^{28} may regulate the expression of a number of genes involved in flagellar synthesis in *B. subtilis* based on its homology to such a positive regulatory factor in *E. coli*, the *flbB* gene product (3). Our data would support that hypothesis. If expression of other components of the flagellar organelle is essential for flagellin transport and their synthesis is dependent on a common sigma factor, then the intracellular accumulation of flagellin that we attribute to an increase in the copy number of the *hag* promoter could be a direct consequence of reducing intracellular levels of σ^{28} by titration. This would serve to limit expression of other *fla* gene products that rely on σ^{28} for transcription and thereby restrict flagellin export capabilities. If this were the case, then substitution of the *hag* promoter with a promoter that utilizes a different sigma factor should alleviate this block. In fact, we have observed that replacement of the flagellin σ^{28} promoter with a σ^{43} recognition sequence on a multicopy plasmid results in normal flagellin expression, localization, and cell morphology.

We describe five deletion mutations that were analyzed for their ability to direct expression and transport of flagellin from a *hag* mutant host. All four of the deletions that produced detectable levels of protein were able to transport flagellin to some degree. The three deletions that removed interior portions of the protein appeared to be fully transported, with no detectable intracellular accumulation. Interestingly, these mutants were nonmotile. It should be noted that the focus of these experiments was to study the transport of flagellin; as a result, the method used for localizing flagellin in these experiments does not differentiate between polymerized and unpolymerized flagellin in the supernatant fraction. Therefore, these mutants may produce a flagellin molecule that is unable to polymerize to form a filament, or the modified flagellin may form a filament which is unable to function effectively. In any event, these results indicate that the central portion of the molecule, which is the least conserved region among bacterial flagellins, does not play a role in transport of the protein. In fact, we have observed that removal of amino acids 187 through 391 in the divergent central portion of *E. coli* flagellin (Fig. 6) results in a truncated molecule that is fully transported to the culture media (unpublished results). Kuwajima (27) reported smaller deletions within this same region of *E. coli* flagellin; these deletions do not affect flagellar synthesis and motility but, instead, alter the antigenic properties of the resulting flagellar filament. Thus, it appears that the divergent central regions of bacterial flagellins are not involved in transport of the protein but, instead, are responsible for antigenic diversity.

Total removal of the highly conserved C terminus in the fla235 mutation results in significant intracellular accumulation of flagellin and a nonmotile phenotype. This result appears to be consistent with the observations of Homma et al. (19), who found that of 26 export-deficient *S. typhimurium* flagellin mutants (19 H1 and 7 H2), virtually all mapped to the distal end of the flagellin genes. However, our C-terminal deletion still transported some material to the exterior of the cell, and examination of total supernatant protein content has enabled us to rule out cell lysis as a possible explanation (Fig. 7a). Furthermore, the flaΔ234-241 construct removes a segment of the C-terminal end of flagellin that shares a high degree of homology to the same region of *E. coli* and *S. typhimurium* flagellin (Fig. 6), yet the protein is transported normally. The Δ145-266 deletion removes even more of the conserved portion of the C terminus, and it, too, is transported. Therefore, it appears that although a critical region for flagellin transport resides in the last 38 amino acids of the polypeptide (the region remaining in the Δ145-266 mutation), another, as yet undefined region of flagellin can at least partially effect export of the protein.

ACKNOWLEDGMENTS

We thank Christine Norton for antiserum preparation; John Brown and Darlene Vanstone for oligonucleotide synthesis; Paul Schendel for his advice and encouragement; and John McCoy, David Rogers, and Ed Fritsch for critically reviewing the manuscript.

LITERATURE CITED

1. **Amann, E., J. Brosius, and M. Ptashne.** 1983. Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. Gene **25**:167–178.
2. **Anagnostopoulos, C., and J. Spizizen.** 1961. Requirements for transformation in *Bacillus subtilis*. J. Bacteriol. **81**:741–746.
3. **Bartlett, D. H., B. B. Frantz, and P. Matsumura.** 1988. Flagellar transcriptional activators FlbB and FlaI: gene sequences and 5′ consensus sequences of operons under FlbB and FlaI control. J. Bacteriol. **170**:1575–1581.
4. **Beaucage, S. L., and M. H. Caruthers.** 1981. Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. Tetrahedron Lett. **22**:1859–1862.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
6. **Burnette, W. N.** 1981. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. **112**:195–203.
7. **Campbell, D. H., J. S. Garvey, N. E. Cremer, and D. H. Sussdorf.** 1970. Bacterial antigens, p. 125–131. *In* Methods of immunology, 2nd ed. W. A. Benjamin, Inc., Reading, Mass.
8. **Dagert, M., and S. D. Ehrlich.** 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli*

cells. Gene 6:23–28.
9. **Dalbadie-McFarland, G., L. W. Cohen, A. D. Riggs, C. Morin, K. Itakura, and J. H. Richards.** 1982. Oligonucleotide-directed mutagenesis as a general and powerful method for studies of protein function. Proc. Natl. Acad. Sci. USA **79**:6409–6413.
10. **Delange, R. J., J. Y. Chang, J. H. Shaper, and A. M. Glazer.** 1976. Amino acid sequence of flagellin of *Bacillus subtilis* 168. J. Biol. Chem. **254**:705–711.
11. **Ferrari, F. A., A. Nguyen, D. Lang, and J. A. Hoch.** 1983. Construction and properties of an integrable plasmid for *Bacillus subtilis*. J. Bacteriol. **154**:1513–1515.
12. **Gill, P. R., and N. Agabian.** 1982. A comparative structural analysis of the flagellin monomers of *Caulobacter crescentus* indicated that these proteins are encoded by two genes. J. Bacteriol. **150**:925–933.
13. **Gill, P. R., and N. Agabian.** 1983. The nucleotide sequence of the Mr=28,000 flagellin gene of *Caulobacter crescentus*. J. Biol. Chem. **258**:7395–7401.
14. **Gilman, M. Z., J. L. Wiggs, and M. J. Chamberlin.** 1981. Nucleotide sequences of two *Bacillus subtilis* promoters used by *Bacillus subtilis* sigma-28 RNA polymerase. Nucleic Acids Res. **9**:5991–6000.
15. **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.
16. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
17. **Helmann, J. D., and M. J. Chamberlin.** 1987. DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternate σ factor. Proc. Natl. Acad. Sci. USA **84**:6422–6424.
18. **Helmann, J. D., L. M. Marquez, and M. J. Chamberlin.** 1988. Cloning, sequencing, and disruption of the *Bacillus subtilis* σ^{28} gene. J. Bacteriol. **170**:1568–1574.
19. **Homma, M., H. Fujita, S. Yamaguchi, and T. Iino.** 1987. Regions of *Salmonella typhimurium* flagellin essential for its polymerization and transport. J. Bacteriol. **169**:291–296.
20. **Hung, M.-C., and P. C. Wensink.** 1984. Different restriction enzyme-generated sticky DNA ends can be joined *in vitro*. Nucleic Acids Res. **12**:1863–1874.
21. **Iino, T.** 1977. Genetics of structure and functions of bacterial flagella. Annu. Rev. Genet. **11**:161–182.
22. **Joys, T. M.** 1965. Correlation between susceptibility to bacteriophage PBS1 and motility in *Bacillus subtilis*. J. Bacteriol. **90**:1575–1577.
23. **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.
24. **Joys, T. M., and V. Rankis.** 1972. The primary structure of the phase-1 flagellar protein of *Salmonella typhimurium*. I. The tryptic peptides. J. Biol. Chem. **247**:5180–5193.
25. **Kawamura, F., and R. H. Doi.** 1984. Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. J. Bacteriol. **160**:442–444.
26. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**: 1–45.
27. **Kuwajima, G.** 1988. Flagellin domain that affects H antigenicity of *Escherichia coli* K-12. J. Bacteriol. **170**:485–488.
28. **Kuwajima, G., J. Asaka, T. Fujiwara, T. Fujiwara, K. Node, and E. Kondo.** 1986. Nucleotide sequence of the *hag* gene encoding flagellin of *Escherichia coli*. J. Bacteriol. **168**:1479–1483.
29. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
30. **Lawn, R. M., J. Adelman, S. C. Bock, A. E. Franke, C. M. Houck, R. C. Najarian, P. H. Seeburg, and K. L. Wion.** 1981. The sequence of human serum albumin cDNA and its expression in *E. coli*. Nucleic Acids Res. **9**:6103–6114.
31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 382–389. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208–218.
33. **Martinez, R. J.** 1963. A method for the purification of bacterial flagella by ion exchange chromatography. J. Gen. Microbiol. **33**:115–120.
34. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Initiation factor-independent translation of mRNAs from gram-positive bacterial Proc. Natl. Acad. Sci. USA **78**:4912–4916.
35. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene. J. Biol. Chem. **256**:11282–11291.
36. **Murray, C. L., and J. C. Rabinowitz.** 1982. Nucleotide sequences of transcription and translation initiation regions in *Bacillus subtilis* phage φ29 early genes. J. Biol. Chem. **257**: 1053–1062.
37. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
38. **Pooley, H. M., and D. Karamata.** 1984. Genetic analysis of autolysin-deficient and flagellaless mutants of *Bacillus subtilis*. J. Bacteriol. **160**:1123–1129.
39. **Randall, L. L., and S. J. S. Hardy.** 1984. Export of protein in bacteria. Microbiol. Rev. **48**:290–298.
40. **Richardson, C. C.** 1971. Polynucleotide kinase from *Escherichia coli* infected with bacteriophage T4. Prog. Nucleic Acid Res. **2**:815–828.
41. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
42. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
43. **Shine, J., and L. Dalgarno.** 1974. The 3'-terminal sequence of *E. coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**: 1342–1346.
44. **Silhavy, T. J., S. A. Benson, and S. D. Emr.** 1983. Mechanisms of protein localization. Microbiol. Rev. **47**:313–344.
45. **Silverman, M., and M. Simon.** 1977. Bacterial flagella. Annu. Rev. Microbiol. **31**:397–419.
46. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
47. **Spizizen, J.** 1958. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proc. Natl. Acad. Sci. USA **44**:1072–1078.
48. **Stahl, M. L., and E. Ferrari.** 1984. Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro-derived deletion mutation. J. Bacteriol. **158**:411–418.
49. **Suzuki, T., and Y. Komeda.** 1981. Incomplete flagellar structures in *Escherichia coli* mutants. J. Bacteriol. **145**:1036–1041.
50. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
51. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
52. **Wells, J. A., E. Ferrari, D. J. Henner, D. A. Estell, and E. Y. Chen.** 1983. Cloning, sequencing and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. **11**:7911–7925.
53. **Wickner, W.** 1979. The assembly of proteins into biological membranes: the membrane trigger hypothesis. Annu. Rev. Biochem. **48**:23–45.
54. **Wiggs, J. L., J. W. Bush, and M. J. Chamberlin.** 1979. Utilization of promoter and termination sites on bacteriophage T7 DNA by RNA polymerase from a variety of different bacterial orders. Cell **16**:97–109.
55. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide se-

quences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.

56. **Yang, M., A. Galizzi, and D. J. Henner.** 1983. Nucleotide sequence of the amylase gene from *Bacillus subtilis*. Nucleic Acids Res. **11**:237–249.
57. **Yang, M. Y., E. Ferrari, and D. J. Henner.** 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. J. Bacteriol. **160**:15–21.
58. **Zeig, J., and M. Simon.** 1980. Analysis of the nucleotide sequence of an invertable controlling element. Proc. Natl. Acad. Sci. USA **77**:4196–4200.