JOURNAL OF BACTERIOLOGY, July 1983, p. 74–81
0021-9193/83/070074-08$02.00/0

Vol. 155, No. 1

DNA Sequence Adjacent to Flagellar Genes and Evolution of Flagellar-Phase Variation

ELIZABETH SZEKELY AND MELVIN SIMON*

Department of Biology, University of California, San Diego, La Jolla, California 92093

Received 28 December 1982/Accepted 8 April 1983

A variety of factors, including phase variation, are involved in the regulation of flagellin gene expression in *Salmonella* sp. Flagellar-phase variation refers to the alternate expression of two different flagellin genes, *H1* and *H2*. Site-specific inversion of a DNA segment adjacent to the *H2* gene is responsible for switching expression. The segment includes the *H2* promoter as well as the *hin* gene, which is required to mediate the inversion. Sequences in this region have homology with the corresponding sequences adjacent to the *H1* flagellin gene in *Salmonella* sp. and the *hag* flagellin gene in *Escherichia coli*. The *hin* gene has also been shown to be homologous to the *gin* gene, which is found on bacteriophage Mu. To understand gene expression and the origin of these relationships, we have compared the DNA sequence adjacent to all three flagellin genes. The sequence data suggest a mechanism for the evolution of the *hin-H2* locus.

The flagellar filaments of gram-negative bacteria are composed of repeating subunits of a protein, flagellin. *Salmonella typhimurium* possesses two homologous flagellin genes, *H1* and *H2*, that map far apart from each other on the chromosome (15). Only one of these genes is expressed at any time, and expression is determined by a site-specific recombination event which inverts a segment of DNA that lies adjacent to the *H2* gene (28, 29). This 995-base pair (bp) invertible region contains the promoter for the *H2* gene (30). In one orientation of the invertible segment, the promoter is coupled to the *H2* operon and *H2*-type flagellin is made. In addition, the *rhl* gene adjacent to *H2* is expressed and it encodes a repressor of *H1* gene transcription (7, 24). When the invertible region is in the opposite orientation, the *H2* and *rhl* genes are not transcribed and the *H1* gene is expressed. Recombination occurs within 14-bp inverted repeat nucleotide sequences that flank the region, and site-specific recombination requires the presence of an intact *hin* gene. This gene lies within the invertible region and encodes a 19,000-molecular-weight polypeptide (23, 24).

The alternate expression of flagellin types has thus far been found only in *Salmonella* species. *Escherichia coli* possesses a single flagellin gene, *hag*, which maps at a site on the *E. coli* chromosome that corresponds to the site where *H1* maps on the *Salmonella* sp. chromosome (16). Whereas *hag*, *H2*, and *H1* all encode antigenically distinct flagellin proteins, these flagellin genes have been shown to be homologous, both by genetic complementation (6, 8) and by DNA-DNA hybridization (26). Partial homology between the *hin* gene and genes that are associated with certain transposable elements has been found. Bacteriophage Mu, for example, has a site-specific inversion system that regulates the expression of genes which encode host range determinants (27). The inversion of a 3,000-bp DNA segment, called the G-loop, is mediated by the product of a Mu gene, *gin*, which maps directly outside of the G-loop (11). Bacteriophage P1 also possesses an invertible DNA segment called the C-loop, which is homologous to the G-loop of Mu (4). In addition, P1 contains a gene, *cin*, which is required for inversion of this DNA segment (S. Iida, J. Meyer, K. Kennedy, and W. Arber, Mol. Gen. Genet., in press). From genetic complementation experiments, it is clear that *hin*, *gin*, and *cin* all encode homologous inversion functions (14; D. Kamp and R. Kahmann, in press). Moreover, the DNA sequences at the sites required for G-loop inversion in Mu show a clear homologous relationship to the sequences required for site-specific recombination in *Salmonella* sp. (R. Kahmann and D. Kamp, in press). Finally, the transposons Tn*3* and gamma delta have been shown to encode a protein, TnpR, which is required for cointegrate resolution; it mediates site-specific deletion between directly repeated 19-bp sequences (1, 13, 19). The amino acid sequence of this protein shows 36% identity with the amino acid sequence of the *hin* gene product (25). Thus far, however, genetic experiments have failed to demonstrate complementation be-

tween the TnpR and the Hin functions.

Taken together, these observations suggest that there exists a "family of genes" associated with transposable elements that function to mediate site-specific recombination. Flagellar-phase variation could have evolved from the interaction of one of these genes on a mobile element and a resident flagellin gene. With the aim of tracing the antecedents of the phase variation system, we compared the DNA sequences found in the corresponding regions of the *hag* and *H1* genes. In addition, we mapped the region of homology between the DNA sequences corresponding to *hin* and sequences present in the invertible region of bacteriophage Mu. In this paper, we present these results and suggest a specific hypothesis to explain the evolution of flagellar-phase variation.

MATERIALS AND METHODS

Enzymes and radioactive isotopes. All restriction endonuclease enzymes were obtained from New England Biolabs and Bethesda Research Laboratories. Enzymes used for radioactive labeling of the DNA probes in the hybridization experiments included DNA polymerase I (Boehringer Mannheim) and DNase I (Worthington Diagnostics). Enzymes used for labeling DNA fragments in nucleotide sequence determination included DNA polymerase large fragment (Boehringer Mannheim), bacterial alkaline phosphatase (Worthington), and T4 polynucleotide kinase (P-L Biochemicals). [γ-^{32}P]ATP (>5,000 Ci/mmol) and γ-^{32}P-labeled deoxynucleotide triphosphates (300 Ci/mmol) were obtained from Amersham Corp.

Plasmid purification and isolation of DNA fragments. Plasmid DNAs were prepared by Sarkosyl lysis (2). Slab gels containing 0.7 to 1% agarose (Seakem) or 9% acrylamide (BDH) were prepared in Tris-acetate buffer (40 mM Tris, pH 7.4, 20 mM NaAc). DNA fragments were isolated from 0.7% agarose gels by a modification of the freeze-squeeze method (28) or from acrylamide gels by the squash elution method as previously described (30).

DNA fragments in 1% agarose gels were denatured and transferred to BA85 nitrocellulose filters (Schleicher & Schuell). ^{32}P-labeled DNA probes were prepared by nick translation (20). The details of the procedure for DNA filter hybridization were previously described (20). The normal hybridization conditions were at 65°C in 0.24 M phosphate buffer; the reduced stringency conditions were at 50°C in the same buffer.

Nucleotide sequence determination. Labeling of 5′ ends with polynucleotide kinase and [γ-^{32}P]ATP after bacterial alkaline phosphatase treatment and labeling of 3′ ends with DNA polymerase large fragment and γ-^{32}P-labeled deoxynucleotide triphosphates were accomplished by the procedure of Maxam and Gilbert (17). Single-end-labeled DNA fragments were obtained by secondary cleavage with a restriction enzyme. All sequence determination was done by the method of Maxam and Gilbert.

Plasmids. The plasmids carrying the *Salmonella* flagellin gene *H1b* (pJZ1) and the *H2^{enx}* gene (pJZ100) have been described (29). Plasmid pJZ110 (28) has also been described, and the *Pst* fragment derived from the plasmid which includes the *hin* gene sequence and part of the *H2* flagellin sequence has been analyzed extensively (30). The partial sequence of this fragment has been reported. pLC24-16 is a plasmid obtained from the Clarke and Carbon (5) collection. It includes a large fragment of *E. coli* DNA that carries the sequences corresponding to the *hag* gene as well as *flaD*, *flaC*, *flaB*, and *flaN*. Plasmid pGM1 was kindly sent to us by Dietmar Kamp.

RESULTS

Homology between flagellin gene regions. All three flagellin genes have been cloned onto plasmid vehicles: the *H1* gene on pJZ2 (29), the *hag* gene on pLC24-16 (5), and the *H2* gene on pJZ100 (30). pJZ110f, a *Pst*I restriction endonuclease fragment containing the invertible region and the beginning of the *H2* gene (Fig. 1), was obtained from pJZ100, and its DNA sequence was determined. We found that when pJZ2 and pLC24-16 were radioactively labeled and used as probes (26), they hybridized only to subfragments of pJZ110f that contained sequences corresponding to the N-terminal region of the *H2* gene. By performing the reciprocal hybridization, we were able to localize the beginning of *H1* and *hag* on pJZ2 and pLC24-16, respectively. The pJZ110 insert fragment was radioactively labeled and hybridized to Southern blots containing pJZ2 and pLC24-16 DNA fragments. The results of this hybridization are shown in Fig. 1. A 0.5-kilobase *Eco*RI-*Pst*I fragment derived from pJZ2 and a 1.8-kilobase *Bam*HI fragment derived from pLC24-16 are the only fragments which demonstrate strong homology with the probe. The 1.8-kilobase *Ham*HI fragment was subsequently recloned onto the plasmid vector pMK2004 (10) and the recombinant plasmid was named pES30.

To further define the regions of homology among the flagellin gene sequences, hybridization experiments were done with PES30 and pJZ2. There is only one fragment in which each of these shows strong homology with the other DNA fragments (Fig. 2). This is presumably the same sequence that hybridizes with the pJZ110 fragment and must therefore correspond to the proximal portion of the gene, i.e., the sequence that specifies the N-terminal region of the protein.

Comparison of DNA sequences. The fragments containing the 5′ region of the *hag* and *H1* genes were isolated and mapped with respect to several other restriction endonuclease cleavage sites. Figure 3 shows this map and the strategy adopted to determine the DNA sequence in this region. The sequences corresponding to the beginning of the *H2*, *H1*, and *hag* genes are shown in Fig. 4. They are aligned with respect to their

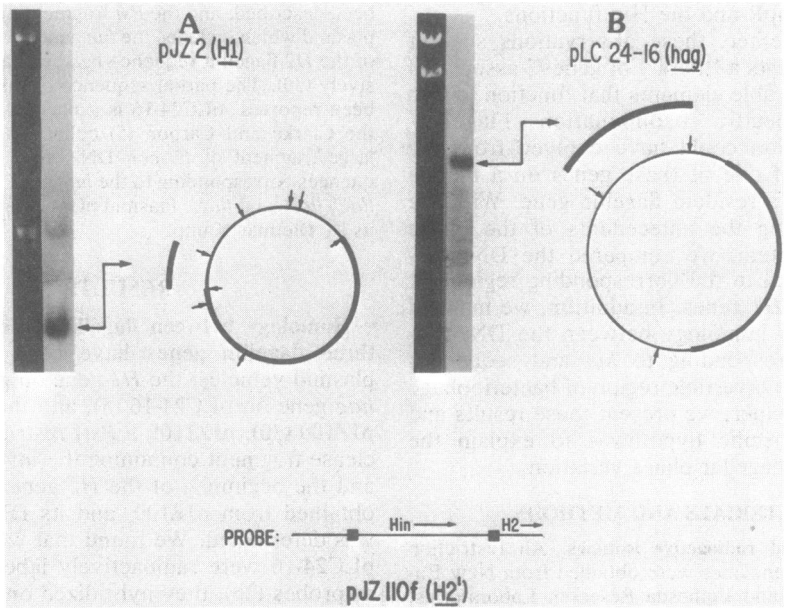

FIG. 1. Identification of DNA fragments containing the 5′ regions of the *H1* and *hag* genes. The map of restriction endonuclease sites on the *H2*-containing plasmid pJZ2 is shown in (A). The inner arrows indicate *Eco*RI cleavage sites and the outer arrows indicate the *Pst*I cleavage sites. The map of the *hag*-containing plasmid pLC24-16 is shown in (B). The positions of the *Bam*HI restriction endonuclease cleavage sites are indicated by arrows. In both maps, the vehicle is distinguished from the insert with a heavier line. An *Eco*RI-*Pst*I digest of pJZ2 and a *Bam*HI digest of pLC24-16 were electrophoresed in 1% agarose, and the ethidium bromide-stained gel profiles are shown in the left lanes of (A) and (B), respectively. The gels were blotted onto a nitrocellulose filter. pJZ110 fragment DNA, containing the first part of the *H2* gene, was radioactively labeled and hybridized to these filters. The map of pJZ110 is shown below. The approximate positions of the *H2* and *hin* genes are shown, and the 14-bp inverted repeat sequences flanking the inversion region are represented by boxes. The autoradiograms obtained after hybridization are shown in (A) and (B) adjacent to the gel profiles. A bar over the pJZ2 and pLC24-16 plasmid maps identifies the fragments which showed homology to the probe.

ATG initiation codons, and the numbering system used here is the same as that used when the sequence of the invertible region was reported (26). At approximately nucleotide position 750 in the sequences corresponding to the *H1* and *hag* regions, there appears to be the start of another gene that could be transcribed in the opposite direction from the flagellin genes. Subsequent experiments have shown that the open reading frame does, in fact, correspond to the N-terminal region of a gene which we have tentatively called *rfs*. In the *H2* gene sequence there is no apparent homolog of *rfs*; instead the sequence specifying the C-terminal region of the *hin* gene is found at nucleotide position 650.

Seven "conserved regions," designated I through VII, are underlined. These regions were identified by the following criteria: (i) they contained at least 5 bp of continuous homology; (ii) they contained <20% overall sequence divergence; and (iii) they were present in the same relative locations upstream from the flagellin genes. The relative spacing of these sequences is shown in Fig. 5. In all three flagellin genes there is a great deal of sequence homology that is continuous. Furthermore, the region corresponding to the *rfs* gene in *E. coli* and *Salmonella* sp. also shows marked homology. However, the sequences between the *rfs* and the flagellin genes are related, but they do not show continuous homology. Instead, an alternating pattern of conserved and divergent sequences is found. In addition, there is a clear difference in the 5′ noncoding region between *H2* and other flagellin gene regions. Whereas all seven of the conserved regions are found adjacent to *H1* and *hag*, only regions I, II, and III are found upstream from the *H2* gene (see Fig. 5). Finally, part of the 14-bp inverted repeat sequence associated with *H2* (RIR) is conserved in the regions 5′ to *H1* and *hag*. This 5-bp sequence, ATCAA, is part of a larger region of homology adjacent to *H1* and *hag* (i.e., region I; see Fig. 4 and 5).

Homology between *hin* and *gin*. Previous results have shown that *hin*-containing plasmids have homology with a region of bacteriophage Mu containing *gin* and the G-loop. This homology was detected by DNA-DNA hybridization

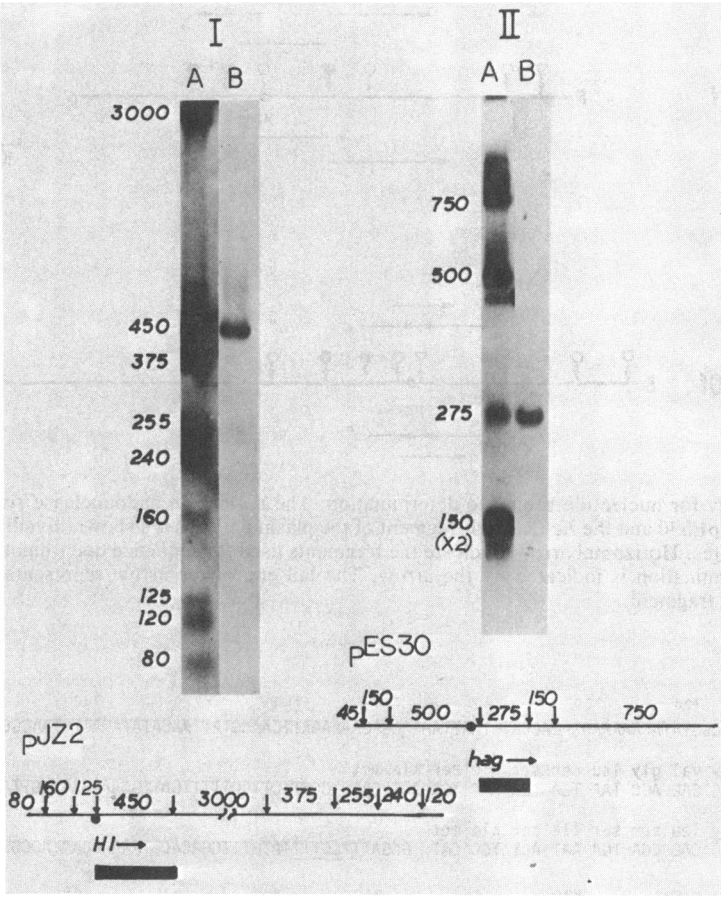

FIG. 2. Determination of sequence homology between pES30 and pJZ2. The insert fragments from plasmids pES30 and pJZ2 were digested with *Sau*3A and separated by electrophoresis in 0.9% acrylamide. The fragments were transferred to nitrocellulose paper and hybridized with radioactively labeled pES30 or pJZ2. The autoradiograms are shown. (I) Lane A, digest of pJZ2 hybridized with radioactive pJZ2; lane B, digest of pJZ2 hybridized with radioactive pES30. (II) Lane A, digest of pES30 hybridized with radioactive pES30; lane B, digest of pES30 hybridized with radioactive pJZ2. The *Sau*3A restriction maps of the two fragments are shown. The closed circles represent the start of the *HI* and *hag* coding regions. The bars below the map indicate the regions of sequence homology identified in these experiments.

under conditions of reduced stringency, and the amount of mismatch between these homologous sequences was estimated by melting experiments to be approximately 20% (26). The following experiment was performed to show that the homology detected previously reflected homology between *hin* and *gin* sequences exclusively. pGM1 (Fig. 6), a plasmid containing both the G-loop and *gin* sequences from Mu, was digested with *Hin*cII restriction endonuclease, and the fragments were separated by electrophoresis in 1% agarose and transferred to nitrocellulose filters. The *hin*-containing plasmid pJZ110, was radioactively labeled and hybridized to the filter blot under conditions of reduced stringency. The results are shown in Fig. 6. Hybridization is found with four bands, A, D, F, and G. Bands A and D are attributed to hybridization between plasmid vehicles, whereas bands F and G represent homology with sequences present on the pJZ110 insert. Since fragments F and G contain sequences from the *gin* gene (11) and since no hybridization is observed with fragments C and B, the homology with pJZ110 must be restricted to sequences outside of the G-loop containing *gin*. These hybridization experiments were repeated with *Alu*I fragments of pJZ110 as probes. Only *Alu*I fragment K of pJZ110 (see Fig. 6, III) shows strong homology with pGM1; again, only fragments F and G from pGM1 hybridized. Weak hybridization to these same fragments was also observed when fragment L was used as

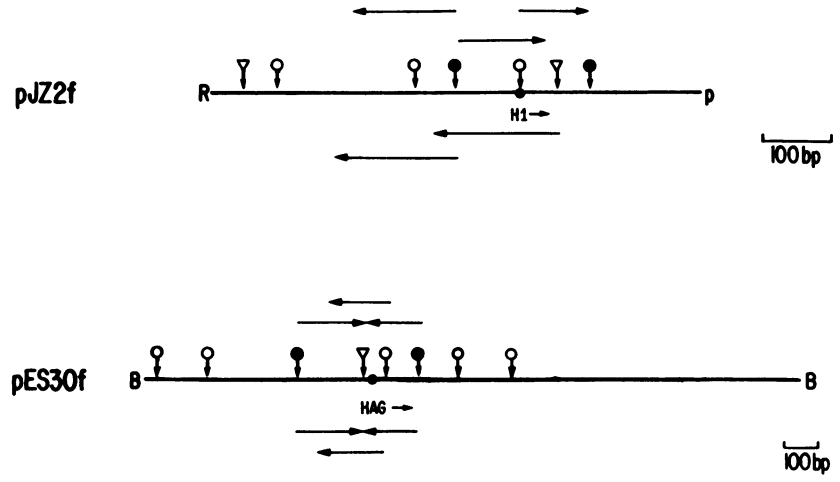

FIG. 3. Strategy for nucleotide sequence determination. The restriction endonuclease sites on the *Bam*HI insert fragment of pES30 and the *Eco*RI-*Pst*I fragment of the plasmid pJZ2Z are shown. Symbols: ○, *Sau*3A; ▽, *Hin*cII; ●, *Taq*I sites. Horizontal arrows indicate the fragments used for sequence determination. The direction of sequence determination is indicated by the arrow. The tail end of the arrow represents the radioactively labeled end of the fragment.

```
                 720        730        740        750        760        770        780        790        800
hin    CCGAAAGGGTATACAAAAAAGACAGCATCTAATTAAAAAGAGAAAAAATTCAACGTATTAACATATATAGTGTAACGCGCTCACGATAAGG

       ser gly val gly leu ser ser ile ser ala met
rfs    TGA TCC CAC ACC TAA TGA TGA AAT TGA AGC CAT  GCCTTCTTCCTTTTTGATTGCAAACAGTAGTTAAGCGCGTTATCGG

       val gly leu ser ser ile ser ala met
rfs    GAC TCC CAG CGA TGA AAT ACT TGC CAT  GCGATTTCCTTTTATCTTTCGACACGTAAAACGAATACCGGGGTTATCGGTCTG
                                                                                      VII
       810        820        830        840        850        860        870        880        890
CCTATGTTACATCCAGCTATAGACGACATCGCTCAAAACACTACCAGACACAGTATTCACCTGGAAAGGCTTTTTAATCAAAATGTTAGATGTAAGCA

CAATCTGGAGGCAAAGTTTAATGATAATTTTGCAAAAATAATGCGCGGAATAATGATGCATAAAGCGGCTATTTCGCCGCCTAAGAAAAAGATCGGGG

AATTGCGCAAAGTTTACGTTTAATTGTTTTTTTTAATAGCGGGAATAAGGGGCAGAGAAAAGAGTATTTCGGCGACTAACAAAAAATGGCTGTTTTTG
       VI                                 V                                  IV
900        910        920        930        940        950        960        970        980        990
ATTACGGACAGAAAAAATAGTAAAGTTTATGCCTCAAGTGTCGATAACCTGGATGACACAGGTAAGCCTGGCATAACATTGGTTATCAAAAACCTTCC

GAAGTGAAAAATTTTCTAAAGTTCGAAATTCAGGTGCCGATACAAGGGTTACGGTGAGAAACCGTGGGGAACAGCCCAATAACATCAAGTTGTAATTG

AAAAAATTCTAAAGGTTGTTTTACGACAGACGATAACAGGGTTGACGGCGATTGAGCCGACGGGTGGAAACCCAATACGTAATCAACGACTTGCAATA
    III                                II                                  I
1000       1010       1020       1030       1040       1050       1060       1070
AAAAGGAAAATTTT ATG GCA CAA GTA ATC AAC ACT AAC AGT CTG TCG CTG CTG ACC CAG AAT AAC CTG AAC AAA  H2
               met ala gln val ile asn thr asn ser leu ser leu leu thr gln asn asn leu asn lys

ATAAGGAAAAGATC ATG GCA CAA GTC ATT AAT ACA AAC AGC CTG TCG CTG TTG ACC CAG AAT AAC CTG AAC AAA  H1

TAGGATAACGAATC ATG GCA CAA GTC ATT AAT ACC AAC AGC CTC TCG CTG ATC ACT CAA AAT AAT ATC AAC AAG Hag
```

FIG. 4. Nucleotide sequences in the regions adjacent to the flagellin genes $H2^{enx}$, $H1^{b}$, and *hag*. The sequences are aligned so that the initial coding nucleotide of each sequence will correspond to 1,012 (30). Homologous regions are underlined and identified by Roman numerals (except for the ribosome-binding region). The amino acid sequence implied for the *H2* gene is shown below the nucleotide sequence. The amino acid sequences for the *rfs* genes in *E. coli* and *Salmonella* sp. are shown above the nucleotide sequences.

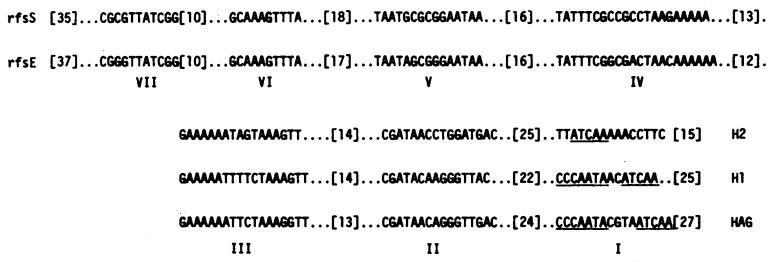

FIG. 5. Comparison of conserved sequences present upstream from the *H2*, *H1*, and *hag* genes. The homologous sequences are shown, with the number of nucleotides that separate it from its neighbor indicated in brackets.

the probe. Fragments K and L are known to both contain sequences corresponding to parts of the *hin* gene (30). Furthermore, when bacteriophage Mu DNA was used as a hybridization probe against *Hpa*II digests of pJZ110 DNA, only the fragments that map within the *hin* gene hybridized (data not shown).

DISCUSSION

There is a great deal of homology between the genes that encode flagellin in *E. coli* and in *Salmonella* sp. Genetic experiments have suggested (9) that the homology is confined to one region and that this region is highly conserved. Our results (Fig. 2) show that approximately the N-terminal one-third of the *hag* and *H2* genes show a great deal of homology and cross-hybridization. The middle and C-terminal regions of the gene did not show hybridization under stringent conditions.

The regulation of the rate of flagellin formation is complex. It involves the products of

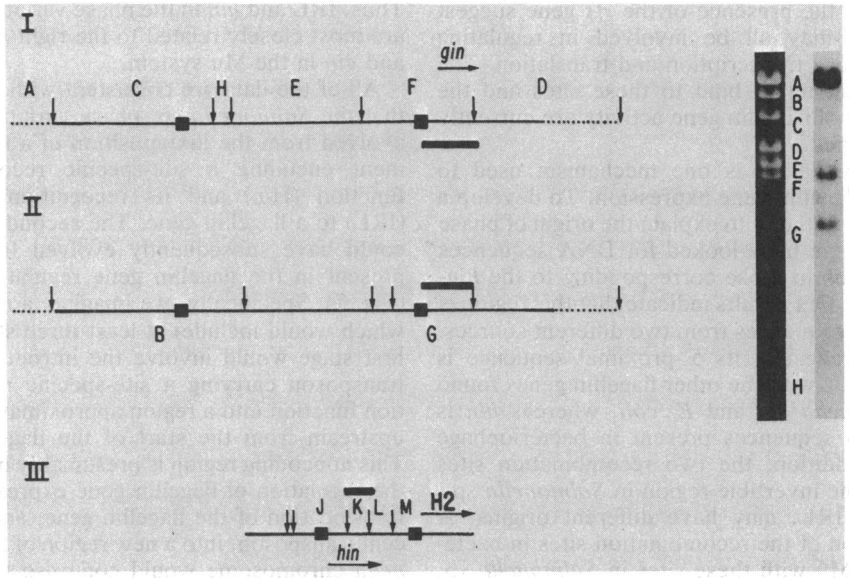

FIG. 6. Homology between the *hin* gene region in *Salmonella* sp. and the *gin* gene region in Mu. I and II show the restriction endonuclease map of pGM1 with the G-loop in the two possible orientations. The dotted line represents the plasmid vehicle. The positions of *Hin*cII restriction endonuclease cleavage sites are indicated by arrows, and the 34-bp inverted repeat sequences flanking the G-loop are represented as boxes. The approximate position of the *gin* gene is shown. A *Hin*cII restriction digest of pGM1 was electrophoresed in 1% agarose and transferred to nitrocellulose filters. The ethidium bromide-stained profile obtained from the gel is shown in the left lane. Radioactively labeled pJZ110 DNA was hybridized to the filter blot, and the autoradiogram obtained is shown in the right lane. Bars are drawn over the plasmid map in I and II to indicate approximate regions of homology between sequences contained on the pGM1 insert and pJZ110. III shows the *Alu*I restriction map of the pJZ110 DNA fragments. The boxes represent the 14-bp inverted repeat sequences flanking the inversion region, and the approximate positions of *hin* and *H2* are indicated. A bar is drawn to indicate the region of *hin* that hybridized with radioactively labeled pGM1.

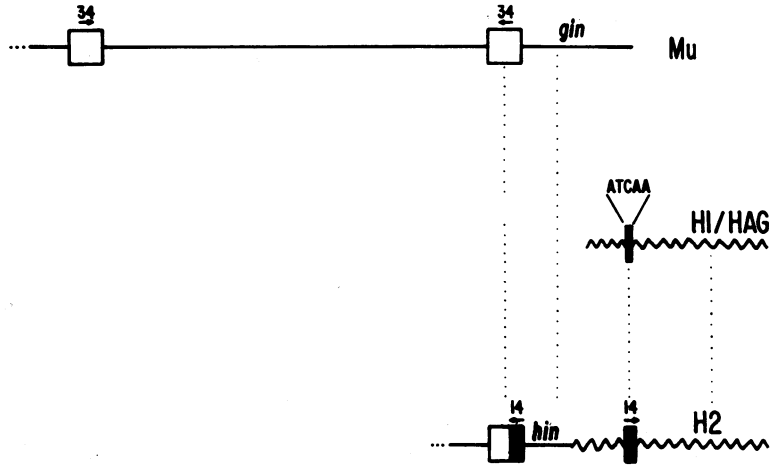

FIG. 7. Model summarizing the apparent origin of sequences found in the *H2* gene region. Dotted vertical lines are drawn to indicate that a relationship has been demonstrated between the connected sequences either by hybridization or directly by nucleotide sequence determination.

genes in the flagellar pathway as well as other gene products that have no direct relationship to the flagellar structure. The large number of conserved sites upstream from the flagellin genes and the presence of the *rfs* gene suggest that they may all be involved in regulating flagellin gene transcription and translation. The factors that might bind to these sites and the regulation of flagellin gene activity are currently being studied.

Phase variation is one mechanism used to regulate flagellin gene expression. To develop a specific hypothesis to explain the origin of phase variation, we have looked for DNA sequences homologous to those corresponding to the *hin-H2* region. Our results indicate that this region is related to sequences from two different sources. The *H2* gene and its 5'-proximal sequence is homologous with the other flagellin genes found in *Salmonella* sp. and *E. coli*, whereas *hin* is related to sequences present in bacteriophage Mu. In addition, the two recombination sites flanking the invertible region in *Salmonella* sp., IRR and IRL, may have different origins. A comparison of the recombination sites in bacteriophage Mu with these sites in *Salmonella* sp. indicates that IRR and IRL are different. Kahmann and Kamp (in press) have recently determined that the G-loop is flanked by 34-bp inverted repeat sequences which bear a close resemblance to the 14-bp inverted repeat sequences in *Salmonella* sp. The 34-bp sequence in Mu contains as a subset the 14-bp IRR and IRL phase variation sequences with only two mismatches. In addition, the 20-bp sequence adjacent to IRL, but not to IRR, has striking homology with the rest of the Mu sequence. The sequence next to IRL in the phase variation system has 15 of 20 bp that are identical with the Mu sequence, whereas the sequence next to IRR matches the Mu sequence in only 3 of 20 bp. Thus, IRL and *hin* in the phase variation system are most closely related to the right-side repeat and *gin* in the Mu system.

All of the data are consistent with the notion that the *Salmonella* sp. phase variation system evolved from the juxtaposition of a mobile element encoding a site-specific recombination function (Hin) and its recognition sequence (IRL) to a flagellin gene. The second site, IRR, could have subsequently evolved from a site present in the flagellin gene regulatory region (Fig. 7). Specifically, we imagine (a mechanism which would include) at least three stages. The first stage would involve the introduction of a transposon carrying a site-specific recombination function into a region approximately 100 bp upstream from the start of the flagellin gene. This noncoding region is presumably involved in the regulation of flagellin gene expression. The translocation of the flagellin gene, and its adjacent transposon, into a new region of the *Salmonella* chromosome would comprise the second step. This step could have occurred by a variety of mechanisms; it may have been mediated by the inserted transposon itself or else by another transposable element. More than one origin is also possible for the *H2* gene precursor. It could have evolved from *H1*, as originally suggested by Iino (8), or it may derive from a flagellin gene in a related bacterium. Finally, an aberrant recombination event mediated by the site-specific recombination system encoded by the transposon could result in the "capture" of the

flagellin gene promoter. A chromosomal sequence adjacent to the promoter which has partial homology with the normally recognized site could act as a new recombination site. As a consequence of this step, recombination between the IRL site carried on the transposon and the new chromosomal site (derived from a sequence present in the flagellin regulatory region) would invert the DNA segment containing the flagellin gene promoter. A reasonable candidate for this chromosomal site is the conserved ATCAA sequence, which is found approximately 25 bp upstream from each of the flagellin genes *hag*, *H1*, and *H2* and is found within the inverted repeated sequence of the phase variation system.

The notion that mobile genetic elements could participate in the evolution of regulatory pathways was first suggested by Barbara McClintock (18) on the basis of her work in maize. The evolution of the phase variation system in *Salmonella* sp., we believe, is an example of this kind of process. The recruitment by recombination of new sites may be a general phenomenon. It allows for the formation of new genetic configurations by exchange of specific functions between transposable elements and chromosomal genes. The resulting mechanisms may be important in cell development, evolution, and disease. Although site-specific recombination functions have, thus far, only been associated with transposons and viruses in procaryotes, it seems probable that such functions are also contained on mobile elements in eucaryotes. Hence, site-specific recombination may play a role in many of the processes involving the regulation of gene activity in both procaryotes and eucaryotes.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation.

We thank D. Kamp and R. Kahman for communicating the results of their determination of the sequences of the Mu inversion sites before publication.

LITERATURE CITED

1. **Arthur, A., and D. Sherratt.** 1979. Dissection of the transposition process: a transposon encoded site-specific recombination system. Mol. Gen. Genet. **175**:267–271.
2. **Bazaral, M., and D. Helinski.** 1968. Circular DNA forms of colcinogenic factors E1, E2, and E3 from *Escherichia coli*. J. Mol. Biol. **36**:185–194.
3. **Campbell, A.** 1981. Evolutionary significance of accessory DNA elements in bacteria. Annu. Rev. Microbiol. **35**:55–83.
4. **Chow, L., and A. Bukhari.** 1976. The invertible DNA segments of coliphages Mu and P1 are identical. Virology **51**:242–248.
5. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
6. **Enomoto, M., and B. Stocker.** 1975. Integration at *hag* or elsewhere of *H2* genes transduced from *Salmonella* to *Escherichia coli*. Genetics **81**:595–614.
7. **Fujita, H., S. Yamaguchi, and T. Iino.** 1973. Studies of H-O variants in *Salmonella* in relation to phase variation. J. Gen. Microbiol. **76**: 127–134.
8. **Iino, T.** 1961. Anomalous homology of flagellar phases in *Salmonella*. Genetics **46**:1471–1474.
9. **Iino, T.** 1977. Genetics of structure and function of bacterial flagella. Annu. Rev. Genet. **11**:161–182.
10. **Kahn, M., R. Kolter, C. Thomas, D. Figurski, R. Meyer, E. Remaut, and D. Helinski.** 1979. Plasmid cloning vehicles derived from plasmids ColR1, F, R6K, and RK2. Methods Enzymol. **68**:268.
11. **Kamp, D., L. Chow, T. Broker, D. Kwoh, D. Zipser, and R. Kahmann.** 1979. Site-specific recombination in phage Mu. Cold Spring Harbor Symp. Quant. Biol. **43**:1159–1167.
12. **Kleckner, N.** 1981. Transposons in bacteria. Annu. Rev. Genet. **15**:341–404.
13. **Kostriken, R., C. Monita, and F. Heffron.** 1981. The transposon Tn3 encodes a site-specific recombination system: identification of essential sequences, genes, and the actual site of recombination. Proc. Natl. Acad. Sci. U.S.A. **78**:4041–4045.
14. **Kutsukake, D., and T. Iino.** 1980. Inversions of specific DNA segments in flagellar phase variation of *Salmonella* and inversion systems of bacteriophage P1 and Mu. Proc. Natl. Acad. Sci. U.S.A. **77**:7338–7341.
15. **Lederberg, J., and T. Iino.** 1956. Phase variation in *Salmonella*. Genetics **41**:743–757.
16. **Makela, P.** 1964. Genetic homologies between flagellar antigens of *Escherichia coli* and *Salmonella abony*. J. Gen. Microbiol. **35**:513–510.
17. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **76**:1760–1764.
18. **McClintock, B.** 1957. Controlling elements and the gene. Cold Spring Harbor Symp. Quant. Biol. **21**:197.
19. **Reed, R.** 1981. Resolution of co-integrates between transposons and Tn3 defines the recombination site. Proc. Natl. Acad. Sci. U.S.A. **78**:3428–3432.
20. **Rigby, P., J. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
21. **Shimada, K., R. A. Weisberg, and M. Gottesman.** 1972. Prophage lambda at unusual chromosomal locations. I. Location of the secondary attachment sites and the properties of the lysogens. J. Mol. Biol. **63**:483–503.
22. **Silverman, M.** 1980. Building bacterial flagella. Q. Rev. Biol. **55**:395–407.
23. **Silverman, M., and M. Simon.** 1980. Phase variation: genetic analysis of switching mutants. Cell **19**:845–854.
24. **Silverman, M., J. Zieg, and M. Simon.** 1979. Flagellar phase variation: isolation of the *rhl* gene. J. Bacteriol. **137**:517–521.
25. **Simon, M., J. Zieg, M. Silverman, G. Mandel, and R. Doolittle.** 1980. Phase variation: evolution of a controlling element. Science **209**:1370–1374.
26. **Szekely, E., and M. Simon.** 1981. Homology between the invertible DNA sequence which controls flagellar phase variation in *Salmonella* and DNA sequences in other organisms. J. Bacteriol. **148**:829–836.
27. **Van de Putte, P., S. Cramer, and M. Giphart-Gassler.** 1980. Invertible DNA determines host specificity of bacteriophage Mu. Nature (London) **286**:218–222.
28. **Zieg, J., M. Hilmen, and M. Simon.** 1977. Recombination switch for gene expression. Science **196**:170–172.
29. **Zieg, J., M. Silverman, M. Hilmen, and M. Simon.** 1977. Recombination switch for gene expression. Science **196**:170–172.
30. **Zieg, J., and M. Simon.** 1980. Analysis of the nucleotide sequence of an invertible controlling element. Proc. Natl. Acad. Sci. U.S.A. **77**:4196–4260.