JOURNAL OF BACTERIOLOGY, Sept. 1993, p. 5395–5402
0021-9193/93/175395-08$02.00/0

Vol. 175, No. 17

Comparative Analysis of Flagellin Sequences from *Escherichia coli* Strains Possessing Serologically Distinct Flagellar Filaments with a Shared Complex Surface Pattern

GARY SCHOENHALS† AND CHRIS WHITFIELD*

Department of Microbiology, College of Biological Sciences, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Received 16 March 1993/Accepted 13 June 1993

Escherichia coli **morphotype E flagellar filaments have a characteristic surface pattern of short-pitch loops when examined by electron microscopy. Seven of the 50 known *E. coli* H (flagellar antigen) serotypes (H1, H7, H12, H23, H45, H49, and H51) produce morphotype E filaments. Polymerase chain reaction was used to amplify flagellin structural (*fliC*) genes from *E. coli* strains producing morphotype E flagellar filaments and from strains with flagellar filaments representing other morphotypes. A single DNA fragment was obtained from each strain, and the size of the amplified DNA correlated with the molecular mass of the corresponding flagellin protein. This finding and hybridization data suggest that these bacteria are monophasic. *fliC* genes from three *E. coli* serotypes (H1, H7, and H12) possessing morphotype E flagellar filaments were sequenced in order to assess the contribution of conserved flagellin primary sequence to the characteristic filament architecture. The H1 and H12 *fliC* sequences were identical in length (1,788 bp), while the H7 *fliC* sequence was shorter (1,755 bp). The deduced molecular masses of the FliC proteins were 60,857 Da (H1), 59,722 Da (H7), and 60,978 Da (H12). The H1, H7, and H12 flagellins demonstrated 98 to 99% identity over the amino-terminal region (190 amino acid residues) and 89% (H7) to 99% (H1 and H12) identity in the carboxy-terminal region (100 amino acid residues). The complete primary amino acid sequences for H1 and H12 flagellins differed by only 10 amino acids, accounting for previously reported serological cross-reactions. However, the central region of H7 flagellin had only 38% identity with H1 and H12 flagellins. The characteristic morphology of morphotype E flagellar filaments is therefore not dependent on a highly conserved primary sequence within the exposed central region. Comparison of morphotype E *E. coli* flagellins with those from *E. coli* K-12, *Serratia marcescens*, and several *Salmonella* serovars supported the established concept of highly conserved terminal regions flanking a variable central region.**

Flagella are the locomotory organelles of many prokaryotes. In *Escherichia coli*, the flagellum is a complex structure consisting of three main structural regions: the basal body, the hook, and the filament (reviewed in reference 30). The eubacterial flagellar filament is composed of many thousands of copies of a single protein subunit, flagellin (reviewed in reference 17). The flagellin molecule carries the antigenic determinant for the H (flagellar) antigen, which can be identified by using H antigen-specific antibodies in slide agglutination tests. In *E. coli*, there are at least 50 known H serotypes (3, 29). Most H serotypes can be classified into one of six flagellar filament morphotypes (A, B, C, D, E, and F) by the surface architecture of intact flagellar filaments viewed by electron microscopy (28, 29). *E. coli* H serotypes H1, H7, H12, H23, H45, H49, and H51 are included as members of morphotype E, since flagellar filaments possessing these H antigens display the surface pattern of short-pitch loops characteristic of morphotype E filaments (29, 44, 60).

Immunological analysis of the flagellins from members of the family *Enterobacteriaceae* has demonstrated the presence of both serotype-specific determinants (3, 16, 19, 20, 39, 43, 60) and cross-reactive determinants (1, 3, 5, 12, 13, 26, 43). The immunological relationships among morphotype E flagellins have been analyzed by using monoclonal antibodies (43, 60). In addition to serotype-specific and broadly cross-reactive epitopes, other epitopes were common among subgroups of closely related flagellin serotypes within morphotype E. Immunogold labeling studies demonstrated that serotype-specific determinants tend to be surface exposed in intact flagellar filaments, while cross-reactive determinants tend to be hidden (43, 60).

DNA sequence analysis of flagellin structural genes from *E. coli* K-12 (24), *Serratia marcescens* (10), and several *Salmonella* serovars (15, 48, 58, 59) demonstrated that the termini of flagellin proteins are conserved. The central region of the flagellin molecule is variable and gives rise to serotype-specific epitopes (6, 16, 18, 19, 23, 36, 45, 47, 48, 58). The flagellin protein therefore possesses several distinct structural domains (34, 46, 51–54, 56). These have been mapped on the primary structure of the flagellin molecule and localized within the intact flagellar filament by electron density mapping studies (57). The conserved, terminal regions of flagellin were found to be associated with the inner structure of intact flagellar filaments. The outer surface-exposed architecture is formed from the folding outward of

* Corresponding author.
† Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511.

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Serotype or relevant genotype	Flagellar filament morphotype[a]	Reference or source
E. coli			
Su 1242	O2:K2:H1	E	F. Ørskov
U5-41	O1:K1:H7	E	F. Ørskov
Bi 316-42	O9:K9:H12	E	F. Ørskov
K42	O45:K1:H23	E	F. Ørskov
4106-54	O52:K?:H45	E	F. Ørskov
2147-59	O6:K13:H49	E	F. Ørskov
C218-70	O8:K50:H51	E	F. Ørskov
U9-41	O2:K1:H4	A	F. Ørskov
U4-41	O4:K3:H5	B	F. Ørskov
P4	O16:K?:H48	C	F. Ørskov
U11a-44	O8:K49:H21	D	F. Ørskov
5306-56	O26:K?:H46	F	F. Ørskov
DH5α	K-12 F$^-$ ϕ80d *lacZΔM15 endA1 recA1 hsdR17* (r_K^- m_K^-) *supE44 thi-1 gyrA96 relA1 Δ(lacZYA-argF)U169*		41
Plasmids			
pGEM-7Zf(+)	Cloning vector; Ap^r		Promega Biotech
pKS93 (pBR322/hag93)	pBR322 derivative containing *E. coli* K-12 *fliC* gene; Ap^r		24, 25
pWQ701	pGEM-7Zf(+) derivative containing *E. coli* H1 *fliC* gene; Ap^r		This study
pWQ707	pGEM-7Zf(+) derivative containing *E. coli* H7 *fliC* gene; Ap^r		This study
pWQ712	pGEM-7Zf(+) derivative containing *E. coli* H12 *fliC* gene; Ap^r		This study

[a] Data are from references 29 and 60.

the central variable region of the flagellin molecule (4, 14, 23, 34, 35, 51–53, 55).

The unique properties of flagellins, including recognition for export and self-assembly (25, 55, 58), place structural constraints on these proteins. The structural constraints on morphotype E flagellins are potentially more severe because they possess antigenically unique determinants while conserving the complex structural features (surface patterns) of the flagellar filaments seen in this group. The objective of this work was to investigate the relationships between the FliC proteins from morphotype E flagellar filaments to determine whether primary sequence conservation is essential for the identical flagellar filament morphology seen in the morphotype E group.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. Cultures were grown in Luria broth at 37°C. Ampicillin (100 μg/ml) was added when appropriate. Plasmid pGEM-7Zf(+) (Promega Biotech) was used for cloning experiments. Recombinant plasmids were maintained and amplified in *E. coli* DH5α.

DNA manipulation. Chromosomal DNA was prepared by the method of Hull et al. (11). Plasmid DNA was purified by an alkaline lysis method, and transformations were performed with $CaCl_2$-treated competent cells (41). DNA sequencing was performed by the dideoxy chain termination method (42) with Sequenase version 2.0 (USB, Cleveland, Ohio); both strands were sequenced. T7 and SP6 oligonucleotide primers were obtained from Promega Biotech. Other synthetic oligonucleotides for DNA sequencing and polymerase chain reactions (PCRs) were obtained from Vetrogen Corp. (London, Ontario, Canada). Protein sequences were aligned with PALIGN (33), part of the PC Gene software package (Intelligenetics, Mountain View, Calif.). Enzymes were purchased from either Bethesda Research Laboratories Inc. (Gaithersburg, Md.) or Boehringer-Mannheim Canada (Laval, Quebec, Canada) and used as recommended by the manufacturer.

PCR amplification and cloning of *fliC* sequences. PCR amplification reactions were performed with a Coy Temp-Cycler model 60 thermocycler. Oligonucleotide primers for PCR amplification were designed from the previously published sequence of the *E. coli* K-12 *fliC* gene (24). Primer A (5′-CCGAATTCATGGCACAAGTCATTAATAC-3′) contained the first 20 nucleotides of *E. coli fliC* preceded by an *Eco*RI restriction site (underlined) and a 5′-terminal CC clamp sequence. The reverse primer, primer B (5′-CC GAATTCTTAACCCTGCAGTAGAGACA-3′), also contains an *Eco*RI site and clamp sequence but is followed by the complementary nucleotide sequence for the 3′-terminal 20 nucleotides of *fliC*. PCR amplification was preceded by an initial denaturation step at 95°C for 10 min. Each cycle consisted of a denaturing step at 95°C for 1.5 min, an annealing step at 55°C for 1.5 min, and a polymerization step at 72°C for 2 min. Amplification was carried out for 35 to 40 cycles. *Taq* DNA polymerase was obtained from Boehringer-Mannheim. PCR amplification products were purified by phenol-chloroform extraction, digested with *Eco*RI, and ligated to linearized pGEM-7Zf(+).

Nucleotide sequence accession numbers. The nucleotide sequences for the *fliC* genes of *E. coli* H1, H7, and H12 have been entered into GenBank under accession numbers L07387, L07388, and L07389, respectively.

RESULTS AND DISCUSSION

PCR amplification of *E. coli fliC* sequences. Comparison of *fliC* nucleotide sequence data for *E. coli* K-12 and several *Salmonella* species demonstrates high similarity in the 5′ and 3′ regions of the *fliC* genes. The conservation is as high as 80% at the 5′ end of the gene and 60% at the 3′ end (24). In preliminary experiments, a gene probe from the conserved 5′ region of *E. coli fliC* hybridized to a single DNA fragment in *Eco*RI digests of chromosomal DNA from serotypes H1, H7, and H12 (44). The *E. coli* K-12 sequence data were therefore

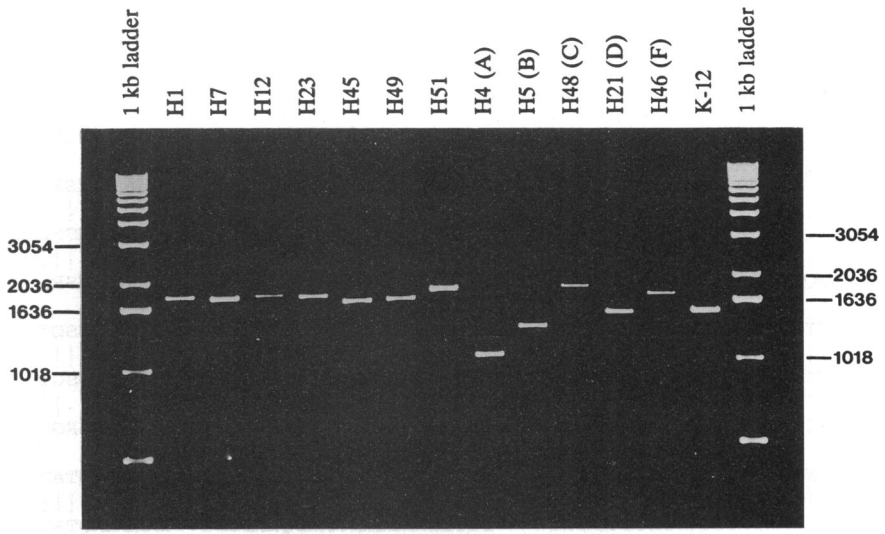

FIG. 1. Agarose gel showing PCR-amplified *fliC* DNA from *E. coli* strains. The outer lanes show size markers (Bethesda Research Laboratories), and the sizes (in base pairs) of relevant markers are indicated. The representative H serotype is indicated above each lane, and the strains are described in Table 1. Lanes 2 to 8 demonstrate amplification products from the *fliC* gene of all seven morphotype E *E. coli* strains. Lanes 9 to 13 demonstrate amplification products from the *fliC* gene of the representatives of the other five *E. coli* morphotypes (indicated in parentheses). Lane 14 demonstrates PCR-amplified DNA from plasmid pKS93 (formerly pBR322/hag93), which contains the cloned *fliC* sequence for *E. coli* K-12 (24).

used to design oligonucleotide primers for PCR amplification of *fliC* sequences from different H serotypes. As shown in Fig. 1, these oligonucleotide primers were used to amplify *fliC* sequences from *E. coli* strains with flagellar filaments representing all seven morphotype E H serotypes. In addition, *fliC* sequences were amplified from representatives of the other *E. coli* flagellar morphotypes. This result indicates that the termini of *fliC* genes from different *E. coli* serotypes are conserved. As expected, the sizes of the *fliC* PCR-amplified products correlate with the size profile of the respective FliC flagellin proteins in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Table 2).

TABLE 2. Relative sizes of PCR-amplified *fliC* DNA sequences and molecular masses of the corresponding flagellins

E. coli serotype or strain	Morphotype	Size of PCR product (kb)	Flagellin size (kDa)
H1	E	1.8	67[a] (60,857 Da)[b]
H7	E	1.8	67[a] (59,722 Da)[b]
H12	E	1.8	66[a] (60,978 Da)[b]
H23	E	1.8	67[a]
H45	E	1.7	61[a]
H49	E	1.7	60[a]
H51	E	1.9	65[a]
H4	A	1.1	37[c]
H5	B	1.4	46[c]
H48	C	1.9	67[d]
H21	D	1.5	56[c]
H46	F	1.7	56[c]
K-12		1.5	51[e]

[a] Apparent size from SDS-PAGE (60).
[b] Deduced from DNA sequence (this study).
[c] Apparent size from SDS-PAGE (29).
[d] Apparent size from SDS-PAGE (44).
[e] Deduced from DNA sequence (24).

The sizes of the FliC proteins in morphotypes A, B, C, D, E, and F also correlate with the diameter and surface pattern complexity of flagellar filaments from each of these morphotypes (28, 29).

E. coli K-12 has been reported to be serotype H48, a representative of morphotype C (29). PCR amplification of the *E. coli* K-12 *fliC* sequence from plasmid pKS93 (25), formerly known as pBR322/hag93 (24), resulted in production of the expected fragment of approximately 1.5 kb (Fig. 1). However, the PCR-amplified *fliC* DNA from the reference strain *E. coli* O16:K?:H48 (morphotype C) was approximately 1.9 kb in size. This size is expected because the flagellin from the H48 reference strain shows a molecular mass of 67 kDa in SDS-PAGE (Table 2). The basis for the anomaly between the H48 and K-12 flagellins is unclear.

H antigens in *Salmonella* strains are often diphasic because of phase variation between two chromosomal flagellin structural genes, which give rise to serologically distinct flagellins (reviewed in reference 30). An additional plasmid-encoded flagellin structural gene is responsible for the triphasic properties of some *Salmonella* serovars (49). *Salmonella* flagellin genes representing different phases show conservation in the 5′ and 3′ regions (37, 49). Ratiner (40) has suggested that many *E. coli* strains also have diphasic H antigens. In the strains tested here, we found no evidence for multiple flagellin genes; we have not examined the strains studied by Ratiner. Some sequence variation in the terminal regions of alternate-phase flagellin genes could result in the inability to amplify one of two flagellins. However, hybridization experiments with a gene probe from the predicted conserved 5′ region also detected only a single fragment in digests of chromosomal DNA from *E. coli* serotypes H1, H7, and H12 (44).

Analysis of amino acid sequences of flagellins from morphotype E filaments. To investigate relationships among FliC sequences, H1, H7, and H12 serotypes were chosen for detailed analysis based on immunological data (43). Previ-

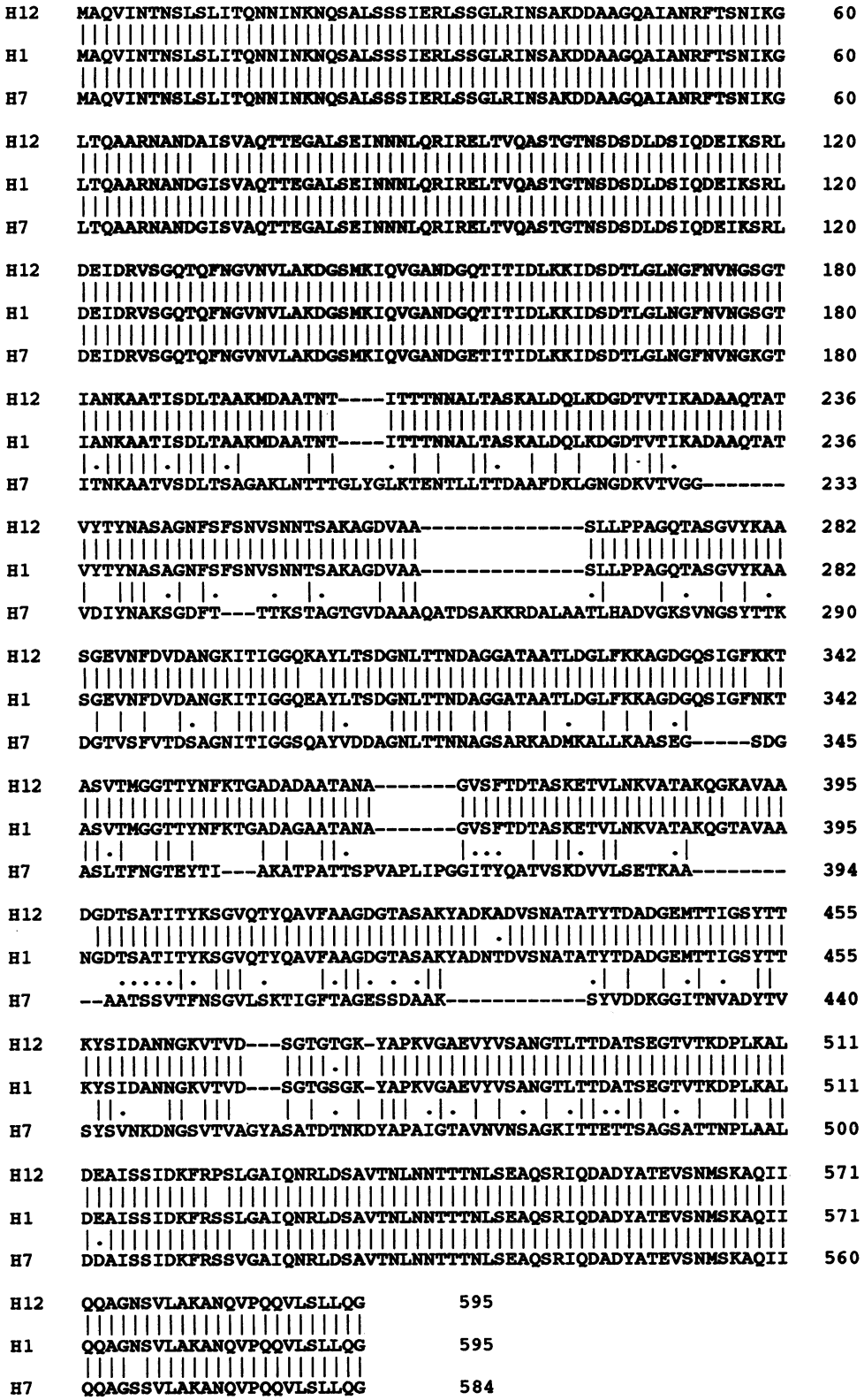

FIG. 2. Amino acid sequence comparison of the FliC proteins from *E. coli* serotypes O1:K2:H1, O1:K1:H7, and O9:K9:H12. H1 is used as the prototype, and the H7 and H12 sequences are compared with that of H1. Alignments were made with PALIGN (33). Identical amino acids are denoted by vertical bars, and similar amino acids are indicated by dots. The amino acids which are considered similar are A, S, and T; D and E; N and Q; R and K; I, L, M, and V; and F, Y, and W.

```
Escherichia coli H1          MAQVINTNSLSLITQNNINKNQSALSSSIERLSSGLRINSAKDDAAGQAIANRFTSNIKGLTQAARNAND   70
Escherichia coli H12         ************************************************************************
Escherichia coli H7          ************************************************************************
Escherichia coli K-12        ************************************************************************
Salmonella typhimurium (i)   ***********L****L**S****GTA*****************************A********S*****
Salmonella muenchen (d)      ***********L****L**S****GTA*****************************A********S*****
Salmonella rubislaw (r)      ***********L****L**S****GTA*****************************A********S*****
Salmonella paratyphi A (a)   ***********L****L**S****GTA*****************************A********S*****
Salmonella choleraesuis (c)  ***********L****L**S****GTA*****************************A********S*****
Serratia marcescens          ***********MA***L**S**S*GTA*********************S****A********S*****

Escherichia coli H1          GISVAQTTEGALSEINNNLQRIRELTVQASTGTNSDSDLDSIQDEIKSRLDEIDRVSGQTQFNGVNVLAK   140
Escherichia coli H12         A***********************************************************************
Escherichia coli H7          ************************************************************************
Escherichia coli K-12        ********************V*******T******E***S*****************************
Salmonella typhimurium (i)   ***I********N********V***A**SANS***Q*******A**TQ**N*****N*****S**K***Q
Salmonella muenchen (d)      ***I********N********V***A**SAN****Q*******A**TQ**N**************K***Q
Salmonella rubislaw (r)      ***I********N********V***A**SANS***Q*******A**TQ**N**************K***Q
Salmonella paratyphi A (a)   ***I********N********V***A**SANS***Q*******A**TQ**N**************K***Q
Salmonella choleraesuis (c)  ***I********N********V***A**SANS***Q*******A**TQ**N**************K***Q
Serratia marcescens          ***L********N*V*D***N**R*****QN*S**T***K******TQ**S**N*I*E**D****K**SS

Escherichia coli H1          DGSMKIQVGANDGQTITIDLKKIDSDTLGLNGFNVNGSGTIANKAATISD     190
Escherichia coli H12         ************************************************
Escherichia coli H7          ************E*********************K***T*****V**
Escherichia coli K-12        N**********N********Q**AK****D**S*KNND*VTTS*PVTAF
Salmonella typhimurium (i)   *NTLT********E**D****Q*N*Q****DTL**QQKYKVSDTA**VTG
Salmonella muenchen (d)      *NTLT********E**D****E*S*K****DKL**QDAY*PKET*V*VDK
Salmonella rubislaw (r)      *NTLT********E**D****Q*N*Q****DTL**QQKYKVSDT***VTG
Salmonella paratyphi A (a)   *NTLT********E**D****Q*N*Q****DTL**QKKYDVKSEVTPSAT
Salmonella choleraesuis (c)  *NTLT********E**D****Q*N*Q****DTL**QKKYDVSDT*VAA*Y
Serratia marcescens          *QKLT********E*TD*******AKQ**MDT*D*TTKSAK*GAEIATGT
```

FIG. 3. Alignment of the first 190 amino acids of the protein sequences of FliC from *E. coli* serotypes O2:K2:H1, O1:K1:H7, and O9:K9:H12 and *E. coli* K-12 (24). The FliC sequence from *E. coli* O2:K2:H1 was used as the template for comparison. Identity with the O2:K2:H1 FliC sequence is indicated by an asterisk. Included in the comparison are the predicted amino acid sequences for flagellins from *Salmonella typhimurium fliC*(*i*) (from reference 15, as corrected in references 21 and 47), *Salmonella muenchen fliC*(*d*) (58), *Salmonella rubislaw fliC*(*r*) (59), *Salmonella paratyphi* A *fliC*(*a*) (58), *Salmonella choleraesuis fliC*(*c*) (58), and *S. marcescens* (10).

ously, we reported that H1 and H12 flagellar filaments showed extensive cross-reaction with anti-H12 polyclonal antiserum and shared an epitope (recognized by a monoclonal antibody) that was absent in other morphotype E flagellins. In contrast, serotype H7 showed no significant cross-reaction with H1 and H12 with the anti-H7 polyclonal serum. All *E. coli* flagellins have at least one common epitope (5, 43).

To facilitate DNA sequence analysis, the H1, H7, and H12 *fliC* sequences were amplified by using primers designed with *Eco*RI restriction sites at the 5′ ends and cloned in pGEM7Zf(+). The absence of *Eco*RI sites within the *fliC* genes was previously established by Southern blotting (44). The H1 and H12 *fliC* genes were identical in size (1,788 bp). The deduced H1 and H12 FliC proteins were very similar in primary structure, with only 10 amino acid differences (Fig. 2). Both showed comparable predicted molecular masses of 60,857 Da (H1) and 60,987 Da (H12). The overall similarities in these proteins are reflected in the immunological cross-reactions reported previously (43). Any of the specific amino acid differences in the central regions of the H1 and H12 FliC proteins could be responsible for the serotype-specific epitopes.

In contrast, the H7 *fliC* sequence is smaller (1,755 bp) than those of H1 and H12, as is the predicted protein (59,722 Da). The deduced H7 FliC protein shows near identity to the H1 and H12 flagellins at the termini. The N-terminal conserved region extends over the first 190 amino acid residues (Fig. 3). The C-terminal conserved domain occupies approximately 100 amino acids (Fig. 4). The central variable region of the H7 flagellin is poorly conserved. It was necessary to introduce breaks in the FliC central region sequences of H1 (residues 190 to 517) and H7 (residues 190 to 506) to generate an optimal alignment (Fig. 2). This analysis resulted in 38% identity and 14.6% similarity in the central regions of H1 and H7. It is therefore clear that the typical and complex morphotype E surface pattern on H1, H7, and H12 flagellar filaments is not dependent on a highly conserved primary sequence in the central region of FliC.

H7 flagellin was 11 amino acids shorter than both H1 and H12, and the size difference did not affect filament diameter or architecture (29, 60). Kuwajima (22, 23) showed that the central variable region of the flagellin from *E. coli* K-12 can be deleted without detrimental effects on flagellin function. However, deletion of this region does affect H antigenicity and might be expected to affect filament morphology, since the central region is surface exposed in intact filaments (4, 14, 23, 34, 35, 51–53, 55). It is conceivable that the morpho-

```
Escherichia coli H1                    495   DATSEGTVTKDPLKALDEAISSIDKFRSSLGAIQNRLDSAVTNLNNTTTNLSEAQSRIQD
Escherichia coli H12                   495   ****************************P*********************************
Escherichia coli H7                    484   ET**A*SA*TN**A***D**********V*********************************
Escherichia coli K-12                  398   LTAVANGK*T*******D**A*V*********V*********************************
Salmonella typhimurium (i)             394   L*EAAA*T*EN**QKI*A*LAQV*TL**D***V***FN**I***G**VN***S*R***E*
Salmonella muenchen (d)                405   LKEVNTDK*EN**QKI*A*LAQV*TL**D***V***FN**I***G**VN***S*R***E*
Salmonella rubislaw (r)                393   L*EAAA*T*EN**QKI*A*LAQV*TL**D***V***FN**I***G**VN**TS*R***E*
Salmonella paratyphi A (a)             394   L*EAAAAT*EN**AKI*A*LAQV*AV**D***V***FN**I***G**VN***S*R***E*
Salmonella choleraesuis (c)            389   L*ERAA*T*EN**QKI*A*LAQV*AL**D***V***FN**I***G**VN***S*R***E*
Serratia marcescens                    351   TDSGTDAGV*N**AT**K*LAQV*GL******V***F**VIN***S*VN***AS******

Escherichia coli H1                    ADYATEVSNMSKAQIIQQAGNSVLAKANQVPQQVLSLLQG   595
Escherichia coli H12                   ****************************************   595
Escherichia coli H7                    ********************S*******************   584
Escherichia coli K-12                  ****************************************   498
Salmonella typhimurium (i)             S**********R***L****T****Q******N*****R   494
Salmonella muenchen (d)                S**********R***L****T****Q******N*****R   505
Salmonella rubislaw (r)                S**********R***L****T****Q******N*****R   493
Salmonella paratyphi A (a)             S**********R***L****T****Q******N*****R   494
Salmonella choleraesuis (c)            S**********R***L****T****Q******N*****R   489
Serratia marcescens                    **********R*N*L****T****Q***ST*N*****R   351
```

FIG. 4. Alignment of the last 100 amino acids of the protein sequences for FliC from *E. coli* serotypes O2:K2:H1, O1:K1:H7, and O9:K9:H12 and *E. coli* K-12 (24). The FliC sequence from *E. coli* O2:K2:H1 was used as the template for comparison. Identity with the O2:K2:H1 FliC sequence is indicated by an asterisk. Included in the comparison are the predicted amino acid sequences for flagellins from *Salmonella typhimurium fliC(i)* (from reference 15 as corrected in references 21 and 47), *Salmonella muenchen fliC(d)* (58), *Salmonella rubislaw fliC(r)* (59), *Salmonella paratyphi* A *fliC(a)* (58), *Salmonella choleraesuis fliC(c)* (58), and *S. marcescens* (10).

type E filament pattern is more dependent on the size of the variable flagellin region than a specific primary sequence. The increased size would provide a larger domain extending out from the surface of the flagellar filament. This could cause localized trapping of negative stain and give rise to the features characteristic of morphotype E seen in electron micrographs (29, 60).

A single point mutation in a *fliC* sequence can potentially result in the generation of novel H antigens (19, 21, 36, 45, 58), and this mechanism may have contributed to the origin of serotypes such as H1 and H12, which differ by only a few amino acids. However, Smith and Selander (48) showed that the sequences of the central regions of *fliC(i)* genes from *Salmonella* strains with different genotype backgrounds were invariant. These authors suggested that the highly variable central region of the *Salmonella fliC* sequence may also be under a great deal of selective pressure and is probably not subject to mutation rates higher than the mutation rate found in parts of other genes, such as the *trp* genes of *E. coli*. It has been proposed that the great diversity of H antigens may be less dependent on accumulation of mutations in the central variable region and that new serotypes more likely arise from lateral gene transfer and recombination of DNA within *fliC* genes (47, 48). Whether the large differences seen between the H1/H12 and H7 flagellins reflect a similar phenomenon remains to be established.

Comparison of morphotype E FliC proteins with other enterobacterial flagellins. The central regions of *E. coli* H1, H7, and H12 flagellins could not be aligned with those from other enteric bacteria. In contrast, the FliC terminal regions showed substantial identity (Fig. 3 and 4). Conservation of the terminal regions of enterobacterial flagellin proteins has been reported previously (10, 15, 24, 58, 59). Wei and Joys (58) suggested that conservation in the ends of FliC molecules reflected constraints imposed on flagellins by essential assembly and regulatory roles. This is consistent with the observation that a single amino acid change in FliC resulted in improper assembly or export of unassembled flagellin into the surrounding growth medium (21). Furthermore, other terminal-region mutations have been reported to result in straight, nonfunctional filaments (21, 31).

An alignment of the N-terminal 190 amino acids of known enterobacterial FliC proteins (Fig. 3) extends these observations to morphotype E flagellins. The highest degree of identity with morphotype E flagellins was observed with *E. coli* K-12 FliC, with greater than 87% identity. When the *Salmonella* flagellins were aligned, 74 to 76% identity was detected over the 190-amino-acid N-terminal region. The conservation at the N terminus breaks down after amino acid 170. *S. marcescens* was less related to H1 and H12, with 69% identity. Other bacterial flagellin sequences, including those from *Bacillus* (27, 32), *Caulobacter* (8), *Campylobacter* (9), *Pseudomonas aeruginosa* (50), *Rhizobium* (2, 38), and *Halobacterium* (7) spp., all showed less than 75% identity with *E. coli* morphotype E flagellins in the N-terminal region (44).

A C-terminal conserved region of 100 amino acids was also evident (Fig. 4). The *E. coli* H1 and H12 flagellins had 87% identity with the C-terminal region of the flagellin from *E. coli* K-12 and approximately 64 and 62% identity with those from the *Salmonella* spp. and *S. marcescens*, respectively.

The relative sizes of these conserved terminal regions reflect the relatedness of the flagellins. On this basis, the H1 and H12 flagellins are more closely related to each other than they are to H7 flagellin, and as expected, H1, H12, and H7 flagellins appear to be more closely related to each other than to flagellins of other members of the *Enterobacteriaceae*, such as *Salmonella* spp. (Fig. 3 and 4). It is clear from

the terminal alignments that the *S. marcescens* FliC is more closely related to *Salmonella* FliC than to *E. coli* FliC.

In conclusion, we have presented data indicating that the *E. coli* strains examined here are monophasic with respect to the flagellar H antigen. The flagellins which give rise to morphotype E filaments show highly conserved termini compared with flagellins from other *Enterobacteriaceae*. However, the characteristic surface pattern of the morphotype E filaments shows no requirement for a highly conserved primary sequence in the central exposed domain of the flagellin.

ACKNOWLEDGMENTS

We thank Frits and Ida Ørskov for generously providing bacterial strains, G. Kuwajima for graciously providing the cloned *E. coli* K-12 *fliC*, and Christine Dodgson for skilled technical assistance. We are especially grateful to R. Macnab for helpful discussions and for critically reading the manuscript.

This work was funded by a Natural Sciences and Engineering Research Council grant to C.W.

REFERENCES

1. **Austin, C. M., and G. J. V. Nossal.** 1966. Mechanism of induction of immunological tolerance. III. Cross-tolerance amongst flagellar antigens. Aust. J. Exp. Biol. Med. Sci. **44**:341–354.
2. **Bergman, K., E. Nulty, and L. H. Su.** 1991. Mutations in the two flagellin genes of *Rhizobium meliloti*. J. Bacteriol. **173**:3663–3672.
3. **Ewing, W. H. (ed.).** 1986. Edwards and Ewing's identification of *Enterobacteriaceae*, 4th ed. Elsevier Science Publishers, Amsterdam.
4. **Fedorov, O. V., A. S. Kostyukova, and M. G. Pyatibratov.** 1988. Architectonics of a bacterial flagellin filament subunit. FEBS Lett. **241**:145–148.
5. **Feng, P., R. J. Sugasawara, and A. Schantz.** 1990. Identification of a common enterobacterial flagellin epitope with a monoclonal antibody. J. Gen. Microbiol. **136**:337–342.
6. **Frankel, G., S. M. C. Newton, G. K. Schoolnik, and B. A. D. Stocker.** 1989. Intragenic recombination in a flagellin gene: characterization of the *H1-j* gene of *Salmonella typhi*. EMBO J. **8**:3149–3152.
7. **Gerl, L., and M. Sumper.** 1988. Halobacterial flagellins are encoded by a multigene family. J. Biol. Chem. **263**:13246–13251.
8. **Gill, P. R., and N. Agabian.** 1983. The nucleotide sequence of the Mr = 28,500 flagellin gene of *Caulobacter crescentus*. J. Biol. Chem. **258**:7395–7401.
9. **Guerry, P., R. A. Alm, M. E. Power, S. M. Logan, and T. J. Trust.** 1991. Role of two flagellin genes in *Campylobacter* motility. J. Bacteriol. **173**:4757–4764.
10. **Harshey, R. M., G. Estepa, and H. Yanagi.** 1989. Cloning and nucleotide sequence of a flagellin-coding gene (*hag*) from *Serratia marcescens* 274. Gene **79**:1–8.
11. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33**:933–938.
12. **Ibrahim, G. F., G. H. Fleet, M. J. Lyons, and R. A. Walker.** 1985. Immunological relationships between *Salmonella* flagella and their potential application for salmonellae detection by immunoassay. Med. Microbiol. Immunol. **174**:87–99.
13. **Ibrahim, G. F., G. H. Fleet, M. J. Lyons, and R. A. Walker.** 1985. Immunological relationships between *Salmonella* flagellins and between these and flagellins from other species of Enterobacteriaceae. Med. Microbiol. Immunol. **174**:101–113.
14. **Iino, T.** 1977. Genetics of structure and function of bacterial flagella. Annu. Rev. Genet. **11**:161–182.
15. **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.
16. **Joys, T. M.** 1976. Identification of an antibody binding site in the phase-1 flagellar protein of *Salmonella typhimurium*. Microbios **15**:221–228.
17. **Joys, T. M.** 1988. The flagellar filament protein. Can. J. Microbiol. **34**:452–458.
18. **Joys, T. M., and J. F. Martin.** 1973. Identification of amino acid changes in serological mutants of the *i* flagellar antigen of *Salmonella typhimurium*. Microbios **7**:71–73.
19. **Joys, T. M., and F. Schödel.** 1991. Epitope mapping of the *d*-flagellar antigen of *Salmonella muenchen*. Infect. Immun. **59**:3330–3332.
20. **Joys, T. M., and B. A. D. Stocker.** 1966. Isolation and serological analysis of mutant forms of flagellar antigen *i* of *Salmonella typhimurium*. J. Gen. Microbiol. **44**:121–138.
21. **Kanto, S., H. Okino, S.-I. Aizawa, and S. Yamaguchi.** 1991. Amino acids responsible for flagellar shape are distributed in terminal regions of flagellin. J. Mol. Biol. **219**:471–480.
22. **Kuwajima, G.** 1988. Construction of a minimum-size functional flagellin of *Escherichia coli*. J. Bacteriol. **170**:3305–3309.
23. **Kuwajima, G.** 1988. Flagellin domain that affects H antigenicity of *Escherichia coli* K-12. J. Bacteriol. **170**:485–488.
24. **Kuwajima, G., J.-I. 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.
25. **Kuwajima, G., I. Kawagishi, M. Homma, J.-I. Asaka, E. Kondo, and R. M. Macnab.** 1989. Export of an N-terminal fragment of *Escherichia coli* flagellin by a flagellum-specific pathway. Proc. Natl. Acad. Sci. USA **86**:4953–4957.
26. **Langman, R. E.** 1972. The occurrence of antigenic determinants common to flagella of different *Salmonella* strains. Eur. J. Immunol. **2**:582–586.
27. **LaVallie, E. R., and M. L. Stahl.** 1989. Cloning of the flagellin gene from *Bacillus subtilis* and complementation studies of an in vitro-derived deletion mutation. J. Bacteriol. **171**:3085–3094.
28. **Lawn, A. M.** 1977. Comparison of the flagellins from different flagellar morphotypes of *Escherichia coli*. J. Gen. Microbiol. **101**:121–130.
29. **Lawn, A. M., I. Ørskov, and F. Ørskov.** 1977. Morphological distinction between different H serotypes of *Escherichia coli*. J. Gen. Microbiol. **101**:111–119.
30. **Macnab, R. M.** 1992. Genetics and biosynthesis of bacterial flagella. Annu. Rev. Genet. **26**:129–156.
31. **Martinez, R. J., A. T. Ichiki, N. P. Lundh, and S. R. Tronick.** 1968. A single amino acid substitution responsible for altered flagellar morphology. J. Mol. Biol. **34**:559–564.
32. **Mirel, D. B., and M. J. Chamberlin.** 1989. The *Bacillus subtilis* flagellin gene (*hag*) is transcribed by the σ^{28} form of RNA polymerase. J. Bacteriol. **171**:3095–3101.
33. **Myers, E. W., and W. Miller.** 1988. Description of the alignment method used in program PALIGN. CABIOS **4**:11–17.
34. **Namba, K., I. Yamashita, and F. Vonderviszt.** 1989. Structure of the core and central channel of bacterial flagella. Nature **342**:648–654.
35. **Newton, S. M. C., C. O. Jacob, and B. A. D. Stocker.** 1989. Immune response to cholera toxin epitope inserted in *Salmonella* flagellin. Science **244**:70–72.
36. **Newton, S. M. C., R. D. Wasley, A. Wilson, L. T. Rosenberg, J. F. Miller, and B. A. D. Stocker.** 1991. Segment IV of a *Salmonella* flagellin gene specifies flagellar antigen epitopes. Mol. Microbiol. **5**:419–425.
37. **Okazaki, N., S. Matsuo, K. Saito, A. Tominaga, and M. Enomoto.** 1993. Conversion of the *Salmonella* phase 1 flagellin gene *fliC* to the phase 2 gene *fliB* on the *Escherichia coli* K-12 chromosome. J. Bacteriol. **175**:758–766.
38. **Pleier, E., and R. Schmitt.** 1989. Identification and sequence analysis of two related flagellin genes in *Rhizobium meliloti*. J. Bacteriol. **171**:1467–1475.
39. **Qadri, A., S. Ghosh, S. Upadhyay, and G. P. Talwar.** 1989. Monoclonal antibodies against flagellar antigen of *Salmonella typhimurium*. Hybridoma **8**:353–360.
40. **Ratiner, Y. A.** 1987. Different alleles of the flagellin gene *hagB* in *Escherichia coli* standard H test strains. FEMS Microbiol. Lett. **48**:97–104.
41. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular

cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
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. **Schoenhals, G., and C. Whitfield.** 1990. Monoclonal antibodies against serotype specific and conserved epitopes in morphotype E *Escherichia coli* flagellins. FEMS Microbiol. Lett. **72**:117–122.
44. **Schoenhals, G., and C. Whitfield.** Unpublished results.
45. **Selander, R. K., N. H. Smith, J. Li, P. Beltran, K. E. Ferris, D. J. Kopecko, and F. A. Rubin.** 1992. Molecular evolutionary genetics of the cattle-adapted serovar *Salmonella dublin*. J. Bacteriol. **174**:3587–3592.
46. **Shirakihara, Y., and T. Wakabayashi.** 1977. Three-dimensional image reconstruction of straight flagella from a mutant *Salmonella typhimurium*. J. Mol. Biol. **131**:485–507.
47. **Smith, N. H., P. Beltran, and R. K. Selander.** 1990. Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. J. Bacteriol. **172**:2209–2216.
48. **Smith, N. H., and R. K. Selander.** 1990. Sequence invariance of the antigen-coding central region of the phase 1 flagellar filament gene (*fliC*) among strains of *Salmonella typhimurium*. J. Bacteriol. **172**:603–609.
49. **Smith, N. H., and R. K. Selander.** 1991. Molecular genetic basis for complex flagellar antigen expression in a triphasic serovar of *Salmonella*. Proc. Natl. Acad. Sci. USA **88**:956–960.
50. **Totten, P. A., and S. Lory.** 1990. Characterization of the type *a* flagellin gene from *Pseudomonas aeruginosa* PAK. J. Bacteriol. **172**:7188–7199.
51. **Trachtenberg, S., and D. J. DeRosier.** 1987. Three-dimensional structure of the frozen-hydrated flagellar filament: the left-handed filament of *Salmonella typhimurium*. J. Mol. Biol. **195**:581–601.
52. **Trachtenberg, S., and D. J. DeRosier.** 1988. Three-dimensional reconstruction of the flagellar filament of *Caulobacter crescentus*. A flagellin lacking the outer domain and its amino acid sequence lacking an internal segment. J. Mol. Biol. **202**:787–808.
53. **Trachtenberg, S., and D. J. DeRosier.** 1991. A molecular switch: subunit rotations involved in the right-handed to left-handed transitions of *Salmonella typhimurium* flagellar filaments. J. Mol. Biol. **220**:67–77.
54. **Trachtenberg, S., and D. J. DeRosier.** 1992. Conformational switching in the flagellar filament of *Salmonella typhimurium*. J. Mol. Biol. **226**:447–454.
55. **Vonderviszt, F., S. I. Aizawa, and K. Namba.** 1991. Role of the disordered terminal regions of flagellin in filament formation and stability. J. Mol. Biol. **221**:1461–1474.
56. **Vonderviszt, F., S. Kanto, S.-I. Aizawa, and K. Namba.** 1989. Terminal regions of flagellin are disordered in solution. J. Mol. Biol. **209**:127–133.
57. **Vonderviszt, F., H. Uedaira, S. I. Kidokoro, and K. Namba.** 1990. Structural organization of flagellin. J. Mol. Biol. **214**:97–104.
58. **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.
59. **Wei, L.-N., and T. M. Joys.** 1986. The nucleotide sequence of the H-1*r* gene of *Salmonella rubislaw*. Nucleic Acids Res. **14**:8227.
60. **Whitfield, C., S. G. Walker, C. F. Atkinson, J. S. Lam, L. A. MacDonald, T. J. Beveridge, I. Ørskov, and F. Ørskov.** 1988. Serotype-specific monoclonal antibodies against the H12 flagellar antigen of *Escherichia coli*. J. Gen. Microbiol. **134**:1747–1753.