JOURNAL OF BACTERIOLOGY, Sept. 1993, p. 5359–5365
0021-9193/93/175359-07$02.00/0

Vol. 175, No. 17

Molecular Analyses of the *Salmonella g. . .* Flagellar Antigen Complex

BARBARA J. MASTEN AND TERENCE M. JOYS*

Department of Microbiology, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas 79430

Received 25 February 1993/Accepted 29 April 1993

Salmonella **flagellar filaments are polymers of a highly antigenic protein, termed flagellin. Eight main subfactors have been identified in the *Salmonella* phase-1 *g. . .* series flagellar antigen. To determine the molecular basis for expression of the epitopes by which the *g. . .* family subfactors are distinguished, 10 members of this series were selected and their *fliC* (the structural gene for phase-1 flagellin) genes were sequenced. Comparative analyses of the inferred primary structures of these flagellins did not allow the identification of linear epitopes responsible for the antigen subfactors. This suggests that conformational aspects are involved in determining the antigenic specificity in these cases. A phylogenetic analysis of the flagellin sequences showed that members of the *g. . .* series do not form a single coherent unit.**

Most bacteria of the genus *Salmonella* are motile by means of flagella. DePamphilis and Adler (3) defined three regions of the flagellar organelle. The complex basal body, curved hook, and helical filament function as a rotary motor, universal joint, and propeller, respectively (19). The filament is a self-assembling polymer of thousands of molecules (26) of a single protein, termed flagellin (1), specified by either the *fliC* (phase-1) gene or the *fljB* (phase-2) gene (10). Previous reports from our laboratory have shown that *Salmonella* flagellins with antigens *a*, *c*, and *d* (41), *i* (11), and *r* (42) consist of extremely conserved terminal regions and a variable center section. The central area embodies a hypervariable segment, region IV, of ca. 120 amino acids (41). Wei and Joys (41) found that the predicted amino acid sequence of region IV showed no greater than 30% homology for any pairwise comparison between alleles and suggested that this region was not involved in flagellar localization or function, but contributed to antigenic specificity. By chemical and immunological analyses on *f,g* antigen flagellin, Parish et al. (30) demonstrated that all the antigenic specificities reside in the central region of the flagellin polypeptide. From a purely genetic approach, Iino (9) provided evidence that the central area of the molecule contained the antigenically variable regions. A previously identified epitope of flagellar antigen *i* (12) and two major epitopes of antigen *d* (13) were located within region IV. These properties led others (27, 28, 43) to suggest that flagella might be useful in vaccine development if part of the hypervariable region could be substituted with a medically important epitope-specifying oligonucleotide, in the correct orientation and reading frame, resulting in exposure of the epitope at the surface of the flagella. To date, such substitutions have been limited to a unique restriction site in the structural gene for the flagellin antigen *i* of *S. typhimurium* (27) and to a 48-bp segment in region IV (41) of the structural gene for the flagellar antigen *d* of *S. muenchen* (2, 20, 24, 28, 29, 38, 39, 43), which is readily excised by restriction enzyme *Eco*RV. These attempts have been only partially successful. The primary problem seems to be that substitutions in flagellin result in an inability of the molecules to polymerize into flagellar filaments. It occurred to one of us (T.M.J.) that an examination of other flagellins may reveal more suitable molecules for interchange with known epitopes. The *d* and *i* flagellins investigated to date are both defined (4, 15, 17) by their production of a single major flagellar antigen. Other flagellar antigens have been divided into subfactors (4, 15, 17), and it would seem a priori that the existence of such natural variation might reveal a site at which manipulation could be attempted with less effect on the properties of the molecule. In the Kauffmann-White scheme (15), eight major subfactors (*f, g, m, p, q, s, t,* and *u*) of the *Salmonella* phase-1 *g* antigen were described, and it has been shown that the factor *g* itself is a complex composed of two or more of at least five subfactors, *g*1 to *g*5 (44). On the basis of amino acid composition, McDonough (22) determined that members of the *g. . .* series (selected serovars with flagellar antigens *g,m*; *g,p*; and *g,s,t*) showed greater variability than single-factor flagellins. For this reason, we selected flagellins from the *g. . .* series of *Salmonella* flagellar antigens in the expectation that the subfactors that separate the different serovars would be located in a common region that would make an efficient site for directed substitutions. Here we report the comparison of the amino acid sequences for the products of 10 *fliC* genes that specify the phase-1 antigens: *f,g,t* (*S. berta*); *g,t* (*S. budapest*); *f,g* (*S. derby*); *g,p* (*S. dublin*); *g,m* (*S. enteritidis*); *g,m,s* (*S. montevideo*); *g,q* (*S. moscow*); *m,t* (*S. oranienberg*); *g,p,u* (*S. rostock*); and *g,s,t* (*S. senftenberg*). Although *S. oranienberg* flagellin does not express the *g* antigen, this serovar is grouped with the *g. . .* series due to its expression of the common *m* and *t* antigens. As can be readily seen from the data presented, the locations of the disparities detected differed from our expectations in that they spanned the whole central portion (regions IV, V, VI, and VII) of the flagellin and were not localized to one common area.

(Portions of this research were presented at the 1993 Keystone Symposia Conference on Emerging Principles for Vaccine Development: Antigen Processing and Presentation. The total work will be submitted by Barbara Masten to Texas Tech University Health Sciences Center in partial fulfillment of the requirements for the degree of Ph.D.)

* Corresponding author.

TABLE 1. Nucleotide sequences of the *fliC g. . .* series primers used[a]

Primer	Sequence	5′ Position in bp	Reference or source
Sense			
PCR1S	5′-TGCAAACAGTAGTTAAGCGC-3′	−241	40
BM1S	5′-TGAGAAACCGTGGCAACAGC-3′	−61	40
BM2S	5′-TTGAGCGTCTGTCCTCTGGTCTGC-3′	83	This study
BM3S	5′-CCGCTTCACTTCTAATATCAAAGG-3′	153	This study
BM4S	5′-TTCCGATCTGAAATCTATCCAGGA-3′	315	This study
BM5S	5′-ACAGTGGGTGATCTGAAATCCAGC-3′	526	This study
BM6S	5′-GTGATCTGAAATCCAGCTTCAAGA-3′	533	This study
BM7S	5′-TACTGCTGGTACCGCTGAAGCCAA-3′	738	This study
BM8S	5′-CTGATTTGGAAGCAAACAATGCTG-3′	1,031	This study
PCR2S	5′-CUACUACUACUATGCAAACAGTAGTTAAGCGC-3′	−241	40
Antisense			
PCR1AS	5′-CATTACACAGGCTCCGGAAT-3′	−84	42a
BM1AS	5′-TTGATTGTGTACCACGTGTCGGT-3′	−41	42a
BM2AS	5′-CGTATTGCCAAGGTTGGTAATGGC-3′	157	This study
BM3AS	5′-TGCCAGCTAAGGTGACCTTATCAC-3′	351	This study
BM4AS	5′-GGTGACCTTATCACCCGCGGCGTT-3′	361	This study
BM5AS	5′-CTGACCGTTCACTACAGATGTATA-3′	493	This study
BM6AS	5′-GTGTTGTTCTCCGCATCGTCAGTT-3′	773	This study
BM7AS	5′-GTGTTATTTTGCGCATCGGCAGTT-3′	773	This study
BM8AS	5′-AATGGTTTCACCATCGTTAGCACC-3′	1,016	This study
PCR2AS	5′-CUACUACUACUACATTACACAGGCTCCGGAAT-3′	−84	42a

[a] 5′ Positions refer to the equivalent 5′ base positions previously reported for the *fliC* flagellin with antigen *a* (41).

MATERIALS AND METHODS

Bacterial strains and plasmids. *S. berta* ATCC 8392 (*fliC*^{f,g,t}), *S. derby* ATCC 6960 (*fliC*^{f,g}), *S. dublin* ATCC 15480 (*fliC*^{g,p}), *S. enteritidis* ATCC 13076 (*fliC*^{g,m}), *S. montevideo* ATCC 8387 (*fliC*^{g,m,s}), *S. oranienberg* ATCC 9239 (*fliC*^{m,t}) and *S. senftenberg* ATCC 8400 (*fliC*^{g,s,t}) were obtained from the American Type Culture Collection (Rockville, Md.). *S. budapest* CDC 23 (*fliC*^{g,t}), *S. moscow* CDC 67 (*fliC*^{g,q}) and *S. rostock* CDC 66 (*fliC*^{g,p,u}) were obtained from the Centers for Disease Control (Atlanta, Ga.). *Escherichia coli* DH5α (F$^-$ *endA1 hsdR17 supE44* λ$^-$ *thi-1 gryA96 relA1*) (GIBCO BRL, Grand Island, N.Y.) was used as the initial host for recombinant plasmids. Plasmid pAMP^{TM}1 (Gibco BRL) was used as the vector in transformation experiments involving salmonellar flagellin gene cloning. *E. coli* C600 *hag*, labeled LC2a in our laboratory, was a generous gift from M. Simon (California Institute of Technology, Pasadena, Calif.) and was used as a second host for recombinant plasmids.

Media. *Salmonella* serovars were cultured in 2YT broth (1.6% [wt/vol] Bacto tryptone (Difco Laboratories, Detroit, Mich.), 1.0% (wt/vol) Bacto yeast extract (Difco), and 1.0% NaCl (Sigma Chemical Co., St. Louis, Mo.). Cells from the transformation experiments were plated on 2YT plates (1.5% [wt/vol] Bacto agar [Difco] in 2YT broth) containing 100 μg of ampicillin (Sigma) per ml. LC2a cells transformed with pAMP vector-annealed polymerase chain reaction (PCR) product were selected on the basis of motility in semisolid medium (0.2% [wt/vol] Bacto agar and 100 μg of ampicillin per ml in 2YT broth). The serological identities of the *fliC* antigens were confirmed with semisolid medium (0.4% [wt/vol] Bacto agar in 2YT broth) containing factor-specific antisera (for *Salmonella* serovars) or factor-specific antisera and 100 μg of ampicillin per ml (for clones).

Chromosomal amplification by the PCR. Chromosomal DNA, as a substrate for the PCR (31, 32), was prepared from cultures grown overnight in 2YT medium by use of the GENOME DNA Isolation Kit (BIO 101 Inc., La Jolla, Calif.) followed by phenol-chloroform extraction. The DNA samples (1 μl) were quantified by electrophoresis in a 0.75% agarose gel along with a lamda DNA *Hin*dIII fragment (GIBCO BRL) sample and pUC18 *Hin*dIII/bacterial alkaline phosphatase (Pharmacia, Piscataway, N.J.) standards. For amplification of the *fliC* gene for sequencing, the GeneAmp PCR Reagent Kit (Perkin-Elmer Cetus, Norwalk, Conn.) was used along with 20-mer primers (Table 1), synthesized by Midland Certified Reagent Co. (Midland, Tex.), derived from the upstream and downstream regions of the *fliC* gene: sense primer PCR1S, −241 bp upstream (40), and antisense primer PCR1AS, −84 bp downstream (42a). For amplification of the *fliC* gene for cloning, primers PCR2S and PCR2AS, which contain the same sequences as primers PCR1S and PCR1AS, respectively, but with the addition of CUACUACUACUA to their 5′ ends, were used. Conditions for amplification were as recommended by the manufacturer. The amplified PCR products (5 μl) were analyzed by electrophoresis in a 0.75% agarose gel. After amplification, each sample was purified with the Magic PCR Preps DNA Purification System (Promega, Madison, Wis.) by following the manufacturer's instructions. The purified PCR products were quantified by electrophoresis in a 0.75% agarose gel.

Cloning PCR products. Unless indicated otherwise, all recombinant DNA techniques used were those described by Sambrook et al. (33). The *fliC* gene was originally cloned by using the CloneAmp System (GIBCO BRL) with Subcloning Efficiency DH5α competent cells (GIBCO BRL) according to the manufacturer's instructions. Cells from the transformation experiment were plated on 2YT plates containing 100 μg of ampicillin per ml. Colonies were selected and analyzed for plasmid insert size by rapid plasmid preparation by using the Miniprep Kit Plus (Pharmacia) followed by double digestion with *Sma*I (Pharmacia) and *Bam*HI (Pharmacia). Plasmid DNA containing the insert was isolated by using the Miniprep Kit Plus (Pharmacia) and transformed into LC2a, a

nonflagellated strain of *E. coli*. The transformed cells were plated on semisolid 2YT plates containing 2 mg of agar per ml and 100 μg of ampicillin per ml. LC2a cells containing the vector-annealed PCR product were selected on the basis of motility in the semisolid medium.

Screening motile clones. The identity of each flagellar antigen was confirmed by the motility inhibition test (7), in which bacterial motility is inhibited in semisolid medium by the incorporation of factor-specific antiserum. Antisera were raised (Berkeley Antibody Co., Richmond, Calif.) in rabbits with formalinized broth cultures of flagellated strains by using standard techniques (5) and titrated with a microagglutination test (14). To obtain factor-specific antisera, antisera were absorbed by established methods (5) and titrated.

Sequencing of PCR products and clones. Purified PCR products were directly sequenced by using the fmol DNA Sequencing System (Promega) with walking primers (Table 1) end-labeled with [γ-^{32}P]ATP (New England Nuclear, Boston, Mass.). Clone DNA for sequencing was isolated from transformed LC2a by using the QIAGEN Plasmid Kit (Qiagen, Chatsworth, Calif.). Clones were sequenced by using the dsDNA Cycle Sequencing System (GIBCO BRL) with end-labeled primers (Table 1) in accordance with the manufacturer's instructions. The nucleotide sequences of each PCR product and clone DNA were determined in both orientations. The following starting primers (Midland Certified Reagent Co.) were derived from the upstream and downstream regions of the *fliC* gene: sense primer BM1S, −61 bp upstream (40), and antisense primer BM1AS, −41 bp downstream (42a). Progressive walking primers (Midland Certified Reagent Co.), BM2S to BM8S and BM2AS to BM8AS, were derived as sequences were determined.

Data analysis. The sequences, sense and antisense, obtained from the autoradiograms were entered and stored in spreadsheet form by using Microsoft Excel (Microsoft Corp., Redmond, Wash.). The Mount and Conrad (25) program was used to make a complementary DNA strand from the antisense strand data and to make a translation product from the sense strand data. By using the Pearson HMATCH executable file, which is part of the Mount and Conrad program, the complementary strand obtained from the antisense data was compared to the sense strand data to confirm agreement. By using the sense strand data, nucleic acid alignments were accomplished with the Pearson HMATCH executable file. Amino acid alignments were accomplished by using the Lipman and Pearson FASTP (18) program. An MTRANS (21) program was used to transfer files (sense strand data or translation products) from a Mount and Conrad format to an Excel spreadsheet format for nucleotide or amino acid comparisons based on obtained alignment data. By using the PHYLIP (phylogeny inference package) version 3.4 PROTPARS (protein sequence parsimony method) (6) program, an unrooted phylogeny tree was obtained from protein sequences.

Nucleotide sequence accession numbers. The AUTHORIN (Intelligenetics/GenBank, Los Alamos, N.M.) program was used to submit the nucleotide sequences of the *fliC* flagellar proteins to GenBank (Los Alamos, N.M.). The sequences were ultimately deposited in the EMBL data library (Heidelberg, Germany) with the following accession numbers: *S. berta*, Z15064 (BERTA); *S. budapest*, Z15065 (BUDA); *S. derby*, Z15066 (DERBY); *S. dublin*, Z15067 (DUB); *S. enteritidis*, Z15068 (ENTER); *S. montevideo*, Z15069 (MONT); *S. oranienberg*, Z15070 (ORAN); *S. rostock*, Z15071 (ROS); *S. senftenberg*, Z15072 (SENF); and *S. moscow*, Z15086 (MOS).

RESULTS

PCR amplification, cloning, and sequencing of *fliC*. Genomic DNA from the 10 *g. . .* series members listed was used as a substrate for PCR reactions with primers (Table 1) derived from the the upstream (40) and downstream (42a) regions of the *fliC* gene. Both PCR primers are specific for the *fliC* gene, do not amplify the *fljB* gene, and amplify a single PCR product. Purified PCR products primed with PCR1S and PCR1AS were directly sequenced (in both directions) by using progressive walking primers (Table 1) derived as sequences were determined. Purified PCR products primed with PCR2S and PCR2AS were cloned by using the plasmid pAMP1 as the vector and DH5α cells as the host. Plasmid DNA from colonies containing the pAMP1 with insert was transformed to LC2a, a nonflagellated strain of *E. coli* which has most of its *fliC* (previously termed *hag*) gene deleted (45) so that recombination between the *fliC* gene of the recipient and the *fliC* gene of the donor does not occur. Clone DNA was isolated from LC2a and sequenced (in both directions) by using the sequencing primers listed in Table 1.

Serological identities of *fliC* antigens. The serological identities of the *fliC* antigens of *Salmonella* serovars and recombinant clones expressing the *fliC* gene were confirmed by the motility inhibition test (7) using factor-specific antisera (data not shown).

Comparative analyses of *Salmonella g. . .* series serovars. An amino acid comparison based on the *g. . .* series sequencing data is presented in Table 2. Position numbers are equivalent to those previously reported for the flagellin encoded by the *fliC*^{a} gene (41). Amino acid comparisons and alignments were accomplished by using the Lipman and Pearson FASTP (18) program, along with an MTRANS (21) program and an Excel spreadsheet. The FASTP program of Lipman and Pearson (18) permitted the alignment of two protein sequences with allowance for gaps and insertions. The MTRANS program of Masten (21) allowed transfer of the Mount and Conrad translation files to an Excel spreadsheet for sequence comparisons based on the FASTP alignment data. Comparative analyses of the inferred covalent structures of the 10 *Salmonella g. . .* series serovars showed that both ends of the flagellin molecule (regions I, II, and III [amino acids 1 to 180] and region VIII [amino acids 421 to 493]) were conserved in the amino acid sequences. Region VI, embodying amino acids 323 to 359, showed complete amino acid conservation for all serovars except *S. oranienberg*, which showed differences at amino acid positions 327 and 333. Region VII, embodying amino acids 360 to 420, showed a high degree of amino acid homology, with differences at amino acid positions 368 and 372. Region IV (amino acids 181 to 299) and region V (amino acids 300 to 322), which are located in the central portion of the flagellin molecule, showed the greatest diversity among the 10 serovars, with *S. oranienberg* being the most divergent. While the other nine *g. . .* series serovars demonstrated 92 to 98% amino acid homology in region IV and 91 to 96% amino acid homology in region V, *S. oranienberg* showed only 80% amino acid homology in region IV and 65% amino acid homology in region V (data not shown). The inferred flagellin amino acid sequence of 8 of the 10 *g. . .* series serovars showed a 9-bp deletion at amino acid positions 213 to 215, located in region IV. The other two serovars, *S. berta* and *S. oranienberg*, showed threonine, valine, and proline, respectively, at amino acid positions 213 to 215.

Comparative analyses of *Salmonella fliC a*, *g. . .* series, *c*, *d*,

TABLE 2. Comparison of 10 *Salmonella* phase-1 *g*. . . series flagellins[a]

Region	Amino acid position	*Salmonellae g...* series serovars (phase-1 flagellin antigens)									
		be (*f,g,t*)	bu (*g,t*)	de (*f,g*)	du (*g,p*)	en (*g,m*)	mt (*g,m,s*)	mo (*g,q*)	or (*m,t*)	ro (*g,p,u*)	se (*g,s,t*)
IV	195	val	ala	ala	ala	ala	ala	ala	val	ala	ala
IV	197a	asn	asp	asp	asp	asp	asn	asp	asn	asp	asp
IV	200	val	ile	ile	ile	ile	ile	ile	val	ile	ile
IV	209	thr	ala	ala	ala	ala	asp	ala	thr	ala	asp
IV	210	thr	ala	val	val	ala	ala	ala	thr	ala	ala
IV	213	thr	-	-	-	-	-	-	thr	-	-
IV	214	val	-	-	-	-	-	-	val	-	-
IV	215	pro	-	-	-	-	-	-	pro	-	-
IV	216	asp	asp	asn	asp	asp	asp	asp	asp	asp	asp
IV	230	asp	asp	asp	asp	asp	asp	asp	ala	asp	asp
IV	233	glu	glu	glu	glu	glu	glu	glu	gln	glu	glu
IV	239	asp	asn	asp	asp	asp	asn	asp	asp	asp	asn
IV	243	thr	thr	thr	thr	thr	thr	thr	ser	thr	thr
IV	247	thr	thr	thr	thr	thr	ala	thr	ala	thr	thr
IV	251	ala	asp	ala	ala	ala	asp	ala	asp	ala	asp
IV	252	glu	glu	glu	glu	glu	glu	glu	asp	glu	glu
IV	258	gly	gly	gly	gly	gly	gly	gly	thr	gly	ser
IV	259	ala	ala	ala	ala	ala	ala	ala	ser	ala	ala
IV	265	glu	glu	glu	glu	glu	glu	glu	val	glu	glu
IV	273	thr	ser	thr	thr	thr	thr	thr	ser	thr	ser
IV	280	thr	ala	thr	thr	thr	thr	thr	ala	thr	ala
IV	282	asn	asn	asn	asn	asp	asp	asp	asp	asp	asn
IV	283	asp	asp	asp	asp	asp	asp	gly	asp	asp	asp
IV	287	lys	thr	lys	lys	lys	lys	lys	thr	lys	thr
IV	301	val	val	val	val	val	val	val	ile	val	val
IV	302	ala	ala	ala	ala	ala	ala	ala	ser	ala	ala
IV	305	thr	thr	thr	thr	ala	ala	ala	gly	gly	thr
V	306	ala	ala	gly	gly	thr	thr	thr	ala	ile	ala
V	307	gly	gly	gly	gly	gly	gly	gly	ser	gly	gly
V	309	ala	ala	ala	ala	thr	thr	thr	thr	ala	ala
V	310	asn	asn	asn	asn	asp	asp	asn	asp	asp	asn
V	312	asp	asn	asp	asp	asn	asn	asn	asn	asn	asn
V	313	ala	asp	ala	ala	ala	ala	ala	ser	ala	asp
V	315	thr	thr	thr	thr	thr	thr	thr	lys	thr	thr
V	315a	leu	leu	leu	leu	leu	leu	leu	ile	leu	leu
V	320	asn	asn	asn	asn	asn	asn	asn	asp	asn	asn
VI	327	asn	asn	asn	asn	asn	asn	asn	ser	asn	asn
VI	333	asp	asp	asp	asp	asp	asp	asp	ala	asp	asp
VII	368	ala	ala	thr	thr	thr	ala	thr	ala	thr	ala
VII	372	val	val	val	val	ile	val	ile	val	ile	val

[a] Only residues showing differences between the proteins are included. Dashes indicate no equivalent residue. Boxes indicate nonconsensus amino acids. Position numbers are equivalent to those previously reported for the *fliC* flagellin with antigen *a* (41). Regions were previously classified by Wei and Joys (41). Abbreviations: be, *S. berta*; bu, *S. budapest*; de, *S. derby*; du, *S. dublin*; en, *S. enteritidis*; mt, *S. montevideo*; mo, *S. moscow*; or, *S. oranienberg*; ro, *S. rostock*; se, *S. senftenberg*.

TABLE 3. Comparison of 10 members of the *Salmonella fliC* flagellin *g. . .* series with *fliC* flagellin *a* on the basis of amino acid (aa) and nucleic acid (na) homologies[a]

Region	Position in flagellin *a*	Sequence	% Homology for the following flagellin:														
			Salmonella g. . . series serovar										*c*	*d*	*i*	*r*	*E. coli*
			be	bu	de	du	en	mt	mo	or	ro	se					
I	1–100	aa	93	93	93	93	93	93	93	93	93	93	100	100	99	100	86
	1–300	na	91	91	91	91	91	92	91	91	91	90	99	99	99	99	84
II	101–160	aa	78	78	78	78	78	78	78	78	78	78	100	98	97	100	72
	301–480	na	76	76	76	76	76	76	76	76	76	76	97	94	95	97	71
III	161–180	aa	40	40	40	40	40	40	40	40	40	40	100	60	85	90	45
	481–540	na	55	55	55	55	55	55	55	55	55	55	97	62	90	92	55
IV	181–299	aa	18	15	17	17	17	17	17	16	17	16	18	24	33	29	26
	541–897	na	38	35	36	36	36	36	36	39	36	36	42	45	47	46	44
V	300–322	aa	4	0	4	4	0	0	0	0	0	0	83	61	74	70	26
	898–966	na	33	32	33	33	29	29	29	32	30	32	72	67	81	74	43
VI	323–359	aa	19	19	19	19	19	19	19	22	19	19	41	38	65	70	32
	967–1077	na	42	42	42	42	42	42	42	45	42	42	48	48	68	70	48
VII	360–420	aa	41	41	41	41	39	41	39	41	39	41	90	57	85	85	46
	1078–1260	na	48	48	50	49	48	48	48	48	48	48	88	62	84	84	50
VIII	421–493	aa	88	88	88	88	88	88	88	88	88	88	100	100	100	99	75
	1261–1479	na	83	83	83	83	83	83	83	83	83	83	99	98	97	96	60

[a] Also included for contrast are comparisons of *fliC* flagellin *a* with *fliC* flagellins *c*, *d*, *i*, and *r* and the *E. coli fliC* flagellin. Regions were previously classified by Wei and Joys (41). Abbreviations: *a*, *S. paratyphi* (41); be, *S. berta*; bu, *S. budapest*; de, *S. derby*; du, *S. dublin*; en, *S. enteritidis*; mt, *S. montevideo*; mo, *S. moscow*; or, *S. oranienberg*; ro, *S. rostock*; se, *S. senftenberg*; *c*, *S. cholerae-suis* (41); *d*, *S. muenchen* (41); *i*, *S. typhimurium* (11); *r*, *S. rubislaw* (42); *E. coli*, *E. coli* K-12 (16).

***i*, *r*, and *E. coli fliC* flagellins.** Based on nucleotide and amino acid homology with flagellin *a*, *Salmonella* flagellins from the *g. . .* series were compared with the previously reported flagellins *c*, *d*, *i*, and *r* and *E. coli fliC* (Table 3). By using the sense strand data, nucleic acid alignments and comparisons were accomplished with the Pearson HMATCH executable file included in the Mount and Conrad (25) program, along with an MTRANS (21) program and an Excel spreadsheet. Amino acid alignments and comparisons were accomplished as stated above. The *g. . .* series flagellins showed more diversity than the single-factor *Salmonella* flagellins and the *E. coli* flagellin. In regions I and VIII, where flagellins *a*, *c*, *d*, *i*, and *r* were highly conserved, flagellins of the *g. . .* series were less homologous with flagellin *a*. Although the central area (regions III, IV, V, VI, and VII) of flagellin is known to be variable between serovars and species, the *g. . .* series flagellins demonstrated more variability than single factor flagellins (*c*, *d*, *i*, and *r*) and *E. coli* flagellin when compared to flagellin *a*.

Evolutionary relatedness of *Salmonella fliC a*, *c*, *d*, *i*, *r*, and *g. . .* series and *E. coli* flagellin. An unrooted phylogenetic tree based on the flagellin sequences is presented in Fig. 1. Evolutionary relatedness was analyzed by using the PROTPARS program of the PHYLIP version 3.4 (6) analysis package. A total of 11 trees were found, with the one in Fig. 1 being representative. Four major divisions are detected. All of the *g. . .* series serovars, except *S. berta* and *S. oranienberg*, form one offshoot of the tree. *S. berta* and *S. oranienberg* each constitute their own branch. The fourth section is formed by the *Salmonella* strains exhibiting single-factor antigen flagellins and the *E. coli* flagellin. The two

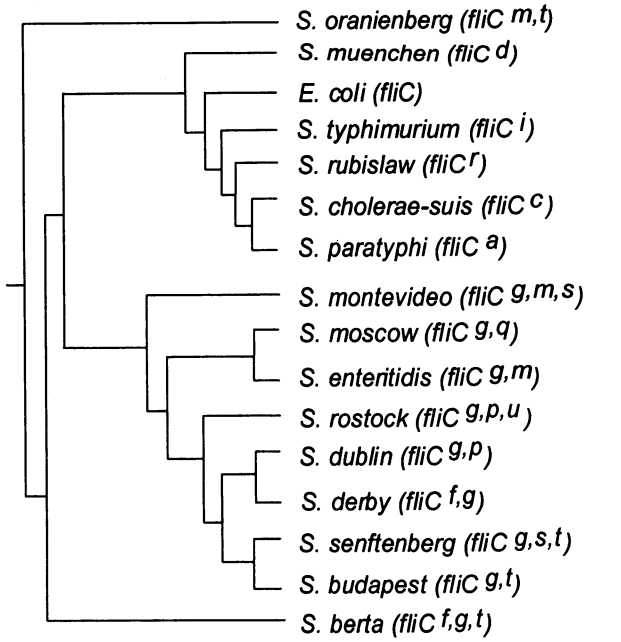

FIG. 1. Evolutionary relatedness of *Salmonella* phase-1 *g. . .* series, *a*, *c*, *d*, *i*, and *r* and *E. coli* flagellins. This is an unrooted tree, with *S. oranienberg* as the outgroup species. Flagellin antigens are shown in parentheses.

offshoots that contain the majority of the flagellins branch from *S. berta*, which branches from *S. oranienberg*.

DISCUSSION

In this work, we have sequenced the *Salmonella fliC* genes, encoding the phase-1 flagellins from 10 members of the *g*. . . series. A comparison of the flagellin amino acid sequences (Table 2) showed complete homology in the N-terminal (regions I, II, and III) and C-terminal (regions VIII) segments of the proteins. Differences in amino acids were found throughout (regions IV, V, VI, and VII) the central portion of the flagellins. No localized area specifying subfactor epitopes could be identified, suggesting that the subfactors of the *g*. . . series are conformational at the molecular level. The amino acids constituting each of the subfactor epitopes were not definable by sequence analysis. Whether the major antigen *g* can be located by available techniques (as has been achieved for the *d* antigen [13]) was not addressed in this report.

During the analyses of our data, Selander et al. (35) published data showing the variations in the sequences of *fliC* flagellin genes from isolates of *S. enteritidis* (En 1: $fliC^{g,m}$) and *S. dublin* (Du 1 and Du 3: $fliC^{g,p}$; "Du 2": $fliC^{g,p,m}$). The *fliC* sequence of *S. dublin* ($fliC^{g,p}$) reported by us differs from that reported by Selander et al. (35) in having 9 synonymous (amino acid positions 52, 54, 120, 239, 261, 307, 360, 382, and 388) and 6 nonsynonymous (amino acid positions 282, 305, 306, 310, 312, and 372) substitutions compared to Du 1 and Du 3, and in having 11 synonymous (amino acid positions 52, 54, 59, 120, 137, 239, 261, 307, 360, 382, and 388) and 7 nonsynonymous (amino acid positions 210, 282, 305, 306, 310, 312, and 372) substitutions compared to "Du 2". Note that all sequence numbering is based on equivalent residues in flagellin *a* (41). On the basis of their *fliC* comparison, Selander et al. (35) deduced that the presence of alanine at amino acid position 309 is responsible for the expression of the phase-1 flagellin *p* subfactor, while alanine at amino acid position 210 and threonine at amino acid position 306 contributes to the *m* subfactor. Interestingly, our phase-1 flagellin *g*. . . series comparison (Table 2) showed that at amino acid position 309, alanine is present not only in *S. dublin* ($fliC^{g,p}$), but also in five other *g*. . . series serovars, of which only one, *S. rostock*, manifests the *p* subfactor. At amino acid position 210, alanine is present in *S. enteritidis* ($fliC^{g,m}$) and *S. montevideo* ($fliC^{g,m,s}$) as well as in four other non-*m* subfactor serovars. Also, alanine is not present at amino acid position 210 in *S. oranienberg*, although the flagellin has *m* and *t* antigens. At amino acid position 309, threonine is present in *S. enteritidis* ($fliC^{g,m}$), *S. montevideo* ($fliC^{g,m,s}$) and *S. oranienberg* ($fliC^{m,t}$), but it is also present in *S. moscow* ($fliC^{g,q}$), a non-*m* subfactor serovar. Selander et al. (35) also stated that the *fliC* sequence of *S. rostock* ($fliC^{g,p,u}$) differed from that of Du 1 ($fliC^{g,p}$) only at amino acid position 305, where glycine was present in *S. rostock*. Selander et al. (35) suggested that this glycine was responsible for the expression of the μ subfactor in the *S. rostock* phase-1 flagellin. However, we found glycine at amino acid position 305 in *S. oranienberg* ($fliC^{m,t}$) as well. In addition, the comparison of our *S. dublin* ($fliC^{g,p}$) and *S. rostock* ($fliC^{g,p,u}$) *fliC* sequences showed six nonsynonymous differences between the two serovars.

Our expectations were that the linear epitopes which composed the subfactors of the *g*. . . series would be localized to one region, providing an efficient site for directed substitution with medically important epitopes. Although our expectations were not borne out, a potential site for epitope substitution was observed upon comparison (Table 3) of the *g*. . . series flagellins with flagellins *a*, *c*, *d*, *i*, and *r*, and *E. coli* flagellin. Only in region V with the *g*. . . series flagellins did we find extremely low (4%) or absolutely no homology with flagellin *a*, suggesting that this region in the *g*. . . series has potential for site-directed substitutions. Joys and Schödel (13) located a major epitope of antigen *d* within region V. Interestingly, in region V, *S. oranienberg*, which has no *g* antigen but is a member of the *g*. . . series because of its *m* and *t* antigens, demonstrated only 65% amino acid homology (data not shown) with other members of the *g*. . . series. Other members of the *g*. . . series exhibited 91 to 96% homology (data not shown) in region V. This information suggests that the *fliC* gene of *S. oranienberg* may be the gene of choice for substitution(s), since it may tolerate amino acid interchanges in region V better than the *fliC* gene of other members of the *g*. . . series.

Based upon phylogenetic analyses (Fig. 1) and comparative analyses of Salmonella *fliC a*, *g*. . . series, *c*, *d*, *i*, and *r*, and *E. coli* flagellins (Table 3), the *S. oranienberg* flagellin with antigens *m* and *t* shows the greatest diversity from all salmonellar flagellins analyzed to date. The *S. berta* flagellin shows some similarity to the *S. oranienberg* flagellin and also differs from the other flagellins. The remaining 14 flagellins divide into two groups, one consisting of the *g*. . . series and the other consisting of the single-factor flagellins. Interestingly, the flagellin of *E. coli* is grouped with the single-factor flagellins, although serological (5) and molecular weight (23) analyses show *E. coli* flagellins to be even more diverse than *Salmonella* flagellins.

Both the accumulation of spontaneous mutations (41) and zenologous (8) replication (34, 36, 37) has been suggested to explain serotypic variation in flagellin. Based upon the data presented here, it appears that a combination of both mechanisms is likely to have occurred.

ACKNOWLEDGMENTS

This work was supported by the Texas Advanced Technology Program under grant numbers 010674-006 and 010674-010.

REFERENCES

1. **Astbury, W. T., E. Beighton, and C. Weibull.** 1955. The structure of bacterial flagella. Symp. Soc. Exp. Biol. **9**:282–305.
2. **Brey, R. N., G. S. Bixler, and J. P. Fuiginiti.** 1991. Oral delivery of antigens in live bacterial vectors. Adv. Exp. Med. Biol. **303**:169–184.
3. **DePamphilis, M. L., and J. Adler.** 1971. Purification of intact flagella from *Escherichia coli* and *Bacillus subtilis*. J. Bacteriol. **105**:376–383.
4. **Edwards, P. R., and W. H. Ewing.** 1972. The genus *Salmonella*, p. 146–207; Antigenic schema for *Salmonella*, p. 208–258. *In* P. R. Edwards and W. H. Ewing (ed.), Identification of *Enterobacteriaceae*, 3rd ed. Burgess Publishing Co., Minneapolis, Minn.
5. **Ewing, W. H.** 1986. Antigenic schema for *Salmonella*, p. 247–318. *In* P. R. Edwards and W. H. Ewing (ed.), Identification of *Enterobacteriaceae*, 4th ed. Burgess Publishing Co., Minneapolis, Minn.
6. **Felsenstein, J.** 1991. PHYLIP (phylogeny inference package) version 3.4. University of Washington, Seattle, Wash.
7. **Gard, S.** 1938. Ein neuer *Salmonella*-Typ (*S. abortus canis*). Z. Hyg. **121**:139–142.
8. **Gray, G. S., and W. Fitch.** 1983. Evolution of antibiotic resistant genes: the DNA sequence of a kanamycin resistant gene from *Staphylococcus aureus*. Mol. Biol. Evol. **1**:57–66.
9. **Iino, T.** 1977. Genetics of structure and function of bacterial flagella. Annu. Rev. Genet. **11**:161–182.

10. **Iino, T., Y. Komeda, K. Kutsukake, R. M. Macnab, P. Matsumura, J. S. Parkinson, M. I. Simon, and S. Yamaguchi.** 1988. New unified nomenclature for the flagellar genes of *Escherichia coli* and *Salmonella typhimurium*. Microbiol. Rev. **52:**533–535.
11. **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.
12. **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.
13. **Joys, T. M., and F. Schödel.** 1991. Epitope mapping of the *d* flagellar antigen of *Salmonella muenchen*. Infect. Immun. **59:** 3330–3332.
14. **Joys, T. M., and B. A. D. Stocker.** 1969. Recombination in *H1*, the gene determining the flagellar antigen-*i* of *Salmonella typhimurium*; mapping of *H1* and *fla* mutations. J. Gen. Microbiol. **58:**267–275.
15. **Kauffmann, F.** 1964. Das Kauffmann-White-Scheme, p. 21–66. *In* E. van Oye (ed.), The world problem of salmonellosis. W. Junk Publishers, The Hague, The Netherlands.
16. **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.
17. **Le Minor, L., and M. Y. Popoff.** 1987. Antigenic formulas of the *Salmonella* serovars. WHO Collaborating Centre for Reference and Research on *Salmonella*, Institut Pasteur, Paris.
18. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227:**1435–1441.
19. **MacNab, R. M.** 1987. Flagella, p. 70–83. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
20. **Majarian, W. R., S. J. Kasper, and R. N. Brey.** 1989. Expression of heterologous epitopes as recombinant flagella on the surface of attenuated *Salmonella*, p. 277–281. *In* R. A. Lerner et al. (ed), Vaccines 89: modern approaches to new vaccines including prevention of AIDS. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Masten, G. B.** 1992. Unpublished data.
22. **McDonough, M. W.** 1965. Amino acid composition of antigenically distinct *Salmonella* flagellar proteins. J. Mol. Biol. **12:**342–355.
23. **McDonough, M. W., and S. E. Smith.** 1976. Molecular weight variation among bacterial flagellins. Micobios **16:**49–53.
24. **McEwen, J., R. Levi, R. J. Horwitz, and R. Arnon.** 1992. Synthetic recombinant vaccine expressing influenza haemagglutinin epitope in *Salmonella* flagellin leads to partial protection in mice. Vaccine **10:**405–411.
25. **Mount, D. W., and B. Conrad.** 1984. Microcomputer programs for graphic analysis of nucleic acid and protein sequences. Nucleic Acids Res. **12:**811–817.
26. **Namba, K., I. Yamashita, and F. Vonderviszt.** 1989. Structure of the core and central channel of bacterial flagella. Nature (London) **342:**648–654.
27. **Newton, S. M., W. C. Manning, M. Hovi, and B. A. D. Stocker.** 1990. Aromatic-dependent *Salmonella* with foreign epitope insert in flagellin as live vaccine, p. 439–445. *In* R. A. Lerner et al. (ed.), Vaccines 90: modern approaches to new vaccines including prevention of AIDS. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **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.
29. **Newton, S. M. C., M. Kotb, T. P. Poirier, B. A. D. Stocker, and E. H. Beachey.** 1991. Expression and immunogenicity of a streptococcal M protein epitope inserted in *Salmonella* flagellin. Infect. Immun. **59:**2158–2165.
30. **Parish, C. R., R. Wistar, and G. L. Ada.** 1969. Cleavage of bacterial flagellin with cyanogen bromide: antigenic properties of the protein fragments. Biochem. J. **113:**501–506.
31. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:**487–491.
32. **Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230:**1350–1354.
33. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Selander, R. K., and N. H. Smith.** 1990. Molecular population genetics of *Salmonella*. Rev. Med. Microbiol. **1:**219–228.
35. **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:**3587–3692.
36. **Smith, M. J.** 1991. The population genetics of bacteria. Proc. R. Soc. London Ser. B **245:**37–41.
37. **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.
38. **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.
39. **Stocker, B. A. D.** 1990. Aromatic-dependent *Salmonella* as live vaccine presenters of foreign epitopes as inserts in flagellin. Res. Microbiol. **141:**787–796.
40. **Szekely, E., and M. Simon.** 1983. DNA sequence adjacent to flagellar genes and evolution of flagellar-phase variation. J. Bacteriol. **155:**74–81.
41. **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.
42. **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.

42a. **Wei, L.-N., and T. M. Joys.** Unpublished data.

43. **Wu, J. Y., S. M. C. Newton, A. Judd, B. A. D. Stocker, and W. S. Robinson.** 1989. Expression of immunogenic epitope of hepatitis B surface antigen with hybrid flagellin protein by a vaccine strain of *Salmonella*. Proc. Natl. Acad. Sci. USA **86:**4726–4730.
44. **Yamaguchi, S., and T. Iino.** 1969. Genetic determination of the antigenic specificity of flagellar protein in *Salmonella*. J. Gen. Microbiol. **55:**59–74.
45. **Zieg, J., M. Silverman, M. Hilmen, and M. Simon.** 1977. Recombinational switch for gene expression. Science **196:**170–172.