JOURNAL OF BACTERIOLOGY, Apr. 1990, p. 1905–1915
0021-9193/90/041905-11$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 172, No. 4

Sequence and Molecular Characterization of a Multicopy Invasion Plasmid Antigen Gene, *ipaH*, of *Shigella flexneri*

ANTOINETTE B. HARTMAN,[1]* MALABI VENKATESAN,[2] EDWIN V. OAKS,[3] AND JERRY M. BUYSSE[2]

Department of Biologics Research,[1] Department of Bacterial Immunology,[2] and Department of Enteric Infections,[3] Division of Communicable Diseases and Immunology, Walter Reed Army Institute of Research, Washington, D.C. 20307-5100

Received 2 August 1989/Accepted 22 December 1989

A λgt11 expression library of Tn*5*-tagged invasion plasmid pWR110 (from *Shigella flexneri* serotype 5, strain M90T-W) contained a set of recombinants encoding a 60-kilodalton protein (designated IpaH) recognized by rabbit antisera raised against *S. flexneri* invasion plasmid antigens (J. M. Buysse, C. K. Stover, E. V. Oaks, M. M. Venkatesan, and D. J. Kopecko, J. Bacteriol. 169:2561–2569, 1987). Southern blot analysis of wild-type *S. flexneri* serotype 5 invasion plasmid DNA (pWR100) digested with various combinations of five restriction enzymes and hybridized with defined *ipaH* probes showed complex hybridization patterns resulting from multiple copies of the *ipaH* gene on pWR100. DNA sequence analysis of a 2.9-kilobase (kb) *Eco*RI fragment directing IpaH antigen synthesis in plasmid recombinant pWR390 revealed an open reading frame coding for a 532-amino-acid protein (60.8 kilodaltons); this size matched well with the estimated size of IpaH determined by Western blot analysis of M90T-W cells and maxicell analysis of *Escherichia coli* HB101(pWR390) transformants. Examination of the amino acid sequence of IpaH revealed a hydrophilic protein with six evenly spaced 14-residue (L-X_2-L-P-X-L-P-X_2-L-X_2-L) repeat motifs in the amino-terminal end of the molecule. Southern blot analysis of *Hin*dIII-digested pWR100 DNA probed with defined segments of the pWR390 2.9-kb insert demonstrated that the multiple band hybridization pattern resulted from repeats of a significant portion of the *ipaH* structural gene in five distinct *Hin*dIII fragments (9.8, 7.8, 4.5, 2.5, and 1.4 kb). Affinity-purified IpaH antibody, used to monitor the expression of the antigen in M90T-W cells grown at 30 and 37°C, showed that IpaH synthesis was not regulated by growth temperature.

The pathogenesis of bacillary dysentery requires the coordinate expression of a number of components that control the epithelial cell invasion, intracellular replication, and intercellular spreading phenotypes characteristic of the expression of virulence in *Shigella* species and enteroinvasive *Escherichia coli*. Genes encoding virulence-associated elements are located on the chromosome and on a 120- to 140-megadalton (MDa) plasmid found in all virulent *Shigella* and enteroinvasive *E. coli* strains (12, 16, 31, 42). At least eight unique polypeptides are encoded by this invasion plasmid (12, 13); five of these (VirG and invasion plasmid antigens [Ipa] A, B, C, and D) are immunogens consistently recognized by serum and mucosal antibodies in convalescent humans and primates (26). Molecular cloning and nucleotide sequence determination for *Shigella flexneri ipaB*, *ipaC*, *ipaD*, and *virG* genes have been previously described (2, 3, 9, 19, 32, 40). However, phenotypes and functions associated with the expression of these antigens are only broadly defined. The expression of IpaB, IpaC, and IpaD is consistently associated with the attachment and invasion steps of dysentery pathogenesis (25). The VirG protein (encoded by the *virG* or *icsA* locus), along with the product of the chromosomal *kcpA* gene, has recently been implicated in the intercellular spread of the bacteria once they have invaded target epithelial cells and escaped the phagosome (4, 21, 27, 32). How the action of these proteins contributes to specific phenotypes at the molecular level remains to be elucidated.

Invasion of colonic epithelial cells by the *Shigella* bacillus demands close interactions between surface structures on the bacteria and target host cells. It is likely that these surface components are also recognized by the host immune system in an attempt to neutralize the pathogen, as appears to be the case for VirG, IpaB, IpaC, and IpaD, which are all immunogenic to the host. To clarify the mechanisms of invasion and to identify potential protective epitopes that can be utilized by the host immune system to counteract invasion, it is important to characterize all *Shigella* invasion plasmid antigens. In an earlier report (9), we described the isolation of an additional invasion plasmid antigen, designated IpaH. The IpaH protein (60 kilodaltons [kDa]) was similar in size to the IpaB protein (62 kDa) but was distinct from the latter antigen both immunologically and at the DNA level. In this report, we describe the further characterization and DNA sequence analysis of the *ipaH* gene.

In contrast to the *ipaBCDA* loci (23, 24, 32, 39, 40), the *ipaH* gene is reiterated on the invasion plasmid and the expression of the IpaH antigen is not temperature regulated. Analysis of the deduced amino acid sequence of IpaH indicated the presence of a unique 14-residue motif which is repeated six times in the amino-terminal end of this hydrophilic molecule. The multicopy nature of the *ipaH* gene and its unique amino acid sequence may reflect an essential, though as yet undefined, role for this antigen in *Shigella* virulence or in the genetic instability of the invasion plasmid (33).

MATERIALS AND METHODS

Bacterial strains, culture conditions, and recombinant DNA techniques. A Tn*5*-tagged invasion plasmid (pWR110) of *S. flexneri* serotype 5 (strain M90T-W) was used as the source of insert DNA for the construction of λgt11 *ipaH* recombinants (9). *E. coli* Y1090 cells (Δ*lacU169 proA*$^+$ Δ*lon araD139 rpsL supF trpC*::Tn*10 hsdR hsdM*$^+$ *lacI*^{q}) were used for the production of high-titer λgt11 *ipaH* lysates and in the isola-

* Corresponding author.

tion of lysogens. Unless noted otherwise, all strains were routinely cultured in LB broth or on L agar plates at 37°C. Y1090::λgt11 *ipaH* lysogens and *E. coli* JM109(pWR390) transformants were selected on L agar supplemented with 100 μg of ampicillin per ml.

Construction of the λgt11 expression library of invasion plasmid pWR110 has been described previously (9, 25); this library was used to isolate several recombinant bacteriophage carrying the *ipaH* gene on the basis of their reaction with plasmid antigen-specific rabbit screening antisera. Recombinant pWR390 was prepared by ligating the 2.9-kilobase (kb) insert DNA of λgt11 *ipaH* S39 (9) (see Table 1) into the *Eco*RI site of pUC12 by standard techniques for vector preparation, insert ligation, and identification of recombinant plasmids (22). The recombinant plasmid was then transformed into *E. coli* JM109 [*recA1 endA1 gyr96 thi hsdR17* (r_K^- m_K^+) *supE44 relA1* Δ*lac-proAB* (F′ *traD36 proAB lacI*^{q}Δ*M15*)] cells. The 2.9-kb *Eco*RI fragment was later subcloned into pBR322, generating pWR391, to facilitate electrophoretic purification of the insert fragment from the plasmid vector.

Affinity purification of IpaH-specific antibodies and immunoblotting procedures. Antibodies directed against IpaH antigen epitopes were affinity purified from polyvalent rabbit antisera by using protein expressed from λgt11 *ipaH* recombinants as the affinity matrix (9). In this procedure, IpaH-specific antibodies were bound to IpaH antigen immobilized on a nitrocellulose membrane and were then eluted from the filter with 0.2 M glycine—0.15 M NaCl (pH 2.8). After neutralization to pH 7.0, the selected antibodies were diluted 1:3 before use in the appropriate Western blot assay. Polypeptides of whole-cell sodium dodecyl sulfate (SDS) lysates obtained from wild-type *S. flexneri* serotype 5 or from *E. coli* strains harboring recombinant *ipaH* plasmid or phage were separated on 13% acrylamide cross-linked with *N,N*′-diallyltartardiamide in a discontinuous SDS-polyacrylamide gel electrophoresis (PAGE) system with Laemmli buffers as described previously (9, 25). After separation, the proteins were electroblotted to nitrocellulose and probed with rabbit polyvalent antisera or affinity-purified antibodies to the IpaH antigen, using a previously described protocol (25, 26).

Plasmid DNA preparation and DNA hybridizations. Invasion plasmid DNA was isolated by the method of Cassie et al. (10) and purified by cesium chloride-ethidium bromide density gradient ultracentrifugation. λgt11 *ipaH* phage DNA was isolated by the glycerol step-gradient procedure of Silhavey et al. (34). pWR390, pWR391, and λgt11 *ipaH* phage insert DNA were excised by digestion with *Eco*RI and purified twice by agarose gel electrophoresis with 0.8% agarose. Invasion plasmid DNA was digested with the appropriate restriction endonuclease enzyme(s), and the fragments were electrophoresed on 0.6% agarose gels with 0.5× TBE buffer (121.1 g of Tris base, 51.34 g of boric acid, 3.72 g of EDTA disodium salt, per liter, pH 7.8). The separated fragments were transferred to nitrocellulose or Nytran filters (Schleicher & Schuell, Inc., Keene, N.H.) by the method of Southern (35). Hybridizations were done in 50% formamide–5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7.2)–5× Denhardt solution (1× Denhardt solution consists of 0.02% each bovine serum albumin, Ficoll, and polyvinylpyrrolidone)–5% dextran sulfate–50 μg of denatured salmon serum DNA per ml–2 × 10^6 cpm of α-^{32}P-labeled *ipaH* probe per ml for 12 h at 37°C. Insert fragments from pWR390, pWR391, or λgt11 *ipaH* recombinant phage were radiolabeled by nick translation (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) with [α-^{32}P]dCTP from Amersham Corp. (Arlington Heights, Ill.). Filters were washed for 15 min at room temperature with 2× SSC–0.1% SDS, 30 min at 37°C with 0.1× SSC–0.1% SDS, and 30 min at 65°C with 0.1× SSC–0.1% SDS before autoradiography.

Maxicell analysis to identify plasmid-encoded proteins. The plasmids pWR390 and pUC12 were transformed into *E. coli* HB101. Plasmid-encoded proteins were then identified by a modified maxicell procedure (29, 37). Briefly, a culture of HB101 containing either pWR390 or pUC12 was grown to an A_{600} of 0.6 and harvested. After suspension, the cells were irradiated, collected, and grown in M9 minimal medium (22) with the addition of 1% Casamino Acids (Difco Laboratories, Detroit, Mich.). Cycloserine (200 μg/ml) was added 2 h after irradiation and 2 h before harvesting of the cells. After harvest, cells were labeled with [^{35}S]methionine (Dupont, NEN Research Products, Boston, Mass.) during a 1-h incubation at 37°C. Cells were collected and analyzed by SDS-13% PAGE and autoradiography as described previously (37).

DNA sequence analysis of *ipaH*. Overlapping fragments isolated from the 2.9-kb *Eco*RI insert of pWR390 were subcloned into M13mp18 and M13mp19. The fragments were then sequenced by the dideoxynucleotide termination method (30). Both strands of the 2.9-kb fragment were sequenced to ensure accuracy. Restriction site analysis of the fragment, protein translation of the DNA sequence, open reading frame (ORF) searches, hydropathy plot, and antigenic index analyses were done by using the MacGene Plus application on a MacIntosh SE microcomputer and the International Biotechnologies, Inc. (IBI)/Pustell Sequence Database Manager. To look for possible nucleotide and protein sequence homologies to the *ipaH* primary sequence, GenBank Nucleic Acid and National Biomedical Research Foundation (NBRF) PRI Protein Databases were searched as part of the IBI/Pustell Sequence Database Manager and University of Wisconsin GCG DNA sequence analysis packages.

RESULTS

Hybridization of λgt11 *ipaH* insert DNA to wild-type *S. flexneri* invasion plasmid (pWR100). In a previous study (9), a λgt11 expression library of invasion plasmid pWR110 (a Tn*5* derivative of pWR100) was probed with plasmid antigen-specific rabbit sera, and several clones, subsequently characterized as λgt11 *ipaB*, λgt11 *ipaC*, and λgt11 *ipaD* recombinants, were identified. One set of 17 recombinants, however, did not correspond to any of the known *ipa* gene loci, and members of this group, designated λgt11 *ipaH*, were found to encode the synthesis of a 60-kDa protein. Because of the similarity in size of the antigens encoded by *ipaB* (62 kDa) and *ipaH* (60 kDa), insert DNA prepared from λgt11 *ipaB* and λgt11 *ipaH* recombinants was cross-hybridized in a Southern blot analysis to detect any DNA sequence homology between the antigen genes; no cross-hybridization was found. Furthermore, affinity-purified antibodies prepared against representative λgt11 *ipaB* and λgt11 *ipaH* recombinants reacted with a 60-kDa protein present in whole-cell lysates of strain M90T-W but did not recognize polypeptides synthesized by the heterologous Y1090 lysogen (i.e., IpaH affinity-purified antibodies did not recognize IpaB antigen and vice versa). These results proved that recombinants λgt11 *ipaB* and λgt11 *ipaH* represented unique clones of two distinct but similarly sized invasion plasmid antigens.

We began the current investigation by isolating DNA from

TABLE 1. Polypeptide products and insert DNA size of λgt11 *ipaH* clones

λgt11 *ipaH* recombinant	Polypeptide synthesized by lysogen (kDa)[a]	*Eco*RI-cleaved insert DNA (bp)
S39	Non-I (60)	2,900
S25	Non-I (60)	1,950
S31	Non-I (60)	2,300[b], 950
S52	Non-I (60)	2,800[b], 1,000
S63	Non-I (60)	3,500[b], 950
S16	Non-I (60)	2,650[b], 780, 600[c]
S53	Non-I (60)	2,300[b], 680
S66	Non-I (60)	2,300[b], 550
S67	Non-I (60)	2,300[b], 780
S48	Non-I (60)	2,000[b], 1,150
W71	Non-I (60)	1,950
W20	Non-I (60)	2,600
S40	Non-I (60)	2,300
S42	Non-I (60)	1,850
S46	Non-I (60)	2,600[b], 700
W7	I (>116)	950
S54	I (60)	1,800[b], 1,100

[a] Non-I and I denote recombinant protein synthesis that was noninducible or inducible, respectively, with isopropyl-β-D-thiogalactopyranoside.
[b] *ipaH*-containing fragment identified by cross-hybridization with S39 insert DNA.
[c] Insert fragment obtained from the 2.9- or 2.1-MDa cryptic ColE1-derived plasmids of *S. flexneri* 5, as determined by hybridization with ColE1 DNA.

each of the 17 λgt11 *ipaH* clones so that insert DNA size and the number of *Eco*RI fragments carried by the recombinants could be determined (Table 1). One such recombinant, λgt11 *ipaH* S39, contained a single insert fragment that encoded the synthesis of the 60-kDa IpaH antigen; therefore, this fragment was isolated and used as a probe to define homologous *ipaH*-containing sequences in λgt11 *ipaH* recombinants that carried more than one *Eco*RI insert fragment (Table 1). To determine a restriction map for the region of the *ipaH* gene on pWR100, insert fragments from several *ipaH* clones were used as probes against Southern-blotted *S. flexneri* 5 invasion plasmid DNA (pWR100) digested to completion with various combinations of the restriction enzymes *Eco*RI, *Hin*dIII, *Bam*HI, *Bgl*II, and *Pst*I (Fig. 1). λgt11 *ipaH* recombinants containing only one *Eco*RI insert fragment (e.g., S25, S39; Table 1) gave a complex pattern when used to probe endonuclease-digested pWR100 DNA, hybridizing multiple restriction fragments (1 to >23 kb) with various intensities (Fig. 1A). Particularly noteworthy were the five restriction fragments detected with the *ipaH* probe in *Hin*dIII-digested pWR100 DNA (9.8, 7.8, 4.5, 2.5, and 1.4 kb). Probes derived from λgt11 *ipaH* recombinants containing more than one insert fragment (e.g., S52, S63; Table 1) showed that the IpaH-encoding fragment again gave the same complex repeated hybridization pattern, while the accompanying contiguous fragment detected only one or two bands in a pattern more amenable to the construction of a restriction map (Fig. 1A and B). These results suggested that the *ipaH* gene (or immediate flanking DNA) was repeated on the pWR100 plasmid.

Examination of IpaH expression in *S. flexneri*. One of the distinguishing characteristics of the invasion plasmid antigens IpaB, IpaC, IpaD, and IpaA is the stringent regulation of their expression by temperature, allowing synthesis of the antigens at 37°C but not at 30°C (13, 23, 24, 40). We wanted to determine whether IpaH protein synthesis in the native *S. flexneri* host was temperature regulated as well. Affinity-

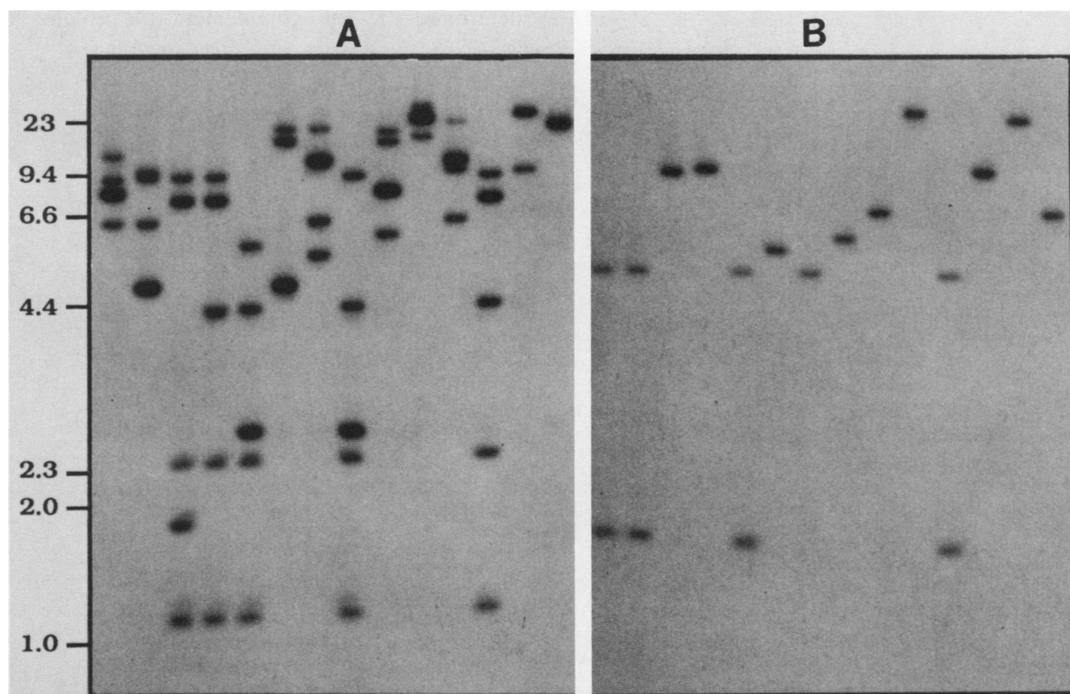

FIG. 1. Southern blot analysis of *S. flexneri* invasion plasmid DNA (pWR100) probed with insert DNA obtained from λgt11 *ipaH* S52. Hybridization patterns obtained with the 2,800-bp *ipaH*-containing fragment are shown in panel A, while panel B depicts the results obtained with the contiguous 1,000-bp fragment of S52 that does not carry *ipaH* sequences. Lanes in each panel from left to right are pWR100 digested with *Pst*I-*Bgl*II, *Pst*I-*Bam*HI, *Hin*dIII-*Bam*HI, *Bgl*II-*Hin*dIII, *Pst*I-*Hin*dIII, *Bam*HI-*Eco*RI, *Pst*I-*Eco*RI, *Hin*dIII-*Eco*RI, *Bgl*II-*Eco*RI, *Bam*HI, *Pst*I, *Hin*dIII, *Bgl*II, and *Eco*RI. Lambda *Hin*dIII DNA molecular weight standards (in kilobases) are indicated to the left of panel A. Hybridization conditions were as described in Materials and Methods.

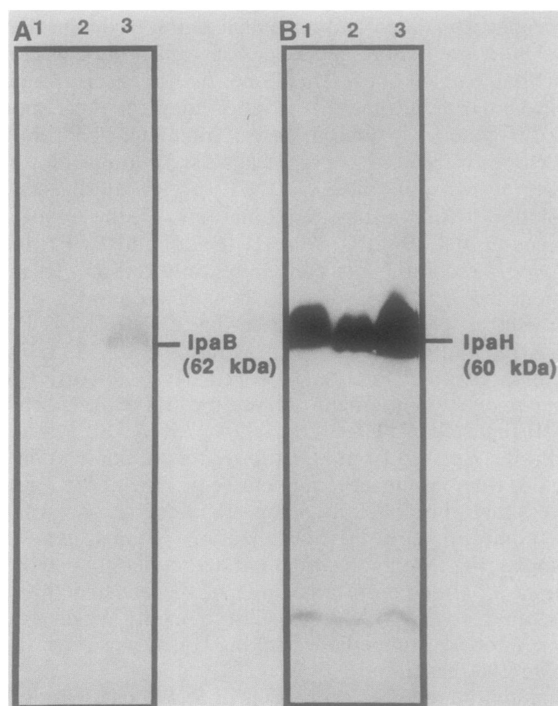

FIG. 2. Affinity-purified antibodies prepared against IpaB (panel A) and IpaH (panel B) were used to monitor antigen expression in M90T-A_3 (lanes 1), M90T-W at 30°C (lanes 2), and M90T-W at 37°C (lanes 3). Positions of the IpaB (62-kDa) and IpaH (60-kDa) antigens are indicated to the right of each panel. Preparation of whole-cell lysates, SDS-PAGE parameters, and blotting procedures are described in Materials and Methods.

purified antibodies were prepared from λgt11 *ipaH* S39- and λgt11 *ipaB* S12-encoded-antigens (9) and were used in a Western blot (immunoblot) analysis of whole-cell lysates of strains M90T-W and M90T-A_3 (a 65-kb deletion derivative of M90T-W lacking the *ipaBCDA* and *virG* genes [39, 41]) grown at 30 and 37°C (Fig. 2). IpaB-selected antibody detected the 62-kDa IpaB antigen only in the sample prepared from M90T-W grown at 37°C and did not react with protein prepared from M90T-W grown at 30°C or M90T-A_3 (Fig. 2A). In contrast, the IpaH-selected antibody reacted with all three samples, including the M90T-W sample grown at 30°C (Fig. 2B). Additionally, Northern (RNA) blot analysis of total RNA prepared from M90T-W cells grown at 30 and 37°C demonstrated that *ipaH* transcription was not temperature regulated (data not shown). These findings indicated that expression of *ipaH* is neither temperature dependent nor affected by the deletion carried on the M90T-A_3 invasion plasmid.

IpaH expression by recombinant plasmid pWR390. To further study the nature of the *ipaH* gene and its product, we selected an *ipaH* clone, λgt11 *ipaH* S39, that (i) contained only one insert fragment, (ii) synthesized the complete IpaH antigen, and (iii) gave the characteristic mixed intensity, five-band hybridization pattern when probed against *Hin*dIII-digested pWR100 DNA (Fig. 1A). The 2.9-kb *Eco*RI insert fragment of λgt11 *ipaH* S39 was subcloned into pUC12, producing recombinant plasmid pWR390, which was used to transform *E. coli* JM109. Western blot analysis of JM109(pWR390) transformants showed production of a 60-kDa antigen that reacted with the plasmid antigen-specific rabbit antisera (Fig. 3A) and with antibodies affinity purified against recombinant IpaH antigen (data not shown). This 60-kDa protein was also recognized by monkey and human convalescent antisera to *S. flexneri* (unpublished data). To determine the full complement of proteins encoded by

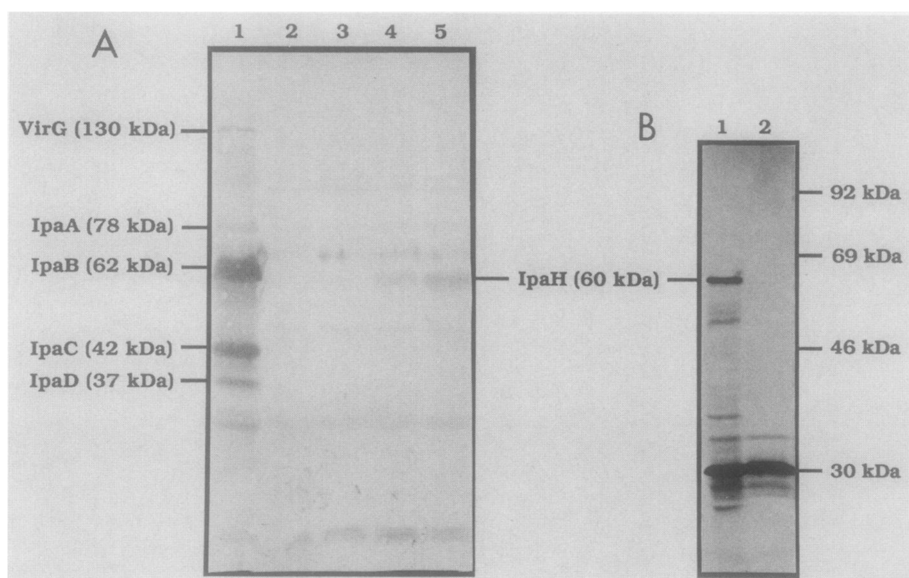

FIG. 3. Immunoblot and maxicell analysis of proteins encoded by pWR390. (A) JM109(pWR390) cells were grown to the log phase in LB broth plus 100 μg of ampicillin per ml at 30°C (lane 4) and 37°C (lane 5) and analyzed by Western blotting with rabbit antisera specific for *S. flexneri* invasion plasmid antigens (9). M90T-W (lane 1), JM109 (lane 2), and JM109(pUC12) (lane 3) controls are also shown. (B) pWR390 and pUC12 were transformed into strain HB101, and the transformants were analyzed by the maxicell technique as described in Materials and Methods. An autoradiograph of a representative gel is shown. Lane 1, HB101(pWR390) transformant; lane 2, HB101(pUC12) transformant. The position of the IpaH protein is indicated to the left, and molecular mass markers are shown on the right. The major 30-kDa band seen in both the control pUC12 transformant and the pWR390 transformant is β-lactamase.

pWR390, maxicell analysis of HB101(pWR390) transformants was performed (Fig. 3B). The major plasmid product was the IpaH polypeptide (60 kDa); however, three other polypeptide products appeared in minor quantities (52, 38, and 17 kDa). These minor peptides might represent specific degradation products of the IpaH antigen or distinct polypeptides produced by other ORFs located within the insert fragment (see below).

Nucleotide sequence of 2.9-kb insert fragment of pWR390. The complete DNA sequence of the 2.9-kb *Eco*RI insert of pWR390 and the deduced amino acid sequence of the IpaH protein obtained from the nucleotide sequence are presented in Fig. 4. pWR390 contained three ORFs, one of which encoded a 60.8-kDa protein (pI 5.9) that matched well with the estimated size of the IpaH protein determined by SDS-PAGE Western blots (9). The IpaH ORF extended from the translation initiation codon at position 251 to the TAA stop codon at position 1847, corresponding to a 532-amino-acid protein (1,596 nucleotides). A Shine-Dalgarno ribosome-binding site (GAGAA) was located 12 base pairs (bp) upstream from the ATG codon; potential −10 and −35 promoter regions and transcription terminator structures were also noted. In a different reading frame on the IpaH sense strand, two additional ORFs, ORF2 and ORF3, schematically represented in Fig. 5A, were also found. The initiation codons for these proteins were located at positions 1177 and 2277, respectively. Both ORF2 and ORF3 had potential transcription initiation −10 and −35 elements and Shine-Dalgarno ribosome-binding sites, but ORF3 did not contain a stop codon in the 2.9-kb insert, suggesting that only a portion of the protein encoded by ORF3 was found in this fragment. This truncated ORF3 product may be represented by the 38-kDa minor protein product detected in the maxicell analysis of JM109(pWR390) cells (Fig. 3B).

Hydropathy analysis of the IpaH amino acid sequence with the Kyte-Doolittle algorithm at a residue resolution of 15 (17) revealed that IpaH has a predominantly hydrophilic structure with small regions of hydrophobic residues interspersed in the protein (Fig. 5B). Hydrophilic peaks in this profile may reflect antigenic sites on the protein, as has been noted previously for IpaB and IpaC (40), and the preponderance of these sites was expected in view of the demonstrated immunogenicity of the IpaH protein. When the antigenic index of IpaH was calculated by the algorithm of Jameson and Wolf (15), results indicated that the most likely antigenic sites were located in the region between amino acids 140 and 320. This region overlaps the first large hydrophilic section of the protein shown in Fig. 5B. A hydrophobic stretch with the characteristics of known signal-peptide sequences (43) was not found in IpaH. A search for similarity between *ipaH* and sequences recorded in the National Institutes of Health-GenBank Nucleic Acid or EMBL databases did not produce any striking homologies. In addition, no strong homologies were found between the IpaH protein sequence and sequences found in the NBRF database. Both observations were in agreement with the demonstrated *Shigella* species-enteroinvasive *E. coli*-specific nature of the *ipaH* gene and protein (41). Analysis of the amino acid sequence of the IpaH protein revealed six evenly spaced 14-residue repeat motifs consisting of Leu-X_2-Leu-Pro-X-Leu-Pro-X_2-Leu-X_2-Leu (where X represents any amino acid) located between amino acid residues 39 and 149 in the amino-terminal end of the molecule (Fig. 4). Each repeat of this 14-residue motif was separated from the next element by six amino acids, and the fifth amino acid in this intervening sequence was a conserved asparagine residue. The only variation in this scheme was the substitution of an isoleucine for a leucine residue (both of which are nonpolar amino acids) in positions 11 and 14 of the first 14-residue motif and position 4 of the third motif. In addition, it was noted that the amino acid residues located immediately after these repeat motifs (residues 145 through 155) produce an amphipathic region which corresponds to the beginning of the amino acid stretch most likely to contain antigenic sites (residues 140 to 320, see above). Seven of the repeat motifs were also detected in the amino-terminal end of the protein encoded by ORF3 (Fig. 4 and 5).

The nucleotide sequence of pWR390 contained a number of perfect 8- to 11-bp inverted repeats and an additional number of longer imperfect inverted repeats (with 80% or greater match) located near the boundaries of the *ipaH* coding sequence; eight of the longest repeats are marked in Fig. 4. Several of these (perfect repeats 1 to 3 and imperfect repeat a) bracket the entire *ipaH* coding region, while the other repeats (4 to 6 and b) are positioned after the region encoding the 14-residue repeat motifs and at the 3′ end of the *ipaH* coding region.

A major portion of the *ipaH* structural gene is repeated on the pWR100 invasion plasmid. When *Hin*dIII-digested pWR100 DNA was hybridized with *ipaH*-specific probes, five distinct bands were detected, comprosing a characteristic *ipaH* signature pattern for this plasmid (9.8, 7.8, 4.5, 2.5, and 1.4 kb; Fig. 1A). Since the nucleotide sequence of IpaH encoded on plasmid pWR390 does not contain a *Hin*dIII restriction site, the five bands detected with probe S39 represent five distinct copies of the *ipaH* gene. To determine whether complete copies of the *ipaH* gene were present in each of these five *Hin*dIII fragments and to delineate the portion of the 2.9-kb pWR390 insert DNA that was present in these fragments, we subdivided the insert DNA into seven smaller segments which were then used to probe *Hin*dIII-digested pWR100 DNA (Fig. 6). Three of these segments overlapped the IpaH coding region: (i) the *Pvu*II-*Ava*I 406-bp segment (PA406); (ii) the *Ava*I-*Sal*I 507-bp segment (AS507); and (iii) the *Sal*I-*Ava*I 568-bp segment (SA568). All five *Hin*dIII fragments of the pWR100 *ipaH* signature pattern hybridized these three *ipaH* ORF probes, suggesting that each of the fragments contained significant parts of the *ipaH* coding region. However, for at least the three smallest *Hin*dIII fragments, the hybridization intensity produced by probe SA568 was noticeably less than that shown by probes PA406 and AS507, indicating that these *ipaH* copies may contain only portions of the SA568 sequence. Restriction mapping of pWR100 DNA with *Sal*I, *Hin*dIII, and *Sal*I-*Hin*dIII digests of pWR100 probed with both SA568 and the entire 2.9-kb pWR390 insert also indicated some truncation in sequences 5′ to the *Sal*I site in the three smallest *Hin*dIII fragments (data not shown). In contrast to the hybridization patterns observed when segments overlapping the *ipaH* ORF were used to probe *Hin*dIII-digested pWR100 DNA, flanking region probes hybridized single *Hin*dIII fragments (Fig. 6). Promoter (AE294) and transcription terminator (HP710) region probes hybridized a single 7.8-kb pWR100 *Hin*dIII band, indicating that this *Hin*dIII fragment was the source of the *ipaH* gene cloned and sequenced in pWR390. Accordingly, this sequence was designated $ipaH_{7.8}$ and its counterparts were named $ipaH_{9.8}$, $ipaH_{4.5}$, $ipaH_{2.5}$, and $ipaH_{1.4}$, respectively. We noted that probe EH253, which overlaps ORF3, only hybridized the $ipaH_{4.5}$ sequence, suggesting that ORF3 is part of the $ipaH_{4.5}$ gene copy. This was confirmed by restriction mapping with the entire 2.9-kb *ipaH* probe (Fig. 1A and 6) and EH253, which showed that a

0 CATAGAAAACCTCCATAAATAAATTACAACTAACTTCTGTTATGTGTAAAATGGAAACTATTAAAACTTAATATCGGAAATGGTAAGTGAAATTTGTATAAATATACAATTTTAAATATT 120

121 TATTCTCACAAATATAAGGTTGACCTAGCATTATGTTCTCTGTAAATAATACACACTCATCAGTTTCTTGCTCCCCCTCTATTAACTCAAACTCAACCAGTAATGAACATTATCTGAGAA 240

241 TCCTGACTGA ATG GGA AAA GAA CTC TTC TCC CGG GAA GAG CGA GGC ATT GCT TTT AAC AGA CTC TCC CAG TGC TTT CAG AAT CAA GAA GCA 331
M G K E L F S R E E R G I A F N R L S Q C F Q N Q E A

332 GTA TTA AAT TTA TCA GAC CTA AAT TTG ACG TCT CTT CCC GAA TTA CCA AAG CAT ATT TCT GCT TTG ATT GTA GAA AAT AAT AAA TTA ACA 421
V L N L S D L N L T S L P E L P K H I S A L I V E N N K L T

422 TCA TTG CCA AAG CTG CCT GCA TTT CTT AAA GAA CTT AAT GCT GAT AAT AAC AGG CTT TCT GTG ATA CCA GAA CTT CCT GAG TCA TTA ACA 511
S L P K L P A F L K E L N A D N N R L S V I P E L P E S L T

512 ACT TTA AGT GTT CGT TCT AAT CAA CTG GAA AAC CTT CCT GTT TTG CCA AAC CAT TTA ACA TCA TTA TTT GTT GAA AAT AAC AGG CTA TAT 601
T L S V R S N Q L E N L P V L P N H L T S L F V E N N R L Y

602 AAC TTA CCG GCT CTT CCC GAA AAA TTG AAA TTT TTA CAT GTT TAT TAT AAC AGG CTG ACA ACA TTA CCC GAC TTA CCG GAT AAA CTG GAA 691
N L P A L P E K L K F L H V Y Y N R L T T L P D L P D K L E

692 ATT CTC TGT GCT CAG CGC AAT AAT CTG GTT ACT TTT CCT CAA TTT TCT GAT AGA AAC AAT ATC AGA CAA AAG GAA TAT TAT TTT CAT TTT 781
I L C A Q R N N L V T F P Q F S D R N N I R Q K E Y Y F H F

782 AAT CAG ATA ACC ACT CTT CCG GAG AGT TTT TCA CAA TTA GAT TCA AGT TAC AGG ATT AAT ATT TCA GGG AAT CCA TTG TCG ACT CGC GTT 871
N Q I T T L P E S F S Q L D S S Y R I N I S G N P L S T R V

872 CTG CAA TCC CTG CAA AGA TTA ACC TCT TCG CCG GAC TAC CAC GGC CCG CAG ATT TAC TTC TCC ATG AGT GAC GGA CAA CAG AAT ACA CTC 961
L Q S L Q R L T S S P D Y H G P Q I Y F S M S D G Q Q N T L

962 CAT CGC CCC CTG GCT GAT GCC GTG ACA GCA TGG TTC CCG GAA AAC AAA CAA TCT GAT GTA TCA CAG ATA TGG CAT GCT TTT GAA CAT GAA 1051
H R P L A D A V T A W F P E N K Q S D V S Q I W H A F D H E

1052 GAG CAT GCC AAC ACC TTT TCC GCG TTC CTT GAC CGC CTT TCC GAT ACC GTC TCT GCA CGC AAT ACC TCC GGA TTC CGT GAA CAG GTC GCT 1141
E H A N T F S A F L D R L S D T V S A R N T S G F R E Q V A

1142 GCA TGG CTG GAA AAA CTC AGT GCC TCT GCG GAG CTT CGA CAG CAG TCT TTC GCT GTT GCT GCT GAT GCC ACT GAG AGC TGT GAG GAC CGT 1231
A W L E K L S A S A E L R Q Q S F A V A A D A T E S C E D R

1232 GTC GCG CTC ACA TGG AAC AAT CTC CGG AAA ACC CTC CTG GTC CAT CAG GCA TCA GAA GGC CTT TTC GAT AAT GAT ACC GGC GCT CTG CTC 1321
V A L T W N N L R K T L L V H Q A S E G L F D N D T G A L L

1322 TCC CTG GGC AGG GAA ATG TTC CGC CTC GAA ATT CTG GAG GAC ATT GCC CGG GAT AAA GTC AGA ACT CTC CAT TTT GTG GAT GAG ATA GAA 1411
S L G R E M F R L E I L E D I A R D K V R T L H F V D E I E

1412 GTC TAC CTG GCC TTC CAG ACC ATG CTC GCA GAG AAA CTT CAG CTC TCC ACT GCC GTG AAG GAA ATG CGT TTC TAT GGC GTG TCG GGA GTG 1501
V Y L A F Q T M L A E K L Q L S T A V K E M R F Y G V S G V

FIG. 4. DNA sequence of the 2.9-kb insert of pWR390. The translated amino acids for the IpaH ORF and for the partial protein sequence encoded by ORF3 are marked with a single-letter code below the nucleotide sequence. Shine-Dalgarno (SD) sequences and potential −10 and −35 transcription initiation elements for the IpaH ORF and ORF3 are underlined with solid lines and marked. A possible transcription termination site for the IpaH ORF is marked by arrows with dashes over T_n residues. Inverted repeat sequences are underlined with solid lines and marked with numbers 1 to 6 for perfect inverted repeats and letters a and b for imperfect inverted repeats. The 14-residue Leu-X_2-Leu-Pro-X-Leu-Pro-X_2-Leu-X_2-Leu repeat motifs for the IpaH ORF and ORF3 are underlined with dashed lines. The nucleotides in parentheses are not part of the 2.9-kb *Eco*RI insert but were included to show the full repeat motif and were obtained from partial sequencing of a contiguous clone. These data have been submitted to GenBank under accession no. M32063.

```
1502 ACA GCA AAT GAC CTC CGC ACT GCC GAA GCT ATG GTC AGA AGC CGT GAA GAG AAT GAA TTT ACG GAC TGG TTC TCC CTC TGG GGA CCA TGG 1591
      T   A   N   D   L   R   T   A   E   A   M   V   R   S   R   E   E   N   E   F   T   D   W   F   S   L   W   G   P   W

1592 CAT GCT GTA CTG AAG CGT ACG GAA GCT GAC CGC TGG GCG CAG GCA GAA GAG CAG AAG TAT GAG ATG CTG GAG AAT GAG TAC TCT CAG AGG 1681
      H   A   V   L   K   R   T   E   A   D   R   W   A   Q   A   E   E   Q   K   Y   E   M   L   E   N   E   Y   S   Q   R

                     3
1682 GTG GCT GAC CGG CTG AAA GCA TCA GGT CTG AGC GGT GAT GCG GAT GCG CAG AGG GAA GCC GGT GCA CAG GTG ATG CGT GAG ACT GAA CAG 1771
      V   A   D   R   L   K   A   S   G   L   S   G   D   A   D   A   Q   R   E   A   G   A   Q   V   M   R   E   T   E   Q

1772 CAG ATT TAC CGT CAG CTG ACT GAC GAG GTA CTG GCC CTG CGA TTG TCT GAA AAC GGC TCA CGA CTG CAC CAT TCA  TAATCACGTCGCATAAGC 1864
      Q   I   Y   R   Q   L   T   D   E   V   L   A   L   R   L   S   E   N   G   S   R   L   H   H   S

                              4                                              --------->  --       --  <---------        ......
1865 ATAAACCGCAGACCGGATTGACTCCGGAAAAACTGTGACCCGATTACGGACCTTAACAACAACCCGTAAATCCTCGCTCAATACCGGCAGGGATTTACGGCGTGCAACTGACTTTTTTGA 1984
                                                                            a              5

1985 GGGGATAACCAACCAGATCGTTTGCTATGGGAATATCGAGACAGTAATGAGTTAAATGATAAAAATTGTTTGAAAATATAGGGGATAAAGATCAATCCAAACTGGATGAAAGTAGAACTG 2104

          -35          -10
2105 GTCACATTAACATGGGTAGACTGATATAACAATCGACGGTTACTGGAAAGACAGGAACATATTCCTCCAGCCGGAATGAAAACGCCGATAAAGCTCTAGGATTGTTTTTTTAAAGACTTT 2224

                                           SD                                  b
2225 CTCGTTTTATTTGCATTAATAGACCAAGATATGAATAGTGAGGGGTTAATAA   ATG AAA CCG ATC AAC AAT CAT TCT TTT TTT CGT TCC CTT TGT GGC TTA 2324
                                                           M   K   P   I   N   N   H   S   F   F   R   S   L   C   G   L

2325 TCA TGT ATA TCT CGT TTA TCG GTA GAA GAA CAG TGT ACC AGA GAT TAC CAC CGC ATC TGG GAT GAC TGG GCT AGG GAA GGA ACA ACA ACA 2414
      S   C   I   S   R   L   S   V   E   E   Q   C   T   R   D   Y   H   R   I   W   D   D   W   A   R   E   G   T   T   T

                                                               2
2415 GAA AAT CGC ATC CAG GCG GTT CGA TTA TTG AAA ATA TGT CTG GAT ACC CGG GAG CCT GTT CTC AAT TTA AGC TTA CTG AAA CTA CGT TCT 2504
      E   N   R   I   Q   A   V   R   L   L   K   I   C   L   D   T   R   E   P   V   L   N   L   S   L   L   K   L   R   S
                                                                                                               -----------

                                                               6
2505 TTA CCA CCA CTC CCT TTG CAT ATA CGT GAA CTT AAT ATT TCC AAC AAT GAG TTA ATC TCC CTA CCT GAA AAT TCT CCG CTT TTG ACA GAA 2594
      L   P   P   L   P   L   H   I   R   E   L   N   I   S   N   N   E   L   I   S   L   P   E   N   S   P   L   L   T   E
     ------------------------------------------                         ----------------------------------------------------

2595 CTT CAT GTA AAT GGT AAC AAC TTG AAT ATA CTC CCG ACA CTT CCA TCT CAA CTG ATT AAG CTT AAT ATT TCA TTC AAT CGA AAT TTG TCA 2684
      L   H   V   N   G   N   N   L   N   I   L   P   T   L   P   S   Q   L   I   K   L   N   I   S   F   N   R   N   L   S
     ---                         ------------------------------------------------------                              --------

                                                               1
2685 TGT CTG CCA TCA TTA CCA CCA TAT TTA CAA TCA CTC TCG GCA CGT TTT AAT AGT CTG GAG ACG TTA CCA GAG CTT CCA TCA ACG CTA ACA 2774
      C   L   P   S   L   P   P   Y   L   Q   S   L   S   A   R   F   N   S   L   E   T   L   P   E   L   P   S   T   L   T
     ----------------------------------------------                         ----------------------------------------------

2775 ATA TTA CGT ATT GAA GGT AAT CGC CTT ACT GTC TTG CCT GAA TTG CCT CAT AGA CTA CAA GAA CTC TTT GTT TCC GGC AAC AGA CTA CAG 2864
      I   L   R   I   E   G   N   R   L   T   V   L   P   E   L   P   H   R   L   Q   E   L   F   V   S   G   N   R   L   Q
     -------                            ------------------------------------------------------                      -------

2865 GAA CTA CCA GAA TTT CCT CAG AGC(TTA AAA TAT TTG)                                                                        2900
      E   L   P   E   F   P   Q   S(  L   K   Y   L)
     ---------------------------------------------
```

FIG. 4—*Continued.*

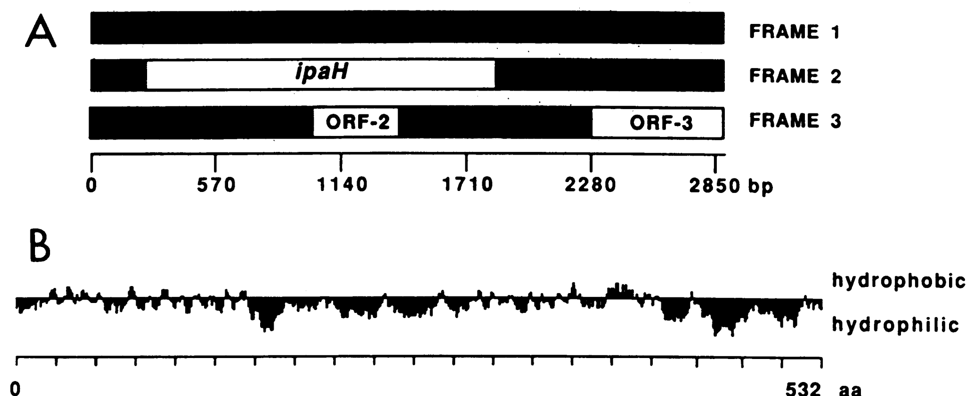

FIG. 5. Schematic representation of major ORFs found in the 2.9-kb insert fragment of pWR390. The ORFs are shown as open boxes. Transcription direction is from left to right. (B) Hydrophobicity profile of IpaH, calculated by the method of Kyte and Doolittle (17) with an amino acid resolution of 15. Hydrophilic regions are found below the base line, and hydrophobic regions are above. The bottom line shows the scale of IpaH in amino acids (aa).

4.6-kb *Eco*RI-*Bam*HI fragment (Fig. 1A, lane 6; Fig. 6) contained both $ipaH_{7.8}$ and $ipaH_{4.5}$ as well as EH253 (unpublished data).

DISCUSSION

In this report, we extended the characterization of Ipa proteins to include a unique 60-kDa antigen, IpaH, produced by *S. flexneri*. It is not known whether IpaH is associated with a particular aspect of the virulence phenotype (i.e., adherence, invasion, intercellular replication, or intracellular dissemination) since IpaH$^-$ mutants have not been isolated, perhaps due, in part, to the reiteration of *ipaH* throughout the *Shigella* invasion plasmid. IpaH protein produced by *E. coli* JM109(pWR390) and HB101(pWR390) cells, however, did not make the bacteria invasive when tested in the HeLa cell invasion assay and did not mediate Congo red dye binding.

Regulation of *ipaH* expression in *S. flexneri* was found to be different from that seen for Ipa antigens B, C, D, and A, all of which are subject to transcriptional control in response to environmental temperature (25, 40) (mediated by the product of the *virR* gene [24]) and also require the products of two positive effectors for their synthesis, *virF* (28) and *virB* (1). In contrast, IpaH synthesis was not temperature responsive since *ipaH* transcription and translation were both demonstrated at 30 and 37°C. Furthermore, during this study, we found that a number of avirulent *S. flexneri* strains, such as M90T-A_3 (which contains a 65-kb deletion in the invasion plasmid encompassing *virG*, *ipaBCDA*, and the *invA* region; see reference 41) and strains that contain *virF* mutations, continue to synthesize the IpaH antigen, indicating that these gene products are not necessary for IpaH synthesis. IpaH$^+$ invasion plasmid deletion strains, such as M90T-A_3, still retain multiple copies of the *ipaH* gene; however, the arrangement of the genes on the invasion plasmid is often altered significantly, as reflected by changes in Southern hybridization patterns (J. M. Buysse, A. B. Hartman, N. Strockbine, and M. M. Venkatesan, manuscript in preparation).

Previous work on the characterization of *S. flexneri* invasion plasmid antigen (*ipa*) genes has shown that these virulence-associated determinants are present as single copies in *Shigella* species and that the corresponding restriction fragments are highly conserved throughout the *Shigella* genus (9, 39). These antigens are also remarkably homogeneous with respect to their antigenic properties and amino acid sequence (2, 13, 25, 39). The distinctive property of the *S. flexneri ipaH* gene is that it is the first recognized multicopy antigen gene of *Shigella* species that is unique to the *Shigella* genus and enteroinvasive *E. coli* (41). Since *ipaH* occurs in multiple copies throughout the *Shigella* invasion plasmid and Southern blot analysis indicated that a major portion of the *ipaH* gene is contained in each copy, it is conceivable that different IpaH antigenic types might be generated if gene conversion occurred between copies that were not completely identical. In fact, preliminary investigations into the structure of the five pWR100 *ipaH* genes, using defined oligonucleotide probes derived from the *ipaH* gene cloned in pWR390, have shown that the copies are not equivalent (M. Venkatesan, A. Hartman, and J. M. Buysse, Abstr. Annu. Meet. Am. Soc. Microbiol. 1989, B92, p. 46; manuscript in preparation). However, no detectable antigenic variation of IpaH in *S. flexneri* has been noted to date. The *ipaH* gene of *S. flexneri* joins a growing list of antigen genes that are carried as multiple copies on their respective bacterial genomes, including the pilus and opacity (protein II) proteins of *Neisseria gonorrhoeae*, the variable major protein of *Borrelia hermsii*, the P1 protein of *Mycoplasma pneumoniae*, and the surface lipoprotein antigen in *Mycoplasma hyorhinis* (6, 7, 36, 38).

Although the role of the IpaH protein, if any, in the expression of the virulence phenotype is unknown at present, its hydrophilic nature is in keeping with its demonstrated immunogenicity in rabbits immunized with *Shigella* antigens (9) and in convalescent humans (unpublished data). The hydrophilic and antigenic nature of IpaH, as well as its presence in water extracts of *Shigella* species that also contain Ipa antigens B, C, and D (26; unpublished data), suggest that IpaH is exposed on the surface of the bacillus. However, IpaH does not have a typical signal peptide in its amino acid sequence (14, 43), a property that it shares with IpaB, IpaC, and IpaD, which are likely located on the surface of the bacillus as well (13, 25, 26, 40). This indicates that an alternative transport mechanism or the action of additional factors may be necessary for the proper positioning of the IpaH molecule on the bacterial cell surface.

The structural implications of the unusual evenly spaced 14-residue repeat motif (six copies) found in the amino-terminal end of IpaH are not known at present. It is noteworthy that seven copies of this motif are found in the

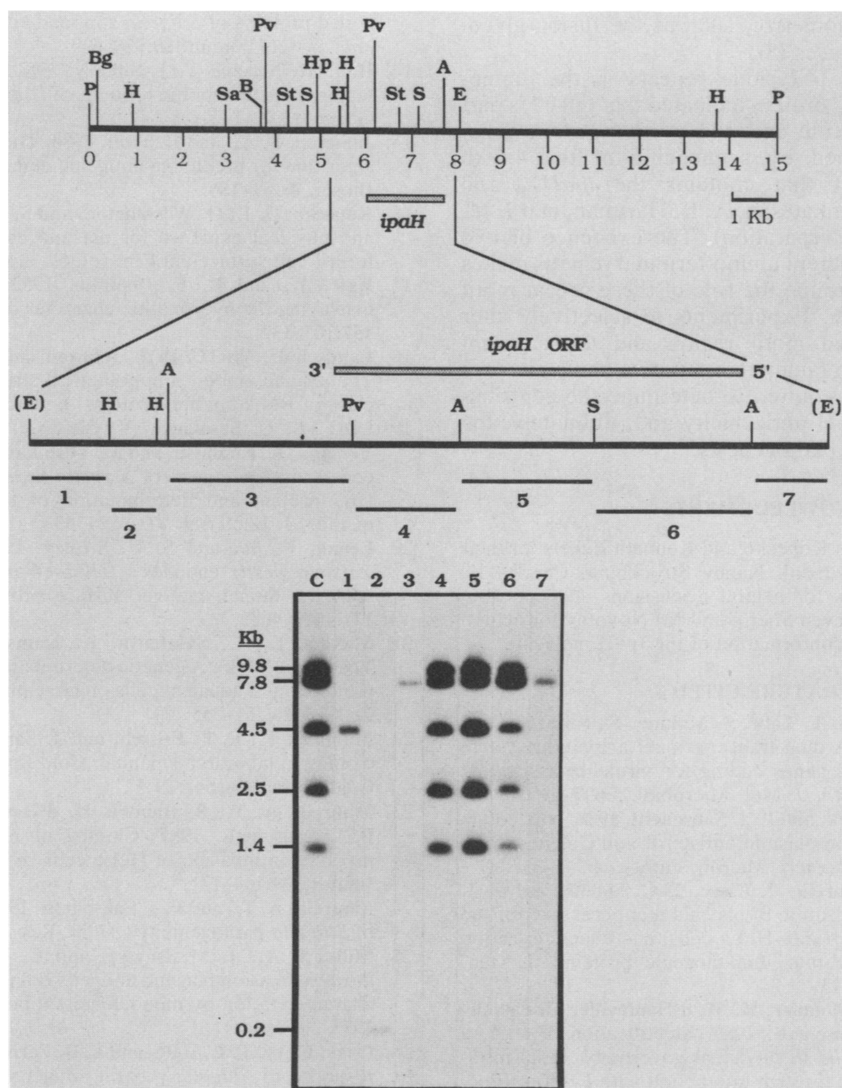

FIG. 6. Restriction map of a 15-kb *Pst*I fragment of pWR100 containing the 2.9-kb *Eco*RI insert of pWR390. The IpaH ORF sequenced from pWR390 is indicated on the enlarged 2.9-kb *Eco*RI insert (the *Eco*RI sites shown on this insert are artificial cloning sites introduced with *Eco*RI linkers and are not found in pWR100). Solid lines at the bottom of the figure indicate the seven individual fragments used in hybridization studies to establish the extent of the repeat region. The fragments are as follows (left to right): (1) *Eco*RI-*Hin*dIII 253-bp fragment (EH253); (2) *Hin*dIII-*Hin*dIII 173-bp fragment (HH173); (3) *Hin*dIII-*Pvu*II 710-bp fragment (HP710); (4) *Pvu*II-*Ava*I 406-bp fragment (PS406); (5) *Ava*I-*Sal*I 507-bp fragment (AS507); (6) *Sal*I-*Ava*I 568-bp fragment (SA568); (7) *Ava*I-*Eco*RI 294-bp fragment (AE294). Lower panel shows hybridization studies to delineate the repeat region of the 2.9-kb fragment. *Hin*dIII-digested pWR100 DNA was hybridized to the total 2.9-kb fragment (lane C) and to fragments 1 to 7 (lanes 1 to 7, respectively). Enzymes used to prepare probe fragments were as follows: A, *Ava*I; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hp, *Hpa*I; P, *Pst*I; Pv, *Pvu*II; Sa, *Sac*I; S, *Sal*I; St, *Stu*I.

amino-terminal end of ORF3 as well as 13 copies in a different bacterial sequence, the YopM protein from *Yersinia pestis* (20). The basic 14-residue repeat motif with six intervening residues containing an asparagine in the fifth position is conserved in all these molecules. A number of models can be proposed to take into account the unique structural features of the IpaH amino-terminal motifs. The regular spacing of the leucine residues (L-X_2-L-P-X-L-P-X_2-L-X_2-L-X_4-N-X) suggests that one structure contains an array of leucine residues that could interdigitate with other IpaH molecules or with other proteins presenting a similar structure, thus facilitating oligomerization, which may be important for the proper conformation and function of IpaH. Alternatively, if turns exist in the regions between the 14-residue repeat motifs, the leucine arrays of the IpaH molecule could interdigitate internally. A third model would contain the leucine residues of the motif arranged on one side of a helix, thus presenting a uniform hydrophobic surface. In fact, the regular spacing of the leucine residues in the α-helical regions between the L-P-X-L-P-X residues supports this last possibility. The six-residue L-P-X-L-P-X portion of the repeat is similar to the polyproline helix found in the avian pancreatic peptide (5, 11); it has been proposed that the hydrophobic surface of this molecule, which is involved in the dimerization of the pancreatic peptide monomer, might be important in receptor binding (5). The regular spacing of the leucine residues also shows some similarity to the leucine heptad repeats (the spacing between the last leucine of one 14-residue repeat and the first leucine of the next motif is L-X_6) found in some DNA-binding proteins (18)

and in proteins that oligomerize, such as the fusion glycoproteins of paramyxoviruses (8).

The presence of the 14-residue repeats in the amino-terminal end of the ORF3 protein indicated that ORF3 is part of a second IpaH molecule encoded by the $ipaH_{4.5}$ gene. This has been confirmed by sequencing of the 4.6-kb *Bam*HI-*Eco*RI fragment that contains the $ipaH_{7.8}$ and $ipaH_{4.5}$ genes (M. M. Venkatesan, A. B. Hartman, and J. M. Buysse, manuscript in preparation). The existence of two IpaH molecules with different amino-terminal regions makes crucial the need to determine the role of these repeat motif regions in IpaH function. Experiments to selectively alter the IpaH amino-terminal motif region and the adjacent amphipathic segment containing the putative antigenic sites of the molecule will be required to determine the contribution of the motifs to IpaH antigenicity and, ultimately, the role of IpaH in *Shigella* pathogenesis.

ACKNOWLEDGMENTS

We are grateful to Dennis Kopecko and Kenneth Eckels for their support in this work. We thank Nancy Strockbine, Charles K. Stover, and Jonathan Mills for helpful discussions and contributions. We are grateful to Steven Sheriff and Jiri Novotny for helpful discussions on the protein conformation of the IpaH molecule.

LITERATURE CITED

1. **Adler, B., C. Sasakawa, T. Tobe, S. Makino, K. Komatsu, and M. Yoshikawa.** 1989. A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. Mol. Microbiol. **3**:627–635.
2. **Baudry, B., M. Kaczorek, and P. J. Sansonetti.** 1988. Nucleotide sequence of the invasion plasmid antigen B and C genes (*ipaB* and *ipaC*) of *Shigella flexneri*. Microb. Pathog. **4**:345–357.
3. **Baudry, B., A. T. Maurelli, P. Clerc, J. C. Sadoff, and P. J. Sansonetti.** 1987. Localization of plasmid loci necessary for the entry of *Shigella flexneri* into HeLa cells, and characterization of one locus encoding four immunogenic proteins. J. Gen. Microbiol. **133**:3403–3413.
4. **Bernardini, M. L., J. Mounier, H. M. d'Hauteville, M. Coquis-Rondon, and P. J. Sansonetti.** 1989. Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. Proc. Natl. Acad. Sci. USA **86**:3867–3871.
5. **Blundell, T. L., J. E. Pitts, I. J. Tickle, S. P. Wood, and C.-W. Wu.** 1981. X-ray analysis (1.4-Å resolution) of avian pancreatic polypeptide: small globular protein hormone. Proc. Natl. Acad. Sci. USA **78**:4175–4179.
6. **Borst, P., and D. R. Greaves.** 1987. Programmed gene rearrangements altering gene expression. Science **235**:658–667.
7. **Boyer, M. J., and K. S. Wise.** 1989. Lipid-modified surface protein antigens expressing size variation within the species *Mycoplasma hyorhinis*. Infect. Immun. **57**:245–254.
8. **Buckland, R., and F. Wild.** 1989. Leucine zipper motif extends. Nature (London) **338**:547.
9. **Buysse, J. M., C. K. Stover, E. V. Oaks, M. Venkatesan, and D. J. Kopecko.** 1987. Molecular cloning of invasion plasmid antigen (*ipa*) genes from *Shigella flexneri*: analysis of *ipa* gene products and genetic mapping. J. Bacteriol. **169**:2561–2569.
10. **Cassie, F., C. Boucher, J. S. Julliot, M. Michael, and J. Denaire.** 1979. Identification and characterization of large plasmids in *Rhizobium meliloti* using agarose gel electrophoresis. J. Gen. Microbiol. **113**:229–242.
11. **Glover, I., I. Haneef, J. Pitts, S. Wood, D. Moss, I. Tickle, and T. Blundell.** 1983. Conformational flexibility in a small globular hormone: X-ray analysis of avian pancreatic polypeptide at 0.98-Å resolution. Biopolymers **22**:293–304.
12. **Hale, T. L., and S. B. Formal.** 1986. Genetics of virulence in *Shigella*. Microb. Pathog. **1**:511–518.
13. **Hale, T. L., E. V. Oaks, and S. B. Formal.** 1985. Identification and antigenic characterization of virulence-associated, plasmid-coded proteins of *Shigella* spp. and enteroinvasive *Escherichia coli*. Infect. Immun. **50**:620–629.
14. **Hall, M. N., and T. J. Silhavy.** 1981. Genetic analysis of the major outer membrane proteins of *Escherichia coli*. Annu. Rev. Genet. **15**:91–142.
15. **Jameson, B. A., and H. Wolf.** 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. Comput. Appl. Biosci. **4**:181–186.
16. **Kopecko, D. J., O. Washington, and S. B. Formal.** 1980. Genetic and physical evidence for plasmid control of *Shigella sonnei* form I cell surface antigen. Infect. Immun. **29**:207–214.
17. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
18. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
19. **Lett, M., C. Sasakawa, N. Okada, T. Sakae, S. Makino, M. Yamada, K. Komatsu, and M. Yoshikawa.** 1989. *virG*, a plasmid-coded virulence gene of *Shigella flexneri*: identification of the *virG* protein and determination of the complete coding sequence. J. Bacteriol. **171**:353–359.
20. **Leung, K. Y., and S. C. Straley.** 1989. The *yopM* gene of *Yersinia pestis* encodes a released protein having homology with the human platelet surface protein GPIb. J. Bacteriol. **171**:4623–4632.
21. **Makino, S., C. Sasakawa, K. Kamata, T. Kurate, and M. Yoshikawa.** 1986. A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in *Shigella flexneri* 2a. Cell **46**:551–555.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Maurelli, A. T., B. Baudry, H. d'Hauteville, T. L. Hale, and P. J. Sansonetti.** 1985. Cloning of plasmid DNA sequences involved in invasion of HeLa cells by *Shigella flexneri*. Infect. Immun. **49**:164–171.
24. **Maurelli, A. T., and P. J. Sansonetti.** 1988. Genetic determinants of *Shigella* pathogenicity. Annu. Rev. Microbiol. **42**:127–150.
25. **Mills, J. A., J. M. Buysse, and E. V. Oaks.** 1988. *Shigella flexneri* invasion plasmid antigens B and C: epitope location and characterization by monoclonal antibodies. Infect. Immun. **56**:2933–2941.
26. **Oaks, E. V., T. L. Hale, and S. B. Formal.** 1986. Serum immune response to *Shigella* protein antigens in rhesus monkeys and humans infected with *Shigella* spp. Infect. Immun. **53**:57–63.
27. **Pal, T., J. W. Newland, B. D. Tall, S. B. Formal, and T. L. Hale.** 1989. Intracellular spread of *Shigella flexneri* associated with the *kcpA* locus and a 140-kilodalton protein. Infect. Immun. **57**:477–486.
28. **Sakai, T., C. Sasakawa, and M. Yoshikawa.** 1988. Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kilodalton *virF* protein. Mol. Microbiol. **2**:589–597.
29. **Sancar, A., A. M. Hack, and W. D. Rupp.** 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. **137**:692–693.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Sansonetti, P. J., T. L. Hale, and E. V. Oaks.** 1985. Genetics of virulence in enteroinvasive *Escherichia coli*, p. 74–77. *In* D. Schlessinger (ed.), Microbiology—1985. American Society for Microbiology, Washington, D.C.
32. **Sasakawa, C., K. Kamata, T. Sakai, S. Makino, M. Yamada, N. Okada, and M. Yoshikawa.** 1988. Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in *Shigella flexneri* 2a. J. Bacteriol. **170**:2480–2484.
33. **Sasakawa, C., K. Kamata, T. Sakai, S. Y. Murayuma, S. Makino, and M. Yoshikawa.** 1986. Molecular alteration of the 140-megadalton plasmid associated with the loss of virulence and Congo red binding activity in *Shigella flexneri*. Infect. Immun. **51**:470–475.

34. **Silhavey, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions, p. 140–141. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
36. **Stoenner, H. G., T. Dodd, and C. Larsen.** 1982. Antigenic variation of *Borrelia hermsii*. J. Exp. Med. **156**:1297–1311.
37. **Stover, C. K., J. Kemper, and R. C. Marsh.** 1988. Molecular cloning and characterization of *supA/newD*, a gene substitution system for the *leuD* gene of *Salmonella typhimurium*. J. Bacteriol. **170**:3115–3124.
38. **Su, C.-J., A. Chavoya, and J. B. Baseman.** 1988. Regions of *Mycoplasma pneumoniae* cytoadhesin Pa structural gene exist as multiple copies. Infect. Immun. **56**:3157–3161.
39. **Venkatesan, M. M., J. M. Buysse, E. Vandendries, and D. J. Kopecko.** 1988. Development and testing of invasion-associated DNA probes for detection of *Shigella* spp. and enteroinvasive *Escherichia coli*. J. Clin. Microbiol. **26**:261–266.
40. **Venkatesan, M. M., J. M. Buysse, and D. J. Kopecko.** 1988. Characterization of invasion plasmid antigen genes (*ipaBCD*) from *Shigella flexneri*. Proc. Natl. Acad. Sci. USA **85**:9317–9321.
41. **Venkatesan, M. M., J. M. Buysse, and D. J. Kopecko.** 1989. Use of *Shigella flexneri ipaC* and *ipaH* gene sequences for the general identification of *Shigella* spp. and enteroinvasive *Escherichia coli*. J. Clin. Microbiol. **27**:2687–2691.
42. **Watanabe, H., and A. Nakamura.** 1986. Identification of *Shigella sonnei* form I plasmid genes necessary for cell invasion and their conservation among *Shigella* species and enteroinvasive *Escherichia coli*. Infect. Immun. **53**:352–358.
43. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5164.