JOURNAL OF BACTERIOLOGY, Apr. 1993, p. 2334–2346
0021-9193/93/082334-13$02.00/0

Vol. 175, No. 8

Eight Genes in Region 5 That Form an Operon Are Essential for Invasion of Epithelial Cells by *Shigella flexneri* 2a

CHIHIRO SASAKAWA,* KEIKO KOMATSU, TORU TOBE, TOSHIHIKO SUZUKI, AND MASANOSUKE YOSHIKAWA

Department of Bacteriology, Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

Received 30 December 1992/Accepted 9 February 1993

The 7-kb region 5 on the large 230-kb plasmid pMYSH6000 in *Shigella flexneri* 2a YSH6000 is one of the virulence-associated DNA segments required for the invasion of epithelial cells (C. Sasakawa, K. Kamata, T. Sakai, S. Makino, M. Yamada, N. Okada, and M. Yoshikawa, J. Bacteriol. 170:2480–2484, 1988). To elucidate the functional organization of region 5 and to determine the virulence-associated genes encoded by region 5, we performed insertion and deletion mutagenesis, DNA subcloning, and complete nucleotide sequencing of region 5 and found that region 5 contained 11 open reading frames (ORFs) named ORF-1 through ORF-11 which could be translated into proteins with molecular masses of 15.1, 47.5, 13.2, 33.0, 33.4, 24.2, 9.4, 28.5, 39.9, 9.1, and 10.4 kDa, respectively. Complementation tests of the 14 Tn*5*-induced noninvasive mutants of region 5 with the above plasmid constructs have indicated that region 5 consists of an operon and that ORF-2 through ORF-9, but not ORF-1, ORF-10, and ORF-11, are essential for invasion, and 7 of 8 ORFs (ORF-2 and ORF-4 through ORF-9) and presumably the remaining ORF (ORF-3) are required for secretion of the Ipa proteins. The transcriptional organization, as determined by a promoter-proving vector, S1 nuclease protection, and primer extension RNA sequencing analysis revealed that region 5 is transcribed from a promoter located 47 bp upstream of the 5′ end of ORF-2 for the 47.5-kDa protein and that the promoter activity identified was regulated by the *virB* gene, the transcriptional activator on the 230-kb plasmid.

The ability of shigellae to invade epithelial cells is a prerequisite for the pathogenesis of bacillary dysentery (19). The large 230-kb plasmid of *S. flexneri* has been shown to encode various classes of virulence determinants such as those required for epithelial cell invasion (3–6, 8, 9, 11, 31, 36, 37), spreading into adjacent cells (2, 10, 20, 23), and regulation of the virulence genes (1, 28, 35).

Among the virulence proteins encoded by the large plasmid of *S. flexneri*, IpaB, IpaC, and IpaD, three immunogenic proteins encoded by the *ipaBCD* genes in the *ipa* operon (8, 11, 31, 36), have been shown to play the most direct role in the invasion of epithelial cells (30). High et al. (16) have recently shown that the IpaB protein is required for contact-mediated hemolysis and lysis of phagocytic vacuoles by bacteria. Several studies have shown that the IpaB and IpaC proteins are secreted onto the bacterial surface and/or into the external medium (3–6, 17, 22, 37). The ability to secrete Ipa proteins has been shown to be crucial for the invasiveness of *S. flexneri*, since mutants incapable of secreting these proteins become noninvasive, even with full production of the IpaB, IpaC, and IpaD proteins within the bacterium (3–6, 17, 22, 37). These studies have indicated that the secretion of Ipa proteins requires the expression of other genes encoded by a 21-kb DNA sequence consisting of regions 3, 4, and 5, which is located upstream of the *ipa* operon on the large plasmid (33). These genes include *mxiA* on pSf2a140 of *S. flexneri* 2a 2457T (5, 6, 17), *mxiHIJMED* (2, 3), and *spa* on pWR100 of *S. flexneri* 5 (37). These genes have been shown to correspond to a part of region 3 (*mxiHIJMED*), region 4 (*mxiA*), and a part of region 5 (*spa*) on pMYSH6000, respectively, by comparing the restriction maps (33). Although the exact roles of the proteins involved in Ipa protein secretion are still obscure, some intriguing characteristics have emerged from analysis of their amino acid sequences. For example, the MxiA protein showed significant homology to transmembrane proteins, such as InvA of *Salmonella typhimurium* (12, 13), LcrD of *Yersinia pestis* (26) and *Yersinia enterocolitica* (12), and FlbF of *Caulobacter crescentus* (27). MxiJ and MxiH showed homology to YscJ and YscF proteins, respectively (3), secretion factors of the *Yersinia* Yop proteins (21). The Spa47 protein (37), one of the five proteins encoded by the *spa* locus, showed significant homology to the subunits of the bacterial proton-translocating F_0F_1 ATPase (15) and to a flagellum export protein, FliI, of *S. typhimurium* (38).

It has been shown that the two contiguous 1.25- and 4.25-kb *Hin*dIII segments containing the *spa* locus on pWR100 possess five genes named *spa15*, *spa47*, *spa32*, *spa33*, and *spa24* (37). A *spa47*::Tn*5* insertion mutant (pHS1059) that was noninvasive (Inv^-) and did not exhibit binding to Congo red (Pcr^-) failed to present the IpaB and IpaC proteins on the bacterial surface (Spa^-) (37). However, because of the polar nature of Tn*5* insertions, which of the *spa* genes on pWR100 are actually involved in the Spa phenotype remains to be elucidated. In addition, the 7-kb region 5 DNA sequence on pMYSH6000 bracketed by Tn*5* insertions of S76 and S9 contains an unknown 2.6-kb region downstream of the *spa* locus (33, 37). We thus undertook analysis of the functional organization of the whole DNA sequence of region 5 and sought the virulence determinants required for the secretion of IpaB, IpaC, and IpaD proteins.

The results presented here indicate that region 5 consists of a single operon transcribed from just downstream of region 4 (*mxiA*) and that region 5 promoter activity depends upon the *virB* gene, the transcriptional activator on the large plasmid (1, 35). Our results provide evidence that region 5 contains eight virulence-associated genes and that the seven

* Corresponding author.

genes tested are involved in the Spa function, the surface presentation of IpaB, IpaC, and IpaD proteins.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are shown in Table 1.

Determination of Tn*5* insertions. The sites of Tn*5* insertions into region 5 were previously assigned to each *Sal*I or *Eco*RI fragment of pMYSH6000 (33). To determine the precise site of each Tn*5* insertion in the nucleotide sequence of region 5, the *Sal*I or *Sal*I-*Eco*RI segment of each Tn*5* insertion mutant of pMYSH6000 was shotgun cloned into pBR322, transformed into *Escherichia coli* MC1061, and selected for Km^r clones. The resulting plasmids were made up of part of Tn*5* (IS*50*L and Km^r gene), IS*50*L-flanking DNA, and pBR322. The nucleotide sequences of the IS*50*L-flanking DNA region were obtained with a primer hybridized to the nucleotides from positions 41 to 62 from the outside end of IS*50*L (7).

DNA sequencing. The sequence of region 5 was determined by the chain termination method of Sanger et al. (29), using a Sequence 7-deaza-dGTP kit (United States Biochemical Corp.), following cloning into pUC118 and pUC119.

CAT assay. Bacteria grown in LN broth (32) at 37°C were harvested, suspended in 0.5 ml of 0.2 M Tris-HCl (pH 7.5), and sonicated (model W-225R; Branson Sonic Power Co.) for 1 min in short bursts (of about 10 s each). The supernatants were collected by centrifugation at 10,000 × *g* for 5 min at 4°C. The extracts were incubated with [^{14}C]chloramphenicol (Amersham Corp., Arlington Heights, Ill.) to determine chloramphenicol acetyltransferase (CAT) activity as previously described by Gorman et al. (14).

Gene product analysis. Protein products specific for the cloned fragment were analyzed by the T7 RNA polymerase-dependent promoter system with strain K38/pGP1-2 (34) harboring various cloned parts of region 5 on pT7-5 or pT7-6.

Primer extension analysis. For primer extension analysis, a ^{32}P-end-labeled synthetic oligonucleotide, 5′-TCGAAAC ATCGCTAAGAC-3′, was mixed with RNA prepared from strain YSH6200 carrying pKK232-8-1.25-kb *Hin*dIII clone with or without pCHR405 (50 ng of primer to 50 μg of RNA), and 50 U of Moloney murine leukemia virus reverse transcriptase was added to a 250-μl reaction mixture containing 50 mM Tris-HCl (pH 8.6), 40 mM KCl, 1 mM MnCl, 1 mM dithiothreitol, 1.0 mM dATP, 1.0 mM dGTP, 1.0 mM dTTP, 1.0 mM dCTP, 1 U of RNasin per μl, and 50 μg of actinomycin D per ml. The reaction was allowed to continue for 2 h at 37°C and terminated by adding 1 μl of 0.5 M EDTA and 5 mg of RNase per ml, and the reaction mixture was incubated for 30 min. A portion of the reaction mixture was analyzed on a 6% polyacrylamide–7 M urea sequencing gel, along with the product of a Sanger sequencing reaction generated with the same primer and the 1.25-kb *Hin*dIII fragment single-stranded template.

Invasion assay. Invasion assays were carried out by performing the contact hemolysis activity assay (30) and the focus plaque-forming assay (33).

Colony immunoblots. Colony immunoblots were carried out by the methods described by Venkatesan et al. (37). Antiserum specific for IpaB, IpaC, or IpaD protein was generated by immunizing rabbits with synthetic IpaB, IpaC, or IpaD peptide conjugated to keyhole limpet hemocyanin. Immunoblots were developed with alkaline phosphatase-labelled protein A as described previously (23).

Nucleotide sequence accession number. The nucleotide sequence of region 5 has been deposited in the DDBJ database under accession number D13663.

RESULTS

Complementation analysis of Tn*5* insertion mutants of region 5. To identify the virulence-associated determinants in region 5 and examine whether region 5 consists of an operon, various restriction fragments containing part or all of region 5 were cloned from pMYSH6000 into restriction sites in the tetracycline resistance gene on pCHR404, and the resulting plasmids were then tested for the ability to restore the Inv^+ phenotype to the 14 Tn*5* insertion mutants of region 5 (Fig. 1 and Table 2). The 15.8-kb *Sal*I fragment clone (pCHR503-2) and the 13.0-kb *Hpa*I-*Sma*I segment clone (pCHR512-2), both containing the entire region 5 DNA sequence, restored the Inv^+ phenotype to all of the Tn*5* insertion mutants. The same was true for the 7.8-kb subclone (pCHR576-1) bracketed by the leftmost Tn*5* insertion (S76) and the *Sma*I site. In contrast, when the DNA sequence on the 7.8-kb subclone in pCHR576-1 was deleted from the *Sma*I site to the *Stu*I site, the resulting deletion derivative pCHR576-4 did not restore the Inv^+ phenotype to any of the Tn*5* insertion mutants (Fig. 1 and Table 2). Similarly, none of the other subclones lacking some DNA region of the 3′ portion of region 5, such as pCHR507-1, pCHR508-1, and pCHR517-1, restored the Inv^+ phenotype. In contrast, the other class of subclones containing at least some of the 3′ portion of region 5, e.g., pCHR552-1, pCHR553-1, and pCHR510-1, restored the Inv^+ phenotype to the Tn*5* insertion mutants as far as the DNA segment used for the complementation tests covered the open reading frame (ORF) inserted by Tn*5*. For example, pCHR510-1 containing the 2.6-kb *Hin*dIII fragment, restored the Inv^+ phenotype to M339, M306, and S104 but not to the others (Fig. 1 and Table 2). These results suggested that region 5 consisted of an operon, which is perhaps transcribed from the left to the right in the restriction map shown in Fig. 1.

To estimate the number of virulence-associated determinants in region 5, a series of fill-in and linker-insertion mutations of region 5 were generated at the restriction sites on pCHR512-2, pCHR509-1, and pCHR510-1 (Fig. 1), and the resulting mutants were tested for the ability to restore the Inv^+ phenotype to each of the Tn*5* insertion mutants. The resulting mutant derivatives (pCHR512-3 and -4, pCHR509-3, -4, -5, and -6, and pCHR510-3), summarized in Table 1, restored the Inv^+ phenotype to those mutants whose Tn*5* insertion sites were located to the 3′ side of the linker (or fill-in) mutation in the introduced plasmid. On the basis of the profiles of Tn*5* insertion mutants positively complemented by the various plasmids constructed above (Fig. 1 and Table 2), at least seven virulence-associated loci were predicted to exist in region 5.

Nucleotide sequence of region 5. The 7,874-bp DNA sequence encompassing the three contiguous 1.25-, 4.25-, and 2.6-kb *Hin*dIII fragments of pMYSH6000 was determined (Fig. 2). Analysis of this 7,874-bp DNA sequence allowed the placement of a total of 11 ORFs in the same orientation (Fig. 2). These ORFs were designated ORF-1 through ORF-11 from the 5′ to the 3′ end of region 5 DNA (Fig. 1 and 2). The nucleotide sequences of the 1.25- and 4.25-kb *Hin*dIII fragments of region 5 determined were the same as that of the *spa* locus on pWR100 (37). Thus, ORF-2, ORF-4, ORF-5, and ORF-6, the four ORFs found in the two contiguous 4.25- and 2.6-kb *Hin*dIII fragments in region 5 corresponded to *spa47*, *spa32*, *spa33*, and *spa24* on pWR100,

TABLE 1. Bacterial strains and plasmids

Bacterial strain or plasmid	Relevant characteristic(s)	Reference or source
Bacterial strains		
S. flexneri 2a		
YSH6000	Virulent strain	33
YSH6200	A virulent strain of YSH6000 cured of the 230-kb plasmid pMYSH6000	28
M223	YSH6000 carrying pMYSH6000, Inv$^-$ (*virB*::Tn*5*)	33
S76	YSH6000 carrying pMYSH6000, Inv$^+$ (region 5::Tn*5*)	33
S279	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S343	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
M39	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
M253	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
N1460	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
N1685	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
N873	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
N410	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S1	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S87	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
M329	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
M300	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S282	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S188	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
M339	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
M306	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
N1345	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S104	YSH6000 carrying pMYSH6000, Inv$^-$ (region 5::Tn*5*)	33
S9	YSH6000 carrying pMYSH6000, Inv$^+$ (region 5::Tn*5*)	33
S343-1	S343 carrying pCHR576-1, Inv$^+$	This study
S343-2	S343 carrying pCHR509-1, Inv$^-$	This study
N873-1	S873 carrying pCHR509-1, Inv$^+$	This study
N873-2	S873 carrying pCHR509-4, Inv$^-$	This study
S87-1	S87 carrying pCHR509-1, Inv$^+$	This study
S87-2	S87 carrying pCHR509-5, Inv$^-$	This study
M300-1	M300 carrying pCHR509-1, Inv$^+$	This study
M300-2	M300 carrying pCHR509-6, Inv$^-$	This study
S188-1	S188 carrying pCHR509-6, Inv$^+$	This study
S188-2	S188 carrying pCHR510-1, Inv$^-$	This study
M339-1	M339 carrying pCHR509-6, Inv$^+$	This study
M339-2	M339 carrying pCHR510-3, Inv$^-$	This study
S104-1	S104 carrying pCHR509-6, Inv$^+$	This study
S104-2	S104 carrying pCHR510-4, Inv$^-$	This study
E. coli K-12 MC1061	*araD139* Δ(*ara-leu*) Δ*lacX74 galU galK hsdR rpsL*	31
Plasmids		
pCHR404	pBR322 derivative with a part of Ap^r region replaced by the 1.2-kb Tp^r fragment	35
pCHR405	pACYC184 carrying P_{tac}-*virB* operon fusion	This study
pKK232-8	A promoter-proving vector	Pharmacia
pCHR503-2	pCHR404-cloned 15.8-kb *Sal*I segment containing a part of region 4 and region 5	This study
pCHR512-2	pCHR404-cloned 13.0-kb *Hpa*I-*Sma*I segment containing a part of region 3 and regions 4 and 5	This study
pCHR512-3	Fill-in derivative at the second left *Hin*dIII site of pCHR512-2	This study
pCHR512-4	Fill-in derivative at the second right *Hin*dIII site of pCHR512-2	This study
pCHR576-1	pCHR404-cloned 10.4-kb *Sma*I segment bracketed by the *Sma*I sites in Tn*5* of S76 and downstream of region 5 (region 5 portion is 7.8 kb)	This study
pCHR576-3	Fill-in derivative at the left *Eco*RI site of pCHR576-1	This study
pCHR576-4	Deletion derivative between the *Stu*I and *Sma*I sites of pCHR576-1	This study
pCHR507-1	pCHR404-cloned 7.0-kb *Bgl*II segment containing a part of regions 3 and 5 and region 4	This study
pCHR517-1	pCHR404-cloned 11.5-kb *Eco*RI fragment containing region 4 and a part of regions 3 and 5	This study
pCHR508-1	pCHR404-cloned 4.25-kb *Hin*dIII fragment of region 5	This study
pCHR509-1	pCHR404-cloned 4.25- and 2.6-kb *Hin*dIII fragments of region 5	This study
pCHR509-3	Fill-in derivative at the left *Eco*RI site of pCHR509-1	This study
pCHR509-4	Linker-insertion derivative at the leftmost *Eco*RV site of pCHR509-1	This study
pCHR509-5	Fill-in derivative at the right *Eco*RI site of pCHR509-1	This study
pCHR509-6	Fill-in derivative at the *Bgl*II site of pCHR509-1	This study

Continued on following page

TABLE 1—*Continued*

Bacterial strain or plasmid	Relevant characteristic(s)	Reference or source
pCHR552-1	pCHR404-cloned 4.5-kb *Eco*RI-*Xho*I fragment of region 5	This study
pCHR553-1	pCHR404-cloned 3.7-kb *Bgl*II-*Xho*I fragment of region 5	This study
pCHR510-1	pCHR404-cloned 2.6-kb *Hin*dIII fragment of region 5	This study
pCHR510-3	Fill-in derivative at the right *Pst*I site of pCHR510-1	This study
pCHR510-4	Fill-in derivative at the left *Afl*II site of pCHR510-1	This study
pCHR510-5	Fill-in derivative at the *Eco*T14 site of pCHR510-1	This study
pCHR510-6	Deletion derivative between the *Bcl*I sites of pCHR510-1	This study
pCHR510-7	Fill-in derivative at the right *Afl*II site of pCHR510-1	This study

respectively (37), although an additional ORF, designated ORF-3, located at bp 2402 to 2740, between ORF-2 and ORF-4, was found in this study. The sites of 17 Tn*5* insertions from S76 to S9 were also determined and assigned to the nucleotide sequence of region 5 (Fig. 1 and 2). Accordingly, ORF-1, whose 5′ end started 12 bp downstream of the 3′ end of *mxiA* (region 4), contained S76, and ORF-10 contained S9 (Inv$^+$), while ORF-11 was outside region 5, implying that these three ORFs were irrelevant to the Inv$^+$ phenotype. Consequently, these results together with the complementation tests of Tn*5* insertion mutants with the fill-in and linker-insertion mutations in region 5 (Table 2) led us to conclude that eight ORFs, ORF-2 through ORF-9, were the virulence-associated genes.

The molecular mass of each gene product deduced from the nucleotide sequence of ORF-2 through ORF-9 showed that those ORFs encoded 47.5-, 13.2-, 33.0-, 33.4-, 24.2-, 9.4-, 28.5-, and 39.9-kDa proteins, respectively. The amino

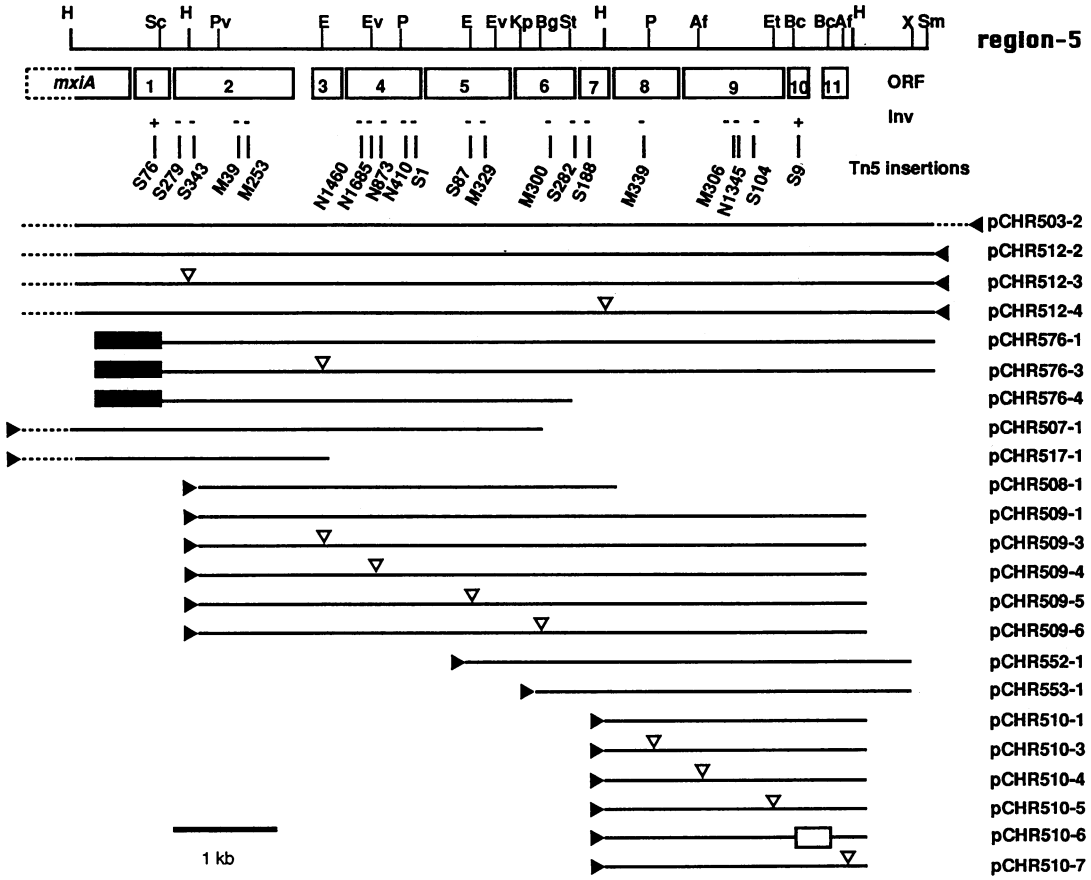


FIG. 1. Physical and genetic map of region 5. The top line represents the restriction sites on the 8.7-kb DNA segment of pMYSH6000 containing region 5. The numbered open boxes indicate the ORFs, assigned on the basis of the nucleotide sequence (see Fig. 2). The vertical bars indicate the sites of Tn*5* insertions, and the + and − signs shown over the bars indicate the Inv$^+$ and Inv$^-$ phenotypes, respectively. The bars at the bottom of the figure represent pMYSH6000 fragments cloned into pCHR404. Symbols: ▶, orientation of the promoter readthrough from pCHR404; ▽, linker-insertion or fill-in mutation; ■, Tn*5* portion; □, deletion region. Restriction site abbreviations: Bc, *Bcl*I; Bg, *Bgl*II; E, *Eco*RI; Et, *Eco*T14; Ev, *Eco*RV; H, *Hin*dIII; Af, *Afl*II; Kp, *Kpn*I; P, *Pst*I; Pv, *Pvu*II; Sc, *Sca*I; Sm, *Sma*I; St, *Stu*I; X, *Xho*I.

TABLE 2. Complementation analysis of region 5

Plasmid	Phenotype[a] conferred by Tn5-inserted avirulent mutant at the following ORF:													
	ORF-2			ORF-4			ORF-5		ORF-6		ORF-7	ORF-8	ORF-9	
	S279	S343	M253	N1460	N873	S1	S87	M329	M300	S282	S188	M339	M306	S104
pCHR503-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pCHR512-2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pCHR512-3	−	−	−	+	+	+	+	+	+	+	+	+	+	+
pCHR512-4	−	−	−	−	−	−	−	−	−	−	−	+	+	+
pCHR576-1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pCHR576-3	−	−	−	+	+	+	+	+	+	+	+	+	+	+
pCHR576-4	−	−	−	−	−	−	−	−	−	−	−	−	−	−
pCHR517-1	−	−	−	−	−	−	−	−	−	−	−	−	−	−
pCHR507-1	−	−	−	−	−	−	−	−	−	−	−	−	−	−
pCHR508-1	−	−	−	−	−	−	−	−	−	−	−	−	−	−
pCHR509-1	−	−	−	+	+	+	+	+	+	+	+	+	+	+
pCHR509-3	−	−	−	−	−	−	+	+	+	+	+	+	+	+
pCHR509-4	−	−	−	−	−	−	+	+	+	+	+	+	+	+
pCHR509-5	−	−	−	−	−	−	−	−	+	+	+	+	+	+
pCHR509-6	−	−	−	−	−	−	−	−	−	−	+	+	+	+
pCHR552-1	−	−	−	−	−	−	−	−	+	+	+	+	+	+
pCHR553-1	−	−	−	−	−	−	−	−	−	−	+	+	+	+
pCHR510-1	−	−	−	−	−	−	−	−	−	−	−	+	+	+
pCHR510-3	−	−	−	−	−	−	−	−	−	−	−	−	+	+
pCHR510-4	−	−	−	−	−	−	−	−	−	−	−	−	−	−
pCHR510-5	−	−	−	−	−	−	−	−	−	−	−	−	−	−
pCHR510-6	−	−	−	−	−	−	−	−	−	−	−	+	+	+
pCHR510-7	−	−	−	−	−	−	−	−	−	−	−	+	+	+

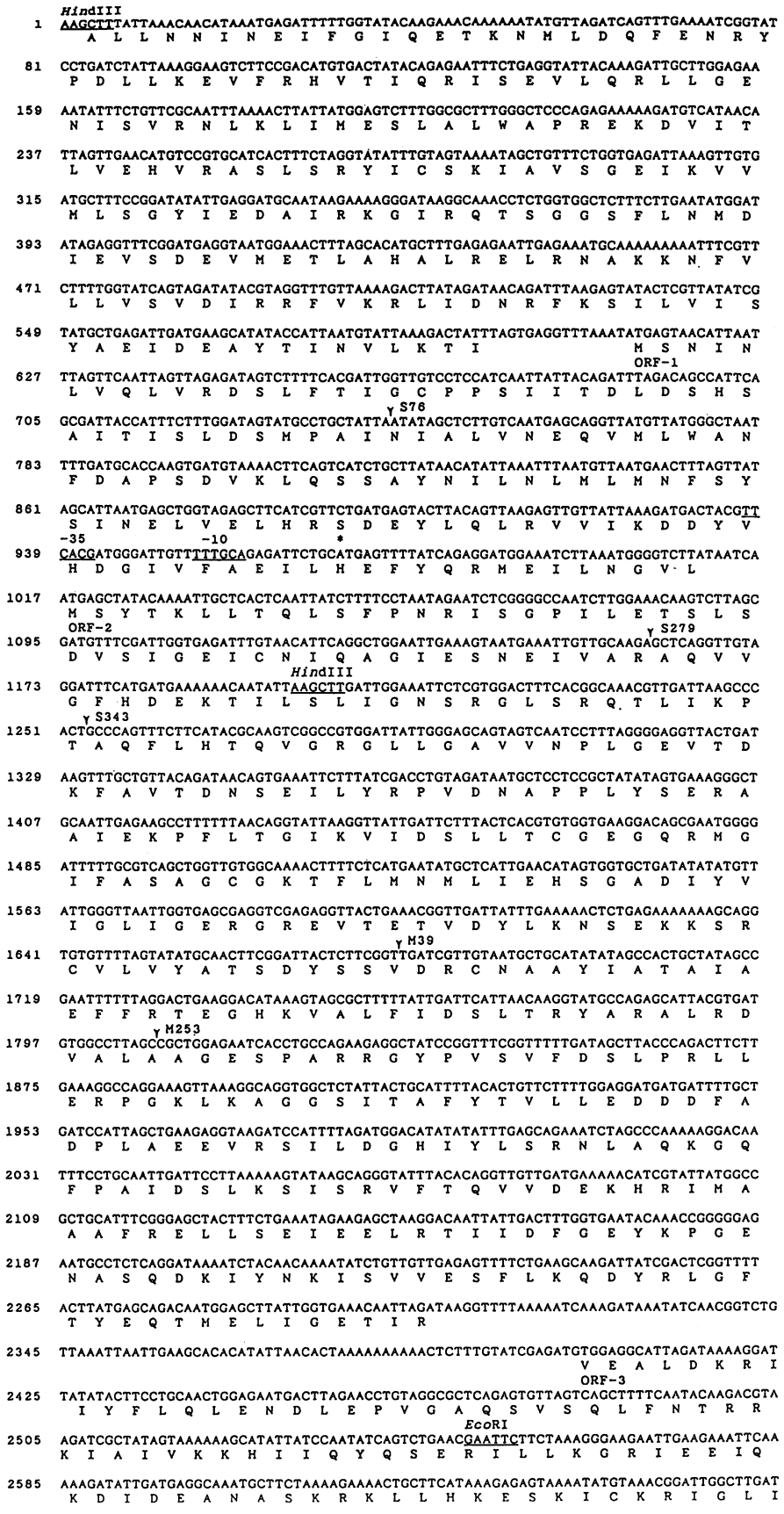
[a] The + and − signs indicate invasion-positive and -negative phenotypes, respectively.

acid sequence of the 47.5-kDa protein encoded by ORF-2 showed striking homology to the proton-translocating F_0F_1 ATPase of bacteria (15), mitochondria (24), and chloroplasts (15), as reported by Venkatesan et al. (37), while the amino acid sequences of the other proteins encoded by ORF-3 to ORF-9 showed no significant homology with the other known proteins in the SWISS-PROT R22.0 data base. Analysis of the hydropathy profiles of the 24.2-, 9.4-, 28.5-, and 39.9-kDa proteins by the method of Kyte and Doolittle (18) revealed a hydrophobic nature. Although none of these proteins possessed potential signal sequences required for the transport of proteins at their N-terminus amino acid sequences, the amino acid sequences of the four proteins did contain putative membrane-spanning domains (data not shown).

Identification of the promoter region for the expression of region 5. To determine the transcriptional organization of region 5, the functional promoter region was sought. Nine contiguous *Hin*dIII fragments containing regions 3, 4, and 5 (Fig. 3) were cloned into a promoter-probing vector, pKK232-8, in both orientations and assayed for the level of CAT production. Since the expression of region 5 and of the *ipa* and region 3.4 operons are regulated by the transcriptional activator *virB* gene (35), each insert was introduced into strain MC1061 with or without the P*tac-virB* fusion plasmid (pCHR405), and the MIC of chloramphenicol was determined in the presence or absence of 1 mM isopropyl-β-D-thiogalactoside (IPTG) at 37°C. The results showed that the 1.9-kb *Hin*dIII fragment containing regions 2 (*ipaBCD*) and 3 and the 1.25-kb *Hin*dIII fragment containing regions 4 and 5 in the correct orientation resulted in MICs of 350 and 250 μg/ml, respectively, in the presence of IPTG. In the absence of IPTG, both MICs decreased >50%. To quantify the activities of the putative promoters, each *Hin*dIII insert in pKK232-8 in the correct orientation was assayed for CAT activity. Although some CAT activity was observed in the 2.3-, 2.7-, 1.6-, and 2.6-kb *Hin*dIII fragments, the levels of CAT production of the 1.9- and 1.25-kb *Hin*dIII fragments were 50 and 20 times higher than those of the other fragments. The levels of CAT activity of only the 1.9- and 1.25-kb *Hin*dIII fragments responded to the activation of the *virB* transcription (Table 3), supporting our premise that regions 2, 3.4, and 5 consist of individual operons (35) (Fig. 3).

To further localize the functional promoter for region 5, the RNA extracted from strains YSH6000 (wild type) and

FIG. 2. Complete nucleotide sequence of region 5. The sequence of the sense strand is shown, as well as the amino acid sequences for the 11 ORFs. The sequence is numbered from the leftmost *Hin*dIII site to the rightmost *Hin*dIII site (Fig. 1). The first 198 amino acid residues in front of ORF-1 represent the 3′ portion of MixA protein (6, 12, 37). The arrowheads over the sequence represent the sites of Tn*5* insertions. The −10 and −35 regions of the possible promoter sequence are placed by the determination of transcription initiation site (asterisk) (see Fig. 4). The lines beneath the nucleotides indicate the putative transcriptional terminator sequence (see Discussion).


HindIII
AAGCTTTATTAAACAACATAAATGAGATTTTTGGTATACAAGAAACAAAAAATATGTTAGATCAGTTTGAAAATCGGTAT
A L L N N I N E I F G I Q E T K N M L D Q F E N R Y
CCTGATCTATTAAAGGAAGTCTTCCGACATGTGACTATACAGAGAATTTCTGAGGTATTACAAAGATTGCTTGGAGAA
P D L L K E V F R H V T I Q R I S E V L Q R L L G E
AATATTTCTGTTCGCAATTTAAAACTTATTATGGAGTCTTTGGCGCTTTGGGCTCCCAGAGAAAAAGATGTCATAACA
N I S V R N L K L I M E S L A L W A P R E K D V I T
TTAGTTGAACATGTCCGTGCATCACTTTCTAGGTATATTTGTAGTAAAATAGCTGTTTCTGGTGAGATTAAAGTTGTG
L V E H V R A S L S R Y I C S K I A V S G E I K V V
ATGCTTTCCGGATATATTGAGGATGCAATAAGAAAAGGGATAAGGCAAACCTCTGGTGGCTCTTTCTTGAATATGGAT
M L S G Y I E D A I R K G I R Q T S G G S F L N M D
ATAGAGGTTTCGGATGAGGTAATGGAAACTTTAGCACATGCTTTGAGAGAATTGAGAAATGCAAAAAAAAATTTCGTT
I E V S D E V M E T L A H A L R E L R N A K K N F V
CTTTTGGTATCAGTAGATATACGTAGGTTTGTTAAAAGACTTATAGATAACAGATTTAAGAGTATACTCGTTATATCG
L L V S V D I R R F V K R L I D N R F K S I L V I S
TATGCTGAGATTGATGAAGCATATACCATTAATGTATTAAAGACTATTTAGTGAGGTTTAAATATGAGTAACATTAAT
Y A E I D E A Y T I N V L K T I M S N I N
ORF-1
TTAGTTCAATTAGTTAGAGATAGTCTTTTCACGATTGGTTGTCCTCCATCAATTATTACAGATTTAGACAGCCATTCA
L V Q L V R D S L F T I G C P P S I I T D L D S H S
S78
GCGATTACCATTTCTTTGGATAGTATGCCTGCTATTAATATAGCTCTTGTCAATGAGCAGGTTATGTTATGGGCTAAT
A I T I S L D S M P A I N I A L V N E Q V M L W A N
TTTGATGCACCAAGTGATGTAAAACTTCAGTCATCTGCTTATAACATATTAAATTTAATGTTAATGAACTTTAGTTAT
F D A P S D V K L Q S S A Y N I L N L M L M N F S Y
AGCATTAATGAGCTGGTAGAGCTTCATCGTTCTGATGAGTACTTACAGTTAAGAGTTGTTATTAAAGATGACTACGTT
S I N E L V E L H R S D E Y L Q L R V V I K D D Y V
-35 -10 *
CACGATGGGATTGTTTTTGCAGAGATTCTGCATGAGTTTTATCAGAGGATGGAAATCTTAAATGGGGTCTTATAATCA
H D G I V F A E I L H E F Y Q R M E I L N G V L
ATGAGCTATACAAAATTGCTCACTCAATTATCTTTTCCTAATAGAATCTCGGGGCCAATCTTGGAAACAAGTCTTAGC
M S Y T K L L T Q L S F P N R I S G P I L E T S L S
ORF-2
S279
GATGTTTCGATTGGTGAGATTTGTAACATTCAGGCTGGAATTGAAAGTAATGAAATTGTTGCAAGAGCTCAGGTTGTA
D V S I G E I C N I Q A G I E S N E I V A R A Q V V
HindIII
GGATTTCATGATGAAAAAACAATATTAAGCTTGATTGGAAATTCTCGTGGACTTTCACGGCAAACGTTGATTAAGCCC
G F H D E K T I L S L I G N S R G L S R Q T L I K P
S343
ACTGCCCAGTTTCTTCATACGCAAGTCGGCCGTGGATTATTGGGAGCAGTAGTCAATCCTTTAGGGGAGGTTACTGAT
T A Q F L H T Q V G R G L L G A V V N P L G E V T D
AAGTTTGCTGTTACAGATAACAGTGAAATTCTTTATCGACCTGTAGATAATGCTCCTCCGCTATATAGTGAAAGGGCT
K F A V T D N S E I L Y R P V D N A P P L Y S E R A
GCAATTGAGAAGCCTTTTTTAACAGGTATTAAGGTTATTGATTCTTTACTCACGTGTGGTGAAGGACAGCGAATGGGG
A I E K P F L T G I K V I D S L L T C G E G Q R M G
ATTTTTGCGTCAGCTGGTTGTGGCAAAACTTTTCTCATGAATATGCTCATTGAACATAGTGGTGCTGATATATATGTT
I F A S A G C G K T F L M N M L I E H S G A D I Y V
ATTGGGTTAATTGGTGAGCGAGGTCGAGAGGTTACTGAAACGGTTGATTATTTGAAAAACTCTGAGAAAAAAAGCAGG
I G L I G E R G R E V T E T V D Y L K N S E K K S R
M39
TGTGTTTTAGTATATGCAACTTCGGATTACTCTTCGGTTGATCGTTGTAATGCTGCATATATAGCCACTGCTATAGCC
C V L V Y A T S D Y S S V D R C N A A Y I A T A I A
GAATTTTTTAGGACTGAAGGACATAAAGTAGCGCTTTTTATTGATTCATTAACAAGGTATGCCAGAGCATTACGTGAT
E F F R T E G H K V A L F I D S L T R Y A R A L R D
M253
GTGGCCTTAGCCGCTGGAGAATCACCTGCCAGAAGAGGCTATCCGGTTTCGGTTTTTGATAGCTTACCCAGACTTCTT
V A L A A G E S P A R R G Y P V S V F D S L P R L L
GAAAGGCCAGGAAAGTTAAAGGCAGGTGGCTCTATTACTGCATTTTACACTGTTCTTTTGGAGGATGATGATTTTGCT
E R P G K L K A G G S I T A F Y T V L L E D D D F A
GATCCATTAGCTGAAGAGGTAAGATCCATTTTAGATGGACATATATATTTGAGCAGAAATCTAGCCCAAAAAGGACAA
D P L A E E V R S I L D G H I Y L S R N L A Q K G Q
TTTCCTGCAATTGATTCCTTAAAAAGTATAAGCAGGGTATTTACACAGGTTGTTGATGAAAAACATCGTATTATGGCC
F P A I D S L K S I S R V F T Q V V D E K H R I M A
GCTGCATTTCGGGAGCTACTTTCTGAAATAGAAGAGCTAAGGACAATTATTGACTTTGGTGAATACAAACCGGGGGAG
A A F R E L L S E I E E L R T I I D F G E Y K P G E
AATGCCTCTCAGGATAAAATCTACAACAAAATATCTGTTGTTGAGAGTTTTCTGAAGCAAGATTATCGACTCGGTTTT
N A S Q D K I Y N K I S V V E S F L K Q D Y R L G F
ACTTATGAGCAGACAATGGAGCTTATTGGTGAAACAATTAGATAAGGTTTTAAAAATCAAAGATAAATATCAACGGTCTG
T Y E Q T M E L I G E T I R
TTAAATTAATTGAAGCACACATATTAACACTAAAAAAAAAACTCTTTGTATCGAGATGTGGAGGCATTAGATAAAAGGAT
V E A L D K R I
ORF-3
TATATACTTCCTGCAACTGGAGAATGACTTAGAACCTGTAGGCGCTCAGAGTGTTAGTCAGCTTTTCAATACAAGACGTA
I Y F L Q L E N D L E P V G A Q S V S Q L F N T R R
EcoRI
AGATCGCTATAGTAAAAAAGCATATTATCCAATATCAGTCTGAACGAATTCTTCTAAAGGGAAGAATTGAAGAAATTCAA
K I A I V K K H I I Q Y Q S E R I L L K G R I E E I Q
AAAGATATTGATGAGGCAAATGCTTCTAAAAGAAAACTGCTTCATAAAGAGAGTAAAATATGTAAACGGATTGGCTTGAT
K D I D E A N A S K R K L L H K E S K I C K R I G L I

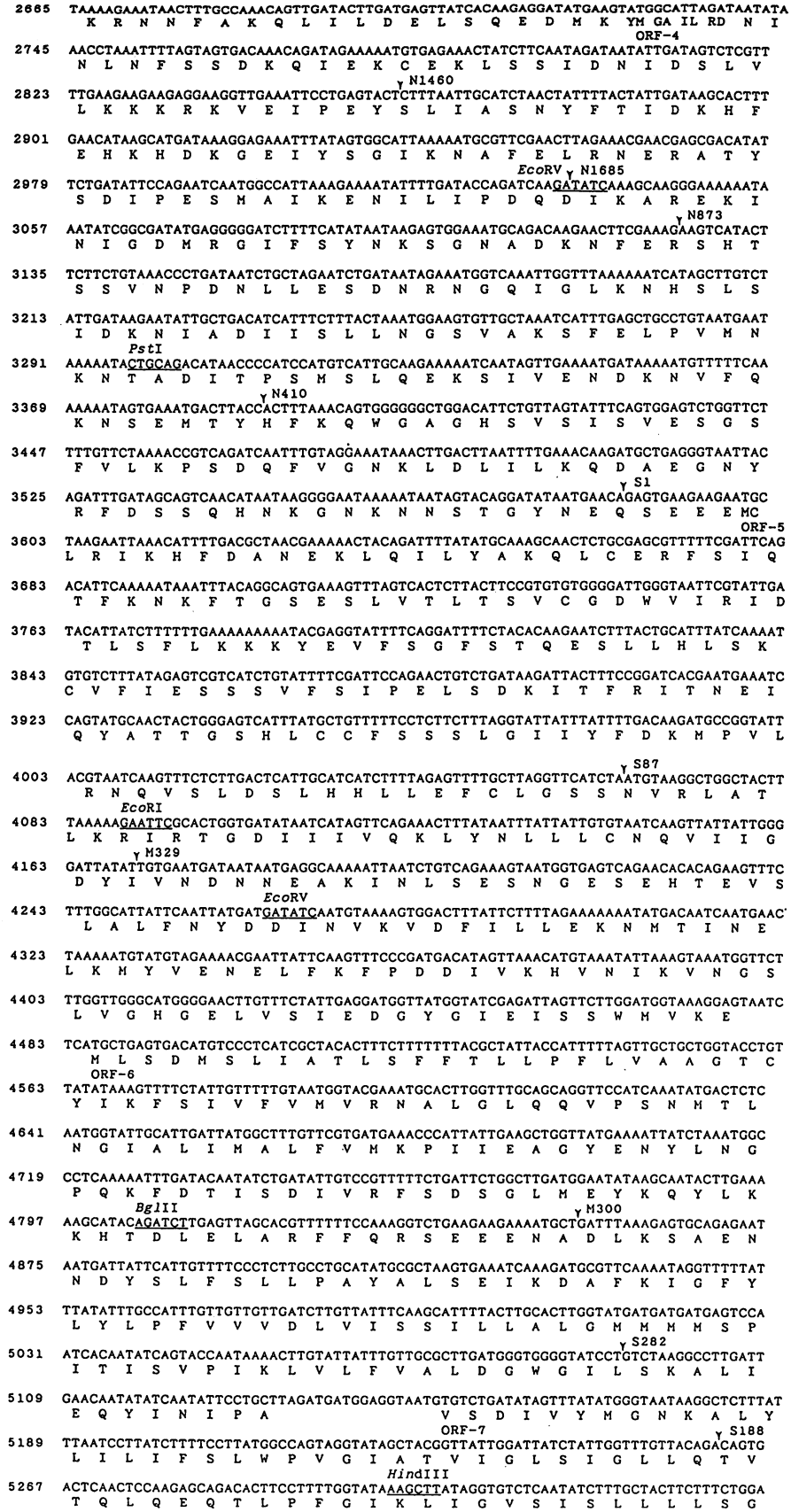


FIG. 2—*Continued.*

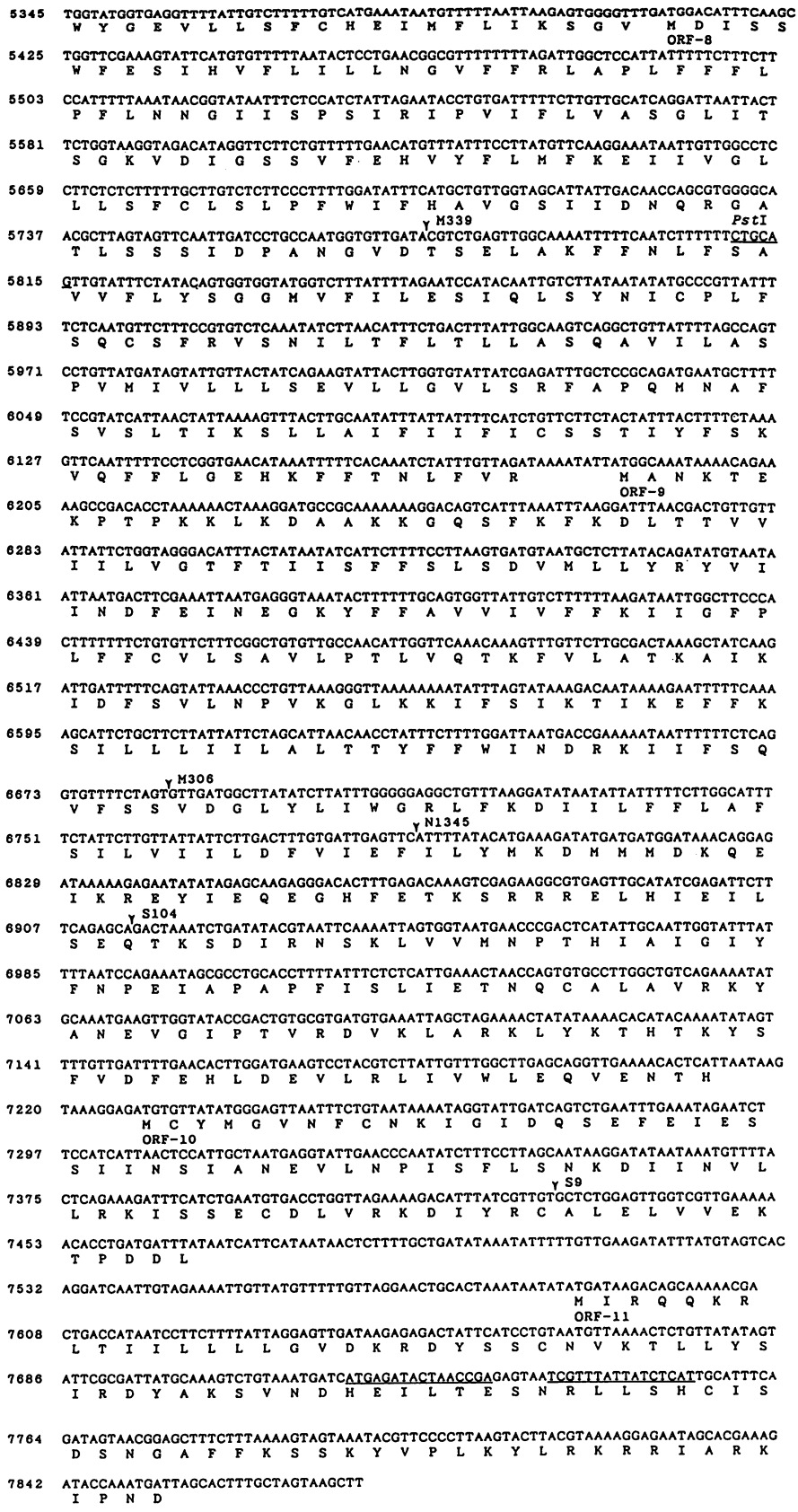


FIG. 2—*Continued.*

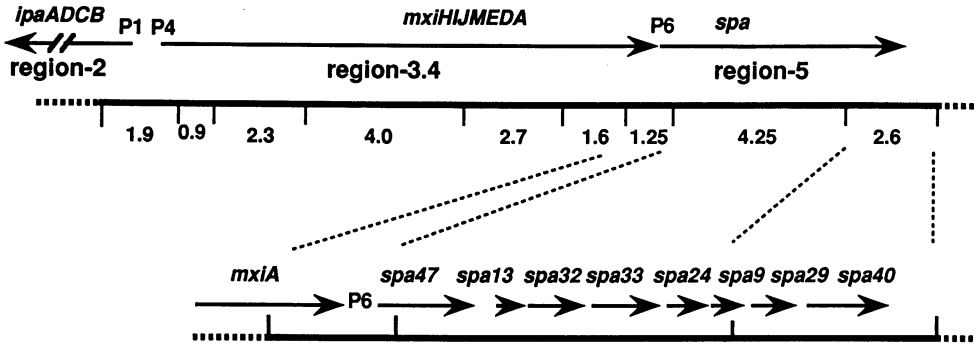


FIG. 3. Transcriptional organization of regions 3, 4, and 5 on pMYSH6000 and placement of the virulence determinants in region 5. The transcriptional organization proposed was deduced from two of our previous studies (31, 35) and this study. The arrows indicate the transcripts. The thick bar represents the *Hin*dIII restriction fragments of regions 3, 4, and 5 of pMYSH6000. The values beneath the bar indicate the sizes (in kilobases) of each *Hin*dIII fragment used for the CAT assay. The placement of the virulence-associated genes *ipaADCB* (8, 11, 31, 36), *mxiHIJMEDA* (2, 3, 5, 6, 17), and *spa* (37), over *ipa*, region 3.4 and 5 operons, respectively, are not to scale. P1, P4, and P6 represent the promoters identified by previous studies (31, 35) and this study. The bottom arrows represent the eight virulence determinants in region 5 and *mxiA* (6). As *spa47*, *spa32*, *spa33*, and *spa24* were shown as the components of *spa* locus on pWR100 (37), we propose calling ORF-2 through ORF-9 in region 5 *spa47*, *spa13*, *spa32*, *spa33*, *spa24*, *spa9*, *spa29*, and *spa40*, respectively.

M223 (*virB*::Tn*5*) were hybridized to the 994-bp *Hae*II-*Hin*dIII fragment of the 1.25-kb *Hin*dIII fragment and digested with S1 nuclease, and the protected DNAs were electrophoresed. A band corresponding to about 250 nucleotides was detected with the restriction fragment in YSH6000 but not M223 (data not shown). The precise transcriptional start site for region 5 was determined by performing primer extension RNA sequencing on total RNA prepared from YSH6200 harboring pKK232-8-1.25-kb *Hin*dIII and pCHR405. For this purpose, a primer containing the 18 bases from positions 1087 to 1104 that hybridized to mRNA sequences within ORF-2 was used. Although at least two extension products were detected, one of the extension products which ended at the A residue at position 965, 47 bp upstream of the 5′ end of ORF-2, was determined to be the 5′ end of transcription (Fig. 4).

Identification of protein products. The protein products expressed from ORF-7, -8, and -9 were determined, since the other protein products expressed from the contiguous 1.2- and 4.25-kb *Hin*dIII segments in region 5 had been previously reported as the Spa proteins of pWR100 (37). The 3.7-kb *Bgl*II-*Xho*I segment of region 5, the 2.6-kb *Hin*dIII segment and its fill-in mutations at the *Pst*I site in ORF-8, and the *Afl*II site in ORF-9 were prepared from pCHR553-1 and pCHR510-1, respectively, and placed downstream of the T7 RNA polymerase-dependent promoter, φ10 (34). The polypeptides expressed from those constructs were then subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 5). At least nine polypeptides, which migrated at apparent molecular masses of 38, 31, 27, 25, 18, 16, 12, 10, 8, and 6 kDa were expressed from the 3.7-kb *Bgl*II-*Xho*I segment (Fig. 5, lane 1). The fill-in mutation at the *Pst*I site in ORF-8 lacked the 27-kDa protein (lane 3), while that at *Afl*II in ORF-9 lacked the 38- and 31-kDa proteins (lane 4). The 10-kDa protein was expressed only from the 3.7-kb *Bgl*II-*Xho*I segment (lane 1), indicating that it is a product of ORF-7. The 12-kDa protein disappeared only when the *Bcl*I segment containing ORF-10 and ORF-11 in the 2.6-kb *Hin*dIII segment was deleted (data not shown), suggesting that the 12-kDa protein is expressed from either one or both of the ORFs. Both the 31- and 38-kDa proteins disappeared when the *Afl*II site was filled in (Fig. 3, lane 4) (Fig. 1). Since the *Afl*II site was located just 7 bp upstream of the second largest ORF coding for the 31-kDa protein, the disappearance of the protein may be due to a change in the putative ribosome-binding sequence required for the translation of the 31-kDa protein. The other smaller 25-, 18-, 16-, 8-, and 6-kDa polypeptides seen would be the breakdown products of the 38-kDa protein or internally translated products of ORF-9, since ORF-9 contains 12 initiation codons. These results indicated that the 10-, 27-, and 38-kDa proteins expressed were encoded by ORF-7, -8, and -9, respectively.

TABLE 3. CAT assay for the functional promoter region in regions 3, 4, and 5

Size (kb) of *Hin*dIII fragment[a]	CAT activity[b]		Ratio[c]
	With IPTG	Without IPTG	
1.9	2,136.3	539.1	4.0
0.9	7.2	6.6	1.1
2.3	25.4	28.6	0.9
4.0	17.3	19.8	0.9
2.7	42.5	51.5	0.8
1.6	25.5	28.8	0.8
1.25	1,048.2	540.0	1.9
4.25	7.4	8.9	0.8
2.6	35.8	40.0	0.9

[a] See Fig. 3.
[b] Shown as the dilution rate which gave rise to 10% of the acetylated products of [^{14}C]chloramphenicol in the presence of *virB*, either activated or not with IPTG. The radioactivity of acetylated products was directly quantified by the AMBIS radioanalytic imaging system following thin-layer chromatography (see Materials and Methods).
[c] Ratio of CAT activity in the presence and absence of 1 mM IPTG, which activates the *virB* gene.

Involvement of the virulence genes in region 5 in the secretion of IpaB, IpaC, and IpaD proteins. As described in the introduction, Tn*5* insertion mutation of the 4.4-kb *spa* locus on pWR100 or in region 5 blocked surface presentation of the IpaB and IpaC proteins, although the levels of these proteins in the mutants, as determined by whole-cell extracts, were the same as those in the wild type (37; our unpublished data). Hence, we undertook to investigate whether the virulence-associated genes identified in region 5 were involved in surface presentation of Ipa proteins (Spa phenotype) (37). To test this hypothesis, pCHR509-1, -4, -5,

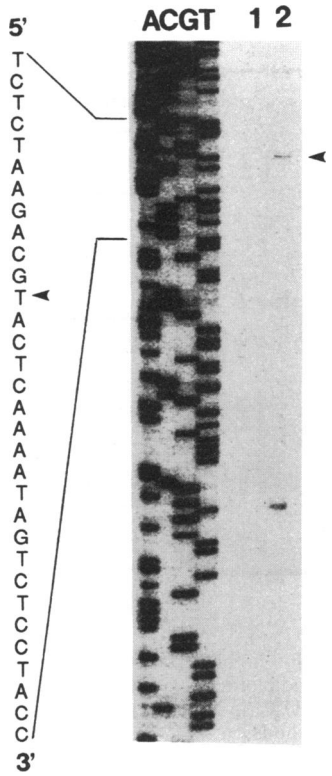


FIG. 4. Determination of the transcription initiation site of region 5 mRNA by primer extension. A ^{32}P end-labelled oligonucleotide corresponding to positions 1087 to 1104 in Fig. 2 was used for reverse transcription. The product was run on a 6% polyacrylamide gel together with a sequencing ladder (ACGT) produced with the same primer. Template from YSH6200 carrying pKK232-8-1.25-kb *Hin*dIII clone without (lane 1) or with (lane 2) pCHR405 was used. The arrowhead indicates the 5′ end for region 5 mRNA.

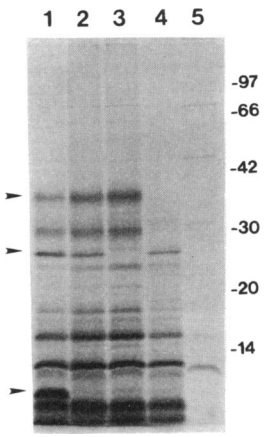


FIG. 5. Protein products expressed from ORF-7, ORF-8, and ORF-9 with the T7 RNA polymerase-dependent promoter system. Lane 1, the 3.5-kb *Bgl*II-*Xho*I segment encoding ORF-7, -8, -9, -10, and -11; lane 2, the 2.4-kb *Hin*dIII segment encoding ORF-8 and ORF-9; lane 3, same as lane 2 but with a fill-in mutation at the *Pst*I site in ORF-8; lane 4, same as lane 2 but with a fill-in mutation at the *Afl*II site in ORF-9 (Fig. 1 and Table 1). The three arrowheads indicate the positions of the 38-, 27-, and 12-kDa proteins. The values to the right show the positions of molecular mass markers (in kilodaltons).

and -6 and pCHR510-1, -3, and -4 were introduced into S343 (ORF-2::Tn*5*), N873 (ORF-4::Tn*5*), S87 (ORF-5::Tn*5*), M300 (ORF-6::Tn*5*), S188 (ORF-7::Tn*5*), M339 (ORF-8::Tn*5*), and S104 (ORF-9::Tn*5*), respectively, and the resulting merodiploids were designated S343-2, N873-2, S87-2, M300-2, S188-2, M339-2, and S104-2, respectively (Table 1). Consequently, each construct was able to produce all of the gene products except that disrupted by the Tn*5* insertion, since the blockage of expression of the genes downstream from the Tn*5* insertion was supplemented in *trans* by the plasmids introduced. By exploiting the set of mutants constructed, we performed colony immunoblots of the whole cells with antibodies specific for IpaB, IpaC, or IpaD proteins and examined the Spa phenotype. As shown in Fig. 6, all of the merodiploid strains, except for N873-2, ceased to secrete the IpaB, IpaC, and IpaD proteins. For N873-2, the surface levels of IpaB and IpaD, but not IpaC, were much lower than those of the wild type. As expected, the seven Tn*5* insertion mutants, S343, N873, S87, M300, S188, M339, and S104, but not the Inv$^+$ derivatives, S343-1, N873-1, S87-1, M300-1, S188-1, M339-1, and S104-1, blocked secretion of the IpaB, IpaC, and IpaD proteins (Fig. 6). These results indicated that the virulence-associated proteins encoded by ORF-2 and ORF-5 through ORF-9 were required for secretion of the three Ipa proteins and that ORF-4 was required only for secretion of IpaB and IpaD.

DISCUSSION

Random insertion mutagenesis of the large plasmid pMYSH6000 of *S. flexneri* 2a YSH6000 has identified five contiguous virulence regions that are required for invasion of epithelial cells (33). These regions, which contain a 31-kb DNA segment on the large plasmid, include *virB* (region 1) (1), *ipaBCDA* and *icsB* (region 2) (2, 8, 9, 11, 31, 37), *mxiHIJMED* (region 3) (3, 4), *mxiA* (region 4) (5, 6), and *spa* (region 5) (37), are highly conserved among shigellae and enteroinvasive *E. coli* (9, 33). The virulence-associated loci so far identified in regions 3, 4, and 5 are involved in the secretion of Ipa proteins and thus account for the Inv$^-$ phenotype of Tn*5* insertion mutants of the three virulence regions (33). Restriction enzyme analysis indicated that a portion of region 5 on pMYSH6000 contained the *spa* locus, which was recently found on pWR100, and is essential for secretion of IpaB and IpaC proteins (37), and that the remaining 3′ portion downstream of *spa* also contains additional virulence loci involved in the Inv$^+$ phenotype (33). In this study, we have analyzed the genetic and transcriptional organization of region 5 and found additional virulence determinants included in the *spa* locus of pWR100.

Complementation tests of the 14 Tn*5* insertion mutants of region 5 with various portions of region 5 or the fill-in and linker-insertion derivatives, as well as the complete nucleotide sequence, indicated that ORF-2 through ORF-9 were the essential determinants for the Inv$^+$ phenotype of *S. flexneri*. Although we found no differences in the nucleotide sequences of the *spa* locus (37) and region 5 (Fig. 2), some differences have been noted in this study. For example, an additional ORF, named ORF-3, was found between ORF-2 and ORF-4; the 5′ end of ORF-3 initiated 93 bp downstream from the 3′ end of ORF-2 (*spa47*) with a GTG (valine) initiation codon and terminated 14 bp downstream from the 5′ end of ORF-4 (*spa32*) (Fig. 2). Since the fill-in mutation at the *Eco*RI site in ORF-3 (pCHR576-3) did not restore the Inv$^+$ phenotype to the upstream Tn*5* insertion mutants, such

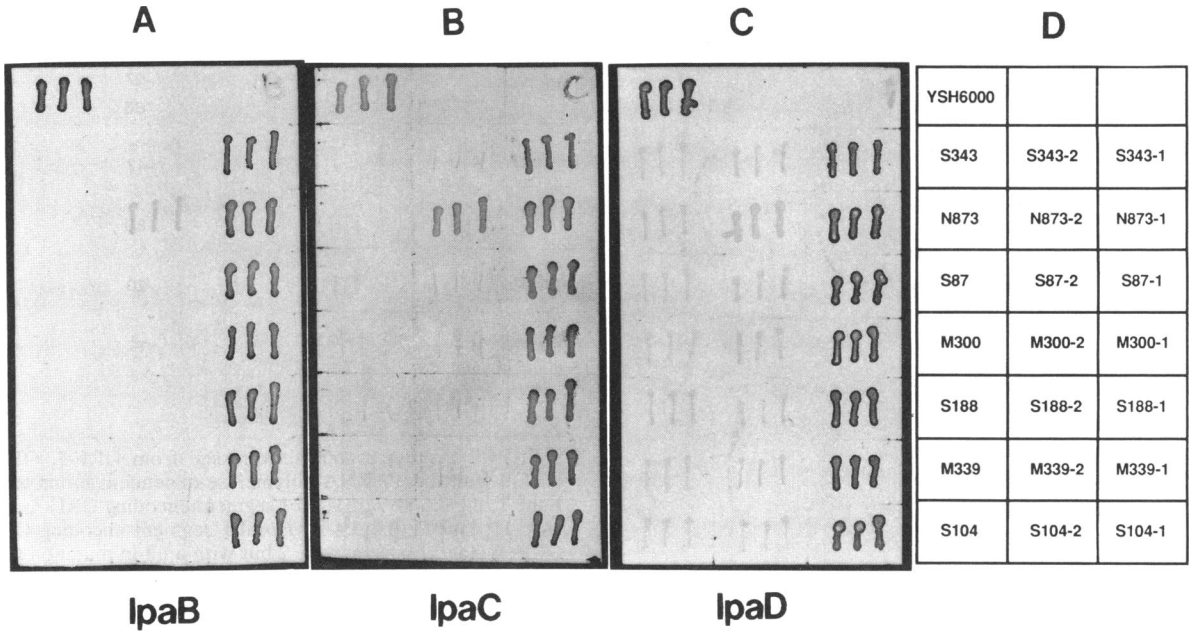


FIG. 6. Involvement of the protein products encoded by ORF-2 and ORF-4 through ORF-9 in the surface expression of IpaB, IpaC, and IpaD proteins. Whole-cell colonies of YSH6000, Tn*5* insertion mutants, and the various merodiploid derivatives were developed by streaking colonies onto nitrocellulose filters placed over LN agar plates and the resulting colonies at 37°C (in triplicate) were immunologically stained with anti-oligopeptide antibodies to IpaB, IpaC, or IpaD as described in Materials and Methods.

as S279, S343, and M253 (ORF-2::Tn*5*), we concluded that ORF-3 should be one of the determinants required for the Inv$^+$ phenotype of *S. flexneri*. Furthermore, a Tn*5* insertion mutant in the *spa47* gene on pWR100 (pHS1059) was shown to be complemented by the *spa* clone pWR266, consisting of the 1.25- and 4.25-kb *Hin*dIII fragments of pWR100, but not by the derivative lacking some 3′ portion of the *spa* locus, suggesting that *spa* forms an operon (37). In contrast, in our complementation tests, subclones lacking the 3′ portion of region 5, such as pCHR576-4, could not restore the Inv$^+$ phenotype to Tn*5* insertion mutants in ORF-2 (*spa47*), such as S279, S343, or M253 (Table 2). Although the reason for the different results in the two independent complementation tests done by two groups is not clear, it may be due to different transcriptional organizations between the two large plasmids, pMYSH6000 and pWR100.

The results of the complementation tests of region 5 Tn*5* insertion mutants (Table 2) and the placement of the promoter sequence (Fig. 3) indicated that the 7-kb region 5 consists of a single operon whose expression was controlled by the positive regulator, *virB* (1). This result was further confirmed by making use of the polar nature of Tn*5* insertions in region 5; we compared the level of region 5 mRNA production upstream and downstream of Tn*5* insertions by Northern (RNA) dot blots with various DNA probes prepared from region 5. For example, the RNA dot blots of M300 (ORF-6::Tn*5*) with four different DNA probes, the 1-kb segment from the leftmost *Pvu*II site to the leftmost *Eco*RI site, the 0.7-kb segment from the second left *Eco*RI site to the *Bgl*II site, the 0.4-kb segment from the second right *Hin*dIII site to the rightmost *Pst*I site, and the 0.4-kb *Bcl*I segment (Fig. 1) revealed that the mRNA levels detected by the last two DNA probes were much lower than with the first two DNA probes (data not shown). Interestingly, when the level of mRNA in M300 was assayed with the 0.2-kb segment from the rightmost *Hin*dIII site to the *Xho*I site, the mRNA level was the same as in the wild type, suggesting that transcription of region 5 terminates just before the rightmost *Hin*dIII site. In agreement with this, we noted a thermodynamically stable ($\Delta G = -21.9$ kcal/mol) palindromic sequence, starting at position 7717 and ending at position 7754, which is possibly the transcriptional terminator for the region 5 operon (Fig. 2).

By constructing a series of merodiploid derivatives of the Tn*5* insertion mutants of region 5, we tested whether each of the ORFs identified as being virulence associated was involved in the Spa phenotype. The results of colony immunoblots with three antibodies specific for each of the IpaB, IpaC, and IpaD proteins indicated that all of the ORFs, except for ORF-4, were essential for secretion of the three Ipa proteins. ORF-4 was involved in IpaB and IpaD but not IpaC secretion (Fig. 5). The only notable feature of the amino acid sequence of the ORF-4 product deduced from its nucleotide sequence was its hydrophilic nature, since the other protein products such as those encoded by ORF-2 or ORF-5 to ORF-9 revealed a significant hydrophobic nature, including some hydrophobic amino acids domains predictable as membrane-spanning proteins (data not shown). Alternatively, the Ipa secretion shown by ORF-4::Tn*5* may be due to the different mechanisms of IpaC versus IpaB and IpaD secretion. To explain the different behavior of the 32-kDa protein encoded by ORF-4 from those of the other ORFs in the Ipa secretion, we must await further characterization of all of the proteins involved in Ipa protein secretion. In any case, ORF-2 through ORF-9 have been determined to be virulence-associated genes in region 5, and ORF-2 and ORF-4 to ORF-6 corresponded to *spa47*, *spa32*, *spa33*, and *spa24*, the components of the *spa* locus on pWR100 (37), we thus propose calling ORF-2 to ORF-9

spa47, *spa13*, *spa32*, *spa33*, *spa24*, *spa9*, *spa29*, and *spa40*, respectively (Fig. 3).

In certain cases, such as hemolysin or the Ipa proteins, extracellular secretion occurs without a signal sequence. In other cases, proteins translocate first to the periplasm using the signal sequence-dependent pathway and are subsequently translocated to the outer membrane or external medium with the help of accessory proteins (25). In this regard, extracellular secretion of Ipa proteins by shigellae would share, in part, the common mechanisms with Yop protein secretion by yersiniae, since the secretion of Ipa and Yop proteins takes place without signal sequences but is dependent upon other factors encoded by various genes on the virulence plasmid (21). Indeed, as Allaoui et al. (3, 4) have recently pointed out, significant homology exists between MxiD, MxiH, and MxiJ and YscC, YscF, and YscJ.

We have recently undertaken quantitative immunoblots of the subcellular fractions of YSH6000 grown to exponential phase with specific antiserum specific to IpaB, IpaC, or IpaD protein to determine the distribution of the Ipa proteins (12). The results indicated that the Ipa proteins are distributed mostly in the periplasm and cytoplasm, with some in the inner membrane, but not in the outer membrane or in the culture supernatant. Nevertheless, the potential abilities of the IpaB, IpaC, and IpaD proteins to be secreted onto the bacterial surface when grown on solidified medium (37; this study), into the external medium when grown in liquid medium to stationary phase (12) or into a hypotonic medium in which bacteria have been suspended (5, 17), have been shown. The subcellular distribution of the Ipa proteins in S334 (ORF-2::Tn*5*) was similar to that in the wild type (data not shown), whereas the Tn*5* insertion mutant failed to secrete the IpaB, IpaC, and IpaD proteins onto the bacterial surface (Fig. 6) and into the culture medium (data not shown). We therefore presume that the proteins encoded by region 5 are involved in functions associated with transport of the Ipa proteins from the periplasm to the bacterial surface. To further confirm this hypothesis, we have to elucidate the roles played by each of the gene products encoded by regions 3, 4, and 5 in Ipa excretion. Thus, the functional organization established in this study, together with other reports on the virulence genes in regions 3, 4, and 5 (3–6, 17, 37), will contribute to the elucidation of the mechanisms of Ipa protein secretion, a process crucial for the invasion of epithelial cells by *S. flexneri*.

ACKNOWLEDGMENTS

We are grateful to R. MacSiomoin for critical reading of the manuscript.

This study was supported by grants 01440031, 03304030, 03557022, and 04454194 from the Ministry of Education, Science and Culture of the Japanese government. C.S. was the recipient of a grant from the Naito Foundation.

REFERENCES

1. **Adler, B., C. Sasakawa, T. Tobe, S. Makino, K. Komatsu, and M. Yoshikawa.** 1989. A dual transcriptional activation system for the 230 kilobase plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. Mol. Microbiol. **3**:627–635.
2. **Allaoui, A., J. Mounier, M.-C. Prevost, P. J. Sansonetti, and C. Parsot.** 1992. *icsB*: a *Shigella flexneri* virulence gene necessary for the lysis of protrusions during intercellular spread. Mol. Microbiol. **6**:1605–1616.
3. **Allaoui, A., P. J. Sansonetti, and C. Parsot.** 1992. MxiJ, a lipoprotein involved in secretion of *Shigella* Ipa invasion, is homologous to YscJ, a secretion factor of the *Yersinia* Yop proteins. J. Bacteriol. **174**:7661–7669.
4. **Allaoui, A., P. J. Sansonetti, and C. Parsot.** 1993. MixD: an outer membrane protein necessary for the secretion of the *Shigella flexneri* Ipa invasins. Mol. Microbiol. **7**:59–68.
5. **Andrews, G. P., A. E. Hromockyj, C. Coker, and A. T. Maurelli.** 1991. Two novel virulence loci, *mxiA* and *mxiB*, in *Shigella flexneri* 2a facilitate excretion of invasion plasmid antigens. Infect. Immun. **59**:1997–2005.
6. **Andrews, G. P., and A. T. Maurelli.** 1992. *mxiA* of *Shigella flexneri* 2a, which facilitates export of invasion plasmid antigens, encodes a homolog of the low-calcium-response protein, LcrD, of *Yersinia pestis*. Infect. Immun. **60**:3287–3295.
7. **Auerswald, E. A., G. Ludwig, and H. Schaller.** 1981. Structural analysis of Tn*5*. Cold Spring Harbor Symp. Quant. Biol. **45**:107–113.
8. **Baudry, B., M. Kaczorek, and P. J. Sansonetti.** 1988. Nucleotide sequence of the invasion plasmid antigens B and C genes (*ipaB* and *ipaC*) of *Shigella flexneri*. Microb. Pathog. **4**:345–357.
9. **Baudry, B., A. T. Maurelli, P. Clerc, J. G. 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 polypeptides. J. Gen. Microbiol. **133**:3403–3413.
10. **Bernardini, M. L., J. Mounier, H. D'Hauteville, M. Coquis-Rondon, and P. J. Sansonetti.** 1989. Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacteria intra- and intercellular spread through interaction with F-actin. Proc. Natl. Acad. Sci. USA **86**:3867–3871.
11. **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.
12. **Fukuda, I., N. Okada, K. Komatsu, M. Yoshikawa, and C. Sasakawa.** Unpublished results.
13. **Galan, J., E. Ginocchio, and P. Costeas.** 1992. Molecular and functional characterization of the *Salmonella typhimurium* invasion gene *invA*: homology of InvA to members of a new protein family. J. Bacteriol. **174**:4338–4349.
14. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
15. **Higgins, C. F., I. D. Hiles, G. P. C. Salmond, D. R. Gill, J. A. Downie, I. J. Evans, I. B. Holland, L. Gray, S. D. Buckel, A. W. Bell, and M. A. Hermodson.** 1989. A family of related ATP binding subunits coupled to many distinct biological processes in bacteria. Nature (London) **323**:448–450.
16. **High, N., J. Mounier, M.-C. Prevost, and P. J. Sansonetti.** 1992. IpaB of *Shigella flexneri* causes entry into epithelial cells and escape from the phagocytic vacuole. EMBO J. **11**:1991–1999.
17. **Hromockyj, A. E., and A. T. Maurelli.** 1989. Identification of *Shigella* invasion genes by inoculation of temperature-regulated *inv*::*lacZ* operon fusions. Infect. Immun. **57**:2963–2970.
18. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
19. **LaBrec, E. H., H. Schneider, J. J. Magnani, and S. B. Formal.** 1964. Epithelial cell penetration as an essential step in the pathogenesis of bacillary dysentery. J. Bacteriol. **88**:1503–1518.
20. **Makino, S., C. Sasakawa, K. Kamata, T. Kurata, and M. Yoshikawa.** 1986. A genetic determinant required for continuous reinfection of adjacent cells on a large plasmid in *Shigella flexneri* 2a. Cell **46**:551–555.
21. **Michiels, T., J.-C. Vanooteghem, C. L. deRouvroit, B. China, A. Gustin, P. Boudry, and G. Cornelis.** 1991. Analysis of *virC*, an operon involved in the secretion of Yop proteins by *Yersinia enterocolitica*. J. Bacteriol. **173**:4994–5009.
22. **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.
23. **Nakata, N., C. Sasakawa, N. Okada, T. Tobe, I. Fukuda, T. Suzuki, K. Komatsu, and M. Yoshikawa.** 1992. Identification and characterization of *virK*, a virulence-associated large plasmid gene essential for intercellular spreading of *Shigella flexneri*.

Mol. Microbiol. **6**:2387–2395.
24. **Ohta, S., H. Tamura, K. Matsuda, and Y. Kagawa.** 1988. Gene structure of the human mitochondrial adenosine triphosphate synthase β subunit. J. Biol. Chem. **236**:11257–11262.
25. **Pagsely, A. P., C. d'Enfert, I. Reyss, and M. G. Kornacker.** 1990. Genetics of extracellular protein secretion by Gram-negative bacteria. Annu. Rev. Genet. **24**:67–90.
26. **Plano, G. V., S. S. Barve, and S. Straley.** 1991. LcrD, a membrane-bound regulator of the *Yersinia pestis* low-calcium response. J. Bacteriol. **173**:7293–7303.
27. **Ramakrishnan, G., J.-L. Zhao, and A. Newton.** 1991. The cell cycle-regulated flagellar gene *flbF* of *Caulobacter crescentus* is homologous to a virulence locus (*lcrD*) of *Yersinia pestis*. J. Bacteriol. **173**:7283–7292.
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. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **Sansonetti, P. J., A. Ryter, P. Clerc, A. T. Maurelli, and J. Mounier.** 1986. Multiplication of *Shigella flexneri* within HeLa cells: lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. Infect. Immun. **51**:461–469.
31. **Sasakawa, C., B. Adler, T. Tobe, N. Okada, S. Nagai, K. Komatsu, and M. Yoshikawa.** 1989. Functional organization and nucleotide sequence of virulence region-2 on the large virulence plasmid of *Shigella flexneri* 2a. Mol. Microbiol. **3**:1191–1201.
32. **Sasakawa, C., and D. E. Berg.** 1982. IS*50*-mediated inverse transposition: discrimination between the two ends of an IS element. J. Mol. Biol. **159**:257–271.
33. **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.
34. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
35. **Tobe, T., S. Nagai, N. Okada, B. Adler, M. Yoshikawa, and C. Sasakawa.** 1991. Temperature-regulated expression of invasion genes in *Shigella flexneri* controlled through the transcriptional activation of the *virB* gene on the large plasmid. Mol. Microbiol. **5**:887–893.
36. **Venkatesan, M., J. M. Buysse, and D. J. Kopecko.** 1988. Characterization of invasion plasmid antigen (*ipaBCD*) genes from *Shigella flexneri*: DNA sequence analysis and control of gene expression. Proc. Natl. Acad. Sci. USA **85**:9317–9321.
37. **Venkatesan, M. M., J. M. Buysse, and E. V. Oaks.** 1992. Surface presentation of *Shigella flexneri* invasion plasmid antigens requires the products of the *spa* locus. J. Bacteriol. **174**:1990–2001.
38. **Vogler, A. P., M. Homma, V. M. Irikura, and R. M. Macnab.** 1991. *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F_0F_1, vacuolar, and archaebacterial ATPase subunits. J. Bacteriol. **173**:3564–3572.