

## Nucleotide Sequence and Transcriptional Regulation of a Positive Regulatory Gene of *Shigella dysenteriae*

RUIJIN YAO AND SUNIL PALCHAUDHURI\*

Department of Immunology and Microbiology, Wayne State University  
School of Medicine, Detroit, Michigan 48201

Received 12 August 1991/Accepted 2 December 1991

**A 1,937 bp *Pst*I-*Hind*III fragment containing the *ipaR* locus was cloned from the large invasion plasmid of *Shigella dysenteriae* CG097, and its nucleotide sequence was completely determined. The IpaR protein (35 kDa, calculated from the DNA sequence) was synthesized in *Escherichia coli*  $\chi$ 1411 minicells containing the 1,937-bp *Pst*I-*Hind*III fragment. To determine the regulatory role of *ipaR* for *ipa* genes, we applied genetic complementation experiments using chloramphenicol acetyltransferase (CAT) as reporter. Analyses of CAT activity of the recombinant plasmids containing the 5' flanking sequences of the 24-kDa-protein gene and the *ippI*, *ipaB*, *ipaC*, and *ipaD* genes defined strong promoters upstream of the 24-kDa-protein gene and *ipaD* gene, weak promoters upstream of the *ippI* and *ipaB* genes, and the absence of any promoter activity for the *ipaC* gene. Complementation analyses showed that the CAT activity only under direction of the *ippI* promoter region increased 1.8-fold in the presence of IpaR protein. On the basis of our data, we suggest that an operon comprising *ippI*, *ipaB*, and *ipaC* is positively regulated by IpaR protein which has a *trans* effect on a DNA sequence upstream of the *ippI* promoter.**

*Shigellae* cause bacillary dysentery in humans and monkeys. Invasion of epithelial cells of the colon constitutes one of the earliest steps in *Shigella* infection. Genes required for the full expression of virulence are on both the *Shigella* chromosome and a nontransmissible large plasmid (180 to 230 kb) (9, 19, 20, 23). Invasion plasmid antigens IpaB, IpaC, and IpaD are associated with the initial adhesion and invasion of colonic epithelial cells (3, 6, 11, 14, 21, 29). Expression of virulence genes is temperature regulated, wild-type strains being virulent when grown at 37°C but phenotypically avirulent at 30°C (12). Maurelli and Sansonetti (13) demonstrated that temperature regulation of virulent gene expression in *Shigella flexneri* is negatively controlled by *virR*, a chromosomal gene. Studies of *S. flexneri* YSH6000 and M90T have shown that a plasmid gene, *ipaR* (or *virB*), acts as a positive regulator, at the transcriptional level, for the expression of *ipaB*, *ipaC*, *ipaD*, and *ipaA* (1, 7). An identical regulatory gene, designated *invE*, has also been found in *S. sonnei* (27). However, such a study has not been done with *S. dysenteriae*, which causes bacillary dysentery in children with greater severity than dysentery caused by other *Shigella* species or enteroinvasive *Escherichia coli*. Furthermore, the molecular mechanism by which *ipaR* (or *virB*, or *invE*) activates the transcription of *ipaB*, *ipaC*, *ipaD*, and *ipaA* is not known.

In this work, we report the presence of the *ipaR* gene in the invasion plasmid of *S. dysenteriae* CG097 and describe its positive regulation on the expression of *ippI*, *ipaB* and *ipaC* genes through a *trans* effect on the 5' flanking region of the *ippI* gene.

### MATERIALS AND METHODS

**Bacterial strains.** *E. coli* HB101 was used as the host in recombinant DNA cloning and for chloramphenicol acetyltransferase (CAT) expression. *E. coli* JM103 was used as a host for M13 bacteriophage. *S. dysenteriae* 60R (30) was

used as the host for CAT expression. A minicell-producing strain, *E. coli*  $\chi$ 1411, was used for analysis of plasmid-encoded proteins.

**Cloning vectors and DNA methods.** Plasmids pUC8 (BRL), pKK232-8 (Pharmacia) (5), and pACYC177 (gift of M. P. Jackson) were used for DNA cloning, and vectors M13mp18 and M13mp19 were used for DNA sequencing. Isolation of plasmid DNA, digestion of restriction enzymes, recovery of DNA fragments from agarose gel, ligation, and transformation were performed as described by Sambrook et al. (17).

**Plasmid constructions.** Plasmids pWS102, pWS103, and pWS104 previously described (29), containing *S. dysenteriae ipa* genes, were the source for cloning of *ipa* promoter regions in this study. The 5' flanking sequence of the 24-kDa-protein gene excised from pWS104 by *Bam*HI and *Hind*III digestions (396 bp, -317 to +79, with respect to the initiation codon of the 24-kDa-protein gene), a 772-bp *Hind*III fragment of the 5' flanking sequence of the *ippI* gene (-605 to +167) from pWS104, a 1,707-bp *Pvu*II-*Sal*I DNA fragment of the 5' flanking sequence of the *ipaB* gene from pWS102 (-431 to +1171, also containing 105 bp of pUC8 preceding -431), an 850-bp *Ssp*I-*Bam*HI fragment of the 5' flanking region of the *ipaC* from pWS102 (-406 to +426, also containing 18 bp of pUC8 following +426), and a 632-bp *Hinc*II fragment of the 5' flanking sequence of the *ipaD* gene from pWS103 (-423 to +209) were inserted into the multiple cloning sites upstream of the CAT gene of the pKK232-8 plasmid, to yield recombinant plasmids pWS396, pWS772, pWS1602, pWS832, and pWS632, respectively (Table 1). The direction of each inserted DNA fragment was identified by analysis of the cleavage patterns generated by the specific restriction endonucleases.

To construct the 5' deletion mutants of the *ippI* promoter region, a 562-bp *Bal*I-*Hind*III fragment (-395 to +167, 210-bp deletion from the 5' end of the 772-bp *Hind*III fragment) and a 414-bp *Hae*III-*Hind*III fragment (-247 to +167, 358-bp deletion from the 5' end of the 772-bp *Hind*III fragment) were linked to *Sma*I and *Hind*III sites of

\* Corresponding author.

TABLE 1. Characterization of recombinant plasmids

Plasmid	Vector	Promoter region or gene	Inserted DNA	Source
pWS106	pUC8	<i>ipaR</i> , 9-kDa-protein gene	1,937 bp	pWS100 <sup>a</sup>
pWS396	pKK232-8	24-kDa-protein gene	396 bp, -317 to +79	pWS104 <sup>a</sup>
pWS772	pKK232-8	<i>ippI</i>	772 bp, -605 to +167	pWS104 <sup>a</sup>
pWS210	pKK232-8	<i>ippI</i>	562 bp, -395 to +167	pWS104 <sup>a</sup>
pWS358	pKK232-8	<i>ippI</i>	414 bp, -247 to +167	pWS104 <sup>a</sup>
pWS1602	pKK232-8	<i>ipaB</i>	1,707 bp, -431 to +1171 and 105 bp from pUC8	pWS102 <sup>a</sup>
pWS832	pKK232-8	<i>ipaC</i>	850 bp, -406 to +426 and 18 bp from pUC8	pWS102 <sup>a</sup>
pWS632	pKK232-8	<i>ipaD</i>	632 bp, -432 to +209	pWS103 <sup>a</sup>
pWS1879	pACYC177	<i>ipaR</i> , 9-kDa-protein gene	1,879 bp and 93 bp of pUC8	pWS106, this work
pWS1009	pACYC177	9-kDa-protein gene	731 bp	pWS106, this work

<sup>a</sup> Plasmids pWS100, pWS102, pWS103, and pWS104 were described previously (29).

pKK232-8 to yield pWS210 and pWS358, respectively (Table 1).

To clone the DNA fragment containing the *ipaR* gene into the pACYC177 plasmid, which is compatible with the replicon of pKK232-8 for complementation analyses, a 1,972-bp *Pvu*II fragment from pWS106 (Fig. 1; 1,879 bp from bp 58 to bp 1937 of Fig. 2 and 93 bp of pUC8 following bp 1937) was inserted into the *Hinc*II site of pACYC177 to yield pWS1879 (Table 1).

**Estimation of plasmid copy number.** The plasmid copy number was determined as described by Petersen and Hansen (15). Plasmids pBR322 (15 to 20 copies per chromosome) and pACYC177 (10 to 12 copies per chromosome) (17) were included as controls.

**Analysis of plasmid-encoded proteins.** The pUC8 plasmids, with and without inserts, were transformed into *E. coli*  $\chi$ 1411. Minicells prepared from 250-ml overnight cultures of *E. coli*  $\chi$ 1411 were isolated according to the method of Roozen et al. (16). Plasmid-encoded proteins synthesized in minicells were labeled with [<sup>35</sup>S]-methionine and analyzed by sodium dodecyl sulfate (SDS)-12% polyacrylamide gel electrophoresis (PAGE) (4).

**DNA sequence determination.** The DNA sequences were determined by the chain termination method of Sanger et al.

(18), by using a T7 sequencing kit (Pharmacia), following cloning into M13mp18 and M13mp19.

**Sequence analysis.** GENEPRO (Riverside Scientific) was used to search for open reading frames and restriction sites, to predict protein sequences from nucleic acid data, and to analyze hydropathy of polypeptides.

**CAT enzyme assay.** Bacteria were grown in L broth to an  $A_{660}$  of between 0.5 and 1.0. The harvested bacterial cells were sonicated in an equal volume of 0.25 M Tris-HCl (pH 7.8) for three 10-s treatments with 1-min intervals in an ice-water bath (Fisher Sonic Dismembrator, model 300). After centrifugation (12,000  $\times$  g for 10 min at 4°C) to remove cellular debris, the total protein concentrations of extracts were determined by using a bicinchoninic acid protein assay kit (Pierce Chemical Co., Rockford, Ill.).

Assays for CAT activity were performed as described previously (4), with an incubation time of 45 min at 37°C. CAT reaction was carried out in the following reaction mixture: 0.01 to 3  $\mu$ g of cell extract, 20  $\mu$ l of freshly made acetyl coenzyme A (Sigma) at 3.5 mg/ml, 1  $\mu$ l of [<sup>14</sup>C]chloramphenicol, and 2.5 M Tris-HCl (pH 7.8) in a total final volume of 150  $\mu$ l. After the reaction, the products were separated from unacetylated chloramphenicol on thin-layer chromatography plates (Sigma). Various amounts of lysate were used to ensure proper quantitation. The percentage of acetylation was determined with the AMBIS Radioanalytic Imaging System and converted to units of CAT by using pure CAT (Sigma) as a standard. One unit of CAT activity corresponds to 1 nmol of chloramphenicol acetylated per min at 37°C under optimal assay conditions.

**Nucleotide sequence accession number.** The nucleotide sequence shown in Fig. 2 has been given GenBank accession number X63593.

## RESULTS

**Cloning and DNA sequencing of the *ipaR* region.** In the course of DNA sequencing of *ipaB*, *ipaC*, *ipaD*, and *ipaA* genes from the 9-kb fragment (pWS100) of *S. dysenteriae* CG097 (29), a truncated open reading frame (ORF) downstream of the *ipaA* gene was observed. To identify the extent of this ORF, the relevant restriction fragment (1,937-bp *Pst*I-*Hind*III) from pWS100 was subcloned into pUC8, and the resulting plasmid was designated pWS106 (Table 1). Its restriction map is shown in Fig. 1. Figure 2 presents the nucleotide sequence of this 1.9-kb fragment. The DNA sequence of an open reading frame beginning at bp 597 and ending at bp 1524 shows homology with that of *ipaR* (or *virB* or *invE*) of both *S. flexneri* and *S. sonnei* (1, 7, 27) and was

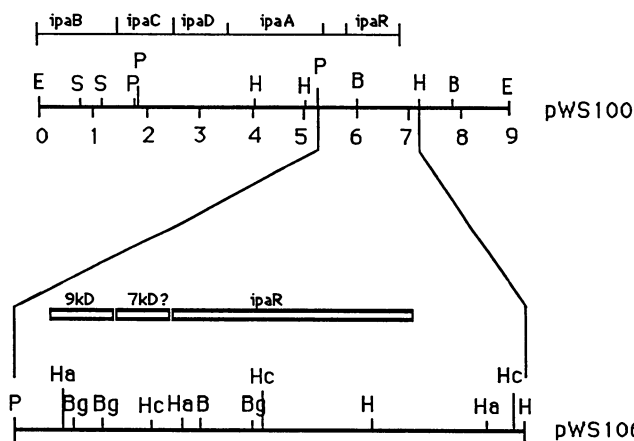


FIG. 1. Restriction map of the 1,937-bp *Pst*I-*Hind*III fragment (pWS106, bottom) derived from the 9-kb *Eco*RI fragment (pWS100, top). The positions of *ipa* genes are also shown. Open boxes represent the 9-kDa, 7-kDa, and *ipaR* ORFs contained in pWS106. E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sal*I; B, *Bam*HI; Bg, *Bgl*II; Ha, *Hae*III; Hc, *Hinc*

CTGCAGAATTACTTACAGATGATATCTGATTTAAAAATAACAATGATATTACAGCTGAAAAACAATAATATATAAAGCAGCAAAAGATGTAACCCAC 100  
 A E L L T D D I S D L K N N N D I T A E N N N I Y K A A K D V T T  
 TTCCEATCAAAGTATTAAGAATCAATAAGGATTAATAATGATAAAAGAAAAATATTATCAATAGTGGCCCTTTGGTATAGCGTACAGC 200  
 S L S K V L K N I N K D \* M I K E K I L S I V A F C Y G I A Y S ↓  
 AAACATCTGAAGAACTAAATTTATTGAAGATCTTTCTGCCGATTCTCTTTCTCTGATTGAAATGCTGGATATGATTAGTTTGGAGTTCAATCTACGTA 300  
 K L S E E T K F I E D L S A D S L S L I E M L D M I S F E F N L R I  
 TAGATGAATCTGCATTAGAACACATTATTACTATTGGAGATCTTATTTCTGTAGTCAAAAATAGTACAAAATCAATATAGTAAATTGATAAGCATTTTTTC 400  
 D E S A L E H I I T I G D L I S V V K N S T K S I \* ←←←←←  
 ATCTATGGAGCTCTCACATCAGAGCTCCACAAGAATATTATTCTTTTATCCAATAAAGATAAATTGCATCAATCCAGCTATTAAAAATAGTAAATTTTAAAG 500  
 ACTACCGTTGACTATCATCAACAGTTACATACAATCTTGCTCACTGCATTTAACTTTTGTCAATATGAATCAATATGAATAAACAGGGTGTGATATGG 600  
 TGGATTGTGCAACGACTTGTAAAGTATAAAGGAAGGCCAAAAGAAAGAGTTTACACTCCATTTCTGGTAATAAAGTTTCTTTTCAAAGCCAAGATTC 700  
 D L C N D L L S I K E G Q K K E F T L H S G N K V S F I K A K I P  
 TCATAAAGGATCCAAAGTTTAACTTCGTCACCAAAAACGAATGTACGCGATCAAGAATCCCTAACAGAAGAATCACTAGCCGATATCATAAAAAC 800  
 H K R I Q D L T F V N Q K T N V R D Q E S L T E E S L A D I I K T  
 ATAAAGCTACAACAATTTCCCTGTAATAGGAAGGGAGATTGATGGTGAATGAAATTTCTGGATGGCACTCGTAGAAGAGCATCTGCAATATATGCAG 900  
 I K L Q Q F F P V I G R E I D G R I E I L D G T R R R A S A I Y A G  
 GAGCAGATCTGAAGTTCTATATTCAAAGAATATATATCTACTCTTGTAGTCCAGAAAACCTAGCAAAAGATATACAAAACAGCAAAAGAGCATAGCATCCG 1000  
 A D L E V L Y S K E Y I S T L C D A R K L A N D I Q T A K E H S I R  
 AGAATTGGTATTGGTCTTAATTTCTGAAAGTATCAGGGATCTCTATAAAGACATAGCCAAAAGAGAATCTGTCTCGCGGAAAGTCACTCGTGCC 1100  
 E L G I G L N F L K V S G M S Y K D I A K K E N L S R A K V T R A  
 TTTCACGCAGCAAGCGTTCCACAGGAAATAATATCTCTATTTCCAATCGCGTCAGAACTTAACTTTAATGACTACAAGATATATTCAATTATTATAAAG 1200  
 F H A A S V P Q E I I S L F P I A S E L N F N D Y K I L F N Y Y K G  
 GACTTGAAAAGGCTAATGAATCTCTTAGTCTACACTACCAATATTAAGGAAGAAAATAAAGACCTTGATACAAAATCTGCCCCGACATATATAAAAA 1300  
 L E K A N E S L S S T L P I L K E E I K D L D T N L P P D I Y K K  
 AGAAATTTTAAACATATAAAGAAAAGCAAAAACAGAAAGCAAAAACCCCTTCGCTAAAAGTTGACTCGTATTATTCTTAAAGCAAAACGGACTTACATA 1400  
 E I L N I K I K K N R K Q N P S L K V D S L F I S K D R T Y I  
 AAAAGAAAAGAGAATAAAAACAAACAGAACTTAATTTTCAATTATCTAAAATAAACAACAGTTCAGAGAGAAAATAGAGCTAATTCGGGATATAA 1500  
 K R K E N K T N R T L I F T L S K I N K T V Q R E I D E A I R D I I  
 TTTCTCGCATCTATCGTCTTCAATAAACATCGAACACTGATGTAAACTGCCAATTCAGATAATAAAAAATCTTCGTTGAACCTTGACTTTCAGCCGGTGGG 1600  
 S R H L S S S \*  
 GTAATTCGCTCGGCTATCTGTTATTCAATCCCACGCACTGTGTTTCCTTAAAATTACGGTGAAAACCTACACCTGCTGACCAGGAGAACATTATCTGG 1700  
 TGAAATTAGTAATATTCATATCATTTGAAAATGAATTGTGCAACTAAGGGATGCGGCATGATGATCAAAATTTTATGTGAAGATGAGTGACTGGCCGAAAA 1800  
 TATTCAGCGTATTTTTAATCGTTTCAATAAAAACAGCGAGGATTAGCAGTCAACGCTTTTACTACTCAATTTTTTTTACATGGTGATAAACAATAATCCA 1900  
 CCATTAACAGACTTCGTTACAGTATCAAAAAGCTT 1937

FIG. 2. Nucleotide sequence of the 1,937-bp *PstI-HindIII* fragment. The deduced amino acids for the *ipaR* ORF (positions 597 to 1524) and 9-kDa ORF (positions 144 to 379) are also shown, as well as the 45 amino acids which overlap the carboxyl terminus of the IpaA protein. The putative -10 and -35 sequences are underlined. Symbols: ●●●●●, Shine-Dalgarno sequence; \*, terminator codon; ↓, cleavage site of the signal peptide of the 9-kDa polypeptide; →← ΔΔΔΔΔΔ, the putative rho-independent terminator; ▼, boundaries of 7 kDa ORF (bp 405 to bp 582).

designated *ipaR*. The amino acid sequence of the IpaR protein deduced from the DNA sequences contains 309 amino acids with a molecular mass of 35 kDa. The putative -10 and -35 promoter regions and ribosome binding sequences are located upstream from the initiation codon of *ipaR*. At the 5' end of the 1,937-bp *PstI-HindIII* fragment, we observed a nucleotide sequence encoding 45 amino acids which overlapped the 3' end of the *ipaA* gene (Fig. 2) (25).

As shown in Fig. 2, the nucleotide sequence of the 1,937-bp *PstI-HindIII* fragment (pWS106) reveals two additional small ORFs located upstream of the *ipaR* gene. The first ORF begins 3 bp downstream from the terminator codon of *ipaA* (bp 144 to 379) and encodes 78 amino acids with a calculated molecular mass of 9 kDa. The second ORF begins 24 bp downstream from the terminator codon of the first ORF (bp 405 to 582) and codes for 59 amino acids (molecular mass, 7 kDa; its amino acids not shown). These two small ORFs were in the same orientation as the *ipaR* ORF. The putative rho-independent terminator was found at the 3' end of the 9-kDa ORF but not at the 3' end of the 7-kDa or the *ipaR* ORF. Interestingly, a typical signal sequence of 19 amino acids was found on the amino-terminal sequence of the 9-kDa polypeptide but not on that of the 7-kDa or IpaR polypeptide. This signal sequence contains nine hydropho-

bic residues preceded by two basic residues, lysines, with serine and alanine at the -1 and -3 positions before the cleavage site (i.e. MIKEKILSIVAFICYGIAYS↓KLS). In addition, the 9-kDa polypeptide contains two hydrophobic transmembrane domains as predicted by our hydropathy analysis. Thus, we speculate that this polypeptide is targeted onto the bacterial membrane by using the same translocation mechanism as those of OmpA, OmpC, OmpF, and alkaline phosphatase of *E. coli* (28).

**Minicell analysis of *ipaR* region products.** Figure 3 shows the result of SDS-PAGE of [<sup>35</sup>S]-methionine-labeled proteins, isolated from the minicell-producing strain of *E. coli* χ1411 carrying the plasmid pUC8 with and without insert. SDS-PAGE analysis of proteins made in *E. coli* χ1411 minicells containing pWS106 showed two polypeptides of 35 and 9 kDa (Fig. 3, lane B). The observed molecular masses of these polypeptides agree with the molecular masses of the products calculated from the nucleotide sequence. However, the 7-kDa polypeptide was not seen in the extract of the minicells containing pWS106. No Shine-Delgarno sequence is found upstream of the 7-kDa ORF (Fig. 2), and this probably accounts for its absence in the minicell analysis (Fig. 3, lane B).

**Characterization of *ipa* promoters.** On the basis of the

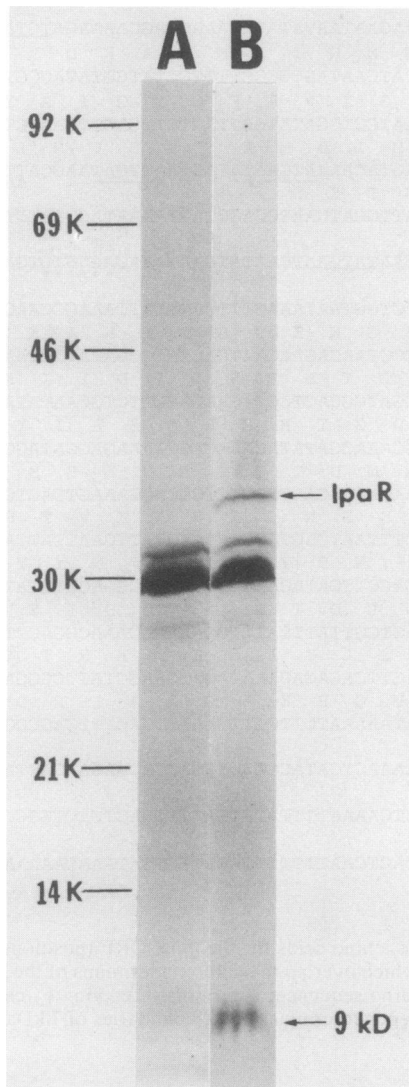


FIG. 3. Minicell analysis of plasmid-coded proteins. The positions of molecular mass markers are indicated by the bars. Lane A, pUC8; lane B, pWS106.

homology of *ipaR* DNA sequences of *S. dysenteriae* CG097 with those of *ipaR* (*virB*, or *invE*) of both *S. flexneri* and *S. sonnei*, which positively regulate the expression of *ipaB*, *ipaC*, *ipaD*, and *ipaA* (1, 7, 27), and in order to determine whether or not IpaR acts on the promoter regions of *S. dysenteriae ipa* genes, we first identified the promoters of *ipa* genes by using CAT as the reporter gene. DNA fragments, containing the 5' flanking region of each *ipa* gene with the promoter region and a portion of amino acid codons from the amino terminus, were inserted upstream of the CAT gene of the promoter cloning vector pKK232-8 plasmid in the same orientation as the CAT gene, resulting in the recombinant plasmids pWS396, pWS772, pWS1602, pWS832, and pWS632 (Table 1). These pKK232-8-derivative plasmids containing the 5' flanking region of *ipa* genes were found to be stable in *E. coli* HB101 and *S. dysenteriae* 60R, and the insertion did not affect the plasmid copy number (10 to 12 copies per chromosome). The function of promoters of the *ipa* genes was determined by measuring CAT activity of these recom-

binant plasmids in *E. coli* HB101 (Fig. 4). Promoterless vector pKK232-8 was included as a negative control (Fig. 4, lane 13). pWS396 (carrying the promoter region of the 24-kDa-protein gene) and pWS632 (carrying the promoter region of the *ipaD*) produced a high level of CAT activity (Fig. 4, lanes 2 and 11). Low CAT activity was detected in pWS772 (carrying the promoter region of *ippI*) and pWS1602 (carrying the promoter region of *ipaB*) (Fig. 4, lanes 4 and 7). CAT was not expressed in pWS832 (carrying the 5' flanking region of *ipaC*) (Fig. 4, lane 9). Similar results were obtained when these recombinant plasmids were introduced into *S. dysenteriae* 60R (data not shown). Strain 60R does not contain the invasion plasmid (30), so it was used as *S. dysenteriae* background. These results suggest that there are strong promoters in the DNA fragments containing the 5' flanking sequences of both the 24-kDa-protein gene and *ipaD* and weak promoters in *ippI* and *ipaB*. Furthermore, it appears that the *ipaC* gene does not have a promoter.

**IpaR protein has a *trans* effect on the 5' flanking sequence of the *ippI* gene.** To determine whether the *ipaR* product has a *trans* effect on the *ipa* promoter regions identified above, we cloned the 1,972-bp *PvuII* DNA fragment containing the *ipaR* locus from pWS106 into the low-copy-number plasmid pACYC177 (P15A of replicon, stringent), yielding pWS1879 (Table 1). pWS1879 containing the *ipaR* locus is maintained in the same copy number as pACYC177 and stably coexists with pWS396, pWS772, pWS1602, or pWS632 in *E. coli* HB101 and *S. dysenteriae* 60R. As shown in Fig. 4, the CAT activity increased 1.8-fold in complementation of pWS1879 (*ipaR*) with pWS772 (lane 5) in HB101 compared with that of pWS772 only (lane 4). No increased CAT activity was detected in complementation of pWS1879 with pWS396, pWS1602, or pWS632 (Fig. 4, lanes 3, 8, and 12, respectively). Yet no CAT activity was detected in HB101 harboring both pWS832 (carrying the 5' flanking region of *ipaC*) and pWS1879 (*ipaR* locus) (Fig. 4, lane 10). Similar results were observed in the *S. dysenteriae* 60R background (data not shown). Since pWS1879 is derived from pACYC177, which is maintained in lower copy number than pBR322 (17) and is not amplified by chloramphenicol, it is unlikely that the *trans* effect of IpaR protein is a result of overexpression of the *ipaR* gene. Thus, we suggested that IpaR protein has a *trans* effect only on the promoter region of *ippI* but not on the promoter regions of the 24-kDa-protein gene, *ipaB*, and *ipaD*.

Nucleotide sequence and minicell analysis of the 1,937-bp *PstI-HindIII* fragment revealed the presence of a 9-kDa polypeptide upstream of the *ipaR* ORF (Fig. 2 and 3). To determine whether the 9-kDa polypeptide is involved in the regulation, a 731-bp *PvuII-EcoRV* fragment, containing only the 9-kDa ORF, was obtained from pWS106 and cloned into the *HincII* site of the pACYC177 plasmid, yielding pWS1009 (Table 1). No increased CAT activity was observed in complementation of pWS1009 with pWS772 (Fig. 4, lane 6, and Fig. 5), indicating that IpaR but not the 9-kDa polypeptide is involved in the regulation of expression of *ipa* genes.

**Functional properties of the DNA sequence located in the 5' flanking region of the *ippI* gene.** The above-described results suggest the presence of a DNA sequence responsible to IpaR protein in the 5' flanking sequence of the *ippI* gene. To further map the location of this DNA sequence between -605 bp and the initiation codon of *ippI*, two deletion mutant plasmids, pWS210 (-395 to +167) and pWS358 (-247 to +167), were constructed (Table 1) and their CAT activity was measured in the presence or absence of pWS1879 (*ipaR* gene). The activation of IpaR was abolished in these two

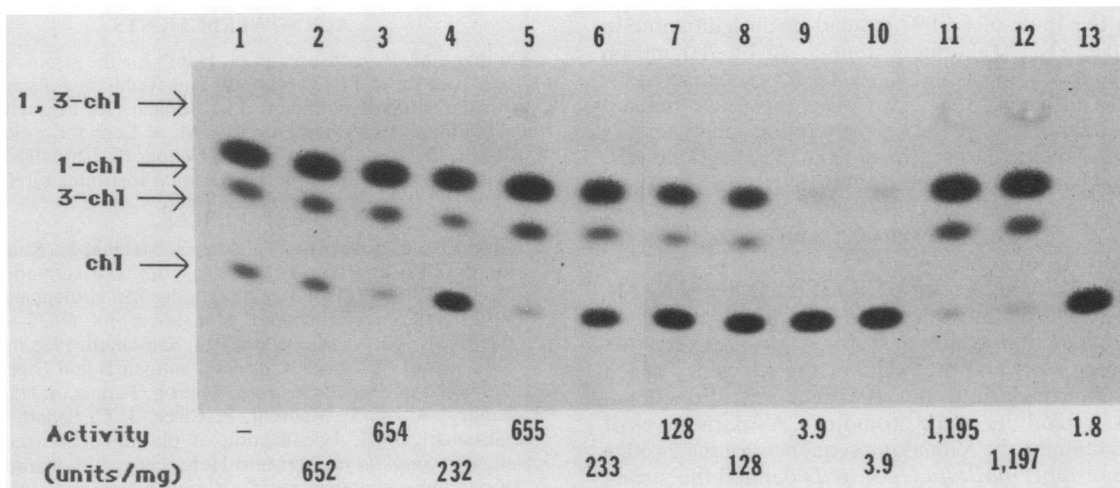


FIG. 4. Fluorography of the products of [<sup>14</sup>C]chloramphenicol obtained in CAT assays (see Materials and Methods) of *E. coli* HB101 containing various recombinant plasmids in the presence or absence of pWS1879 (*ipaR* locus). Lane 1, 0.5 U of chloramphenicol transferase (Sigma); lane 2, pWS396; lane 3, pWS396/pWS1879; lane 4, pWS772; lane 5, pWS772/pWS1879; lane 6, pWS772/pWS1009; lane 7, pWS1602; lane 8, pWS1602/pWS1879; lane 9, pWS832; lane 10, pWS832/pWS1879; lane 11, pWS632; lane 12, pWS632/pWS1879; and lane 13, pKK232-8/pWS1879. Cell extract used: lanes 2 to 8, 11, and 12, 1 μg; lanes 9, 10, and 13, 3 μg. CAT activity was determined as described in Materials and Methods. chl, chloramphenicol; 1-chl, 1-acetyl-chloramphenicol; 3-chl, 3-acetyl-chloramphenicol; 1,3-chl, 1,3-acetyl-chloramphenicol.

mutant plasmids (Fig. 5). However, pWS210 (with deletion of 210 bp, -395 to +167) produced a level of CAT activity (over fivefold) higher than that of pWS772, whereas a lower CAT activity was detected in pWS358 (deletion of 358 bp, -247 to +167) than in pWS772 (Fig. 5).

**DISCUSSION**

Comparison of the nucleotide sequence of the *ipaR* gene of *S. dysenteriae* CG097 with those of *S. flexneri* M90T and YSH6000 and *S. sonnei* revealed a difference of six bases

among these species (1, 7, 27). At position 569, the adenine (A) in the *ipaR* or *virB* gene of *S. flexneri* or the *invE* gene of *S. sonnei* was converted into guanine (G) in the *ipaR* gene of *S. dysenteriae*. This alteration occurred within the putative -10 promoter region (i.e., TATAAAT in M90T, YSH6000, and *S. sonnei*; TATGAAT in CG097). Furthermore, the change of this single base (A to G) produced two direct repeats (CAATATGAATCAATATGAAT) in the *ipaR* promoter region of *S. dysenteriae* CG097. At bp 1106, the change of G in *S. flexneri* M90T or YSH6000 or *S. sonnei* to C in *S. dysenteriae* CG097 resulted in a new amino acid,



FIG. 5. Effect of 5' deletions on CAT expression of pWS772 (containing the promoter region of the *ippI* gene). pWS772, as well as its 5' deletion mutants pWS210 and pWS358, was complemented with (+) or without (-) pWS1879 (*ipaR* locus) or pWS1009 (9-kDa locus). CAT activity was determined as described in Materials and Methods. chl, chloramphenicol; 1-chl, 1-acetyl-chloramphenicol; 3-chl, 3-acetyl-chloramphenicol; 1,3-chl, 1,3-acetyl-chloramphenicol.

histidine, in the IpaR of CG097 instead of the glutamine in IpaR of M90T, YSH6000, and *S. sonnei*. At bp 780 and bp 1278, thymine in YSH6000 and *S. sonnei* was substituted for cytosine in CG097 and M90T. However, these two nucleotide changes did not change the corresponding amino acid; in all cases it was leucine. Downstream from the 3' end of *ipaR*, two additional bases had changed: at bp 1691, A in M90T and YSH6000 was converted into C in CG097 and *S. sonnei*, and at bp 1763, G in M90T, YSH6000, and *S. sonnei* was converted into A in CG097.

The promoter cloning vector pKK232-8 has been used to identify and measure relative activities of various promoters (8, 22, 24), since it has three stop codons between the multiple cloning sites and the AUG of the *cat* gene, which prohibits readthrough into the *cat* gene (5). Functional analyses of CAT activity of the promoter-CAT derivatives of pKK232-8 containing the 5' flanking sequences of the 24-kDa protein gene or *ippI*, *ipaB*, *ipaC*, or *ipaD* defined the strong promoters upstream of the 24-kDa-protein gene and *ipaD* and the weak promoters upstream of the *ippI* and *ipaB* genes (Fig. 4). CAT activity was not detected in plasmid pWS832 (carrying the 5' flanking sequence of *ipaC*), indicating that the *ipaC* gene does not contain a promoter, although a putative promoter sequence was found upstream of initiation codon of the *ipaC* (2, 21, 26, 29). Complementation analyses showed that IpaR positively regulates the expression of CAT under the direction of the *ippI* promoter region but not under the direction of the 24-kDa-protein gene, *ipaB*, or *ipaD* promoter region. Since the 24-kDa-protein gene, *ippI*, *ipaB*, *ipaC*, and *ipaD* of *S. dysenteriae* CG097 are closely linked and arranged clockwise in the large invasion plasmid (29) and the 24-kDa-protein gene and *ipaD* are both expressed independently from their respective promoters (Fig. 4, lanes 2 and 11), we deduce that the *ippI*, *ipaB*, and *ipaC* genes are organized as an operon, and it appears that the *ippI* promoter is activated by IpaR in *trans* to express coordinately these three genes, *ippI*, *ipaB*, and *ipaC*. Although the *ipaB* gene contains a weak promoter (Fig. 4, lane 7), it may be nonfunctional in vivo, because IpaR has no effect on it. This is further supported by the work of Venkatesan et al. (26) who detected a 3.3-kb RNA by Northern (RNA) blotting in *S. flexneri* which encompasses the *ippI*, *ipaB*, and *ipaC* genes. Although the 2.4- and 1.4-kb RNA bands which hybridized with individual *ipaB* and *ipaC* probe were also observed, Venkatesan et al. claimed that these two RNAs could conceivably be derived from the 3.3-kb transcript by an unknown processing scheme (26).

A putative -10 sequence (TATAAA) and a putative -35 sequence (TTGAAA) which almost match with TATAAT (-10 sequence) and TTGACA (-35 sequence) of *E. coli* were found at positions -318 bp and -337 bp, respectively. A sequence of dyad symmetry (TGTTTTAACTTAAACA) is found between -605 and -395. IpaR polypeptide shows significant sequence homology with the bacteriophage P1 partition protein ParB (7, 27). The ParB has been identified as a DNA-binding protein, and its recognition site consists of a 34-bp DNA sequence including a sequence of dyad symmetry (10). Therefore, it is presumed that IpaR protein may bind to TGTTTTAACTTAAACA sequence and render the binding of RNA polymerase to the *ippI* promoter sequences efficient and constitutive. Alternatively, as a positive regulator, IpaR protein may interact either with DNA sequence or with other proteins such as transcriptional factor(s) or repressor. The interaction of IpaR protein with specific DNA sequence or another protein(s) remains to be elucidated.

#### ACKNOWLEDGMENTS

We thank M. P. Jackson and N. Habib for the strains containing pKK232-8 and pACYC177 plasmids and stimulating discussion. We express our sincere gratitude to J. A. Moshier for help with AMBIS Radioanalytic Imaging System and to M. A. Leon for critical reading of this manuscript. This work was supported by Public Health Service grant A1-25104 from the National Institutes of Health.

#### REFERENCES

- 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.* 6:627-635.
- 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.
- Baudry, B., A. T. Maurelli, P. Clerc, J. C. Sadoff, and P. J. Sansonetti. 1987. Localization of plasmid loci necessary for entry of *Shigella flexneri* into HeLa cells, and characterization of one locus encoding four immunogenic polypeptides. *J. Gen. Microbiol.* 113:3403-3413.
- Berger, S. L., and A. R. Kimmel. 1987. Guide to molecular cloning techniques: methods in enzymology. Academic Press, Inc., San Diego, Calif.
- Brosius, J. 1984. Plasmid vectors for the selection of promoters. *Gene* 27:151-160.
- Buysse, J. M., C. K. Stover, E. V. Oaks, M. M. Venkatesan, and D. J. Kopecko. 1987. Molecular cloning of invasive plasmid antigen (*ipa*) genes from *Shigella flexneri*: analysis of *ipa* gene products and genetic mapping. *J. Bacteriol.* 169:2561-2569.
- Buysse, J. M., M. Venkatesan, J. A. Mills, and E. V. Oaks. 1990. Molecular characterization of a trans-acting, positive effector (*ipaR*) of invasion plasmid antigen synthesis in *Shigella flexneri* serotype 5. *Microb. Pathog.* 8:197-211.
- Dillard, J. P., and J. Yother. 1991. Analysis of *Streptococcus pneumoniae* sequences cloned into *Escherichia coli*: effect of promoter strength and transcription terminators. *J. Bacteriol.* 173:5105-5109.
- Formal, S. B., P. Gemski, Jr., L. S. Baron, and E. H. LaBrec. 1971. A chromosomal locus which controls the ability of *Shigella flexneri* to evoke keratoconjunctivitis. *Infect. Immun.* 3:73-79.
- Martin, K., S. A. Friedman, and S. J. Austin. 1987. Partition site of the P1 plasmid. *Proc. Natl. Acad. Sci. USA* 84:8544-8547.
- 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.
- Maurelli, A. T., B. Blackmon, and R. Curtiss III. 1984. Temperature-dependent expression of virulence genes in *Shigella* species. *Infect. Immun.* 43:195-201.
- Maurelli, A. T., and P. J. Sansonetti. 1988. Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. *Proc. Natl. Acad. Sci. USA* 85:2820-2824.
- Pal, T., and T. L. Hale. 1989. Plasmid-associated adherence of *Shigella flexneri* in a HeLa cell model. *Infect. Immun.* 57:2580-2582.
- Petersen, S. K., and F. G. Hansen. 1991. A missense mutation in the *rpoC* gene affects chromosomal replication control in *Escherichia coli*. *J. Bacteriol.* 173:5200-5206.
- Roosen, K. J., R. G. Fenwick, Jr., and R. Curtiss III. 1971. Synthesis of ribonucleic acid and protein in plasmid-containing minicells of *Escherichia coli* K-12. *J. Bacteriol.* 107:21-33.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Sanger, C. F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Sansonetti, P. J., T. L. Hale, G. J. Daurin, C. Kapfer, H. H. Collins, and S. B. Formal. 1983. Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromo-

- somal genes from *Shigella flexneri*. *Infect. Immun.* **39**:1392-1402.
20. Sansonetti, P. J., D. J. Kopecko, and S. B. Formal. 1982. Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect. Immun.* **35**:852-860.
  21. 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 in *Shigella flexneri* 2a. *Mol. Microbiol.* **3**:1191-1201.
  22. Shimizu, I., and A. Kaji. 1991. Identification of the promoter region of the ribosome-releasing factor cistron (*frr*). *J. Bacteriol.* **173**:5181-5187.
  23. Silva, R. M., M. R. F. Toledo, and L. R. Trabulsi. 1982. Plasmid-mediated virulence in *Shigella* species. *J. Infect. Dis.* **146**:99.
  24. Sung, L. M., M. P. Jackson, A. D. O'Brien, and R. K. Holmes. 1990. Transcription of the Shiga-like toxin type II and Shiga-like toxin type II variant operons of *Escherichia coli*. *J. Bacteriol.* **172**:6386-6395.
  25. Venkatesan, M. M., and J. M. Buysse. 1990. Nucleotide sequence of invasion plasmid antigen gene *ipaA* from *Shigella flexneri* 5a. *Nucleic Acids Res.* **18**:1648.
  26. 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.
  27. Watanabe, H., E. Arakawa, K. I. Ito, J.-I. Kato, and A. Nakamura. 1990. Genetic analysis of an invasion region by use of a *Tn3-lac* transposon and identification of a second positive regulator gene, *invE*, for cell invasion of *Shigella sonnei*: significant homology of *InvE* with *ParB* of plasmid P1. *J. Bacteriol.* **172**:619-629.
  28. Watson, M. E. E. 1984. Compilation of published signal sequences. *Nucleic Acids Res.* **12**:5145-5164.
  29. Yao, R., and S. Palchaudhuri. 1991. Nucleotide sequence of the *ipaBCD* structural genes of *Shigella dysenteriae*. *Mol. Microbiol.* **5**:2217-2221.
  30. Yao, R., K. C. Palmer, M. A. Leon, and S. Palchaudhuri. 1991. *Shigella dysenteriae* 60R adheres to and invades tissue culture cells in the absence of virulence plasmid. *FEMS Microbiol. Lett.* **83**:323-328.