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 PstI-HindIII fragment containing the ipaR locus was cloned from the large invasion plasmid of Shigella dysenteriae CG907, and its nucleotide sequence was completely determined. The IpaR protein (35 kDa, calculated from the DNA sequence) was synthesized in Escherichia coli λ1411 minicells containing the 1,937-bp PstI-HindIII 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 ippl, ipaB, ipaC, and ipaD genes defined strong promoters upstream of the ippl 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 ippl promoter region increased 1.8-fold in the presence of IpaR protein. On the basis of our data, we suggest that an operon comprising ippl, ipaB, and ipaC is positively regulated by IpaR protein which has a trans effect on a DNA sequence upstream of the ippl 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 (130 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 invR, 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 CG907 and describe its positive regulation on the expression of ippl, ipaB and ipaC genes through a trans effect on the 5' flanking region of the ippl 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 λ1411, 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 BamHI and HindIII digestions (396 bp, −317 to +79, with respect to the initiation codon of the 24-kDa-protein gene), a 772-bp HindIII fragment of the 5' flanking sequence of the ippl gene (−605 to +167) from pWS104, a 1,707-bp PvuII-SalI 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 SspI-BamHI fragment of the 5' flanking region of the ipaC gene from pWS102 (−406 to +426, also containing 18 bp of pUC8 following +426), and a 632-bp HindII 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 ippl promoter region, a 562-bp BalI-HindIII fragment (−395 to +167, 210-bp deletion from the 5' end of the 772-bp HindIII fragment) and a 414-bp HaeIII-HindIII fragment (−247 to +167, 358-bp deletion from the 5' end of the 772-bp HindIII fragment) were linked to SmaI and HindIII sites of
TABLE 1. Characterization of recombinant plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Vector</th>
<th>Promoter region or gene</th>
<th>Inserted DNA</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pWS106</td>
<td>pUC8</td>
<td>ipaR, 9-kDa-protein gene</td>
<td>1,937 bp</td>
<td>pWS100*</td>
</tr>
<tr>
<td>pWS396</td>
<td>pKK232-8</td>
<td>24-kDa-protein gene</td>
<td>396 bp, -317 to +79</td>
<td>pWS104*</td>
</tr>
<tr>
<td>pWS772</td>
<td>pKK232-8</td>
<td>ipaI</td>
<td>772 bp, -605 to +167</td>
<td>pWS104*</td>
</tr>
<tr>
<td>pWS210</td>
<td>pKK232-8</td>
<td>ippI</td>
<td>562 bp, -395 to +167</td>
<td>pWS104*</td>
</tr>
<tr>
<td>pWS358</td>
<td>pKK232-8</td>
<td>ipaP</td>
<td>414 bp, -247 to +167</td>
<td>pWS104*</td>
</tr>
<tr>
<td>pWS1602</td>
<td>pKK232-8</td>
<td>ipaB</td>
<td>1,707 bp, -431 to +1171 and 105 bp from pUC8</td>
<td>pWS102*</td>
</tr>
<tr>
<td>pWS1832</td>
<td>pKK232-8</td>
<td>ipaC</td>
<td>850 bp, -406 to +426 and 18 bp from pUC8</td>
<td>pWS102*</td>
</tr>
<tr>
<td>pWS632</td>
<td>pKK232-8</td>
<td>ipaD</td>
<td>632 bp, -432 to +209</td>
<td>pWS103*</td>
</tr>
<tr>
<td>pWS1879</td>
<td>pACYC177</td>
<td>ipaR, 9-kDa-protein gene</td>
<td>1,879 bp and 93 bp of pUC8</td>
<td>pWS106, this work</td>
</tr>
<tr>
<td>pWS1009</td>
<td>pACYC177</td>
<td>9-kDa-protein gene</td>
<td>731 bp</td>
<td>pWS106, this work</td>
</tr>
</tbody>
</table>

a 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 replication of pKK232-8 for complementation analyses, a 1.972-bp PvuII 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 HincII 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 X1411. Minicells prepared from 250-ml overnight cultures of E. coli X1411 were isolated according to the method of Rozen et al. (16). Plasmid-encoded proteins synthesized in minicells were labeled with [35S]-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 hydroxylation of polypeptides.

CAT enzyme assay. Bacteria were grown in L broth to an A660 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 x 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 µg of cell extract, 20 µl of freshly made acetyl coenzyme A (Sigma) at 3.5 mg/ml, 1 µl of [14C]chloramphenicol, and 2.5 M Tris-HCl (pH 7.8) in a total final volume of 150 µl. 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 AMBISS 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 PstI-HindIII) 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
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 sites sequences are located upstream from the initiation codon of *ipaR*. At the 5′ end of the 1,937-bp *Pstl*-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 *Pstl*-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 *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 hydrophobic residues preceded by two basic residues, lysines, with serine and alanine at the −1 and −3 positions before the cleavage site (i.e., MIKEKILSVACCGILAY)′s KLS). In addition, the 9-kDa polypeptide contains two hydrophobic transmembrane domains as predicted by our hydrophatyp 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 [35S]-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-Dalgarno 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
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 recombinant 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 ippl) and pWS1602 (carrying the promoter region of ippB) (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 ippl 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 ippl 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 ippl 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 ippl gene. The above-described results suggest the presence of a DNA sequence responsible to IpaR protein in the 5' flanking sequence of the ippl gene. To further map the location of this DNA sequence between −605 bp and the initiation codon of ippl, two deletion mutant plasmids, pWS210 (−395 bp 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. 3. Minicell analysis of plasmid-coded proteins. The positions of molecular mass markers are indicated by the bars. Lane A, pUC8; lane B, pWS106.
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., TATTAAT in M90T, YSH6000, and S. sonnei; TATGAAT in CG097). Furthermore, the change of this single base (A to G) produced two direct repeats (CAATATGAATCAATGAAAT) 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.
histidine, in the IpαR of CG097 instead of the glutamine in IpαR 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 ipαR, 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 ipαl, ipαB, ipαC, or ipαD defined the strong promoters upstream of the 24-kDa-protein gene and ipαD and the weak promoters upstream of the ipαl and ipαB genes (Fig. 4). CAT activity was not detected in plasmid pWS832 (containing the sequence ipαC), implying that the ipαC gene does not contain a promoter, although a putative promoter sequence was found upstream of initiation codon of the ipαC (2, 21, 26, 29). Complementation analyses showed that IpαR positively regulates the expression of CAT under the direction of the ipαl promoter region but not under the direction of the 24-kDa-protein gene, ipαB, or ipαD promoter region. Since the 24-kDa-protein gene, ipαl, ipαB, ipαC, and ipαD of S. dysenteriae CG097 are closely linked and arranged clockwise in the large invasion plasmid (29) and the 24-kDa-protein gene and ipαD are both expressed independently from their respective promoters (Fig. 4, lanes 2 and 11), we deduce that the ipαl, ipαB, and ipαC genes are organized as an operon, and it appears that the ipαl promoter is activated by IpαR in trans to express coordinately these three genes, ipαl, ipαB, and ipαC. Although the ipαB gene contains a weak promoter (Fig. 4, lane 7), it may be nonfunctional in vivo, because IpαR 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 ipαl, ipαB, and ipαC genes. Although the 2.4- and 1.4-kb RNA bands which hybridized with individual ipαB and ipαC 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 (TGTITTTAACTTAA AACA) is found between −605 and −395. IpαR 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 IpαR protein may bind to TGTITTTAACTTAA AACA sequence and render the binding of RNA polymerase to the ipαl promoter sequences efficient and constitutive. Alternatively, as a positive regulator, IpαR protein may interact either with DNA sequence or with other proteins such as transcriptional factor(s) or repressor. The interaction of IpαR 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. Mosher for help with AMBSI Radioanalytic Imaging System and to M. A. Leon for critical reading of this manuscript. This work was supported by Public Health Service grant AI-25104 from the National Institutes of Health.

REFERENCES


