JOURNAL OF BACTERIOLOGY, Sept. 1995, p. 5040–5047
0021-9193/95/$04.00+0

Vol. 177, No. 17

Characterization of the *Salmonella typhimurium* *pagC/pagD* Chromosomal Region

JOHN S. GUNN,§ CELIA M. ALPUCHE-ARANDA,† WENDY P. LOOMIS,‡
WILLIAM J. BELDEN, AND SAMUEL I. MILLER§*

Infectious Disease Unit, Massachusetts General Hospital and
Harvard Medical School, Boston, Massachusetts 02114

Received 23 May 1994/Accepted 30 June 1995

The PhoP/PhoQ two-component system regulates *Salmonella typhimurium* genes that are essential to bacterial virulence and survival within macrophages. The best characterized of these PhoP-activated genes (*pag*) is *pagC*, which encodes a 188-amino-acid envelope protein (W. S. Pulkkinen and S. I. Miller, J. Bacteriol. 173:86–93, 1991). We here report the identification of four genes (*pagD*, *envE*, *msgA*, and *envF*) located 5′ to *pagC*. Each gene is transcribed from its own promoter, two of which (*msgA* and *pagD*) were defined by primer extension analysis. Three of these genes (*pagD*, *envE*, and *envF*) are predicted to encode envelope proteins. The *pagD* gene is transcribed in a direction opposite from that of and adjacent to *pagC* and is positively regulated by PhoP/PhoQ. Transposon insertions within *pagD* and *msgA* attenuate bacterial virulence and survival within macrophages; however, deletion of *pagD* has no effect on virulence. The product of the *envF* gene is predicted to be a lipoprotein on the basis of the presence of a consensus lipid attachment site. The low G+C content of these genes and the homology of *msgA* to *Shigella* plasmid DNA suggest that this region may have been acquired by horizontal transmission.

Recent work with several bacterial virulence systems has demonstrated that virulence factors are coordinately regulated in response to environmental signals (27, 30). One such regulon promotes *Salmonella typhimurium* virulence and is controlled by the PhoP/PhoQ proteins (32). PhoP is a member of the OmpR subclass of transcriptional activators, and PhoQ is similar to other sensor histidine kinases, such as EnvZ, that have a periplasmic domain (32). PhoP and PhoQ both positively and negatively regulate more than 40 gene products that include virulence factors (33). PhoP-activated genes (*pag*) function to promote mouse virulence, survival within macrophages, and resistance to low pH and cationic antimicrobial proteins (31). *pag* transcription is maximally induced within acidified macrophage phagosomes but not after endocytosis by epithelial cells, demonstrating that PhoQ responds to specific signals present within macrophage phagosomes (1). Several loci that require PhoP and PhoQ for expression, including *phoN*, which encodes a periplasmic acid phosphatase, and 16 other PhoP-activated loci arbitrarily designated *pagA-pagP* have been defined (3, 21, 32). Strains with transposon insertions in *pagC*, *pagD*, *pagJ*, *pagK*, and *pagM* have a defect in survival within macrophages as well as a dramatic increase in the mouse 50% lethal dose (LD_{50}) (3). *pagC* is a 188-amino-acid protein that is a member of a family of bacterial outer membrane proteins that are known to be or may be involved in virulence (38).

In many bacteria, virulence genes are located on plasmids or mobile genetic elements or are clustered on the chromosome. Major examples of this genetic organization include the cholera toxin genetic element (26, 44), the *Listeria monocytogenes* *prfA* region (37), and plasmids of *Yersinia* species (9) and *Shigella* species (25). In *S. typhimurium*, a set of genes located on a 60-MDa plasmid promote mouse virulence (15). In addition, multiple genes necessary to induce uptake by epithelial cells are clustered on the chromosome at 58 to 60 min (23) and are similar to plasmid-encoded *Shigella* and *Yersinia* virulence genes (12, 13, 35). Since there are several known examples of clustered bacterial virulence genes and previous work indicated that *pagD* was highly linked to *pagC* (3), we undertook the present study to examine the *pagC* region and to determine if genes linked to *pagC* encode products that promote virulence and survival within macrophages.

MATERIALS AND METHODS

Bacterial strains and media. All *Salmonella* strains are derivatives of ATCC 14028. Bacterial strains and plasmids are described in Table 1. Luria-Bertani broth was used as rich bacterial growth medium. Antibiotics were added to Luria-Bertani broth or agar in the following concentrations (in micrograms per milliliter): ampicillin, 100; chloramphenicol, 25; and kanamycin, 45. Chemicals were analytical grade or better from Sigma Chemical Co. (St. Louis, Mo.). Transposon-encoded gene fusion activity was assayed by using 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) (Bethesda Research Laboratories, Gaithersburg, Md.) or 5-bromo-4-chloro-3-indolyl phosphate (XP) (Amresco, Solon, Ohio) that was added to plates at a final concentration of 40 μg/ml.

Biochemical assays. Alkaline phosphatase (AP) (28) and β-galactosidase (29) assays were performed as previously described and are reported in units as defined by Miller (29).

DNA sequencing and analysis. Double-stranded templates were sequenced by the dideoxy-sequencing method (40) using the Sequenase kit (U.S. Biochemicals, Cleveland, Ohio). Sequencing reaction mixtures were electrophoresed in 6% polyacrylamide–8 M urea gels in an Owl Scientific sequencing apparatus (Cambridge, Mass.). Computer analysis of the DNA sequence was accomplished with the GENEPRO (Riverside Scientific, Riverside, Calif.) and Wisconsin package (Genetics Computer Group) programs.

Bacterial genetic techniques. Mutagenesis of ATCC 14028 with Tn*phoA* was performed by using the broad-host-range plasmids pRK290 and pPH1JI as previously described (43). Strains with AP activity were screened for chromosomal linkage of Tn*phoA* to the Tn*10*Δ*16*Δ*17* of strain AK3233, which is 75% linked to the *pagC* gene. Plasmid mutagenesis was accomplished by using Mu*dJ* (18) or Tn*phoA* (24). Plasmid pWPL17 was mutagenized with λTn*phoA* as previously described (16) or with Mu*dJ* by growth at the permissive temperature for transposition in strain HT10288 (32). Allelic replacement of *msgA* with a

* Corresponding author. Phone: (617) 726-3812. Fax: (617) 726-7416.

† Present address: Departamento de Infectologia, Hospital Infantil de Mexico, Mexico D.F., 06720 Mexico.

‡ Present address: Department of Microbiology, School of Medicine, University of Washington, Seattle, WA 98195.

§ Present address: K116 Biomedical Sciences Research Building, Box 357710, University of Washington, Seattle, WA 98195.

TABLE 1. Plasmids and strains used and their relevant properties

Strain or plasmid	Relevant genotype and/or information	MSI[a]	LD_{50}	Source[b]
S. typhimurium strains				
ATCC14028	Wild type	3.90	<20	ATCC
CS019	*phoN2 zxx*::6251Tn*10*d-Cm			32
JSG318	CS019; Δ*pagD*	4.10	<20	This work
CS585	CS019; *pagD*::Tn*phoA*	0.002	4×10^5	3
CS586	CS585; *phoP*105::Tn*10*d-Tet			3
JSG205	ATCC 14028; *msgA*::Mu*dJ*	0.01	6×10^3	This work
JSG225	JSG205; *phoP105*::Tn*10*d-Tet			This work
CS811	CS019; *envE*::Tn*phoA*			This work
CS812	CS811; *phoP105*::Tn*10*d-Tet			This work
CS100	ATCC 14028; *phoP*105::Tn*10*d-Tet	0.01		Derivitive of TT13208 (gift of Zhu and Roth)
JSG232	JSG205; *envF*::pGPP2			This work
JSG234	CS019; *envF*::pGPP2			This work
JSG352	CS019; *envF*::pGP67		<20	This work
JSG235	JSG234; *phoP105*::Tn*10*d-Tet			This work
JSG244	JSG205; *phoP105*::Tn*10*d-Tet			This work
CS099	ATCC 14028; *zxx*3024::Tn*10*Δ*16*Δ*17 pol-2* (Whitfield *polA* amber)			This work
Other salmonellae				
Ty-2	Vi positive			C. Hardegree, FDA
S. paratyphi A	ATCC 9150			ATCC
S. paratyphi C	ATCC 13428			ATCC
S. enteriditis	Clinical isolate			K. Killeen, VRI
E. coli strains				
SM10λ*pir*	*thi-1 thr-1 leuB6 supE44 tonA21 lacY1 recA*::RP4-2-Tc::Mu			
DH5α	F^- ϕ 80d*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *endA1 recA1 hsdR17 deoR thi-1 supE*44λ^- *gyrA96 relA1*			
Wild type	Clinical isolate			MGH bacteriology laboratory
Other *Enterobacteriaceae*				
Y. enterocolitica	Clinical isolate			MGH bacteriology laboratory
V. cholerae	Clinical isolate			Peruvian epidemic
C. fetus	Clinical isolate			MGH bacteriology laboratory
C. freundii	Clinical isolate			MGH bacteriology laboratory
K. pneumoniae	Clinical isolate			MGH bacteriology laboratory
S. flexneri	Clinical isolate			MGH bacteriology laboratory
2457T	Type 2a virulent strain			Gift of T. Maurelli and R. Sandlin
BS103	2457T virulence plasmid-cured derivative			Gift of T. Maurelli and R. Sandlin
S. sonnei	Clinical isolate			MGH bacteriology laboratory
M. morganii	Clinical isolate			MGH bacteriology laboratory
P. stuartii	Clinical isolate			MGH bacteriology laboratory
Plasmids				
pBCSK$^+$	Cloning vector, Cm^r			Stratagene
pMG11	pBCSK$^+$ containing a 2.4-kb *Pst*I fragment from pWP061			This work
pWPL17	pBR322 containing a 2.8-kb *Hpa*I fragment from pWP061			This work
pCAA9	pWPL17 containing a Tn*phoA* insertion in *envF*			This work
pGP704	*pir*-dependent suicide vector			34
pGP67	pGP704 containing a 550-bp region of the *envF* gene			This work
pGPP2	pGP704 containing the cloned *envF*::*phoA* gene fusion			This work
pWP061	Cosmid clone containing the *pagC* region			38

[a] MSI is calculated by dividing the number of surviving organisms at 24 h postinfection by the number of cell-associated organisms present after the 30-min infection.
[b] MGH, Massachusetts General Hospital; ATCC, American Type Culture Collection; FDA, Food and Drug Administration; VRI, Virus Research Institute.

plasmid copy of the gene containing a Mu*dJ* insertion was accomplished in three steps. A P22 phage lysate containing the mutagenized plasmid was transduced into a *polA* mutant of strain ATCC 14028, and homologous recombination events were identified by selection for the plasmid-encoded ampicillin resistance. Marker exchange events were subsequently identified by screening for loss of ampicillin resistance and maintenance of Mu*dJ*-encoded kanamycin resistance. Replacement of the wild-type copy of the gene with the copy containing the Mu*dJ* insertion was confirmed by Southern blot hybridization. Bacteriophage P22HT*int*-mediated transduction was performed as described by Davis et al. (10). Inactivation of *envF* was accomplished by the following. A 550-bp region internal to the *envF* coding sequence was PCR amplified and cloned into the suicide vector pGP704. The resulting plasmid clone, pGP67, was mated into strain CS019, and transconjugates were selected by the vector antibiotic resistance. The insertion of pGP67 into the *envF* gene was confirmed by Southern blot hybridization.

Construction of a *pagD* deletion. Plasmid clone pMG11, containing *pagD* on a 1.4-kb *Pst*I fragment, was used as a template for PCR amplification. Two products corresponding to the DNA upstream and downstream of *pagD* were amplified with the −20 forward primer and primer JG25 (5′ CCGCTCGAGGAAC GTCATTGAC 3′) or the −48 reverse primer and primer JG26 (5′ CCGCTCG AGAACATAATGTTATTTATC 3′), respectively. These PCR products were digested with *Xho*I (included at the 5′ end of primers JG25 and JG26), ligated,

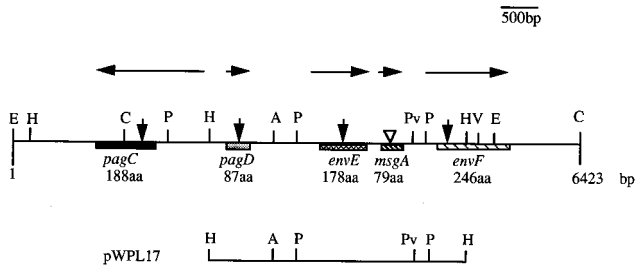


FIG. 1. Partial physical map of the restriction endonuclease sites of the *pagC* chromosomal region. Horizontal arrows demonstrate the directions of transcription. Vertical arrows denote Tn*phoA* insertions, and the hollow triangle indicates a Mu*dJ* insertion. Below the chromosomal map is a representation of the DNA insert in plasmid pWPL17, which was mutagenized with Tn*phoA* and Mu*dJ*. Letter designations: A, *Acc*I; C, *Cla*I; E, *Eco*RI; H, *Hpa*I; P, *Pst*I; and V, *Eco*RV. aa, amino acids.

and then digested with *Xba*I and *Sph*I (located in the DNA corresponding to the vector multiple cloning site at the end of either PCR fragment). This DNA was then ligated to the sucrose selection vector pCVD442, transformed into strain SM10λ Pir, and mated into wild-type *Salmonella* strains. Cells that had deleted the vector were selected on low-salt, 10% sucrose plates (4). The incorporation of the deletion into the chromosome was confirmed by Southern blot hybridization.

Southern and Northern (RNA) blot analysis. Chromosomal DNA was isolated by the method described by Mekalanos et al. (26). DNA was digested, size fractionated in agarose gels, and transferred to GeneScreen Plus membranes (Dupont, Boston, Mass.) by the manufacturer's protocol. High-stringency hybridization (65°C) and detection of hybridization were accomplished with the Genius kit (Boehringer Mannheim, Indianapolis, Ind.). All steps followed the manufacturer's protocol, except that the antibody conjugate was diluted 1:10,000 and 1% dry milk was used as the blocking agent. Northern analysis was accomplished by using radioactive ^{32}P-labeled probes as previously described (35).

Primer extension analysis. RNA was isolated from *Salmonella* strains by the hot-phenol procedure (7). Primers (0.2 pmol) end labeled with [γ-^{32}P]ATP were annealed to *S. typhimurium* RNA (10 μg) and extended with reverse transcriptase. Reaction mixtures were electrophoresed in 8 M urea–6% polyacrylamide gels adjacent to sequencing reactions initiated with the primers used for cDNA synthesis.

Bacterial survival within macrophages. Stationary-phase bacteria were opsonized for 30 min in 10% normal mouse serum. Opsonized bacteria were added at a ratio of 10:1 (bacteria-cells) to bone marrow-derived macrophages harvested from male BALB/c mice (5). Gentamicin (8 μg/ml) was added to macrophage culture medium 30 min postinfection to kill extracellular bacteria. The numbers of surviving organisms were determined at 1, 4, and 24 h after the addition of gentamicin. Assays were done in triplicate on three separate occasions.

Mouse virulence studies. Dilutions of stationary-phase cultures were injected intraperitoneally into 35-day-old (16- to 18-g) female BALB/c mice (Charles River Park Laboratories, Wilmington, Mass.). Experiments strictly adhered to the guidelines of the animal care committee of Massachusetts General Hospital and Harvard Medical School. LD_{50}s were determined by the method described by Reed and Muench (39). A minimum of 10 mice at three different dilutions were tested for each strain examined.

Nucleotide sequence accession number. The DNA sequence presented has been deposited in GenBank and assigned accession number U31849.

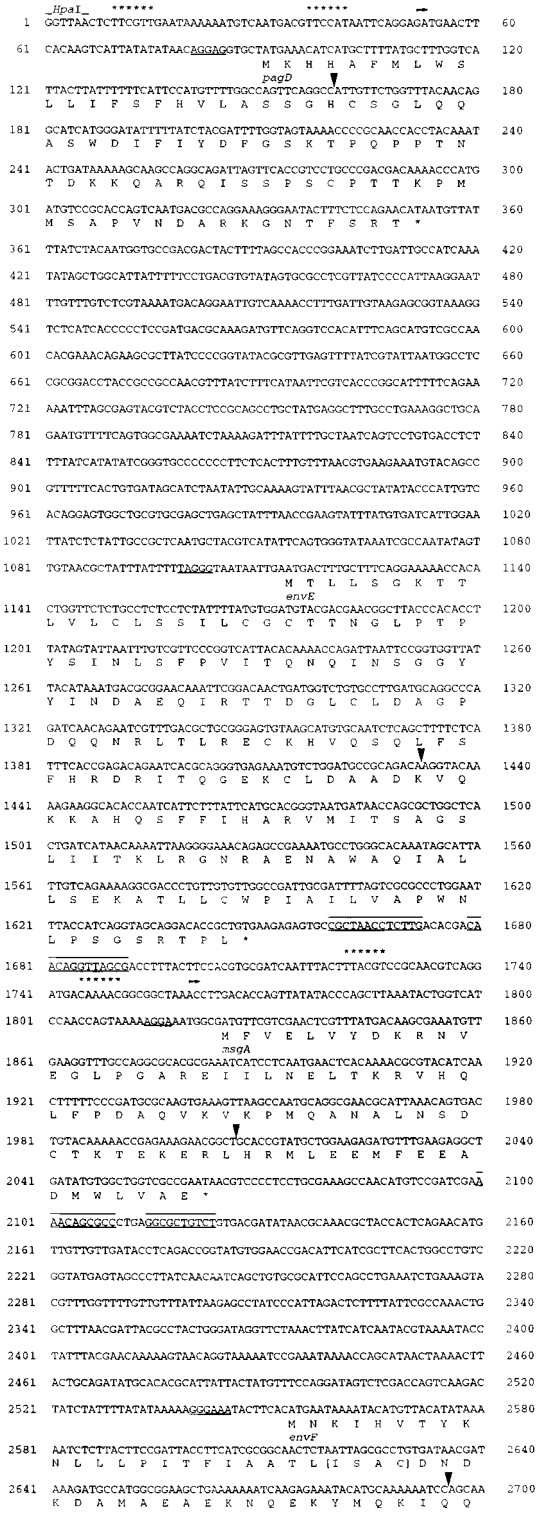
RESULTS

Identification of transcriptional units linked to *pagC*. Previous analysis of *pagC* indicated that transcription likely terminated immediately after the coding sequence (38). In addition, DNA sequence analysis of approximately 1 kb 3′ to this potential terminator did not reveal open reading frames that were likely to encode a protein product. This suggested that no transcriptional units were located in the region immediately downstream of *pagC*. Therefore, our initial effort to identify linked virulence genes was targeted to the DNA upstream of *pagC*.

To help identify new loci upstream of *pagC*, *E. coli* carrying plasmid pWPL17 containing 2.8 kb of DNA 5′ to *pagC* (Table 1 and Fig. 1) was mutagenized with the transposons Mu*dJ* and Tn*phoA*, and strains with AP or β-galactosidase activity were identified on chromogenic substrates. Several strains that contained plasmids with active Mu*dJ*- or Tn*phoA*-generated gene fusions were identified. In addition, previous work demonstrated that a PhoP-activated gene, designated *pagD*, had a linkage to the Tn*10*Δ*16*Δ*17* of strain AK3233 that was similar to that of *pagC* (3). By the same screening method, another strain was identified that contained an active chromosomal Tn*phoA* insertion closely linked to *pagC*. These two strains with chromosomal Tn*phoA* insertions and the plasmids containing transposon insertions that were generated in this study were further characterized. Physical maps of the restriction endonuclease sites surrounding the transposon insertions in strains with active plasmid or chromosomal *lacZ* and *phoA* gene fusions were made to determine the relationship of the transposon insertions to *pagC* (data not shown). This analysis revealed that several regions of the DNA were transcribed in a direction opposite to that of *pagC* (Fig. 1).

Genes linked to *pagC* encode four novel proteins. To further analyze the genes defined by transposon insertions, the DNA sequence of this region was determined (Fig. 2). DNA containing this region was previously cloned in plasmid pWP061 (38). By using various subclones of pWP061 and oligonucleotide primers, 4 kb of DNA was sequenced between the *Hpa*I site 737 bp upstream of the start codon of *pagC* to a *Cla*I site far upstream. The DNA sequences of the fusion junctions of all Tn*phoA* and Mu*dJ* gene fusions were also determined, which produced important information pertaining to the correct reading frames. The DNA sequence data revealed four open reading frames (ORFs) predicted to be transcribed and translated on the basis of the data derived from the Tn*phoA* and Mu*dJ* insertions. All ORFs revealed typical ribosome binding sites 6 to 11 bases from the predicted start of translation (41). The ORF immediately upstream and transcribed in a direction opposite to that of *pagC* is predicted to encode a short envelope protein of 87 amino acids (unprocessed). This ORF contains the transposon generated gene fusion previously shown to be PhoP regulated (*pagD*). It is followed by a second ORF (*envE*) predicted to encode an envelope protein of 178 amino acids (unprocessed). This ORF is followed by a structure that could function as a Rho-independent transcriptional terminator (Fig. 2). The third ORF (later called *msgA* for macrophage survival gene) is predicted to encode a small protein similar in size to that of the first gene product (79 amino acids) and is also followed by a structure that could function as a Rho-independent transcriptional terminator (Fig. 2). The DNA sequence predicts that this protein is composed of several charged residues with a large number of negatively charged amino acids residing at the carboxy terminus. We did not identify any Tn*phoA* insertions within this gene, and the predicted protein product does not contain a structure resembling a signal sequence at its amino terminus or any hydrophobic stretches; therefore, the third ORF is unlikely to encode an envelope protein. The final ORF (*envF*) is predicted to encode an envelope protein of 278 amino acids (unprocessed). A computer search of known protein motifs revealed that EnvF is predicted to contain a consensus prokaryotic membrane lipid attachment site and, therefore, is likely to be a lipoprotein (see Fig. 2 for consensus site location).

The predicted proteins produced by *pagD*, *envE*, and *envF* contain a typical bacterial signal sequence structure (45). In addition, hydrophobic profiles confirmed the hydrophobic nature of the amino termini of these proteins (data not shown). The EnvE and EnvF proteins also contain hydrophobic stretches that could function as membrane spanning domains (22). The G+C contents of the genes in this region are as follows: *pagC*, 43.4%; *pagD*, 42.1%; *envE*, 45.9%; *msgA*, 46.8%;



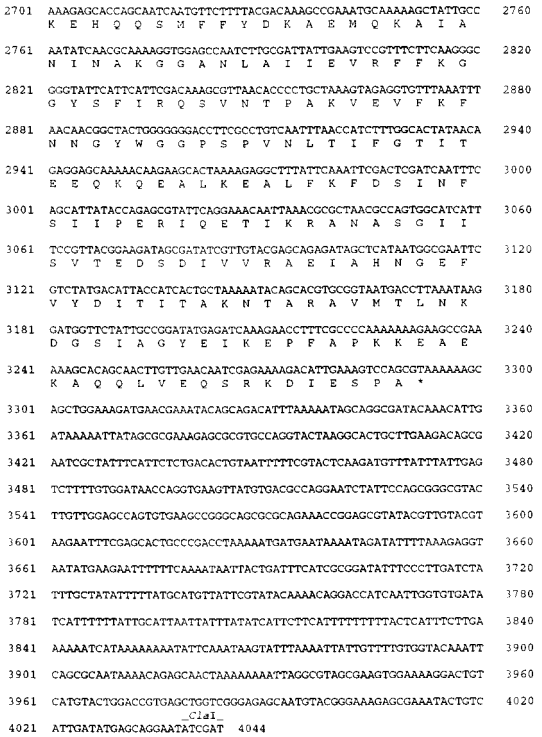


FIG. 2. DNA sequence of the region upstream of *pagC* and the translation of ORFs corresponding to four newly identified genes. The *Hpa*I and *Cla*I sites at the beginning and end of the region are marked. Potential Shine-Delgarno regions are underlined, and stem-loop structures (that could function as Rho-independent terminators) are marked with a line below and above the sequence. Arrowheads denote the locations of the representative transposon insertion within each gene. Horizontal arrows in the *pagD* and *msgA* promoter regions mark the transcriptional start sites, and asterisks mark the predicted −10 and −35 sequences. The consensus lipid attachment site in EnvF is enclosed in brackets.

and *envF*, 40.5%, which is considerably lower than the average G+C content of *S. typhimurium* (52%). A complete search of the database with the predicted protein sequences of these four ORFs showed no significant similarities. Strains containing three distinct Tn*phoA* insertions and one Mu*dJ* insertion, each located in one of the four genes, were chosen for further characterization.

***pagD*, which is divergently transcribed from *pagC*, is positively regulated by PhoP/PhoQ.** Representative strains with transposon insertions were examined to evaluate whether genes transcribed oppositely to *pagC* were increased in synthesis in the presence of PhoP. To accurately determine if these genes were PhoP regulated, it was necessary to recombine plasmid insertions onto the *Salmonella* chromosome. Upon replacement of the wild-type gene with the gene containing the transposon insertion, P22HT*int* lysates made on these strains were transduced into a strain from which PhoP had been deleted (PhoP$^-$) and AP or β-galactosidase levels were monitored. One of these gene fusions (*pagD*), which was also previously examined (3), demonstrated a significant increase in AP activity between PhoP$^-$ and wild-type backgrounds, while the other insertions showed no PhoP regulation (Table 2).

The representative transposon insertion in *envF* was unable to be recombined onto the chromosome, likely because of an insufficient amount of homologous DNA downstream of the transposon. In order to examine the possibility of PhoP regulation of the *envF* gene, a region upstream of this gene through and including the *phoA* gene of the Tn*phoA* transposon was cloned from plasmid pCAA9 as a 3-kb *Pvu*I (blunt-ended)-*Xho*I fragment into the *Eco*RV-*Sal*I sites of the suicide vector pGP704 (34). The *Xho*I site is located within the transposon 3′ to the *phoA* gene. This clone, pGPP2, was mated into *Salmonella* strain CS019, and ampicillin-resistant recombinants

TABLE 2. AP and β-galactosidase gene fusion activities

Strain	Relevant genotype	Gene fusion activity[a]
JSG205	*msgA*::Mu*dJ*	461 ± 16.4 (B)
JSG244	*phoP105*::Tn*10*d-Tet *msgA*::Mu*dJ*	415 ± 13.2 (B)
JSG226	*envE*::Tn*phoA*	50 ± 4.9 (A)
JSG229	*phoP105*::Tn*10*d-Tet *envE*::Tn*phoA*	60 ± 3.5 (A)
JSG204	*pagD*::Tn*phoA*	76 ± 5.6 (A)
JSG225	*phoP105*::Tn*10*d-Tet *pagD*::Tn*phoA*	9 ± 1.9 (A)
JSG234	*envF*::pGPP2	16 ± 1.4 (A)
JSG235	*phoP105*::Tn*10*d-Tet *envF*::pGPP2	19 ± 1.8 (A)
JSG232	*msgA*::Mu*dJ envF*::pGPP2	10 ± 3.0 (A)

[a] A (AP) and B (β-galactosidase) activities were determined as discussed in Materials and Methods and are expressed in units as described by Miller for β-galactosidase (29).

were selected (creating a strain designated *envF*::pGPP2). A *phoP105*::Tn*10*d-Tet mutation was transduced into this strain to create an isogenic pair differing only in the ability to produce a functional PhoP protein. As shown in Table 2, the introduction of the *phoP105*::Tn*10*d-Tet had no effect on the AP levels of these two strains, demonstrating that *envF* is not a PhoP-activated gene.

Assays of virulence and survival within macrophages of strains with insertions in or deletions of *pagD*, *envE*, *msgA*, and *envF*. Since transposon insertions in *pagC* significantly increase the LD_{50} of *S. typhimurium* in BALB/c mice, strains containing transposon insertions linked to *pagC* were evaluated for attenuation of mouse virulence. As can be seen in Table 1, while the transposon insertion in *envE* nor disruption of *envF* by the insertion of a suicide vector into the coding region (see Materials and Methods) had an affect on strain virulence, a Tn*phoA* insertion in *pagD* and the Mu*dJ* insertion 1.8 kb downstream in *msgA* attenuate *S. typhimurium* virulence by greater than 300-fold, compared with wild-type organisms (LD_{50} <20 organisms). This suggested that these two loci are essential to *Salmonella* mouse virulence.

S. typhimurium virulence in mice has been correlated with the ability of the organism to survive within host macrophages (11). To examine the survival capabilities of strains having a virulence defect, *S. typhimurium* cells containing insertions in either *pagD* or *msgA* were used to infect bone marrow-derived macrophages. The results, which are shown in Table 1, demonstrate a macrophage survival defect for these two strains. The survival defect is greater for the *pagD* insertion (macrophage survival index [MSI] = 0.002) than that for the *msgA* insertion (MSI = 0.01), and both defects are equal to or greater than that of the $PhoP^-$ strain (MSI = 0.01).

Since it has previously been observed that Tn*phoA* insertions can cause virulence defects that are not related to the gene inactivated by the transposon, we created a strain deleted specifically for the *pagD* gene (nucleotides 40 to 346). Upon assaying of this strain for virulence, it was found to have an LD_{50} and an MSI comparable to those of wild-type *Salmonella* strains (Table 1). To determine if the virulence defect of the *pagD*::Tn*phoA* strain could be the result of an obvious abnormality in protein secretion caused by the fusion protein, whole cell and membrane protein profiles of the *pagD*::Tn*phoA* strain were compared with those of wild-type *Salmonella* strains. Both protein profiles were shown to be similar to those of wild-type *Salmonella* strains, demonstrating that the virulence defect of the *pagD*::Tn*phoA* strain does not appear to be due to a gross defect in normal protein traffic (data not shown).

To demonstrate that the virulence and macrophage survival defects of *msgA* were not due to a polar effect of the Mu*dJ* insertion on *envF* transcription, a strain with a gene fusion to *envF* was constructed. Plasmid pGPP2 (described above) was recombined into the *msgA*::Mu*dJ* strain, and the AP activity of this strain was compared with that of CS019 containing the recombined pGPP2. These data (shown in Table 2) demonstrate that the transcription of the *envF* gene is unaffected by the *msgA*::Mu*dJ* insertion and that the gene is transcribed from its own promoter. These results were further corroborated by Northern blot analysis. An *envF*-specific PCR fragment was used to probe blots containing wild-type and *msgA*::Mu*dJ* strain RNA. Identical patterns of hybridization with equal intensities were seen between the two samples, again demonstrating that the *msgA*::Mu*dJ* insertion has no effect on downstream *envF* transcription (data not shown). In addition, as mentioned above, the inactivation of *envF* does not attenuate bacterial virulence, as does the Mu*dJ* insertion in *msgA*.

Determination of the *msgA* and *pagD* transcriptional start sites. To define the probable transcriptional start sites of *msgA* and *pagD*, various oligonucleotides complementary to the 5′ end of each ORF or upstream region were used in a primer extension analysis. The results depicted in Fig. 4 show that the *pagD* transcript begins 39 bases upstream of the translational start. The predicted −10 (TTCCAT) and −35 (TTGAAT) regions are similar to the known consensus sequences for *Escherichia coli* promoters (17). The *pagD* transcript is detected in $PhoP^c$ *Salmonella* RNA and not in RNA from $PhoP^-$ *Salmonella* strains (Fig. 3A, lanes 1 and 2). The *msgA* transcriptional start begins 58 bases upstream of the translational start and contains predicted −10 (CAAAAC) and −35 (TTACGT) sequences. These regions do not conform well to consensus −10 and −35 sequences; however, the cDNA from this transcript was easily detected in equal amounts by using various primers in primer extensions of both $PhoP^c$ and $PhoP^-$ RNA.

Distribution of *pagD* and *msgA* genes in members of the family *Enterobacteriaceae* and in two low-G+C-content organisms. The G+C content of the *pagC* chromosomal region is much lower than the average G+C content of *Salmonella* strains. The gene encoding the PhoP-regulated acid phosphatase of *S. typhimurium* (*phoN*) also has a low G+C content (39%), and DNA homologous to *phoN* was found only in two low-G+C-content organisms of several genera tested (14). Therefore, we examined the DNAs of several members of the *Enterobacteriaceae* and two low-G+C-content organisms by blot hybridization for similarity to *pagD* and *msgA*. PCR fragments highly specific to each ORF were labeled and used as probes. Similar to results seen with a *pagC*-specific probe hybridization to the *pagD* probe was seen only in *Salmonella* species at both high and low stringency (Fig. 4A). However, with an *msgA*-specific probe, hybridization was seen at high stringency to all *Salmonella* species examined as well as to *Shigella sonnei*, *Shigella flexneri*, *Klebsiella pneumoniae*, and *Citrobacter freundii* (Fig. 4B). No hybridization was seen to the low-G+C-content organisms *Morganella morganii* or *Providencia stuartii*. Since several virulence genes are known to be located on the 230-kb plasmid of *Shigella* species, we examined if the *Shigella* DNA that hybridized to the *msgA* probe was localized to this plasmid. *S. flexneri* DNA was isolated from a strain with (2457T) and a strain without (BS103) the 230-kb virulence plasmid, electrophoresed, blotted, and probed with the *msgA* probe. Hybridization was observed in identical patterns regardless of the strain from which the DNA was isolated (Fig. 5). Hybridization to plasmid DNA isolated from strain 2457T in patterns identical to those with the chromosomal digests was also observed. The hybridizing bands can be correlated to the bands seen in the ethidium bromide-stained agarose gel. This indicated that the DNA homologous to the

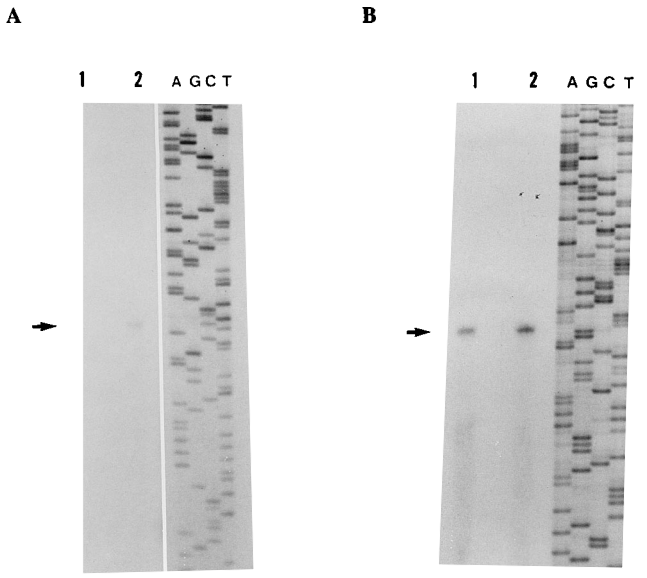


FIG. 3. Primer extension analysis of the *pagD* (A) and *msgA* (B) transcripts. Lanes AGCT in each panel represent di-dioxy sequencing reactions. (A) Lanes: 1, the primer extension reaction initiated with the primer 5′ CCAGAACAAGG CCTGAACTGGCC 3′ with PhoP$^-$ RNA; 2, the extension reaction initiated with the same primer but with PhoP^c RNA. (B) Lanes: 1, the primer extension reaction initiated with the primer 5′GAGGATGATTTCGCGTGCGCCTGGC 3′ with PhoP$^-$ RNA; 2, the extension reaction initiated with the same primer but with PhoP^c RNA.

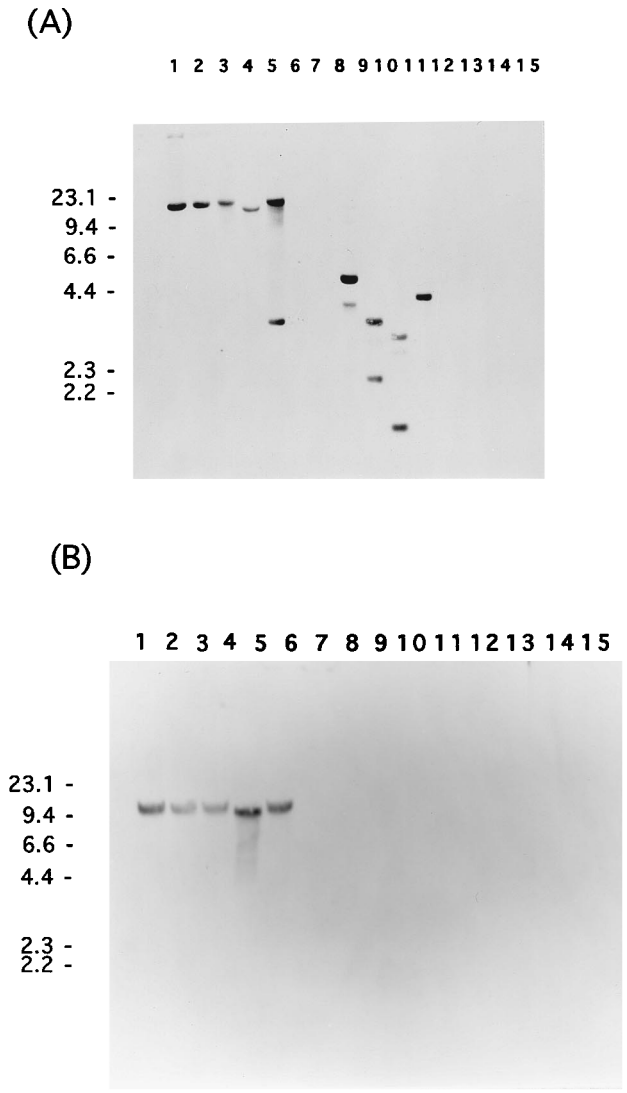


FIG. 4. High-stringency Southern blot hybridization of DNA digested with *Hin*dIII from selected *Enterobacteriaceae* and two low-G+C-content organisms to a *msgA*-specific probe (A) or a *pagD*-specific probe (B). Lanes: 1, *S. typhimurium*; 2, *Salmonella typhi* Ty-2; 3, *Salmonella paratyphi* A; 4, *S. paratyphi* C; 5, *Salmonella enteriditis*; 6, *E. coli*; 7, *Vibrio cholerae*; 8, *S. sonnei*; 9, *S. flexneri*; 10, *Citrobacter freundii*; 11, *K. pneumoniae*; 12, *Campylobacter fetus*; 13, *Yersinia enterocolitica*; 14, *M. morganii*; 15, *P. stuartii*. Molecular size markers are in kilobases.

msgA gene was located on a *Shigella* plasmid but not on the 230-kb virulence plasmid.

DISCUSSION

We have identified four genes upstream and transcribed in a direction opposite to that of the *pagC* gene of *S. typhimurium*. Three genes (*pagD*, *envE*, and *envF*) are predicted to be envelope proteins on the basis of the isolation of active Tn*phoA* insertions in these loci and the presence of a typical signal sequence at the amino-terminus of each protein. None of the four proteins possesses significant homology to any protein in the database.

Transposon-encoded reporter gene activity was examined in wild-type and PhoP$^-$ backgrounds for each of the four genes. Only the gene immediately upstream and transcribed in a direction opposite to that of *pagC* (*pagD*) was determined to be PhoP regulated. Transposon insertions in this gene greatly attenuate virulence and the ability of the organism to survive within murine macrophages; however, deletion of this gene has no effect on the LD_{50} of infected mice. There are several possible explanations for this result. One possibility is that abnormal insertion of the fusion protein into the membrane altered the appropriate cellular location of another virulence factor. Profiles of whole-cell and membrane proteins are similar to that of wild type; therefore, there is no evidence that a major defect in protein secretion exists in the *pagD*::Tn*phoA* strain. Another possible explanation for the observed result is that PagD function is redundant; therefore, its deletion would not have an effect on virulence. There are other examples of bacterial genes with redundant functions that are important to pathogenesis. For example, two *Listeria* phospholipase genes have a small effect and no effect on the LD_{50} alone. However, a combination of the two mutations has a major synergistic effect on the LD_{50} (36a). Four distinct *pag*-containing Tn*phoA* insertions that affect virulence appear to encode proteins of less than 90 amino acids (2a). It is possible that these *pag* have similar functions and that the *pagD*::Tn*phoA* fusion protein has a dominant effect on the proper cellular localizations of these proteins or on the interaction of these proteins with other macromolecules. Analysis of strains carrying single-copy *pagD* promoter-reporter gene fusions has demonstrated this gene to be transcriptionally activated within cultured macrophages (15a), suggesting that this gene product has a function when bacteria are within acidified macrophage phagosomes. This result suggests that *pagD* could be involved in promoting *S. typhimurium* survival within macrophages.

Transposon insertions in the *msgA* gene also were found to have an effect on mouse virulence and macrophage survival. The Mu*dJ* insertion in *msgA* does not result in the formation of a fusion protein; thus, it is unlikely that the virulence pheno-

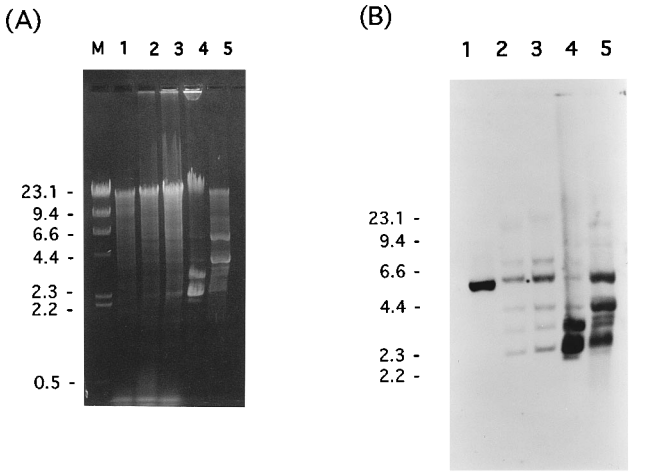


FIG. 5. Southern blot hybridization of *Eco*RI-digested *S. flexneri* DNA to a *msgA*-specific probe. (A) Ethidium bromide-stained agarose gel before blotting; (B) autoradiogram of Southern blot. Lanes: M, lambda *Hin*dIII marker; 1, *S. typhimurium*; 2, *S. flexneri* BS103 (isogenic to strain 2457T but cured of the 230-kb virulence plasmid); 3, *S. flexneri* 2457T; 4, *S. flexneri* 2457T plasmid DNA (undigested); 5, *S. flexneri* 2457T plasmid DNA. In lane 4, the high-molecular-weight DNA that does not hybridize to the probe contains the virulence plasmid, while the smaller-molecular-weight plasmid DNA (major bands seen at 2.4 and 3.5 kb) does hybridize to the *msgA* probe.

types associated with this strain are due to factors other than the disruption of *msgA*. MsgA, like PagD, is predicted to be a small protein. However, in contrast to PagD, MsgA is not likely to be an envelope protein on the basis of the absence of a structure resembling a signal sequence and the fact that no Tn*phoA* insertions were identified in the *msgA* coding sequence.

Transposon insertions can affect the transcription of downstream, cotranscribed genes; however, several lines of evidence suggest that the genes described here do not form an operon. Because none of the genes downstream of *pagD* is PhoP regulated, they are not transcribed from the *pagD* promoter. The presence of a potential transcriptional terminator at the end of the *envE* gene makes it unlikely that *msgA* is cotranscribed with *envE*. The data presented here also demonstrated that the *msgA*::Mu*dJ* insertion is not polar on *envF*, which makes it highly likely that *envF* has its own promoter. Additionally, a potential transcriptional terminator following *msgA* as well as a 493-bp intergenic region makes it unlikely that these genes are cotranscribed.

The other two genes examined in this study, *envE* and *envF*, are predicted to produce membrane proteins that contain potential membrane-spanning regions. The *envF* gene product is likely to be a lipoprotein on the basis of the presence of a consensus lipid attachment site. Since several known or predicted prokaryotic lipoproteins such as the *Neisseria gonorrhoeae* H.8 gene product (6), the TraT product of the *E. coli* F factor (36), the *E. coli nlpD* gene product (19), and the *S. typhimurium prgH* and *prgK* genes (35) have been associated with serum sensitivity or virulence, we examined the possibility that EnvF plays a role in *Salmonella* virulence. Inactivation of this gene has no affect on strain virulence; therefore, it is unlikely that EnvF is a major contributor to *Salmonella* pathogenicity.

Since bacterial genomic DNA G+C content is related to its phylogeny (42), the low G+C contents of the genes in the *pagC* region suggest that they may have been acquired by horizontal transmission. Horizontal or lateral transfer has been suggested as a mode of acquisition of several low-G+C-content genes in *Salmonella* species (14). Southern blot analysis of low-G+C-content organisms probed with the *msgA* or *pagD* gene showed no homology; however, this does not eliminate the possibility that they were acquired from a low-G+C-content organism other than those tested. The possibility also exists that these genes reside on a mobile genetic element acquired from another source. The finding that a region of DNA homologous to *msgA* exists on one of several cryptic plasmids known to exist in *Shigella* species (20) suggests a possible role of extrachromosomal elements in the mode of transfer of this region.

It is of interest that *pagC* and *pagD* are divergently transcribed and are both PhoP activated. There are several examples in the literature of divergently transcribed, regulated genes; however, most require accessory proteins such as CAP in addition to the regulator to activate transcription (2). An example most similar to the *pagC-pagD* arrangement and activation is the *K. pneumoniae pulA-malX* region (8). These two genes are divergently transcribed, and their promoters are arranged back to back. A region of 134 bp exists between the transcriptional start sites of these genes, which is similar to the intergenic region between *pagC* and *pagD* (213 bp between the transcriptional start sites). The *pulA-malX* promoter region is predicted to contain two MalT (the regulatory protein of this system)-binding sites, one for each gene. Other MalT-activated genes require the CAP protein for expression, but the *pulA* and *malX* genes do not, possibly because of the high local concentration of the MalT regulator. Since the intergenic region between the *pagC-pagD* promoters is even larger than that between the *pulA-malX* promoters, we speculate that these two genes are not regulated by PhoP binding to a single site but that binding of one or more phosphorylated PhoP molecules to distinct binding sites positively regulates each gene.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH (AI30479) to S.I.M., by a National Research Service Institution award (AI07061-16) from the NIH (J.S.G.), and by the Hospital Infantil de Mexico (C.A.A.). We thank Brian Seed, MGH Molecular Biology Department, for manufacturing the oligonucleotide primers used in this study and Robin Sandlin for providing the *Shigella* DNAs used for Southern analysis.

REFERENCES

1. **Alpuche-Aranda, C. M., J. A. Swanson, W. P. Loomis, and S. I. Miller.** *Salmonella typhimurium* activates virulence gene transcription within acidified macrophage phagosomes. Proc. Natl. Acad. Sci. USA **89:**10079–10083.
2. **Beck, C. F., and R. A. J. Warren.** 1988. Divergent promoters: a common form of gene organization. Microbiol. Rev. **52:**318–326.
2a. **Belden, W., J. Gunn, and S. Miller.** Unpublished results.
3. **Belden, W. J., and S. I. Miller.** Further characterization of the PhoP regulon: identification of novel PhoP activated loci involved in bacterial virulence. Infect. Immun. **62:**5095–5101.
4. **Blomfield, I. C., V. Vaughn, R. F. Rest, and B. I. Eisenstein.** 1991. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. Mol. Microbiol. **5:**1447–1457.
5. **Buchmeier, N. A., and F. Heffron.** 1989. Intracellular survival of wild-type *Salmonella typhimurium* and macrophage sensitive mutants in diverse populations of macrophages. Infect. Immun. **57:**1–7.
6. **Cannon, J. G.** 1989. Conserved lipoproteins of pathogenic *Neisseria* species bearing the H.8 epitope: lipid-modified azurin and H.8 outer membrane protein. Clin. Microbiol. Rev. **2:**S1–S4.
7. **Case, C. C., S. Roels, J. E. Gonzales, E. L. Simons, and R. W. Simons.** 1988. Analysis of the promoters and transcripts involved in IS10 anti-sense RNA control. Gene **72:**219–236.
8. **Chapon, C., and O. Raibaud.** 1985. Structure of two divergent promoters in front of the gene encoding pullulanase in *Klebsiella pneumoniae* and positively regulated by the *malT* product. J. Bacteriol. **164:**639–645.
9. **Cornelius, G., Y. Laroche, G. Balligand, M. P. Sory, and G. Wauters.** 1987. *Yersinia enterocholitica*, a primary model for bacterial invasivness. Rev. Infect. Dis. **9:**64–87.

10. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacteria genetics: a manual for genetic engineering. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. **Fields, P. I., R. V. Swanson, C. G. Haidaris, and F. Heffron.** 1986. Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. Proc. Natl. Acad. Sci. USA **83:**5189–5193.
12. **Galan, J. E., C. Ginocchio, and P. Costeas.** 1992. Molecular and functional characterization of the *Salmonella* invasion gene *invA*: homology of InvA to members of a new protein family. J. Bacteriol. **174:**4338–4349.
13. **Groisman, E. A., and H. Ochman.** 1993. Cognate gene clusters govern invasion of host epithelial cells by *Salmonella typhimurium* and *Shigella flexneri*. EMBO J. **12:**3779–3787.
14. **Groisman, E. A., M. H. Saier, and H. Ochman.** 1992. Horizontal transfer of the phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. EMBO J. **11:**1309–1316.
15. **Gulig, P. A., L. Caldwell, and V. A. Chiodo.** 1992. Identification, genetic analysis and DNA sequence of a 7.8 kilobase virulence region of the *Salmonella typhimurium* virulence plasmid. Mol. Microbiol. **6:**1395–1411.

15a. **Gunn, J.** Unpublished results.

16. **Gutierrez, C., J. Barondess, C. Manoil, and J. Beckwith.** 1983. The use of transposon Tn*phoA* to detect genes for cell envelope proteins subject to a common regulatory stimulus. Analysis of osmotically regulated genes in *Escherichia coli*. J. Mol. Biol. **185:**51–63.
17. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15:**2343–2357.
18. **Hughes, K. T., and J. R. Roth.** 1988. Transitory *cis* complementation: a method for providing transposon functions to defective transposons. Genetics **119:**9–12.
19. **Ichikawa, J. K., C. Lee, J. Fu, and S. Clarke.** 1994. A gene at 59 minutes on the *Escherichia coli* chromosome encodes a lipoprotein with unusual amino acid repeat sequences. J. Bacteriol. **176:**1630–1638.
20. **Jamieson, A. F., D. A. Bremner, P. L. Bergquist, and H. E. D. Lane.** 1979. Characterization of plasmids from antibiotic-resistant *Shigella* isolates by agarose gel electrophoresis. J. Gen. Microbiol. **113:**73–81.
21. **Kier, L. K., R. M. Weppleman, and B. N. Ames.** 1979. Regulation of non-specific acid phosphatase in *Salmonella*: *phoN* and *phoP* genes. J. Bacteriol. **138:**155–161.
22. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
23. **Lee, C. A., B. D. Jones, and S. Falkow.** 1992. Identification of a *Salmonella typhimurium* invasion locus by selection for hypersensitive mutants. Proc. Natl. Acad. Sci. USA **89:**1847–1851.
24. **Manoil, C., and J. Beckwith.** 1985. Tn*phoA*: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA **82:**8129–8133.
25. **Maurelli, A. T., and P. J. Sansonetti.** 1988. Genetic determinants of *Shigella* pathogenicity. Annu. Rev. Microbiol. **42:**127–150.
26. **Mekalanos, J. J.** 1983. Duplication and amplification of toxin genes in *Vibrio cholerae*. Cell **35:**253–263.
27. **Mekalanos, J. J.** 1992. Environmental signals controlling expression of virulence determinants in bacteria. J. Bacteriol. **174:**1–7.
28. **Michaelis, S., H. Inouye, D. Oliver, and J. Beckwith.** 1983. Mutations that alter the signal sequence of alkaline phosphatase in *Escherichia coli*. J. Bacteriol. **154:**366–374.
29. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Miller, J. H., J. J. Mekalanos, and S. Falkow.** 1989. Coordinate regulation and sensory transduction in the control of bacterial virulence. Science **243:**916–922.
31. **Miller, S. I.** 1991. PhoP/PhoQ: macrophage specific modulators of *Salmonella* virulence? Mol. Microbiol. **5:**2073–2078.
32. **Miller, S. I., A. M. Kurkal, and J. J. Mekalanos.** 1989. A two-component regulatory system (*phoP phoQ*) controls *Salmonella typhimurium* virulence. Proc. Natl. Acad. Sci. USA **86:**5054–5058.
33. **Miller, S. I., and J. J. Mekalanos.** 1990. Constitutive expression of the PhoP regulon attenuates Salmonella virulence and survival within macrophages. J. Bacteriol. **172:**2485–2490.
34. **Miller, V. L., and J. J. Mekalanos.** 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J. Bacteriol. **170:**2575–2583.
35. **Pegues, D. A., M. J. Hantman, I. Behlau, and S. I. Miller.** 1995. PhoP/PhoQ transcriptional repression of *Salmonella typhimurium* invasion genes: evidence for a role in protein secretion. Mol. Microbiol., in press.
36. **Perumal, N. B., and E. G. Minkley, Jr.** 1984. The product of the F sex factor *traT* surface exclusion gene is a lipoprotein. J. Biol. Chem. **259:**5357–5360.

36a. **Portnoy, D.** Personal communication.

37. **Portnoy, D. A., T. Chakaraborty, W. Goebel, and P. Cossart.** 1992. Molecular determinants of *Listeria monocytogenes* pathogenesis. Infect. Immun. **60:**1263–1267.
38. **Pulkkinen, W. S., and S. I. Miller.** 1991. A *Salmonella typhimurium* virulence protein is similar to a *Yersinia enterocolitica* invasion protein and a bacteriophage lambda outer membrane protein. J. Bacteriol. **173:**86–93.
39. **Reed, L. J., and H. Muench.** 1938. A simple method of estimating fifty percent endpoints. Am. J. Hyg. **27:**493–497.
40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
41. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *E. coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
42. **Sueoka, N.** 1961. Variation and heterogeneity of base composition of deoxyribonucleic acids: a compilation of old and new data. J. Mol. Biol. **3:**31–40.
43. **Taylor, R. K., C. Manoil, and J. J. Mekalanos.** 1989. Broad-host-range vectors for delivery of Tn*phoA*: use in genetic analysis of secreted virulence determinants of *Vibrio cholerae*. J. Bacteriol. **171:**1870–1878.
44. **Trucksis, M., J. E. Galen, J. Michalski, A. Fasano, and J. B. Kaper.** 1993. Accessory cholera enterotoxin (*ace*), the third toxin of a *Vibrio cholerae* virulence cassette. Proc. Natl. Acad. Sci. USA **90:**5267–5271.
45. **Von Heinje, G.** 1985. Signal sequences: the limits of variation. J. Mol. Biol. **184:**99–105.