JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 86–93
0021-9193/91/010086-08$02.00/0

Vol. 173, No. 1

A *Salmonella typhimurium* Virulence Protein Is Similar to a *Yersinia enterocolitica* Invasion Protein and a Bacteriophage Lambda Outer Membrane Protein

WENDY S. PULKKINEN AND SAMUEL I. MILLER*

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

Received 20 July 1990/Accepted 3 October 1990

The *phoP-phoQ*-regulated *pagC* locus is essential for full virulence and survival within macrophages of *Salmonella typhimurium*. The protein product, DNA sequence, and transcript of *pagC* were determined. The *pagC* locus encodes a single 188-amino-acid membrane protein that is similar to the *ail*-encoded eucaryotic cell invasion protein of *Yersinia enterocolitica* and the *lom*-encoded protein of bacteriophage lambda. The similarity of PagC and Ail to Lom leads us to hypothesize that Lom is a virulence protein and that bacteriophage gene transfer and lysogeny could have led to the development of proteins essential to survival within macrophages and eucaryotic cell invasion.

Salmonella species are intracellular pathogens that are capable of survival and persistence in mammalian phagocytes (5). Recently, we and other investigators have observed that *Salmonella typhimurium* strains with mutations in the *phoP* locus were avirulent in a BALB/c mouse model of typhoid fever and deficient in survival in cultured macrophages (12, 15, 24). Fields et al. (12) also observed, and we have confirmed (27), that *phoP* mutants are markedly sensitive to cationic proteins, termed defensins, purified from mammalian phagocytes.

The *phoP* locus is composed of two genes, *phoP* and *phoQ*, whose gene products are similar to other bacterial regulatory proteins that activate transcription of a number of unlinked genes in response to environmental stimuli (24, 34, 41). Several unlinked genetic loci that require an intact *phoP* locus for expression have been identified (17, 24). One of these loci, termed *pagC* (*phoP* activated gene C), is at 24 to 25 minutes on the *Salmonella* chromosome. The *pagC* locus was identified as a Tn*phoA* gene fusion with alkaline phosphatase (AP) activity that required the *phoP* and *phoQ* genes for expression (24). *S. typhimurium* strains with the *pagC*::Tn*phoA* insertion were deficient in survival within cultured macrophages and greater than 1,000-fold less virulent in a mouse model of typhoid fever (24). Yet, *pagC* mutants were not found to be more sensitive to defensins than wild-type strains (27). Since the *pagC*::Tn*phoA* fusion protein expressed AP activity, *pagC* was predicted to encode a secreted or envelope protein essential for survival within macrophages. We postulated that the expression of this protein was controlled by the PhoP and PhoQ proteins in response to environmental signals within the macrophage.

We now report the identification of a single *pagC* gene product and determination of the nucleotide sequence of the *pagC* locus. The deduced amino acid sequence of the *pagC* gene product was found to be similar to the *ail* gene product, a membrane protein of *Yersinia enterocolitica* that confers on *Escherichia coli* the ability to invade cultured epithelial cells and whose presence correlates with strain pathogenicity (28–30). The PagC and Ail proteins are also similar to a gene product of bacteriophage lambda, the *lom*-encoded outer membrane protein (8).

MATERIALS AND METHODS

Strains and media. All rich medium was Luria broth (LB), and minimal medium was M9 (9). The construction of *S. typhimurium* CS119 *pagC1*::Tn*phoA phoN2* zxx::6251 Tn*10d-Cam* was previously described (24). *S. typhimurium* ATCC 10428 from the American Type Culture Collection was the wild-type parent of CS119. Other strains derived from ATCC 10428 included CS018, which is isogenic to CS119 except for *phoP105*::Tn*10d* (24), CS022 *pho-24* (26), and CS015 *phoP102*::Tn*10d*-Cam (24). Other wild-type strains used for preparation of chromosomal DNA included *S. typhimurium* LT2 (ATCC 15277), *S. typhimurium* Q1 and *Salmonella drypool* (J. Peterson, University of Texas Medical Branch, Galveston), and *Salmonella typhi* Ty2 (Caroline Hardegree, Food and Drug Administration). pLAFR3 cosmids were mobilized from *E. coli* to *S. typhimurium* by using *E. coli* MM294 containing pRK2013 (14). AP activity was screened on solid media with the chromogenic phosphatase substrate 5-bromo-4-chloro-3-indolyl phosphate. AP assays were performed as previously described (4) and are reported in units as defined by Miller (23).

Protein electrophoresis and Western immunoblot analysis. One-dimensional protein gel electrophoresis was performed by the method of Laemmli (19), and blot hybridization with antibody to AP was performed as previously described (33). Whole-cell protein extracts were prepared from saturated cultures grown in LB at 37°C with aeration by boiling the cells in sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer (19). Two-dimensional gel electrophoresis was performed by the method of O'Farrell (32). Proteins in the 10% polyacrylamide slab gels were visualized by silver staining (22).

DNA analysis and sequencing. Chromosomal DNA was prepared by the method of Mekalanos (21). DNA size fractionated in agarose gels was transfered to nitrocellulose (for blot hybridization) by the method of Southern (40). DNA probes for Southern hybridization analysis were radiolabeled by the random primer method (11). Plasmid DNA

* Corresponding author.

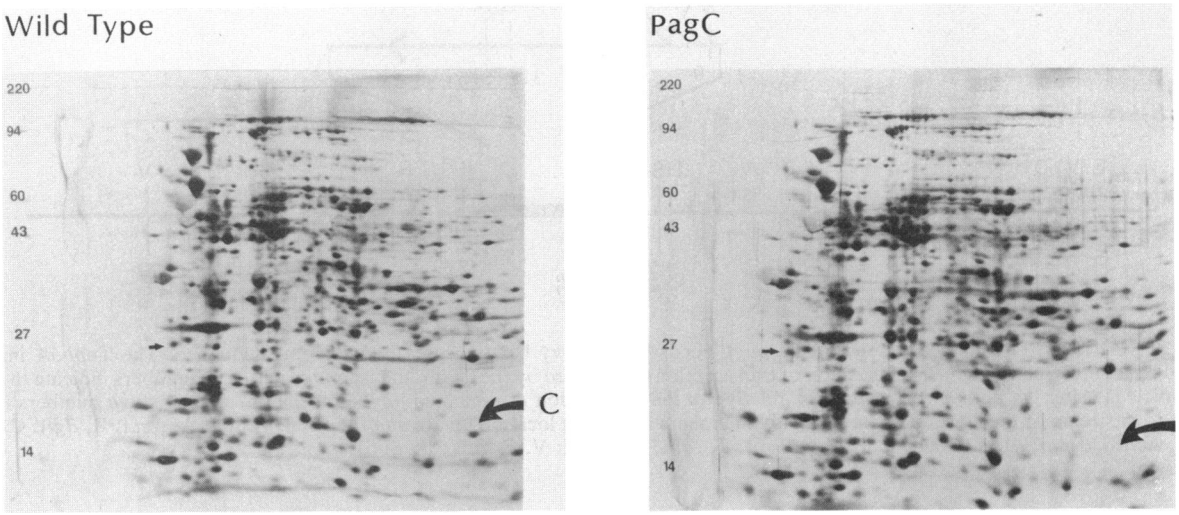

FIG. 1. Two-dimensional gel electrophoresis of proteins from strain CS119 containing the *pagC*::Tn*phoA* insertion and wild-type organisms. The single 18-kDa species missing in the *pagC* mutant strain is identified by the large curved arrow marked C. Molecular mass standards in kilodaltons are indicated by the numbers at the left margins of the gel, and the final tube gel gradient extended from pH 4.1 to 8.1 from right to left. Forty nanograms of vitamin D-dependent calcium-binding protein (pI 5.2) was added to the samples as a standard and is indicated by the small straight arrow on the silver-stained gels. In the PagC mutant at approximately 45 kDa and pI approximately 6.0 (1 cm to the right of the calcium-binding protein standard) are one to two new species that may indicate forms of the PagC-AP fusion protein.

was transformed into *E. coli* and *Salmonella* species by calcium chloride and heat shock (20) or by electroporation with a Genepulser apparatus (Bio-Rad Laboratories, Richmond, Calif.) as recommended by the manufacturer (10). DNA sequencing was performed by the dideoxy-chain termination method of Sanger et al. (36) as modified for use with Sequenase (U.S. Biochemical Corp., Cleveland, Ohio). Oligonucleotides were synthesized on an Applied Biosystems machine and used as primers for sequencing reactions and primer extension of RNA. Specific primers unique to the two ends of Tn*phoA*, one of which corresponds to the AP-coding sequence and the other to the right IS*50* sequence, were used to sequence the junctions of the transposon insertion.

Construction of *S. typhimurium* cosmid gene bank in pLAFR3 and screening for clones containing wild-type *pagC* DNA. DNA from *S. typhimurium* ATCC 10428 was partially digested with the restriction endonuclease *Sau*3A and then size selected on a 10 to 40% sucrose density gradient. T4 DNA ligase was used to ligate chromosomal DNA of 20 to 30 kb into the cosmid vector pLAFR3, a derivative of pLAFR1 (14), that was digested with the restriction endonuclease *Bam*HI. Cosmid DNA was packaged and transfected into *E. coli* DH5-α by using extracts purchased from Stratagene (La Jolla, Calif.). Colonies were screened by blot hybridization analysis.

Analysis of proteins produced from cloned DNA by in vitro transcription-translation assays. In vitro transcription-translation assays were performed with cell extracts (Amersham, Arlington Heights, Ill.) under conditions described by the manufacturer. The resultant radiolabeled proteins were analyzed by SDS-PAGE.

RNA purification and analysis. RNA was purified from early log- and stationary-phase *Salmonella* cultures by the hot phenol method (6) and run in agarose-formaldehyde gels for blot hybridization analysis (42). Primer extension analysis of RNA was performed as previously described (25), using avian myeloblastosis virus reverse transcriptase (Promega Biotec, Madison, Wis.) and synthesized oligonucleotide primers complementary to nucleotides 335 to 350 and 550 to 565 of the *pagC* locus (see Fig. 3).

Nucleotide sequence accession number. The sequence shown in Fig. 3 has been assigned GenBank accession no. M55546.

RESULTS

Identification of 18-kDa protein missing in PagC mutant of *S. typhimurium*. Strain CS119 was analyzed by two-dimensional protein electrophoresis to detect protein species that might be absent as a result of the Tn*phoA* insertion. Only a single missing protein species, of approximately 18 kDa and pI 8.0, was observed when strains that were isogenic except for their transposon insertions were subjected to this analysis (Fig. 1). This 18-kDa species was also missing in a similar analysis of *Salmonella* strains with mutations in *phoP* and *phoQ* (data not shown). Although two-dimensional protein gel analysis might not detect subtle changes of protein expression in strain CS119, this suggested that a single major protein species was absent as a result of the *pagC*::Tn*phoA* insertion.

Additional examination of the two-dimensional gel analysis revealed a new protein species of about 45 kDa that is likely the PagC-AP fusion protein (Fig. 1). The PagC-AP fusion protein was also analyzed by Western immunoblot analysis with antiserum to AP and found to be similar in size to native AP (45 kDa) and not expressed in PhoP$^-$ *S. typhimurium* (data not shown).

Cloning of *pagC*::Tn*phoA* insertion. The *pagC*::Tn*phoA* DNA was cloned to begin a molecular analysis of this essential virulence locus. Chromosomal DNA was prepared from *S. typhimurium* CS119, and a rough physical map of the restriction endonuclease sites in the region of the *pagC*::Tn*phoA* fusion was determined by using a DNA fragment of Tn*phoA* as a probe in blot hybridization analysis. This work indicated that digestion with the restriction

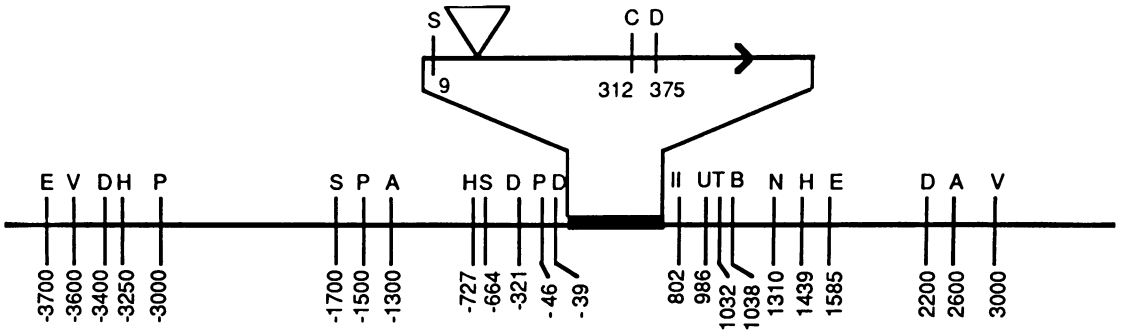

FIG. 2. Restriction endonuclease sites of the *pagC* locus. The heavy bar indicates *pagC* coding sequence. The Tn*phoA* insertion is indicated by an inverted triangle. The direction of transcription is indicated by the arrow and is left to right. The numbers indicate the location of endonuclease sites, in number of base pairs, relative to the start codon of predicted *pagC* translation, with positive numbers indicating location downstream of the start codon and negative numbers indicating location upstream of the start codon. A, *Acc*I; B, *Bgl*I; C, *Cla*I; D, *Dra*I; E, *Eco*RI; H, *Hpa*I; N, *Nru*I; P, *Pst*I; S, *Ssp*I; T, *Stu*I; U, *Pvu*II; V, *Eco*RV; II, *Bgl*II.

endonuclease *Eco*RV yielded a single DNA fragment that included the *pagC*::Tn*phoA* insertion in addition to several kilobases of flanking DNA. Chromosomal DNA from strain CS119 was digested with *Eco*RV (blunt end) and ligated into the bacterial plasmid vector pUC19 that had been digested with the restriction endonuclease *Sma*I (blunt end). This DNA was electroporated into *E. coli* DH5-α, and colonies were plated onto LB agar containing the antibiotics kanamycin (Tn*phoA* encoded) and ampicillin (pUC19 encoded). A single ampicillin- and kanamycin-resistant clone containing a plasmid designated pSM100 was selected for further study.

A radiolabeled DNA probe from pSM100 was constructed and used in Southern hybridization analysis of strain CS119 and its wild-type parent ATCC 10428 to prove that we had cloned the *pagC*::Tn*phoA* fusion. The probe contained sequences immediately adjacent to the transposon at the opposite end of the AP gene (*Hpa*I endonuclease-generated DNA fragment that included 186 bases of the right IS*50* of the transposon and 1,278 bases of *Salmonella* DNA [Fig. 2]). As expected, the pSM100-derived probe hybridized to an 11- to 12-kb *Acc*I endonuclease-digested DNA fragment from the strain containing the transposon insertion, CS119. This was approximately 7.7 kb (size of Tn*phoA*) larger than the 3.9-kb *Acc*I fragment present in the wild-type strain that hybridizes to the probe (data not shown). In addition, a derivative of plasmid pSM100, pSM101 (which did not allow expression of the *pagC-phoA* gene fusion off the *lac* promoter), was transformed into *phoP* (strain CS015) and *phoN* (strain CS019) *Salmonella* strains, and the cloned AP activity was found to be dependent on *phoP* for expression (data not shown). Therefore, we concluded that the cloned DNA contained the *pagC*::Tn*phoA* fusion.

We also tested for the presence of the *pagC* gene in other strains of *S. typhimurium*, in *S. typhi*, and in *S. drypool*. All *Salmonella* strains examined demonstrated similar strong hybridization to an 8.0-kb *Eco*RV fragment and a 3.9-kb *Acc*I restriction endonuclease fragment, suggesting that *pagC* is a virulence gene common to *Salmonella* species (data not shown).

Cloning of wild-type *pagC* locus DNA and its complementation of the virulence defect of an *S. typhimurium pagC* mutant. The same restriction endonuclease fragment described above was used to screen a cosmid gene bank of strain ATCC 10428. A single clone, designated pWP061, contained 18 kb of *S. typhimurium* DNA and hybridized strongly to the *pagC* DNA probe. pWP061 was found to contain *Salmonella* DNA identical to that of pSM100 when analyzed by restriction endonuclease analysis and DNA blot hybridization studies. Probes derived from pWP061 were also used in blot hybridization analysis with DNA from wild-type and CS119 *S. typhimurium*. Hybridization patterns observed were identical to those seen with pSM100 (data not shown). pWP061 was also mobilized into strain CS119, a *pagC* mutant strain. The resulting strain had wild-type virulence for BALB/c mice (a 50% lethal dose of less than 20 organisms when administered by intraperitoneal injection [data not shown]). Therefore, the cloned DNA complements the virulence defect of a *pagC* mutant strain.

Physical mapping of restriction endonuclease sites, DNA sequencing, and determination of *pagC* gene product. Restriction endonuclease analysis of plasmids pSM100 and pWP061 was performed to obtain a physical map of the *pagC* locus and, for pSM100, to determine the direction of transcription (Fig. 2). DNA subclones were generated, and the Tn*phoA* fusion junctions were sequenced, as well as the *Salmonella* DNA extending from the *Hpa*I site 828 nucleotides 5′ to the *phoA* fusion junction to the *Eco*RI site 1032 nucleotides 3′ to the Tn*phoA* insertion (Fig. 2 and 3). The correct reading frame of the DNA sequence was deduced from that required to synthesize an active AP gene fusion. The deduced amino acid sequence of this open reading frame was predicted to encode a 188-amino-acid protein with a predicted pI of 8.2. These data were consistent with the two-dimensional polyacrylamide gel analysis of strain CS119 in which an 18-kDa protein of approximate pI 8.0 was absent. No other open reading frames predicted to encode peptides larger than 30 amino acids were found.

The deduced amino acid sequence of the 188-amino-acid open reading frame contains a methionine start codon 33 amino acids from the fusion of PagC and AP (Fig. 3). This 33-amino-acid *pagC* contribution to the fusion protein was consistent with the size observed in Western immunoblot analysis and contains a hydrophobic N-terminal region, identified by the method of Kyle and Doolittle (18), that is a typical bacterial signal sequence (44). Specifically, amino acid 2 is a positively charged lysine, followed by a hydrophobic domain, and amino acid 24 is a negatively charged aspartate residue. A consensus cleavage site for this leader peptide is predicted to be at an alanine residue at amino acid 23 (43). The DNA sequence also revealed a typical ribosome binding site (37) at 6 to 12 nucleotides 5′ to the predicted start of translation (Fig. 3, nucleotides 717 to 723). This

```
        10          20          30          40          50          60          70
CTTAACCACT CTTAATAATA ATCGGTTTTA TACCGAAATA GACTTTTTTA TCGCGTGTTC AATATTTGCG

        80          90         100         110         120         130         140
TTAGTTATTA TTTTTTTCGA ATGTAAATTC TCTCTAAACA CAGGTGATAT TTATGTTCGA ATTGTGGTGT

       150         160         170         180         190         200         210
TGATTGTATT CTTATAATAT AACAAGAAAT GTTGTAACTG ATAGATATAT TAAAAGATTA AATCGGAGCG
                                *---->

       220         230         240         250         260         270         280
GGAATAAAGC GTGCTAAGCA TGATCGTGAA TATGATTACA GCGGCTGCGA TGGCATATAA CGGTATTCGG

       290         300         310         320         330         340         350
GATGGAGCGT CACGTGAGGA CTGTGAAGCA CAATGCGATA TGTTCTGATT ATATGGCGAG TTTGCTTAAT

       360         370         380         390         400         410         420
GACATGTTTT TAGCCGAACG GTGTCAAGTT TCTTAATGTG GTTGTGAGAT TTTCTCTTTA AATATCAAAA

       430         440         450         460         470         480         490
TGTTGCATGG GTCATTTGTT GTTCTATAGT GGCTAAACAC TTTATGGTTT CTGTTAAATA TATATGCGTG

       500         510         520         530         540         550         560
AGAAAAATTA GCATTCAAAT CTATAAAAGT TAGATGACAT TGTAGAACGG GTTACCTAAA TGAGCGATAG

       570         580         590         600         610         620         630
AGTGCTTCGG TAGTAAAAAT ATCTTTCAGG AAGTAAAGAC ATCAGGAGCG ATAGCGGTGA ATTATTCGTG

       640         650         660         670         680         690         700
GTTTTGTCGA TTCGGCATAG TGGCGATAAC TGAATGCCGG ATCGGTACTG CAGGTGTTTA AACACACCGT

       710         720         728
AAATAATAAG TAGTATTAAG CAGTTGTT

ATG AAA AAT ATT ATT TTA TCC ACT TTA GTT ATT ACT ACA AGC GTT TTG GTT GTA  782
MET LYS ASN ILE ILE LEU SER THR LEU VAL ILE THR THR SER VAL LEU VAL VAL   18

AAT GTT GCA CAG GCG GAT ACT AAC GCC TTT TCC GTG GGG TAT GGA CGG TAT GCA  836
ASN VAL ALA GLN ALA ASP THR ASN ALA PHE SER VAL GLY TYR ALA ARG TYR ALA   36

CAA AGT AAA GTT CAG GAT TTC AAA AAT ATC CGA GGG GTA AAT GTG AAA TAC CGT  890
GLN SER LYS VAL GLN ASP PHE LYS ASN ILE ARG GLY VAL ASN VAL LYS TYR ARG   54

TAT GAG GAT GAC TCT CCG GTA AGT TTT ATT TCC TCG CTA AGT TAC TTA TAT GGA  944
TYR GLU ASP ASP SER PRO VAL SER PHE ILE SER SER LEU SER TYR LEU TYR GLY   72

GAC AGA CAG GCT TCC GGG TCT GTT GAG CCT GAA GGT ATT CAT TAC CAT GAC AAG  998
ASP ARG GLN ALA SER GLY SER VAL GLU PRO GLU GLY ILE HIS TYR HIS ASP LYS   90

TTT GAG GTG AAG TAC GGT TCT TTA ATG GTT GGG CCA GCC TAT CGA TTG TCT GAC 1052
PHE GLU VAL LYS TYR GLY SER LEU MET VAL GLY PRO ALA TYR ARG LEU SER ASP  108

AAT TTT TCG TTA TAC GCG CTG GCG GGT GTC GGC ACG GTA AAG GGG ACA TTT AAA 1106
ASN PHE SER LEU TYR ALA LEU ALA GLY VAL GLY THR VAL LYS ALA THR PHE LYS  126

GAA CAT TCC ACT CAG GAT GGC GAT TCT TTT TCT AAC AAA ATT TCC TCA ACG AAA 1160
GLU HIS SER THR GLN ASP GLY ASP SER PHE SER ASN LYS ILE SER SER ARG LYS  144

ACG GGA TTT GCC TGG GGC GCG GGT GTA CAG ATG AAT CCG CTG GAG AAT ATC GTC 1214
THR GLY PHE ALA TRP GLY ALA GLY VAL GLN MET ASN PRO LEU GLU ASN ILE VAL  162

GTC GAT GTT GGG TAT GAA GGA AGC AAC ATC TCC TCT ACA AAA ATA AAC GGC TTC 1268
VAL ASP VAL GLY TYR GLU GLY SER ASN ILE SER SER THR LYS ILE ASN GLY PHE  180

AAC GTC GGG GTT GGA TAC CGT TTC TGA AAAGC                               1300
ASN VAL GLY VAL GLY TYR ARG PHE                                         188

      1310        1320        1330        1340        1350        1360        1370
ATAAGCTATG CGGAAGGTTC GCCTTCCGCA CCGCCAGTCA ATAAAACAGG GCTTCTTTAC CAGTGACACG

      1380        1390        1400        1410        1420        1430        1440
TACCTGCCTG TCTTTTCTCT CTTCGTCATA CTCTCTTCGT CATAGTGACG CTGTACATAA CATCTCACTA

      1450        1460        1470        1480        1490        1500        1510
GCATAAGCAC AGATAAAGGA TTGTGGTAAG CAATCAAGGT TGCTCAGGTA GGTGATAAGC AGGAAGGAAA

      1520        1530        1540        1550        1560        1570        1580
ATCTGGTGTA AATAACGCCA GATCTCACAA GATTCACTCT GAAAAATTTT CCTGGAATTA ATCACAATGT

      1590        1600        1610        1620        1630        1640        1650
CATCAAGATT TTGTGACCGC CTTCGCATAT TGTACCTGCCG CTGAACGAC TACTGAAAAG TAGCAAGGTA

      1660        1670        1680        1690        1700        1710        1720
TGTATTTTAT CCAGGAGAGC ACCTTTTTTG CGCCTGGCAG AAGTCCCCAG CCGCCACTAG CTCAGCTGGA

      1730        1740        1750        1760        1770        1780        1790
TAGAGCATCA ACCTCCTAA GTTCATGGTGC GAGGTTCGAG GCCTCGGTGG CGGTCCAATG TGGTTATCGT

      1800        1810        1820        1830        1840        1850        1860
ATAATGTTAT TACCTCAGT GTCACGCTGAT GATGTGGGTT CGACTCCCAC TGACCACTTC AGTTTTGAAT

      1870        1880        1890        1900        1910        1920        1930
AAGTATTGTC TCGCAACCC TGTTACAGAAT AATTTCATTT ATTACGTGAC AAGATAGTCA TTTATAAAAA

      1940        1950        1960        1970        1980        1990        2000
ATGCACAAAA ATGTTATTG TCTTTTATTAC TTGTGAGTTG TAGATTTTTC TTATGCGGTG AATCCCCCTT

      2010        2020        2030        2040        2050        2060        2070
TGCGGCGGGG CGTCCAGTC AAATAGTTAAT GTTCCTCGCG AACCATATTG ACTGTGGTAT CGTTCACCGG

      2080        2090        2100        2110        2120        2130        2140
GAGGCACCCG GCACCGCAA TTTTTTATAAA ATGAAATTCA CACCCTATGG TTCAGAGCGG TGTCTTTTTA

      2150        2160        2170        2180        2190        2200        2210
CATCAGGTGG GCAAGCATA ATGCAGGTTAA CTTGAAAGAT ACGATCAATA GCAGAAACCA GTGATTTCGT

      2220        2230        2240        2250        2260        2270        2280
TTATGGCCTG GGGATTTAA CCGCGCCAGAG CGTATGCAAG ACCCTGCGCC GGTTGGCCGG TGATCGTTCA

      2290        2300        2310
ATAGTGCGAA TATGAATGG TTACCAGCCGC TGCGAATTC
```

FIG. 3. DNA sequence and translation of *pagC*::Tn*phoA*. The sequence underlined in boldface indicates a potential ribosome binding site. The single and double lightface underlines indicate sequences in which primers were constructed complementary to these nucleotides for primer extension of RNA analysis. The asterisk indicates the approximate start of transcription. The arrow indicates the direction of transcription. The boxed sequences indicate a region that may function in polymerase binding and recognition. The inverted triangle is the site of the sequenced Tn*phoA* insertion junctions. The arrow indicates a potential site for signal sequence cleavage.

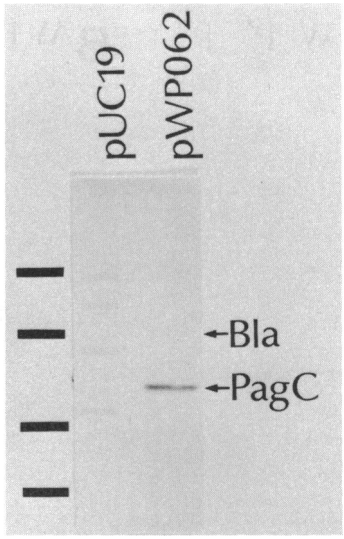

FIG. 4. PAGE analysis of radiolabeled products of an in vitro coupled transcription-translation reaction of cloned *pagC* locus DNA. The molecular size markers shown as bars correspond to 45, 29, 18, and 14 kDa from top to bottom. The location of the products of the β-lactamase gene *bla* on longer autoradiographic exposure is shown by the arrow labeled Bla.

suggested that the open reading frame was, in fact, translated and further supported the assumption that this was the deduced amino acid sequence of the PagC protein interrupted by the Tn*phoA* insertion (Fig. 3).

In vitro synthesis of proteins by cloned *pagC* locus. To detect whether other proteins were encoded by *pagC* and to determine the approximate size of the *pagC* gene product, we performed an in vitro coupled transcription-translation analysis. A 5.3-kb *Eco*RI fragment of pWP061 was inserted into pUC19 so that the *pagC* gene would not be expressed from the *lac* promotor. This plasmid was used in an in vitro coupled transcription-translation assay. A single protein of approximately 22 kDa was synthesized by this cell-free system (Fig. 4). The size was compatible with this being the precursor of the PagC protein containing its leader peptide. This further led us to conclude that we identified the single *pagC* gene product.

Identification of *pagC*-encoded RNA. We purified and analyzed RNA from *S. typhimurium* ATCC 10428 (wild type), CS015 *phoP* 102::Tn*10d*-Cam, and CS022 *pho-24* to determine the size of the *pagC* transcript as well as its dependence on *phoP* for transcription. We also evaluated the expression of *pagC* at different phases of growth. Figure 5 shows that an approximately 1,110-nucleotide RNA is encoded by *pagC*. The *pagC* gene is highly expressed by cells

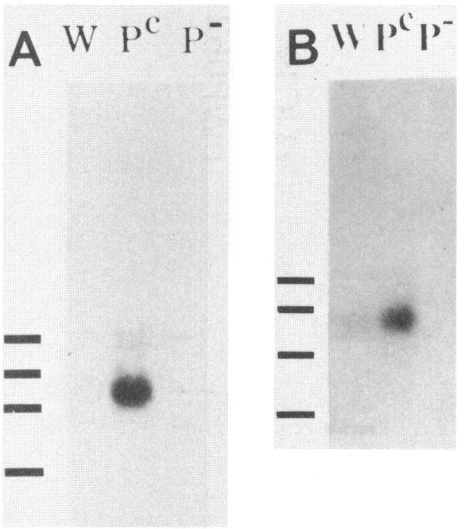

FIG. 5. Identification of the *pagC* transcript. Blot hybridization of a *pagC* gene probe to RNA purified from wild-type (W), *phoP* (P^-), and *phoP* constitutive (P^c) *Salmonella* cells grown to 0.3 optical density units (A) and 2.0 optical density units (B). The bars indicate DNA markers from a φX *Hae*III digest and are 1,353, 1,078, 872, and 603 nucleotides in size from top to bottom.

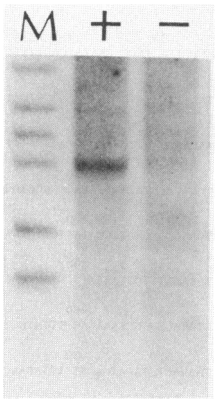

FIG. 6. Primer extension analysis of *Salmonella* RNA by using an oligonucleotide primer corresponding to nucleotides 335 to 350 (Fig. 3) of the *pagC* sequence. Lane M, Radiolabeled markers corresponding to an *Msp*I endonuclease digest of pBR322; the bands seen correspond to 217, 201, 190, 180, 160, and 147 nucleotides from top to bottom. The lane labeled + contains RNA and oligonucleotide primer, and the lane labeled − contains an identical reaction containing RNA alone. The radiolabeled products were analyzed in an 8% polyacrylamide–8 M urea gel.

with a *phoP* constitutive phenotype of *pag* activation, in comparison with expression in wild-type and *phoP* mutant bacteria. In these blot hybridization experiments, *pagC* was detected in wild-type cells grown in rich medium only during stationary growth (Fig. 5). This result, coupled with our previous work (24, 26), demonstrates that *pagC* is transcriptionally regulated by the *phoP* gene products and is only expressed during early logarithmic-phase growth in rich media by cells with a *phoP* constitutive phenotype.

The size of the *pagC* transcript is approximately 500 nucleotides greater than that necessary to encode the 188-amino-acid protein. Therefore, we performed a primer extension analysis of *Salmonella* RNA using oligonucleotide primers specific for *pagC* sequence to determine the approximate start site of transcription and to determine whether these 500 nucleotides might be transcribed 5′ or 3′ to those encoding the 188-amino-acid *pagC* gene product. Primer extension analysis with an oligonucleotide predicted to be complementary to nucleotides 550 to 565 of *pagC* (Fig. 3), 150 nucleotides 5′ to the predicted start codon, resulted in an approximately 300-nucleotide primer extension product (data not shown). Therefore, a primer further upstream was constructed complementary to nucleotides 335 to 350 of *pagC* (Fig. 3) and used in a similar analysis. In Fig. 6, a primer extension product of 180 nucleotides was observed to be primer specific. This is consistent with transcription starting at nucleotide 170 (Fig. 3). Upstream of the predicted transcriptional start, at nucleotides 153 to 160, a classic RNA polymerase-binding site was observed with the sequence TATAAT at −12 nucleotides as well as the sequence TAATAT at −10 nucleotides. No complete matches were observed for the consensus RNA polymerase recognition site (TTGACA) 15 to 21 nucleotides upstream from the −10 region. At −39 (126 to 131) nucleotides (TTGGAA), −38 (127 to 132) nucleotides (TTGTGG), and −25 (135 to 140) nucleotides (TTGATT) are sequences that have matches with the most frequently conserved nucleotides of this consensus sequence.

Based on the above results, transcription was predicted to terminate near the translational stop codon of the 188-amino-acid protein (nucleotide 1295, Fig. 3). Indeed, a stem-loop configuration was found at nucleotides 1309 to 1330 that may function as a transcription terminator. This was consistent with the lack of evidence of open reading frames downstream of the 188-amino-acid protein and the lack of synthesis of other transcription-translation products with the cloned *pagC* DNA. This further suggests that the *pagC*::Tn*phoA* insertion inactivated the synthesis of only a single protein.

Similarity of PagC to Ail and Lom. A computer analysis of protein similarity using the National Biomedical Research Foundation/Protein Identification Resource (16) protein sequence base was conducted to identify other proteins that had similarity to PagC in an attempt to find clues to the molecular function of this protein. Remarkably, PagC was found to be similar to a bacteriophage lambda protein, Lom, that has been localized to the outer membrane in minicell analysis (8) and demonstrated to be expressed by lambda lysogens of *E. coli* (1). Recently, the deduced amino acid sequence of the cloned *ail* gene product of *Y. enterocolitica* was determined and also found to be similar to Lom (28). Therefore, we performed a protein family sequence alignment using a computer algorithm (Fig. 7) that establishes protein sequence families and consensus sequences (39). The formation of this family is indicated by the internal data base values of similarity between these proteins: PagC and Lom, 107.8; PagC and Ail, 104.7; and Ail and Lom, 89.8. These same proteins were searched against 314 control sequences in the data base, and mean values and ranges were 39.3 (7.3 to 52.9) for PagC, 37.4 (7.3 to 52.9) for Ail, and 42.1 (7.0 to 61.9) for Lom. The similarity values for this protein family are all greater than 3.5 standard deviations above the highest score obtained for similarity to the 314 random sequences. We also searched for other members of this family by using the computer-generated consensus sequence, as this method has been shown to be more sensitive for family membership than searching individual sequences (39). No other similarities or other family members were

```
LOM  MRNVCIAVAVFAALAVTVTPARAEGGHGTFTVGYFQ VKPGTLPSLSGGDTGVSHLKGINVKYRYELTDSVGVMASLGF 78
PAGC MKNIILSTLVITTSVLVVNVAQAD  TNAFSVGYA        RYAQSKVQDFKNIRGVNVKYRYEDDSPVSFISSLSY 69
AIL MKKTLLASSLIACLSIASVNVYAAS  ESSISIGYAQHVK       ENGYTLDNDPKGFNLKYRYELDDNWGVIGSFAY 72
      ++   + ++   +   *+++ *      +++*++ ++         +       ** *+++++++ ++++ *+ +

LOM   AASKKSSTVMTGEDTFHYESLRGRYVSVMAGPVLQISKQVSAYAM  AGVAHSRWSGSTMDYRKTEITPGYMKETTT 153
PAGC LYGDRQASGSVEPEGIHYHDKFEVKYGSLMVGPAYRLSDNFSLYAL  AGVGTVKATF           KEHSTQDGDS 135
AIL       THQGYDFFYGSNKFGHGDVDYYSVTMGPSFRINEYVSLYGLLGAAHGKVKASV                     125
            ++    ++        + * *++ **  +++  +*+*++  *+++ ++++

LOM  ARDESAMRHTSVAWSAGIQINPAASVVVDIAYEGSGSGDWRTDGFIVGVGYKF  206
PAGC FSNKISSRKTGFAWGAGVNMNPLENIVVDVGYEGSNISSTKINGFNVGVGYRF  188
AIL  FDESISASKTSMAYGAGVQFNPLPNFVIDASYEYSKLDSIKVGTWMLGAGYRF  178
     +   ++ +++*+ *++*++ **+ + *+*  ***+   + +  ++ +*+++*
```

FIG. 7. Similarity and alignment of PagC, Ail, and Lom predicted protein sequences. Residues conserved across all three proteins are indicated by an asterisk (*). Residues conserved among two proteins are indicated by a plus (+).

found in the data base. As can be seen in Fig. 7, regions of similarity are located not only in the leader peptide transmembrane domains but throughout the protein.

DISCUSSION

We performed a molecular analysis of a strain of *S. typhimurium* with a transposon insertion in the PhoP-PhoQ-regulated *pagC* locus that is essential to the organism's virulence and survival within macrophages. Two-dimensional protein gel electrophoresis indicated that a single 18-kDa protein species with a pI of approximately 8.0 is absent in whole-cell protein extracts as a result of the Tn*phoA* insertion in *pagC*. The DNA surrounding *pagC*::Tn*phoA* and that composing the wild-type *pagC* locus was cloned and its sequence was determined. The deduced amino acid sequence of *pagC*, as well as the results of a coupled transcription-translation assay with the cloned *pagC* DNA, were also consistent with *pagC* encoding a single membrane protein of 188 amino acids and pI 8.2.

The transcript of *pagC* was identified as an approximately 1,100-nucleotide RNA that is synthesized in greater abundance in cells with a *phoP* locus mutation that constitutively expresses *pag*s. The *pagC* transcript is expressed in wild-type cells grown in rich medium only during the stationary growth phase. Although the RNA encoded by this gene is larger than necessary to encode PagC, transcription appears to terminate with the *pagC* gene.

Additionally, a wild-type cosmid containing *pagC* locus DNA was found to complement the virulence defect of a *pagC* mutant *S. typhimurium* strain. We therefore conclude that the PagC protein is a 188-amino-acid (18-kDa) membrane protein essential for survival within macrophages and for virulence of *S. typhimurium*.

The *pagC* transcript contains what appears to be an untranslated leader sequence of approximately 558 nucleotides that could play a role in the stability or regulation of the *pagC* transcript. Typical polymerase-binding sites are present at 10 and 12 nucleotides upstream of the approximate start of transcription, and a stem-loop configuration, suggestive of a transcriptional terminator, is present downstream of the proposed stop of transcription. This further supports the hypothesis that we have correctly defined the extent of *pagC* transcription. Although no classic polymerase recognition sites were observed upstream of the Pribnow box, several possible sites were identified. Further analysis of the *pagC* promoter and other *phoP*-regulated promoters will be required to determine whether a consensus sequence exists for *phoP*-activated promoters.

The molecular basis of the PagC protein's role in survival within macrophages remains to be defined. *S. typhimurium* strains with mutations in *pagC* are equivalent to wild-type organisms in sensitivity to purified rabbit defensins, cationic protein fractions purified from mouse macrophages and intestines, lysosyme from egg and mouse serum, and acid pH (26a, 27).

The *pagC* gene product has extensive similarity to a membrane protein of *Y. enterocolitica*, the gene product of the *ail* locus, which when cloned into laboratory strains of *E. coli* allows these bacteria to invade eucaryotic cells (28, 29). The Ail and PagC proteins are also similar to the predicted protein product of the *lom* gene of bacteriophage lambda. These proteins, by computer similarity analysis, seem to form a family of similar virulence proteins in a manner reminiscent of the bacterial enterotoxins that are present in many members of the family *Enterobacteriaceae* (2).

These proteins are all rich (20 to 22%) in serine and glycine. Four serine and seven glycine residues are conserved among all the members of the family, as are seven tyrosine residues. Five of these tyrosines are located immediately next to a charged hydrophillic amino acid. None of these residues are within consensus sequences for tyrosine or serine kinases (by internal data base; Randy Smith, Molecular Biology Computer Research Resource, Dana Farber Cancer Institute). The most highly conserved region of amino acid sequence among all three proteins (represented by amino acids 47 to 57 of PagC) has the consensus sequence iGXNcKYRYE, where i is K or R, c is V or L, and X is any residue. The remarkable number of close charges, both positive and negative, may mean that these residues form salt bonds. The fact that this is the most conserved region of the family may indicate that this is a functionally important catalytic or binding site. Alternatively, this similarity may indicate a structural similarity such as a site for covalent linkage to lipopolysaccharide. However, searching computer data bases for similarity to this sequence was unrevealing.

The marked similarity of PagC to Ail raises the question of whether PagC$^-$ *Salmonella* species would be defective for invasion of epithelial cells in vitro. Galan and Curtiss (15)

previously observed that PhoP$^-$ *Salmonella* species invaded Henle cells as efficiently as wild-type organisms. We, in collaboration with Virginia Miller, have confirmed this observation (data not shown). Additionally, the reduced virulence of PagC$^-$ *Salmonella* species when administered orally is the same degree as that observed by intraperitoneal inoculation (23a). Therefore, the protein similarity of these two virulence proteins, at least initially, seems inconsistent with their identified function as factors promoting survival within macrophages and epithelial cell invasion.

After invasion of tissue culture cells, enteric bacteria are located in vacuoles that may be similar to the phagosomes of macrophages before lysosomal fusion (13). Therefore, it is possible that tissue culture models of invasion are also testing ability to survive intracellularly. Indeed, these assays are based largely on the detection of intracellular bacteria by counting viable intracellular organisms. Alternatively, the ability to invade cells, including phagocytes, by a specific protein-protein interaction may target bacteria to a different intracellular pathway that allows the organism to survive more easily in macrophages. In macrophage survival experiments, *Salmonella* cells are often opsonized with normal serum to facilitate their uptake by phagocytes. Therefore, a possible role of invasion of these cells may be minimized as a result of opsonization.

The DNA encoding Lom is known to be unessential to bacteriophage lambda functions of lysis, integration, and replication (8). Lom has been localized to the outer membrane of minicells of *E. coli* and was recently shown by Tn*phoA* mutagenesis to be expressed in lambda lysogens of *E. coli* (1). Also, it has recently been observed that another membrane protein of lambda, encoded by the *bor* gene, is similar to a colicin (ColV,I-K94) gene, *iss*, that is essential for serum resistance (3, 7). Recently, Bor has also been shown to be essential for serum resistance of lambda lysogens of *E. coli* (1). These results, combined with the similarity of Lom to PagC and Ail, lead us to postulate that Lom may also be a virulence protein and that lysogeny with lambdoid phages may confer a selective advantage to bacteria in the mammalian host. Perhaps many of the nonessential genes of lambdoid bacteriophages have been retained because they confer a selective advantage to their bacterial hosts. Other lambdoid bacteriophages have been demonstrated to carry the Shiga-like toxin gene, which is expressed in lysogenized *E. coli* (31, 38).

It is interesting to speculate on whether the *pagC* gene was present in some parent *Enterobacteriaceae* or whether it evolved in one genus and then spread to other genera. Given the association of one member of the family (Lom) with a bacteriophage, and the association of another member of the family (Ail) with an IS*3*-like insertion sequence (28), it seems likely that this trait was spread among the enterobacteriaciae by association with extrachromosomal (mobile) genetic elements. These genes may subsequently have evolved different virulence functions.

ACKNOWLEDGMENTS

We especially thank Virginia Miller, who communicated results prior to publication and has collaborated with us on invasion-related experiments. Dr. Miller has also read this paper at various stages of its preparation and has advised as to its content. We also thank Randy Smith for help with computer sequence analysis and Jon Beckwith and Jim Barondess for communication of unpublished results. We also thank John Mekalanos for support and helpful discussions.

This work was supported by Public Health Service grant AI00917-01 to S.I.M. from the NIH. S.I.M. is the recipient of a grant from the Milton Fund of Harvard Medical School.

REFERENCES

1. **Barondess, J. J., and J. Beckwith.** 1990. A bacterial virulence determinant is encoded by lysogenic coliphage λ. Nature (London) **346**:871–874.
2. **Betley, M. J., V. L. Miller, and J. J. Mekalanos.** 1986. Genetics of bacterial enterotoxins. Annu. Rev. Microbiol. **40**:577–605.
3. **Binns, M. M., D. L. Davies, and K. G. Hardy.** 1979. Cloned fragments of the plasmid ColV, I-K94 specifying virulence and serum resistance. Nature (London) **279**:778–781.
4. **Brickman, E., and J. Beckwith.** 1975. Analysis of the regulation of *Escherichia coli* alkaline phosphatase synthesis using deletions and φ80 transducing phage. J. Mol. Biol. **96**:307–316.
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. **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.
7. **Chuba, P. J., M. A. Leon, A. Banerjee, and S. Palchaudhuri.** 1989. Cloning and DNA sequence of plasmid determinant ISS, coding for increased serum survival and surface exclusion, which has homology with lambda DNA. Mol. Gen. Genet. **216**:287–292.
8. **Court, D., and A. B. Oppenheim.** 1983. Phage lambda accessory genes, p. 251–277. *In* R. W. Hendrix et al. (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics, p. 21. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Dower, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16**:6127–6145.
11. **Feinberg, P. I., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restricting endonuclease fragments to high specific activity: addendum. Anal. Biochem. **137**:266–267.
12. **Fields, P. I., E. A. Groisman, and F. Heffron.** 1989. A *Salmonella* locus that controls resistance to microbicidal proteins from phagocytic cells. Science **243**:1059–1062.
13. **Finley, B. B., B. Gumbiner, and S. Falkow.** 1988. Penetration of *Salmonella* through a polarized MDCK epithelial cell monolayer. J. Cell Biol. **107**:221–230.
14. **Friedman, A. M., S. R. Long, S. E. Brown, W. J. Buikema, and F. M. Ausubel.** 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene **18**:289–296.
15. **Galan, J. E., and R. Curtiss III.** 1989. Virulence and vaccine potential of *phoP* mutants of *Salmonella typhimurium*. Microb. Pathog. **6**:433–443.
16. **George, D. G., W. C. Barker, and L. T. Hunt.** 1986. The protein identification resource PIR. Nucleic Acids Res. **14**:11–15.
17. **Groisman, E. A., E. Chiao, C. J. Lipps, and F. Heffron.** 1989. *Salmonella typhimurium* virulence gene *phoP* is a transcriptional activator. Proc. Natl. Acad. Sci. USA **86**:7077–7081.
18. **Kyle, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
19. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
20. **Maclachlan, P. R., and K. E. Sanderson.** 1985. Transformation of *Salmonella typhimurium* with plasmid DNA: differences between rough and smooth strains. J. Bacteriol. **161**:3945–3951.
21. **Mekalanos, J. J.** 1983. Duplication and amplification of toxin genes in Vibrio cholerae. Cell **35**:253–263.
22. **Meril, C. R., D. Goldman, and M. L. Van Keuren.** 1984. Gel protein stains: silver stain. Methods Enzymol. **104**:441.
23. **Miller, J. H.** 1972. Experiments in molecular genetics, p. 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

23a. **Miller, S.** Unpublished observation.
24. **Miller, S. I., A. M. Kukral, and J. J. Mekalanos.** 1989. A two component regulatory system (*phoP* and *phoO*) controls *Salmonella typhimurium* virulence. Proc. Natl. Acad. Sci. USA **86:** 5054–5058.
25. **Miller, S. I., S. M. Landfear, and D. F. Wirth.** 1986. Cloning and characterization of a *Leishmania* gene encoding a RNA spliced leader sequence. Nucleic Acids Res. **14:**7341–7360.
26. **Miller, S. I., and J. J. Mekalanos.** 1990. Constitutive expression of a virulence regulon results in attenuation of *Salmonella typhimurium*. J. Bacteriol. **172:**2485–2490.
26a. **Miller, S. I., A. J. Ouellette, M. E. Selsted, K. Clark, and W. P. Pulkkinen.** Unpublished observations.
27. **Miller, S. I., W. S. Pulkkinen, M. E. Selsted, and J. J. Mekalanos.** 1990. Characterization of defensin resistance phenotypes associated with mutations in the *phoP* virulence regulon of *Salmonella typhimurium*. Infect. Immun. **58:**3706–3710.
28. **Miller, V. L., J. B. Biliska, and S. Falkow.** 1990. Nucleotide sequence of the *Yersinia enterocolitica ail* gene and characterization of the Ail protein product. J. Bacteriol. **172:**1062–1069.
29. **Miller, V. L., and S. Falkow.** 1988. Evidence for two genetic loci in *Yersinia enterocolitica*. Infect. Immun. **56:**1242–1248.
30. **Miller, V. L., J. J. Farmer, W. E. Hill, and S. Falkow.** 1989. The *ail* locus is found uniquely in *Yersinia enterocolitica* serotypes commonly associated with disease. Infect. Immun. **57:**121–131.
31. **Newland, J. W., N. A. Strockbine, S. F. Miller, A. D. O'Brien, and R. K. Holmes.** 1985. Cloning of Shiga-like toxin structural genes from a toxin converting phage of *Escherichia coli*. Science **230:**179–181.
32. **O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250:**4007.
33. **Peterson, K. M., and J. J. Mekalanos.** 1988. Characterization of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonization. Infect. Immun. **56:**2822–2829.
34. **Ronson, C. W., B. T. Nixon, and F. M. Ausabel.** 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell **49:**579–581.
35. **Sanger, F., A. R. Coulson, D. F. Hong, and G. B. Person.** 1982. Nucleotide sequence of bacteriophage lambda DNA. J. Mol. Biol. **162:**729–773.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
37. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *E. coli* 165 ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:** 1342–1346.
38. **Smith, H. W., P. Green, and Z. Parsell.** 1983. Vero cell toxins in *Escherichia coli* and related bacteria: transfer by phage and conjugation and toxic action in laboratory animals, chickens, and pigs. J. Gen. Microbiol. **129:**3121–3137.
39. **Smith, R. F., and T. F. Smith.** 1990. Automatic generation of primary sequence patterns from sets of related protein sequences. Proc. Natl. Acad. Sci. USA **87:**118–122.
40. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–517.
41. **Stock, J. B., A. J. Ninfa, and A. M. Stock.** 1989. Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. **53:**450–490.
42. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77:**5201.
43. **Von Heinje, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173:**243–251.
44. **Von Heinje, G.** 1985. Signal sequences: the limits of variation. J. Mol. Biol. **184:**99–105.