JOURNAL OF BACTERIOLOGY, Mar. 1990, p. 1206–1216
0021-9193/90/031206-11$02.00/0

Vol. 172, No. 3

Nucleotide Sequencing of the *Proteus mirabilis* Calcium-Independent Hemolysin Genes (*hpmA* and *hpmB*) Reveals Sequence Similarity with the *Serratia marcescens* Hemolysin Genes (*shlA* and *shlB*)

TIMOTHY S. UPHOFF AND RODNEY A. WELCH*

Department of Medical Microbiology, University of Wisconsin Medical School, Madison, Wisconsin 53706

Received 16 August 1989/Accepted 27 November 1989

We cloned a 13.5-kilobase *Eco*RI fragment containing the calcium-independent hemolysin determinant (pWPM110) from a clinical isolate of *Proteus mirabilis* (477-12). The DNA sequence of a 7,191-base-pair region of pWPM110 was determined. Two polypeptides are encoded in this region, HpmB and HpmA (in that transcriptional order), with predicted molecular masses of 63,204 and 165,868 daltons, respectively. A putative Fur-binding site was identified upstream of *hpmB* overlapping the −35 region of the proposed *hpm* promoter. In vitro transcription-translation of pWPM110 DNA and other subclones confirmed the assignment of molecular masses for the predicted polypeptides. These polypeptides are predicted to have NH_2-terminal leader peptides of 17 and 29 amino acids, respectively. NH_2-terminal amino acid sequence analysis of purified extracellular hemolysin (HpmA) confirmed the cleavage of the 29-amino-acid leader peptide in the secreted form of HpmA. Hemolysis assays and immunoblot analysis of *Escherichia coli* containing subclones expressing *hpmA*, *hpmB*, or both indicated that HpmB is necessary for the extracellular secretion and activation of HpmA. Significant nucleotide identity (52.1%) was seen between *hpm* and the *shl* hemolysin gene sequences of *Serratia marcescens* despite differences in the G+C contents of these genes (*hpm*, 38%; *shl*, 65%). The predicted amino acid sequences of HpmB and HpmA are also similar to those of ShlB and ShlA, the respective sequence identities being 55.4 and 46.7%. Predicted cysteine residues and major hydrophobic and amphipathic domains have been strongly conserved in both proteins. Thus, we have identified a new hemolysin gene family among gram-negative opportunistic pathogens.

Proteus spp. are second to *Escherichia coli* as the leading cause of urinary tract infections caused by gram-negative bacteria (2). They are often associated with nosocomial infections and urinary tract infections in very young or elderly males (2). *Proteus* infections are often found in the upper urinary tract, resulting in pyelonephritis, calculi formation, and renal impairment (18, 19, 38). Almost 97% of *Proteus* urinary tract infections in humans are caused by *Proteus mirabilis* (37).

A number of putative virulence factors contributing to *Proteus* pathogenesis have been studied. A great deal of work has concentrated on the role of urease as a virulence factor in stone formation by *Proteus* spp. (24, 27, 40). The proposed mechanism of calculi formation involves the secretion of a slimelike glycocalyx around the bacteria which binds struvite and apatite crystals resulting from increased pH due to *Proteus* urease production (24). This process eventually leads to production of an enlarged fossilized bacterial microcolony. *P. mirabilis* adhesins have also been examined. Studies have suggested that *P. mirabilis* has a generalized adhesive capacity for a variety of cells which is widespread among normal flora and pathogenic strains (12, 38). An adhesin capable of binding uroepithelial cells has been isolated from *P. mirabilis* by Wray et al. (42). Studies by Peerbooms et al. demonstrate a direct correlation between the ability of *P. mirabilis* to invade Vero cells and the amount of hemolytic activity produced (29). Koronakis et al. (22) and Welch (41) recently published findings that *Proteus* spp. have two hemolysin determinants. One hemolysin is similar to the *E. coli* alpha-hemolysin (HlyA) and has a Ca^{2+}-dependent activity (11, 22). The other has a Ca^{2+}-independent activity, and to date the genes (*hpm*) encoding that activity have been found only in *Proteus* isolates (41).

We report here our continued investigations of the DNA sequence of *hpm* and the functional characterization of the *hpm* gene products. The DNA sequence revealed two open reading frames (ORFs) encoding polypeptides of 63 kilodaltons (kDa) (HpmB) and 166 kDa (HpmA). Both the DNA and the predicted amino acid sequences showed significant similarity to the *Serratia marcescens* hemolysin sequences (ShlA and ShlB) recently published by Poole et al. (32). Finally, we show that HpmB is necessary for the extracellular secretion and hemolytic activity of the structural hemolysin HpmA.

MATERIALS AND METHODS

Bacteria and bacteriophage strains. *E. coli* DH1 [F^- *recA1 endA1 gyrA96 thi-1 hsdR17* (r^- m^+) *supE44* λ^-] was acquired from Duard Walker, University of Wisconsin (1). The construction of recombinant *E. coli* WPM100 [*E. coli* DH1(pWPM100)] and the origin of *P. mirabilis* clinical isolate 477-12 have been described elsewhere (41). *E. coli* JM101 (F′ *supE traD36 proA*$^+$*B*$^+$ *lacI*^{q} *lacZ*) and bacteriophage M13 vectors mp18 and mp19 were acquired from New England BioLabs, Inc. (Beverly, Mass.) (25). The recombinant vector pACYC184 (5) was acquired from Stanley Falkow, Stanford University.

* Corresponding author.

Media, reagents, and chemicals. LB broth and LB agar were prepared as described by Maniatis et al. (23). *E. coli* JM101 strains for M13 phage production were grown in YT broth and on YT agar plates (26). Antibiotics, 5-bromo-4-chloroindolyl phosphate, salts, and buffers were purchased from Sigma Chemical Co. (St. Louis, Mo.). Restriction endonucleases, sequencing reagents, and other DNA-modifying enzymes were purchased from New England BioLabs, Boehringer Mannheim Biochemicals (Indianapolis, Ind.), Promega Biotec (Madison, Wis.), and United States Biochemical Corp. (Cleveland, Ohio). All radioactive labeled nucleotides and amino acids were obtained from Amersham Corp. (Arlington Heights, Ill.). Defibrinated sheep blood was provided courtesy of the University of Wisconsin Department of Veterinary Sciences.

Molecular cloning of pWPM110 and DNA sequencing. A new clone encoding the entire Ca^{2+}-independent hemolysin determinant (pWPM110) was isolated by the following procedure. Cesium chloride gradient purification of genomic DNA from *P. mirabilis* 477-12 was performed as described by Hull et al. (20). Genomic DNA was digested with *Eco*RI, mixed with appropriately digested pUC19 vector DNA, and ligated by using T4 DNA ligase under conditions recommended by the supplier (New England BioLabs). The mixture was transformed into *E. coli* DH1 and plated onto ampicillin-containing (100 μg/ml) LB agar plates. The resulting transformants were screened for *hpm* sequences by colony hybridization, using GeneScreen membranes (Dupont, NEN Research Products, Boston, Mass.) as specified by the manufacturer. A 0.9-kilobase (kb) *Hin*dIII fragment of pWPM100 was labeled with [α-^{32}P]dATP by nick translation (33) for use as the *hpm* probe.

Appropriate fragments of pWPM100 and pWPM110 were cloned into M13 vectors mp18 and mp19. Chimeric phage DNA were transfected into *E. coli* JM101. The protocols used for the isolation of recombinant templates and the dideoxy-sequencing reactions, using [α-^{32}P]dATP as a label, were those suggested by the commercial suppliers of the DNA polymerase large fragment and Sequenase enzymes (New England BioLabs and United States Biochemical), based on the method of Sanger et al. (34). The labeled reaction mixtures were separated by electrophoresis on 8 M urea–6 or 8% polyacrylamide gels. After electrophoresis, the gels were soaked for 30 min in a 10% acetic acid–12% methanol solution and dried, and autoradiograms were made, using Kodak XAR-5 X-ray film. DNA sequence information was also generated from overlapping deletion sets created in M13 subclones by using methods based on the exonuclease activity of T4 DNA polymerase or exonuclease III previously described by Dale et al. (6) and Henikoff (16) (Erase-a-Base system; Promega), respectively.

In addition, 19 oligonucleotides (20-mers) complementary to specific recombinant templates were used as primers for DNA sequence determinations where overlapping deletions were not found. Oligonucleotides were produced by using an Applied Biosystems DNA synthesizer model 381A as instructed by the manufacturer. After manual deprotection with 30% ammonium hydroxide, the oligonucleotide samples were dried and suspended in TE (10 mM Tris [pH 8.0], 1 mM EDTA). The oligonucleotide sample was then purified by electrophoresis on an 8 M urea–20% polyacrylamide gel. The oligomer band was excised and eluted in 0.5% ammonium acetate at 37°C overnight. The DNA was then precipitated at −20°C with 3 M sodium acetate and 95% ethanol; the pellet was dried, resuspended in TE, and used as a primer in annealing reactions. DNA sequence information was compiled and analyzed by using University of Wisconsin Genetics Computer Group computer programs (8).

In vitro transcription-translation. In vitro transcription-translation was performed on RNase-free CsCl-ethidium bromide-purified plasmid DNA preparations to identify the polypeptide products encoded by recombinant plasmids pWPM99, pWPM100, and pWPM110. A procaryotic DNA-directed translation system (Amersham) based on the method of Zubay (44) was used according to the protocol of the manufacturer. L-[4,5-^{3}H]leucine (Amersham) was used to label plasmid-encoded polypeptides. Approximately half of the final volume of the reaction mixture was added to an equal volume of 2× crack buffer (0.13 M Tris [pH 6.8], 4% sodium dodecyl sulfate [SDS], 20% glycerol, 0.14% bromphenol blue, 2% 2-mercaptoethanol) and loaded onto a 10% SDS-polyacrylamide gel. After electrophoresis, the gel was stained with Coomassie blue (0.125% Coomassie blue, 10% acetic acid, 50% methanol) and destained with a 5% methanol–7% acetic acid solution. The gel was then washed in deionized water for 30 min, soaked in Fluoro-hance (Research Products International Corp., Mount Prospect, Ill.) for 30 min, dried for 2 h at 60°C, and subjected to fluorography.

Amino acid sequence determination. NH_2-terminal amino acid sequence analysis was performed on trichloroacetic acid (TCA)-precipitated (final TCA concentration, 10%) HpmA COOH-terminal truncate from culture supernatants of WPM100 (41) grown in LB broth to an optical density at 600 nm (OD_{600}) of 0.9. Purification was performed by the method of Hunkapillar et al. (21). The TCA precipitate was subjected to SDS-polyacrylamide gel electrophoresis, the 140-kDa HpmA band was excised, and the HpmA truncate was electroeluted. Approximately, 1 nmol of purified HpmA truncate was applied to a gas-phase sequencer (Applied Biosystems model 470A). The phenol thiol hydantoin amino acid from each Edman degradation was analyzed by reverse-phase high-performance liquid chromatography (R. I. Niece, University of Wisconsin Biotechnology Center).

Immunoblotting. *E. coli* cultures were grown in LB broth with constant aeration at 37°C. Samples of 50 μl were removed when the cultures reached an OD_{600} of 0.9 and added to an equal amount of 2× crack buffer. A sample of the culture was also pelleted (5,000 × *g*, 10 min), and the supernatant was then filtered through a 0.2-μm Acrodisk (Gelman Sciences, Inc., Ann Arbor, Mich.) to obtain cell-free supernatant material. This cell-free supernatant was also added to an equal volume of 2× crack buffer. Immunoblotting was performed as previously described (41). Sample volume loaded on polyacrylamide gels did not exceed 20 μl. The primary antiserum used was rabbit anti-HpmA, which was produced against an electrophoretically purified HpmA COOH-terminal truncate encoded by pWPM100 (K. G. Swihart and R. A. Welch, manuscript in preparation). The primary antiserum was diluted 5,000-fold in 0.5% Tween 20–phosphate-buffered saline and was added to the blots. The presence of bound primary antibodies was detected by using alkaline phosphatase-conjugated goat anti-rabbit antiserum (Sigma).

Hemolysis assays. Liquid hemolysis assays were performed on bacterial cultures grown in LB broth and cell-free supernatants of *E. coli* recombinants harvested at an OD_{600} of 0.9. Liquid hemolysis assays were performed as previously described (41), with the following exceptions. A 200-μl sample of whole culture or cell-free supernatant was mixed with 800 μl of 0.85% saline (without 10 mm $CaCl_2$) containing a suspension of washed sheep erythrocytes at a final

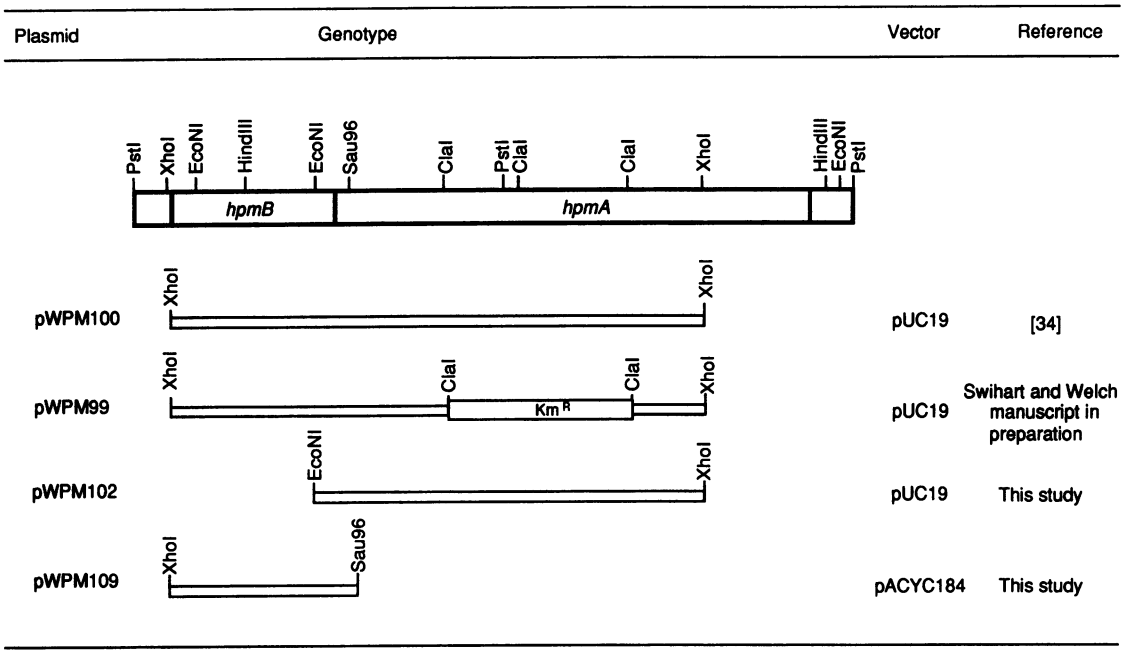

FIG. 1. Recombinant plasmids used. The top map represents the portion of pWPM110 that was sequenced and contains the restriction endonuclease sites used for further subclone construction. Km^r represents the kanamycin resistance gene cassette from Tn*903*.

concentration of 1.0%. This mixture was incubated at 37°C for 30 min and pelleted in a microcentrifuge for 30 s, and the OD_{540} of the hemoglobin present in the supernatant was measured.

RESULTS

Construction of pWPM110. Molecular cloning of the *P. mirabilis* Ca^{2+}-independent hemolysin of pWPM100 and characterization of this plasmid have been previously described (41). This plasmid is a pUC19 derivative containing a 5.3-kb *Xho*I fragment of genomic DNA from *P. mirabilis* 477-12. As discussed below, the DNA sequence analysis of pWPM100 revealed two ORFs. The first ORF (encoding HpmB) began 4 base pairs (bp) downstream of the 5′ *Xho*I site, indicating that the *hpm* genes were being transcribed using the *lac* promoter of pUC19. The downstream ORF (encoding HpmA) did not contain any stop codons within the *Xho*I fragment. Thus, pWPM100 contained only a COOH-terminal truncated version of *hpmA* and none of the transcriptional control region upstream of *hpmB*. We therefore cloned a larger DNA fragment from *P. mirabilis* 477-12 that contained the complete *hpm* determinant by the following procedure.

P. mirabilis 477-12 genomic DNA was digested with *Eco*RI and ligated into pUC19, and the recombinant plasmids were transformed into *E. coli* DH1. Transformants were screened for *hpm* sequences by colony hybridization with a 0.9-kb *Hin*dIII fragment probe from pWPM100. One of several colonies that gave a strong hybridization signal was selected for further study (pWPM110). This plasmid contained a 13.5-kb *Eco*RI fragment of *P. mirabilis* 477-12 DNA. The recombinant plasmids used are shown in Fig. 1.

DNA sequence determination. The majority of DNA sequence information was gathered by analysis of pWPM100 subclones. DNA sequences upstream of *hpmB*, those encoding the COOH terminus of HpmA, and the region downstream of *hpmA* were determined by using subclones of pWPM110. Figure 2 shows a circular physical map of pWPM110 and the sequenced region encoding HpmB and HpmA. The DNA sequence of 8,250 bp was determined from 99% of both strands. We report here the DNA sequence of 7,191 bp that includes *hpmB* and *hpmA* (Fig. 2B and 3); 100% of both DNA strands was directly sequenced in this region. Figure 2C shows a summary of the subclones and indicates the direction of sequencing for each sequence fragment used in determination of the *hpm* DNA sequence reported.

The DNA sequence was searched for ORFs, repeated sequences, and consensus promoter and terminator sites. Two large ORFs encoding 63- and 166-kDa proteins HpmB and HpmA, respectively, were seen. Examination of the codon usage in *hpmA* and *hpmB* showed a strong bias toward A or T in the third position of the codon corresponding to the low G+C content (38%) of the sequence. Putative promoter regions matching *E. coli* consensus sequences were seen at bp 207, which is 100 bp upstream of *hpmB*, and bp 1901, which is upstream of *hpmA* at the 3′ end of *hpmB* (Fig. 3). A putative Rho-independent transcriptional termination sequence was found at bp 6960 just downstream of the 3′ end of *hpmA* (Fig. 3). The *hpm* DNA sequence was compared with itself in a search for direct or indirect repeated sequences; similar searches were also performed on HpmA and HpmB amino acid sequences. However, no significant repeats were detected at the DNA or amino acid level.

Comparison with *S. marcescens* hemolysin sequence. A search of the public data bases (GenBank, EMBL, NBRF, and VecBase) revealed no significant similarities between the *hpm* DNA or amino acid sequence and those in the data bases. Striking similarities in operon structure and protein size led us to compare the DNA and amino acid sequences of this hemolysin with the *S. marcescens* hemolysin sequence published by Poole et al. (32). Although the G+C content of *shl* sequences (65%) was different from that of *hpm* (38%), a

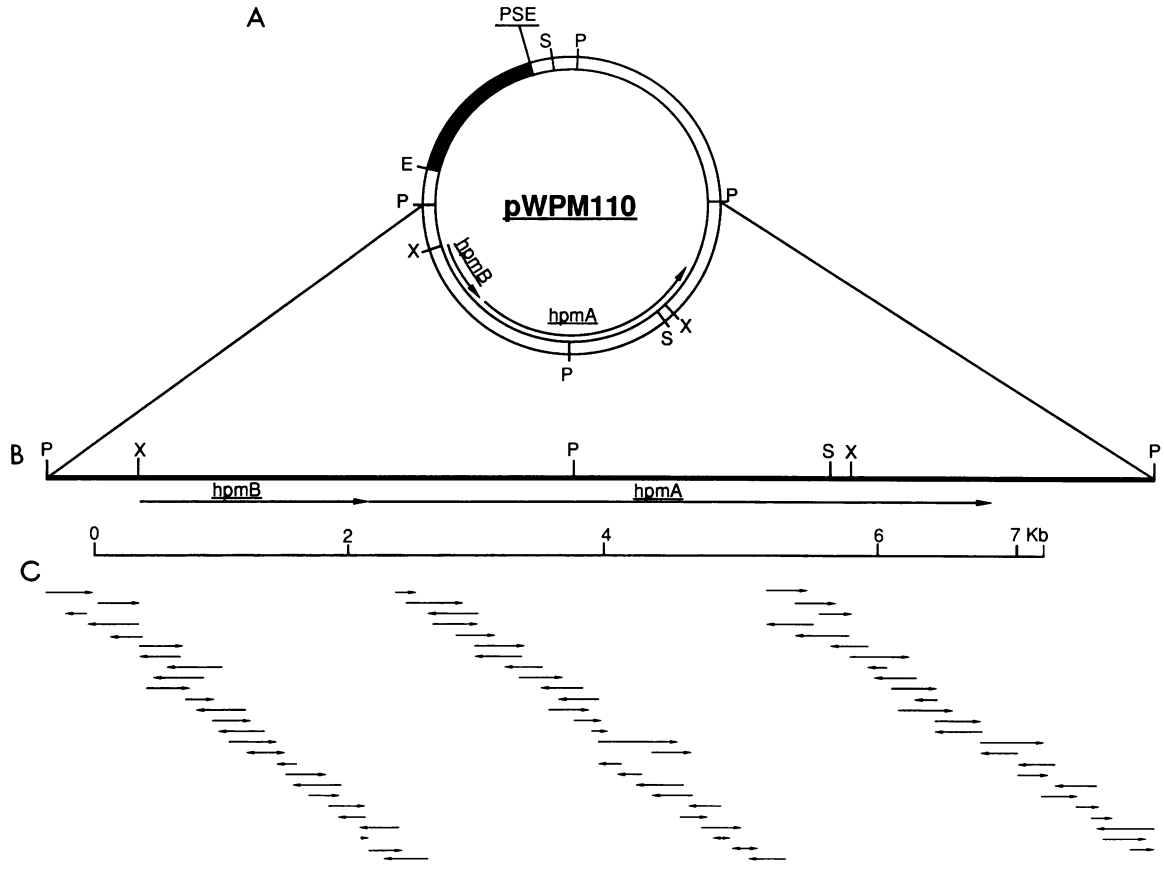

FIG. 2. (A) Map of pWPM110. A 13.5-kb *Eco*RI genomic fragment of *P. mirabilis* 477-12 is inserted into the polylinker of pUC19 (▬). (B) The 8.2-kb region that was sequenced, showing the transcriptional arrangement of *hpmB* and *hpmA*. Beneath the restriction endonuclease map is a line to scale representing the 7,191 bp of *hpm* sequence reported here, containing *hpmB* and *hpmA* as well as putative regulatory regions upstream of *hpmB* and regions downstream of *hpmA*. (C) Strategy for sequencing of pWPM110. Arrows indicate the length and direction of sequencing for each fragment used in assembly of the *hpm* DNA sequence. Abbreviations: E, *Eco*RI; P, *Pst*I; S, *Sal*I; X, *Xho*I.

significant degree of identity was evident at both the DNA and amino acid sequence levels. A summary of the DNA and amino acid sequence alignments is given in Table 1. At the DNA level, 52% overall identity was observed; however, the 5′ terminus (*hpmB* and upstream region) was more highly conserved (56.5%) than the 3′ end (45.8%) encoding HpmA.

Since *shl* gene expression appears to be regulated by the *fur* gene product (31), we searched the *hpm* sequence for *E. coli* consensus Fur-binding sites (4). This search revealed a putative Fur-binding site overlapping the −35 region of the potential promoter upstream of *hpmB* at bp 199 (GTAACT GGTATTTATTATT).

Identification of *hpmB* and *hpmA* gene products. To confirm the production of HpmB and truncated forms of HpmA from pWPM100 and pWPM99, in vitro transcription-translation was performed. Figure 4 is a fluorogram of an SDS-polyacrylamide gel which demonstrates the incorporation of L-[^{3}H]leucine into peptides putatively identified as HpmB and the truncated HpmA encoded by pWPM100 (lane 2). A significant amount of HpmA appeared to be degraded in this system, making it difficult to identify HpmB encoded by pWPM100. A smaller truncate of HpmA (molecular mass, ≈42 kDa) is encoded by pWPM99, which has two *Cla*I fragments (≈1.7 kb) deleted from the middle of *hpmA* and the kanamycin resistance (Km^r) cassette from Tn*903* inserted (Fig. 4, lane 3). This plasmid directs the production of a ≈55-kDa polypeptide (Fig. 4, lane 3) that migrated identically to a polypeptide encoded by pWPM100. Since this band was clearly not a breakdown product of HpmA encoded by pWPM99 and is not encoded by the pUC19 control plasmid or the Km^r cassette (28, 39), we tentatively identified it as HpmB. In addition, synthesis of a 166-kDa peptide corresponding to full-length HpmA from pWPM110 has been seen in this system (data not shown).

Amino acid sequence features of HpmA and HpmB. The proposed translational start site for HpmB was determined to be at bp 313 and is preceded by a potential Shine-Dalgarno site (GAGGT) (13). We identified a putative NH_2-terminal leader peptide of 17 amino acids containing a strongly hydrophobic core and appropriate serine (amino acid [aa] 17) and glycine (aa 18) residues for leader peptide cleavage. We searched HpmB for the consensus ATP- or GTP-binding sites (15). No strong matches to any of these sites was found.

The similarity between HpmB and ShlB (Table 1) was seen throughout the length of the sequences. The predicted isoelectric points and net charges of HpmB and ShlB are very similar (Table 2). Like ShlB, HpmB contains significant hydrophobic domains; we identified five possible transmembrane regions at aa 75 to 100, 250 to 270, 280 to 300, 370 to 390, and 425 to 445 on the basis of a Goldman-Engelman-Steiz hydrophobicity scale over a window of 20 amino acids. The hydrophobic nature and placement of all of these domains was conserved in ShlB except that the COOH-terminal domain was shifted slightly to aa 470 to 490. In

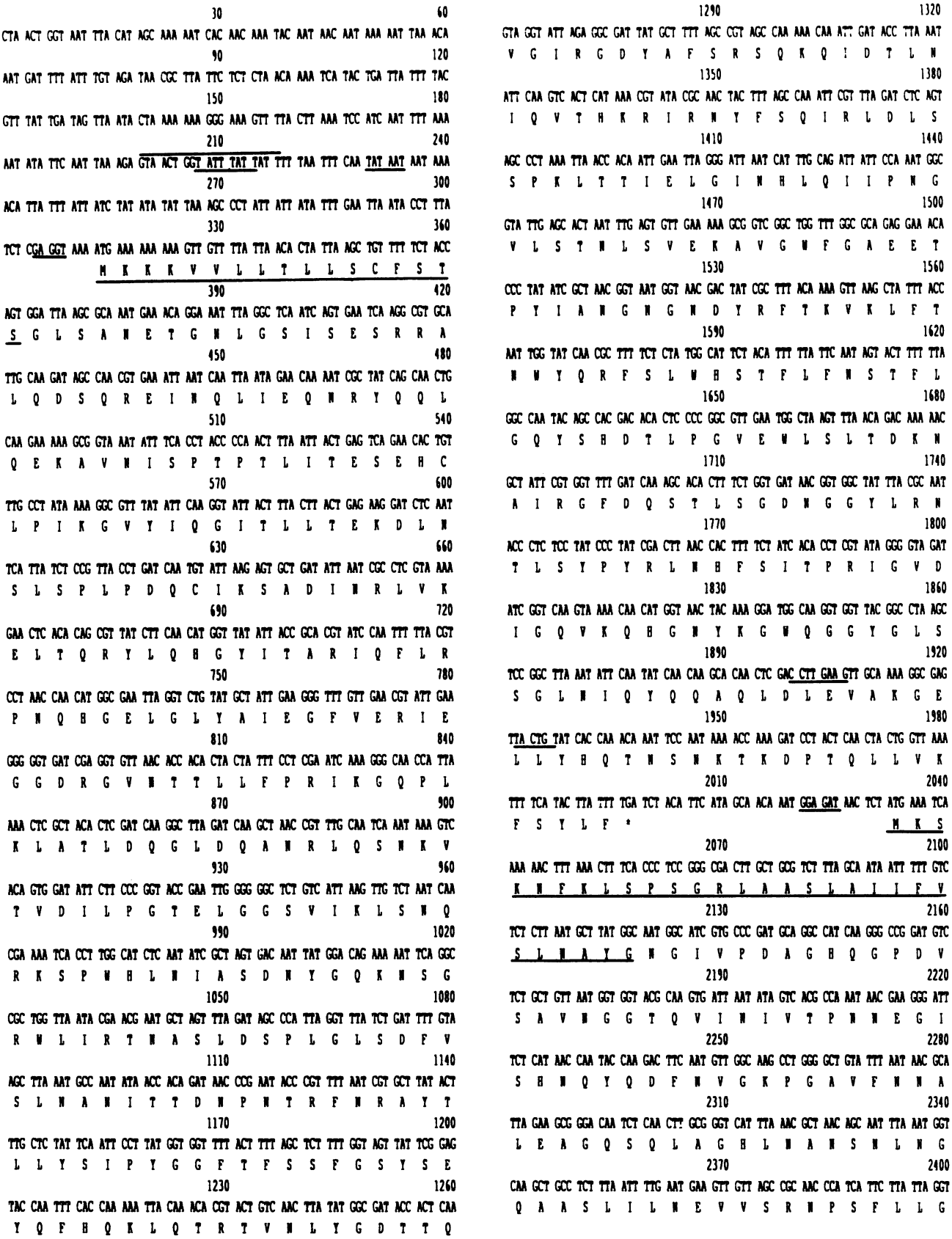

FIG. 3. Nucleotide sequence of the 7,191-bp pWPM110 hemolysin region (GenBank accession no. M30186). Putative promoters are underlined at bp 210 (−35) and 1900 (−35). The proposed Fur-binding site overlapping the −35 region at bp 210 is indicated with lines above the DNA sequence. A putative Rho-independent transcriptional termination site is underlined at bp 6960. Predicted amino acid sequences for HpmB (bp 313 to 1997) and HpmA (bp 2032 to 6765) are shown beneath the DNA sequence. The positions of putative ribosome-binding sites are underlined in the DNA sequence preceding each gene. The NH_2-terminal leader peptide is also underlined for HpmB and HpmA.

```
                                2430                                2460
CAG CAA GAG GTC TTT GGT ATC GCT GCT GAA TAT GTG CTT TCT AAC CCA AAT GGT ATC ACA
 Q   Q   E   V   F   G   I   A   A   E   Y   V   L   S   N   P   N   G   I   T
                                2490                                2520
TGT GAT GGT TGT GGT TTT ATA AAT ACC AGC CGC TCT TCA TTA GTT GTT GGT AAT CCG CTC
 C   D   G   C   G   F   I   N   T   S   R   S   S   L   V   V   G   N   P   L
                                2550                                2580
TTT GAA AAT GGT CAG CTA AAA GGC TAT AGC ACC CTC AAC AAC ACA AAT TTA CTA TCG CTT
 F   E   N   G   Q   L   K   G   Y   S   T   L   N   N   T   N   L   L   S   L
                                2610                                2640
GGT AAA AAT GGC TTA AAT ACA ACA GGG TTG TTA GAC TTA ATT GCT CCT CGT ATT GAT AGT
 G   K   N   G   L   N   T   T   G   L   L   D   L   I   A   P   R   I   D   S
                                2670                                2700
AGA GGG AAA ATC ACT GCT GCT GAA ATT TCA GCC TTT ACT GGA CAA AAC ACC TTC TCA CAA
 R   G   K   I   T   A   A   E   I   S   A   F   T   G   Q   N   T   F   S   Q
                                2730                                2760
CAT TTT GAT ATT CTC TCT TCA CAA AAA CCC GTT TCA GCA TTA GAT AGC TAT TTC TTT GGT
 H   F   D   I   L   S   S   Q   K   P   V   S   A   L   D   S   Y   F   F   G
                                2790                                2820
AGT ATG CAA TCG GGT CGT ATC CGC ATT ATT AAT ACG GCT GAA GGT AGT GGA GTT AAA TTA
 S   M   Q   S   G   R   I   R   I   I   N   T   A   E   G   S   G   V   K   L
                                2850                                2880
GCA GGT AAA TTT ACC GCA GAT AAC GAC CTA AGT GTT AAA GCC GAT AAT ATT CAA ACA GAT
 A   G   K   F   T   A   D   N   D   L   S   V   K   A   D   N   I   Q   T   D
                                2910                                2940
AGT CAA GTC CGT TAT GAC AGT TAC GAT AAA GAT GGC AGT GAA AAT TAC CAA AAC TAT CGT
 S   Q   V   R   Y   D   S   Y   D   K   D   G   S   E   N   Y   Q   N   Y   R
                                2970                                3000
GGC GGG ATC ACG GTT AAT AAT AGT GGC TCT AGT CAA ACA CTC ACT AAA ACC GAA TTA AAA
 G   G   I   T   V   N   N   S   G   S   S   Q   T   L   T   K   T   E   L   K
                                3030                                3060
GGT AAA AAC ATC ACA TTA GTA GCG AGT AGC CAT AAT CAA ATC AAA GCC TCT GAT TTA ATG
 G   K   N   I   T   L   V   A   S   S   H   N   Q   I   K   A   S   D   L   M
                                3090                                3120
GGG GAT GAC ATC ACG TTA CAA GGT GCT GAT TTA ACT ATC GAT GGT AAA CAG CTA CAG CAA
 G   D   D   I   T   L   Q   G   A   D   L   T   I   D   G   K   Q   L   Q   Q
                                3150                                3180
AAA GAG ACC GAT ATT GAT AAT CGC TGG TTC TAC TCG TGG AAA TAC GAT GTG ACT AAA GAG
 K   E   T   D   I   D   N   R   W   F   Y   S   W   K   Y   D   V   T   K   E
                                3210                                3240
AAA GAA CAA ATA CAG CAA ATT GGT AGC CAA ATT GAT GCT AAA AAT AAT GCG ACA TTA ACC
 K   E   Q   I   Q   Q   I   G   S   Q   I   D   A   K   N   N   A   T   L   T
                                3270                                3300
GCA ACT AAA GGA GAT GTT ACC TTA GAC GCG GCT AAA ATT AAT GCG GGG AAT AAC CTT GCA
 A   T   K   G   D   V   T   L   D   A   A   K   I   N   A   G   N   N   L   A
                                3330                                3360
ATT AAT GCC AAT AAA GAT ATC CAT ATC AAT GGA TTA GTT GAG AAA GAA AGT CGT AGT GAA
 I   N   A   N   K   D   I   H   I   N   G   L   V   E   K   E   S   R   S   E
                                3390                                3420
AAT GGC AAT AAA CGT AAT CAT ACT TCT CGC TTA GAA AGT GGT AGT TGG AGT AAC AGC CAC
 N   G   N   K   R   N   H   T   S   R   L   E   S   G   S   W   S   N   S   H
                                3450                                3480
CAA ACT GAG ACG TTG AAA GCC AGT GAA TTA ACG GCA GGT AAA GAT CTT GGT TTA GAT GCT
 Q   T   E   T   L   K   A   S   E   L   T   A   G   K   D   L   G   L   D   A
                                3510                                3540
CAA GGC TCA ATA ACG GCG CAA GGT GCT AAA CTG CAT GCG AAT GAA AAT GTG CTG GTC AAT
 Q   G   S   I   T   A   Q   G   A   K   L   H   A   N   E   N   V   L   V   N
```

```
                                3570                                3600
GCC AAA GAT AAC ATT AAT TTA AAC GTC CAA AAA ACC AAT AAT GAT AAA ACA GTC ACC GAT
 A   K   D   N   I   N   L   N   V   Q   K   T   N   N   D   K   T   V   T   D
                                3630                                3660
AAT CAT GTT ATG TGG GGA GGT ATT GGT GGT GGT CAG AAT AAA AAT AAT AAT AAT CAA CAA
 N   H   V   M   W   G   G   I   G   G   G   Q   N   K   N   N   N   N   Q   Q
                                3690                                3720
CAA GTT AGT CAT GCA ACA CAA TTA ACC GCT GAT GGT CAA TTG CTC CTT GCT GCA GAT AAC
 Q   V   S   H   A   T   Q   L   T   A   D   G   Q   L   L   L   A   A   D   N
                                3750                                3780
AAT GTC AAC ATT ACC GGT AGC CAA GTA AAA GGC AAC CAA GGT GCT TTT GTT AAA ACC ACA
 N   V   N   I   T   G   S   Q   V   K   G   N   Q   G   A   F   V   K   T   T
                                3810                                3840
CAA GGT GAT GTC GTC ATT GAT AAT GCG CTG AGT GAA ACC ATC AGT AAA ATC GAT GAA CGC
 Q   G   D   V   V   I   D   N   A   L   S   E   T   I   S   K   I   D   E   R
                                3870                                3900
ACA GGA ACC GCG TTT AAT ATC ACT AAA AGT TCA CAT AAA AAT GAA ACT AAT AAA CAA ACA
 T   G   T   A   F   N   I   T   K   S   S   H   K   N   E   T   N   K   Q   T
                                3930                                3960
TCT ACA GGT AGT GAA CTC ATT TCC GAT GCA CAA TTG ACC GTT GTC AGT GGC AAT GAT GTT
 S   T   G   S   E   L   I   S   D   A   Q   L   T   V   V   S   G   N   D   V
                                3990                                4020
AAT GTC ATT GGT AGT TTG ATC AAA AGC GCT GAT AAA TTA GGT ATT CAT TCT TTA GGT GAT
 N   V   I   G   S   L   I   K   S   A   D   K   L   G   I   H   S   L   G   D
                                4050                                4080
ATC AAC GTT AAA TCA GCA CAA CAA GTC ACT AAG ATT GAT GAT GAA AAA ACA TCA TTA GCC
 I   N   V   K   S   A   Q   Q   V   T   K   I   D   D   E   K   T   S   L   A
                                4110                                4140
ATT ACT GGG CAC GCC AAA GAA GTT GAA GAC AAG CAA TAT AGT GCG GGC TTT CAT ATC ACT
 I   T   G   H   A   K   E   V   E   D   K   Q   Y   S   A   G   F   H   I   T
                                4170                                4200
CAT ACC ACC AAT AAA AAT ACT AGT ACT GAA ACA GAG CAA GCT AAC TCA ACC ATT AGT GGC
 H   T   T   N   K   N   T   S   T   E   T   E   Q   A   N   S   T   I   S   G
                                4230                                4260
GCC AAT GTT GAT CTG CAA GCG AAT AAA GAT GTG ACT TTT GCG GGA TCT GAT CTA AAG ACT
 A   N   V   D   L   Q   A   N   K   D   V   T   F   A   G   S   D   L   K   T
                                4290                                4320
ACC GCA GGG AAT GCC TCT ATT ACG GGT GAT AAT GTT GCC TTT GTT AGC ACG GAA AAT AAA
 T   A   G   N   A   S   I   T   G   D   N   V   A   F   V   S   T   E   N   K
                                4350                                4380
AAA CAA ACG GAC AAC ACA GAT ACC ACT ATT TCA GGA GGT TTT AGC TAC ACA GGA GGT GTA
 K   Q   T   D   N   T   D   T   T   I   S   G   G   F   S   Y   T   G   G   V
                                4410                                4440
GAT AAA GTC GGT TCT AAA GCC GAT TTT CAA TAT GAC AAA CAA CAT ACA CAA ACA GAG GTA
 D   K   V   G   S   K   A   D   F   Q   Y   D   K   Q   H   T   Q   T   E   V
                                4470                                4500
ACA AAA AAT AGA GGC AGC CAA ACA GAG GTT GCA GGT GAT TTG ACT ATT ACC GCG AAC AAA
 T   K   N   R   G   S   Q   T   E   V   A   G   D   L   T   I   T   A   N   K
                                4530                                4560
GAT TTG CTT CAC GAA GGA GCT TCT CAC CAT GTT GAG GGG CGT TAT CAA GAG TCT GGT GAA
 D   L   L   H   E   G   A   S   H   H   V   E   G   R   Y   Q   E   S   G   E
                                4590                                4620
AAT ATT CAG CAT CTT GCC GTC AAT GAT AGT GAA ACA TCT AAA ACT GAC AGC TTA AAT GTT
 N   I   Q   H   L   A   V   N   D   S   E   T   S   K   T   D   S   L   N   V
                                4650                                4680
GGT ATT GAT GTT GGC GTT AAT CTT GAT TAT AGC GGT GTA ACT AAG CCA GTT AAG AAA GCA
 G   I   D   V   G   V   N   L   D   Y   S   G   V   T   K   P   V   K   K   A
```

```
                                    4710                                    4740
ATC GAA GAT GGT GTT AAC ACA ACC AAA CCG GGT AAC AAT ACT GAT TTA ACT AAA AAA GTT
 I  E  D  G  V  N  T  T  K  P  G  N  N  T  D  L  T  K  K  V
                                    4770                                    4800
ACA GCA AGA GAT GCA ATT GCT AAT TTA GCT AAC CTT AGC AAT TTA GAG ACC CCC AAT GTC
 T  A  R  D  A  I  A  N  L  A  N  L  S  N  L  E  T  P  N  V
                                    4830                                    4860
GGT GTT GAA GTG GGT ATT AAA GGT GGC GGT AGT CAG CAA TCA CAA ACT GAT AGC CAA GCT
 G  V  E  V  G  I  K  G  G  G  S  Q  Q  S  Q  T  D  S  Q  A
                                    4890                                    4920
GTT TCA ACC TCT ATC AAT GCA GGA AAA ATC GAT ATT GAT AGT AAT AAC AAG TTA CAT GAT
 V  S  T  S  I  N  A  G  K  I  D  I  D  S  N  N  K  L  H  D
                                    4950                                    4980
CAA GGT ACT CAC TAT CAA TCA ACC CAA GAG GGA ATT TCT CTC ACA GCG AAT ACT CAC ACA
 Q  G  T  H  Y  Q  S  T  Q  E  G  I  S  L  T  A  N  T  H  T
                                    5010                                    5040
AGT GAA GCA ACC TTA GAT AAA CAT CAA ACG ACA TTC CAT GAA ACA AAA GGT GGT GGG CAA
 S  E  A  T  L  D  K  H  Q  T  T  F  H  E  T  K  G  G  G  Q
                                    5070                                    5100
ATC GGT GTC AGT ACC AAA ACG GGC AGT GAT ATT ACC GTT GCT ATT AAA GGT GAA GGC CAA
 I  G  V  S  T  K  T  G  S  D  I  T  V  A  I  K  G  E  G  Q
                                    5130                                    5160
ACA ACT GAT AAC GCC TTA ATG GAA ACA AAG GCT AAA GGT AGC CAA TTT ACC TCA AAT GGC
 T  T  D  N  A  L  M  E  T  K  A  K  G  S  Q  F  T  S  N  G
                                    5190                                    5220
GAT ATT TCA ATT AAT GTA GGT GAA AAT GCC CAT TAT GAA GGT GCT CAA TTT GAT GCA CAA
 D  I  S  I  N  V  G  E  N  A  H  Y  E  G  A  Q  F  D  A  Q
                                    5250                                    5280
AAG GGC AAA ACA GTT ATC AAT GCG GGG GGT GAT CTC ACT CTT GCA CAG GCG ACT GAT ACT
 K  G  K  T  V  I  N  A  G  G  D  L  T  L  A  Q  A  T  D  T
                                    5310                                    5340
CAT AGC GAA AGT CAA TCT AAT GTT AAT GGT AGT GCA AAC CTG AAA GTG GGC ACC ACG CCA
 H  S  E  S  Q  S  N  V  N  G  S  A  N  L  K  V  G  T  T  P
                                    5370                                    5400
GAG AGT AAA GAC TAT GGT GGA GGT TTC AAT GCA GGG ACA ACT CAT CAC AGC AAA GAG CAG
 E  S  K  D  Y  G  G  G  F  N  A  G  T  T  H  H  S  K  E  Q
                                    5430                                    5460
ACT ACC GCA AAA GTG GGC ACT ATC ACT GGC TCT CAA GGT ATT GAA TTA AAC GCT GGG CAT
 T  T  A  K  V  G  T  I  T  G  S  Q  G  I  E  L  N  A  G  H
                                    5490                                    5520
AAC CTG ACT TTA CAA GGT ACT CAT TTA AGT AGC GAA CAA GAT ATT GCA TTA AAT GCA ACT
 N  L  T  L  Q  G  T  H  L  S  S  E  Q  D  I  A  L  N  A  T
                                    5550                                    5580
AAT AAA GTC GAC CTA CAA TCA GCG AGC TCT GAA CAT ACT GAA AAA GGC AAT AAT TTA TCT
 N  K  V  D  L  Q  S  A  S  S  E  H  T  E  K  G  N  N  L  S
                                    5610                                    5640
GGA GGA GTA CAA GCA GGT TTT GGT AAA AAA ATG ACC GAT GAT GCT TCA TCT GTG AAT GGT
 G  G  V  Q  A  G  F  G  K  K  M  T  D  D  A  S  S  V  N  G
                                    5670                                    5700
TTA GGC AGT GCC CAA TTT GCT ATT GGT AAA CAA GAT GAG AAA AGC GTA TCT CGA GAA GGA
 L  G  S  A  Q  F  A  I  G  K  Q  D  E  K  S  V  S  R  E  G
                                    5730                                    5760
GGA ACG ATT AAT AAC TCA GGA AAC TTA ACG ATT AAT GGC AAT AGC GTC CAT CTT CAA GGA
 G  T  I  N  N  S  G  N  L  T  I  N  G  N  S  V  H  L  Q  G
                                    5790                                    5820
GCA CAA GTT AAT AGT AAA GAT ACT CAG CTA ACA TCA CAA TCC GGT GAT ATT GAG ATC ACA
 A  Q  V  N  S  K  D  T  Q  L  T  S  Q  S  G  D  I  E  I  T
                                    5850                                    5880
TCT GCG CAG TCT ACT GAT TAT AAA AAC AAT TGG GGA ACA GAT ATT GGC TTT AAT GGC AAA
 S  A  Q  S  T  D  Y  K  N  N  W  G  T  D  I  G  F  N  G  K
```

```
                                    5910                                    5940
AAA ACC AAT AAC ACC CCA AAA GAA GTT ACA GAG GAG AAA CCT GCA ACC TCT ATT CAT AAC
 K  T  N  N  T  P  K  E  V  T  E  E  K  P  A  T  S  I  H  N
                                    5970                                    6000
ATA GGG GGT AAA TTA CTG GTT AAT GTG GAA GAT CAA CAA AAA ACG AGC CAC CAA AAT GCA
 I  G  G  K  L  L  V  N  V  E  D  Q  Q  K  T  S  H  Q  N  A
                                    6030                                    6060
ACC TTA GAA ACA GGT ACA TTA ACG ATT AAT AGT AAT AAA GAT CTG ACA CTA TCT GGT GCT
 T  L  E  T  G  T  L  T  I  N  S  N  K  D  L  T  L  S  G  A
                                    6090                                    6120
AAT GTG ACT GCT GAT AGC GTT ACG GGT AAT GTC GGT GGT TCA CTC AAT ATT GCT AGC CAA
 N  V  T  A  D  S  V  T  G  N  V  G  G  S  L  N  I  A  S  Q
                                    6150                                    6180
AAA GAG AGT GAT CGC CAT GTC ACT GTT GGC GTC AAT GTT GGC TAT AAC CAC ACC AAC GAT
 K  E  S  D  R  H  V  T  V  G  V  N  V  G  Y  N  H  T  N  D
                                    6210                                    6240
CCT AAA TCA AGC CAA GTG AAT AAA ACG GCT AAA GCC GGA GGG TCA TTA TTA GAG AAA ACC
 P  K  S  S  Q  V  N  K  T  A  K  A  G  G  S  L  L  E  K  T
                                    6270                                    6300
ATC AAA GAT ACG ATT GAT TCA GGA ATT AAA TCA TCA ACA GAT GCT ATC TCT GAT AAA TAC
 I  K  D  T  I  D  S  G  I  K  S  S  T  D  A  I  S  D  K  Y
                                    6330                                    6360
AAT TCC CTC TCT TCA ACT ATT GCA GAT AAA ACA GGT ATC AGT GAT GAA ACT AAA GCC AAA
 N  S  L  S  S  T  I  A  D  K  T  G  I  S  D  E  T  K  A  K
                                    6390                                    6420
ATT GAT CAA GGT TTT GGT AAA GTT GGT AAT GGT ATT AAG AAT ATC GTC ACA GGT GCT GAG
 I  D  Q  G  F  G  K  V  G  N  G  I  K  N  I  V  T  G  A  E
                                    6450                                    6480
GGT CAT ACT GCT AAT GCA GAT ATC AAA GTC ACT CAT GTA GAT AAT GAT GCT GTC ACT AAA
 G  H  T  A  N  A  D  I  K  V  T  H  V  D  N  D  A  V  T  K
                                    6510                                    6540
ACC ACC TCC TTA ACT AGC AAC AAC GAC CTA TCA TTA AAT GTG AAT GGC TCG ACA AAA CTC
 T  T  S  L  T  S  N  N  D  L  S  L  N  V  N  G  S  T  K  L
                                    6570                                    6600
ACC GGA GCA GAA ATT GTG AGT CAA CAA GGC CAA GTT GAT TTA GGG GGA AGT AGC GTT AAA
 T  G  A  E  I  V  S  Q  Q  G  Q  V  D  L  G  G  S  S  V  K
                                    6630                                    6660
TTA GAA AAT ATT GAA GGT CAT CAT TAT GAA GCC GGC GCC GAT CTT GAT CTG AAA TCC TCT
 L  E  N  I  E  G  H  H  Y  E  A  G  A  D  L  D  L  K  S  S
                                    6690                                    6720
GTA GTG GAT TTA GCA AAA CAA CTG GTA GGT GGA GAT ATT TCT TTT AAA TCC CCC GTT AAA
 V  V  D  L  A  K  Q  L  V  G  G  D  I  S  F  K  S  P  V  K
                                    6750                                    6780
ACG AAT GAA ACC GTC AAT ACA AAA GCA TCT ATT TCT GAA AAA TAG TCG ATA CTT CAT AAA
 T  N  E  T  V  N  T  K  A  S  I  S  E  K  *
                                    6810                                    6840
AGA GTA ACT GCC TAT TTT TTG TAA GCT AAG CCC ACT TGA TTG TCA TTA TTA CAA GTG GGC
                                    6870                                    6900
TTT TTA TTA TCG TAC TTA TGG CAT CTG CTC CCG ATA AAC TTA CAA TTC TTT AAT ATT TAA
                                    6930                                    6960
AAT AAA CCA AGA AAA CAT AAC TCA TTG AAT ATA AAC AAA ATA AAA GCA AAC AAA ATA ATT
                                    6990                                    7020
ATA TAA TCC TTG ACC TTC CCC TAA TGG TAA AGC TTA AGC TTT GTC CCA TAC CAT ATT TAA
                                    7050                                    7080
GAG GGA AAG TTT GAT GAA TAC ACC TAC AAC ATT ATC CTC AGC AAA TAG GCT GAG CTT ACC
                                    7110                                    7140
TGT AGA AGG TAT GAC ATG TGC TTC ATG TGT CGG ACG TGT TGA ACG AGC ATT AAA GGC AGT
                                    7170
ACC TGA AAT AAA AGA TGC TGT CGT GAA TCT TGC AAC AGA ACG TGC TGA TAT
```

TABLE 1. Summary of DNA and amino acid sequence alignments[a]

Alignment	Sequence compared	% Identity	% Similarity	No. of gaps inserted
DNA	*hpm* vs *shl*[b]	52.1		17
Amino acid	HpmA vs ShlA[b]	46.7	65.3	11
	HpmB vs ShlB[b]	55.4	69.9	4

[a] From Bestfit program of the University of Wisconsin Genetics Computer Group software packages.
[b] Determined from the *shl* DNA sequence published by Poole et al. (32).

TABLE 2. Amino acid sequence predictions derived from DNA sequences

Amino acid sequence	No. of amino acids	Molecular mass (Da)	Isoelectric point	Net charged residues
HpmA	1,577	165,868	5.7	−40
HpmB	561	63,204	10.0	+11
ShlA[a]	1,608	165,056	9.2	+6
ShlB[a]	556	61,591	9.9	+6

[a] Determined from the *shl* DNA sequence published by Poole et al. (32).

addition to these potential transmembrane domains, many regions of predicted secondary structure are also highly conserved throughout the two amino acid sequences. It is interesting that two of three cysteine residues in HpmB (aa 76 and 105) were also conserved in ShlB. The nonconserved cysteine is present in the putative leader peptide of HpmB.

The putative translational start site of HpmA was identified at bp 2032 and was preceded by a potential Shine-Dalgarno site, GGAGAT. The NH_2-terminal amino acid sequence ([N/G][N/G]IVPDAGHQGPDV) was determined on HpmA purified from bacterial culture supernatants, and it matched the predicted sequence for aa 30 to 43 of HpmA (Fig. 3). The ambiquities seen at the first two amino acids of secreted HpmA were difficult to resolve between N or G; translation of the *hpmA* DNA sequence predicted NG as the first two amino acids after leader peptide cleavage. This confirmed the cleavage of a 29-amino-acid NH_2-terminal leader peptide in secreted HpmA.

The NH_2-terminal regions of HpmA and ShlA were found to be more similar than the COOH-terminal regions of these proteins (Table 1). Although both proteins are primarily hydrophilic, the predicted isoelectric points and net charges of HpmA and ShlA are very different (Table 2). The difference between the net charge of the two proteins is due mainly to the difference in the number of arginine residues in each protein: 19 for HpmA and 68 for ShlA (32). These additional arginine residues were found throughout ShlA.

The sequence of the region between aa 90 and 300 was very highly conserved between ShlA and HpmA. This region contained the two strongest hydrophobic domains (excluding the core of the leader sequence) which were conserved in each protein. HpmA and ShlA each contained two cysteine residues, both of which were found in this region (Cys-144 and Cys-147 in HpmA; Cys-145 and Cys-148 in ShlA). In addition to these hydrophobic domains, we have identified a number of conserved amphipathic domains in HpmA and ShlA at aa 375 to 385, 505 to 515, 860 to 875, 1040 to 1050, and 1365 to 1382.

Functional characteristics of HpmA and HpmB. Hemolytic activity as well as HpmA were seen in the bacterial culture supernatant (Table 3; Fig. 5). To date we have not detected HpmB in the culture supernatant (data not shown). In addition, HpmA has been shown to bind to sheep erythrocytes (K. G. Swihart and R. A. Welch, unpublished data), suggesting that HpmA is the structural hemolysin analogous to ShlA.

A deletion derivative of pWPM100 (pWPM102) that has a 2.5-kb *Hin*dIII-*Eco*NI fragment encoding HpmB removed was characterized. Liquid hemolysis assays were performed on strains carrying pWPM100, pWPM110, and various subclones, including pWPM102. Whole cells and filtered culture supernatants were used as sources of hemolysin. The form of HpmA produced by pWPM102, HpmA* (HpmA* indicates the inactive hemolysin produced in the absence of HpmB), was not hemolytically active (Table 3). *E. coli* carrying pWPM102 produced large amounts of HpmA*, which was not secreted from the cell (Fig. 5, lane 3). We also observed greater breakdown of intracellular HpmA* than of HpmA (Fig. 5, lane 4). When *hpmB* was cloned into pACYC184 (pWPM109) and expressed in *trans* to pWPM102, HpmA secretion and hemolytic activity were restored (Table 3; Fig. 5, lane 1).

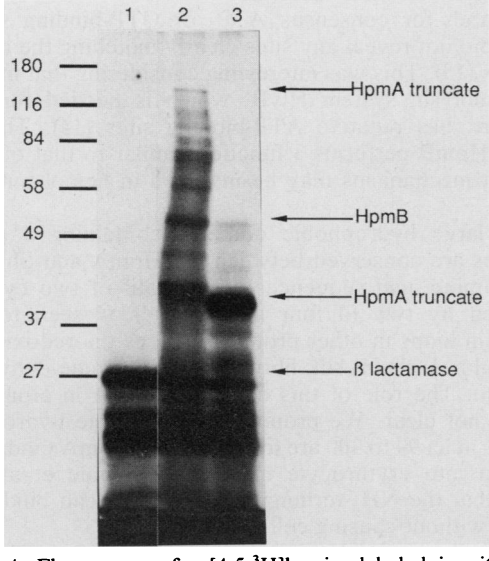

FIG. 4. Fluorogram of L-[4,5-^{3}H]leucine-labeled in vitro transcription-translation polypeptide products of *hpm* recombinant plasmids. Lanes: 1, control DNA (pUC19); 2, pWPM100; 3, pWPM99. Positions of molecular mass markers (in kilodaltons) are shown on the left; positions of HpmB and HpmA truncates and of β-lactamase are shown on the right.

TABLE 3. Hemolytic activity of HpmA expressed in *E. coli*[a]

Recombinant plasmid(s) harbored by *E. coli* DH1	Mean hemolytic activity[b] ± SD (n = 3)	
	Total cell culture	Cell-free supernatant
pUC19	0.029 ± 0.010	0.026 ± 0.004
pWPM100	1.810 ± 0.036	1.908 ± 0.076
pWPM102	0.045 ± 0.018	0.028 ± 0.003
pWPM109	0.022 ± 0.003	0.023 ± 0.002
pWPM102, pWPM109	1.791 ± 0.452	1.486 ± 0.308
pWPM110	1.400 ± 0.382	0.585 ± 0.271

[a] Samples (200 μl) were taken from a bacterial culture harvested at an OD_{600} of 0.9 and incubated at 37°C with a suspension of sheep erythrocytes at a final concentration of 1%.
[b] Expressed as OD_{540} per 30 min of incubation.

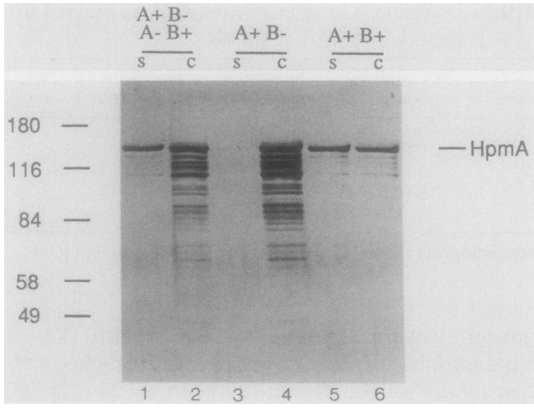

FIG. 5. Immunoblot analysis using rabbit polyclonal anti-HpmA antiserum. Lanes: 1 and 2, *E. coli* DH1 carrying two plasmids, pWPM102 (A+ B−) and pWPM109 (A− B+) in *trans*; 3 and 4, *E. coli* DH1 carrying pWPM102 (A+ B−); 5 and 6, *E. coli* DH1 carrying pWPM100 (A+ B+). An 8-μl sample of either whole-cell culture (c) or filter-sterilized supernatant (s) was added in each lane. Positions of molecular mass markers (in kilodaltons) are shown on the left.

DISCUSSION

We report the DNA sequence of the genes required for the calcium-independent hemolytic activity of *P. mirabilis* clinical isolate 477-12. Two proteins necessary for hemolytic activity are encoded on this determinant: HpmA (166 kDa) and HpmB (63 kDa). Significant identity is seen with the *S. marcescens* hemolysin genes *shlA* and *shlB*. Amino acid sequences of both the A and B proteins are very similar, showing many conserved hydrophobic and amphipathic domains as well as conservation of strong secondary-structure predictions.

Previous *E. coli* minicell analysis of peptides encoded by pWPM100 confirmed the production of HpmA, using L-[^{35}S]methionine as a label (41). As mentioned earlier, pWPM100 actually encodes a COOH-terminal truncated version of HpmA. The fact that HpmB contains only one methionine residue, which is located at the NH_2 terminus of the protein, may explain our previous failure to detect HpmB in these experiments (41). HpmB production was also difficult to demonstrate by using an in vitro transcription-translation system. HpmA is very unstable in this system, resulting in a number of apparent breakdown products that mask the HpmB produced from pWPM100. We used a pWPM100 deletion derivative (pWPM99) that has the Km^r cassette of Tn*903* inserted at the *Cla*I deletion site from bases 3099 to 4889. We have putatively identified HpmB and a COOH-terminal HpmA truncate of ≈42 kDa encoded by pWPM99. With no large molecular mass breakdown products of HpmA, a potential HpmB band is clearly visible. After cleavage of the NH_2-terminal leader peptide, we would predict HpmB to have a molecular mass of ≈60 kDa. The peptide we predict to be HpmB appears to migrate anomalously, with an M_r of 53,000 (Fig. 4). Similar discrepancies have been observed with HlyC and HlyB, which migrate with M_rs of 15,000 and 66,000, respectively, whereas the molecular masses from the predicted amino acid sequences are 20 and 79 kDa (10). The predicted high isoelectric points (9.5 for HlyC, 10.2 for HlyB, and 10.0 for HpmB) may contribute to the faster migration of hemolysin gene products in SDS-polyacrylamide gels. It was surprising that the 42-kDa HpmA truncate band encoded by pWPM99 (Fig. 4, lane 3) appears much more intense than the putative HpmB or the HpmA truncate encoded by pWPM100. This may be due to increased stability of this truncate compared with the longer HpmA or HpmB in this system. Full-length HpmA from pWPM110 also produced a large number of breakdown products (data not shown).

HpmA and ShlA are two of only a few proteins that are secreted extracellularly when expressed in *E. coli* (17). Some proteins secreted extracellularly from *E. coli* require closely linked secretion genes (*Klebsiella* pullulanase [7] and *E. coli* HlyA [10]). Others, e.g., the *Neisseria gonorrhoeae* immunoglobulin A protease (30) and the *S. marcescens* protease (43), require both the NH_2 and COOH-terminal domains of the protein for secretion. HpmA does not require COOH-terminal domains for secretion. HpmB is, however, involved in the secretion of HpmA and is also necessary for conversion of HpmA* to hemolytically active HpmA. Whereas HpmA is secreted from *E. coli* in the presence of HpmB, we have not detected HpmB in culture supernatants. HpmB appears to function analogously to ShlB, since it is necessary for secretion and hemolytic activity of HpmA (35). We therefore hypothesize that HpmB, like ShlB, is located in the outer membrane (32). Schiebel et al. (35) have localized ShlA* to the periplasm in cells not producing ShlB. Given the similarities to HpmA and HpmB, we would predict that HpmA* is also located in the periplasm. Future studies investigating how HpmB facilitates HpmA secretion will allow a better comparison with the Shl system and can provide further insight to the broader question of extracellular protein secretion.

Because the *shl* genes have been shown to be regulated by the *fur* gene product, we searched *hpm* sequences for consensus *E. coli* Fur-binding sites (4). A putative Fur-binding site was identified, with 12 of 19 matches to the consensus. This site has two more matches than seen with the *shl* Fur-binding site (31). Thus far, studies examining the level of HpmA production in the absence of Fe^{2+} or in a *fur* background have been inconclusive (Swihart and Welch, unpublished data). At this point, the extent of Fur regulation of *hpm* genes remains unclear.

A search for consensus ATP- or GTP-binding sites in HpmB did not reveal any sites clearly matching the consensus sites (15). This was interesting considering that in the *E. coli* hemolysin system HlyB, which is needed for HlyA transport, has putative ATP-binding sites (14). Thus, although HpmB performs a function similar to that of HlyB, different mechanisms may be involved in hemolysin transport.

Two large hydrophobic domains containing a pair of cysteines are conserved between the HpmA and ShlA predicted amino acid sequences. This motif of two cysteines separated by two to four amino acids is seen to close intrachain loops in other proteins such as thioredoxin, pepsin, insulin A chain, silk fibroin, and lipoamide dehydrogenase (36). The role of this conserved motif in HpmA and ShlA is not clear. We propose that the large hydrophobic domains at aa 90 to 300 are important for HpmA binding and insertion into erythrocyte membranes. Poole et al. have shown that the NH_2-terminal 43% of ShlA can bind erythrocytes without causing cell lysis (32).

We further speculate that the conserved amphipathic domains (aa 375 to 385, 505 to 515, 860 to 875, 1040 to 1050, and 1365 to 1382) may be responsible for pore formation by these proteins. Our experiments with the HpmA truncate produced by pWPM100 (1,220 aa) and other subclones suggest that some but not all of these domains are necessary

for hemolytic activity (data not shown). Poole et al. have constructed 3′ deletion mutants of *shlA* and studied the hemolytic activity of ShlA truncates produced (32). These truncates appear to lose activity as more of these amphipathic domains are deleted. It is interesting that the original clone pWPM100 encoded only the NH_2-terminal 75% of HpmA yet *E. coli* cells harboring this plasmid produced more hemolytic activity than did those carrying pWPM110, which encodes all of HpmA (Table 3). This result appears to have been due to a cloning artifact. In pWPM100, *hpmB* and *hpmA* are transcribed from the *lac* promoter of pUC19 (the native promoter upstream of *hpmB* was not cloned). In pWPM110, about 800 bp upstream of *hpmB* is present, including the putative native promoter. Poor recognition of this *P. mirabilis* promoter in *E. coli* would explain the reduced expression of HpmA seen from pWPM110, whereas the *lac* promoter in pWPM100 allows for high levels of *hpm* transcription. We are in the process of cloning full-length *hpmA* such that it is also expressed from various promoters. Additional mutants containing in-frame and 3′ deletions of *hpmA* will also be studied to investigate the contribution of different regions to hemolytic activity.

The major differences seen between HpmA and ShlA are the predicted net charges and isoelectric points of the two proteins. A thorough comparison of HpmA and ShlA hemolytic activities has not yet been performed. This comparison will provide information concerning the effects of charge (or other) differences on the overall activity of these cytolysins. To date we have not observed any gross differences in the activity of HpmA compared with that reported by Braun et al. (3) for ShlA.

The most striking aspect of the comparison between *shl* and *hpm* sequences is the conservation of DNA and predicted amino acid sequence similarities despite the large difference seen in G+C contents (*shl*, 65%; *hpm*, 38%) of these genes. The G+C contents of both sequences reflect the overall G+C content of each organism (*S. marcescens*, 58%; *P. mirabilis*, 39%) (9). This finding suggests that the genes diverged from a common ancestral gene long ago. To retain this degree of similarity despite diverging G+C contents would seem to require strong selective pressures resisting functional change. The alternative theory involves convergent vertical evolution of the genes; this seems unlikely given the length of the sequence identity and the conservation of the operon structure.

After the initial cloning of *hpm*, a survey of its dissemination among members of the family *Enterobacteriaceae* and other hemolytic gram-negative bacilli was performed (41). Although *hpm* sequences are common among *Proteus* isolates (Swihart and Welch, in preparation), Southern blot analysis did not reveal sequences similar to the *hpm* sequence in the other genera and species examined (*Providencia*, *Pasteurella*, *Acinetobacter*, *Citrobacter*, *Morganella*, *Serratia*, and *E. coli*) (41). It may be difficult to predict the dissemination of *hpm* sequences in other genera given that the similarity between *shl* and *hpm* sequences is insufficient to allow detection by DNA hybridization. Immunoblot analysis may be useful in further examination of HpmA dissemination in other organisms. On the basis of sequence conservation seen between these two distantly related organisms, we predict that similar sequences will be detected elsewhere. Finally, the comparison of these proteins should be valuable in future structure-function studies of these cytolytic exotoxins.

ACKNOWLEDGMENTS

R.A.W. is a Pew scholar in the biomedical sciences.

We thank Kristin Swihart and Christiane Forestier for their comments on the manuscript. We also thank the University of Wisconsin Genetics Computer Group technical support staff, especially Irv Edelman, for help in the sequence analysis.

LITERATURE CITED

1. **Bachmann, B. J.** 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
2. **Bauernfeind, A., K. Naber, and D. Sauerwein.** 1987. Spectrum of bacterial pathogens in uncomplicated and complicated urinary tract infections. Eur. Urol. **13**(Suppl. 1):9–12.
3. **Braun, V., B. Neuss, Y. Raun, E. Schiebel, H. Schoffler, and G. Jander.** 1987. Identification of the *Serratia marcescens* hemolysin determinant by cloning into *Escherichia coli*. J. Bacteriol. **169**:2113–2120.
4. **Calderwood, S. B., and J. J. Mekalanos.** 1987. Iron regulation of Shiga-like toxin expression in *Escherichia coli* is mediated by the *fur* locus. J. Bacteriol. **169**:4759–4764.
5. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the p15A cryptic miniplasmid. J. Bacteriol. **134**: 1141–1156.
6. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.
7. **d'Enfert, C., A. Ryter, and A. P. Pugsley.** 1987. Cloning and expression in *E. coli* of the *Klebsiella pneumoniae* genes for production, surface localization and secretion of the lipoprotein pullulanase. EMBO J. **6**:3531–3538.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
9. **Falkow, S., I. R. Ryman, and O. Washington.** 1962. Deoxyribonucleic acid base composition of *Proteus* and *Providencia* organisms. J. Bacteriol. **158**:1037–1040.
10. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
11. **Felmlee, T., and R. A. Welch.** 1988. Alterations of amino acid repeats in the *E. coli* hemolysin affect cytolytic activity and secretion. Proc. Natl. Acad. Sci. USA **85**:5369–5273.
12. **Garcia, E., A. M. Hamers, H. E. N. Bergmans, B. A. M. van der Zeijist, and W. Gaastra.** 1988. Adhesion of canine and human uropathogenic *E. coli* and *P. mirabilis* strains to canine and human epithelial cells. Curr. Microbiol. **17**:333–337.
13. **Glass, R. E.** 1982. Gene function: *E. coli* and its heritable elements, p. 95–96. University of California Press, Berkeley.
14. **Gros, P., J. Croop, and D. Housman.** 1986. Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell **47**:371–380.
15. **Harlow, E., and D. Lane.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**: 351–359.
17. **Hirst, T. R., and R. Welch.** 1988. Mechanisms for secretion of extracellular proteins by gram-negative bacteria. Trends Biochem. Sci. **13**:265–269.
18. **Hjelm, E., E. Lundell-Etherden, and P. A. Mardh.** 1987. Ascending urinary tract infections in rats induced by *Staphylococcus saprophyticus* and *Proteus mirabilis*. Acta Pathol. Microbiol. Scand. Sect. B **95**:347–350.
19. **Holmgren, K., B. G. Donielson, and B. Fellstrom.** 1987. Infection induced urinary calculi and renal failure. Scand. J. Urol. Nephrol. **21**:219–223.

20. **Hull, R. A., R. Gill, P. Hsu, B. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33**:933–938.
21. **Hunkapillar, M. W., E. Lujan, F. Ostrander, and L. E. Hood.** 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. Methods Enzymol. **91**:227–236.
22. **Koronakis, V., M. Cross, B. Senior, E. Koronakis, and C. Hughes.** 1987. The secreted hemolysins of *Proteus mirabilis*, *Proteus vulgaris*, and *Morganella morganii* are genetically related to each other and the alpha-hemolysin of *Escherichia coli*. J. Bacteriol. **169**:1509–1515.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **McClean, R. J. C., J. C. Nickel, V. C. Noakes, and J. W. Costerton.** 1985. An in vitro ultrastructural study of infectious kidney stone genesis. Infect. Immun. **49**:805–811.
25. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
26. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Musher, D. M., D. P. Griffith, D. Yawn, and R. D. Rossen.** 1975. The role of urease in pyelonephritis resulting in urinary tract infection with *Proteus*. J. Infect. Dis. **131**:177–181.
28. **Oka, A., H. Sugisaki, and M. Takanami.** 1981. Nucleotide sequence of the kanamycin resistance transposon T*n903*. J. Mol. Biol. **147**:217–226.
29. **Peerbooms, P. G. H., A. Marian, J. J. Verweij, and D. M. MacLaren.** 1984. Vero cell invasiveness of *Proteus mirabilis*. Infect. Immun. **43**:1068–1071.
30. **Pohner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *N. gonorrhoeae* IgA protease. Nature (London) **325**:458–462.
31. **Poole, K., and V. Braun.** 1988. Iron regulation of *Serratia marcescens* hemolysin gene expression. Infect. Immun. **56**: 2967–2971.
32. **Poole, K., E. Schiebel, and V. Braun.** 1988. Molecular characterization of the hemolysin determinant of *Serratia marcescens*. J. Bacteriol. **170**:3177–3188.
33. **Rigby, P. W. J., M. Dreckmann, C. Rhodes, and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schiebel, E., H. Schwarz, and V. Braun.** 1989. Subcellular location and unique secretion of the hemolysin of *Serratia marcescens*. J. Biol. Chem. **264**:16311–16320.
36. **Schulz, G. E., and R. H. Schirmer.** 1979. Principles of protein structure, p. 55. Springer-Verlag, New York.
37. **Senior, B. W.** 1979. The special affinity of particular types of *P. mirabilis* for the urinary tract. J. Med. Microbiol. **12**:1–18.
38. **Svanborg-Eden, C., P. Larsson, and H. Lomberg.** 1980. Attachment of *Proteus mirabilis* to human urinary sediment epithelial cells in vitro is different from that of *Escherichia coli*. Infect. Immun. **27**:804–807.
39. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
40. **Walz, S. E., S. K. Wray, S. I. Hull, and R. A. Hull.** 1988. Multiple proteins encoded within the urease gene complex of *Proteus mirabilis*. J. Bacteriol. **170**:1027–1033.
41. **Welch, R. A.** 1987. Identification of two different hemolysin determinants in uropathogenic *Proteus* isolates. Infect. Immun. **55**:2183–2190.
42. **Wray, S. K., S. I. Hull, R. G. Cook, J. Barrish, and R. A. Hull.** 1986. Identification and characterization of a uroepithelial cell adhesin from a uropathogenic isolate of *Proteus mirabilis*. Infect. Immun. **54**:43–49.
43. **Yanagida, N., T. Uozumi, and T. Beppu.** 1986. Specific excretion of *S. marcescens* protease through the outer membrane of *Escherichia coli*. J. Bacteriol. **166**:937–944.
44. **Zubay, G.** 1973. *In vitro* synthesis of protein in microbial systems. Annu. Rev. Genet. **1**:267–287.