INFECTION AND IMMUNITY, Sept 1990, p. 3042–3049
0019-9567/90/093042-08$02.00/0

Vol. 58, No. 9

Vibrio cholerae O395 *tcpA* Pilin Gene Sequence and Comparison of Predicted Protein Structural Features to Those of Type 4 Pilins

CAROLYN E. SHAW AND RONALD K. TAYLOR*

Department of Microbiology and Immunology, The University of Tennessee, Memphis, Tennessee 38163

Received 22 February 1990/Accepted 25 June 1990

***Vibrio cholerae* O1 expresses a pilus that is coordinately regulated with cholera toxin production and hence termed TCP, for toxin-coregulated pilus. Insertion of Tn*5* IS*50*$_L$::*phoA* (Tn*phoA*) into the major pilin subunit gene, *tcpA*, has previously been shown to render the strain avirulent as a result of its inability to colonize. One such insertion was isolated and used as a probe to screen for clones containing the intact *tcpA* gene. The DNA sequence of *tcpA* was determined by using the intact gene and several *tcpA-phoA* gene fusions. The deduced protein sequence agreed completely with that previously determined for the TcpA N terminus and with the size of the mature pilin protein. The reported homology with *N*-methylphenylalanine (type 4) pilins near the N terminus was extended and shown to include components of the atypical leader peptide as well as overall predicted structural similarities in other regions of the pilins. In contrast to the modified N-terminal phenylalanine residue found in all characterized type 4 pilins, the corresponding position in *tcpA* contains a Met codon, thus implying that the previously uncharacterized amino acid corresponding to the N-terminal position of the mature TcpA pilin is a modified form of methionine. Except for this difference, mature TcpA has the overall predicted structural motifs shared among type 4 pilins.**

The molecular interactions required for the initiation of colonization of mucosal surfaces by a number of pathogenic bacteria are typically thought to be mediated through bacterial surface appendages termed pili, or fimbriae, that recognize specific host cell receptors. *Vibrio cholerae* elaborates several such structures (6, 11, 48, 51) that could possibly have roles in the colonization mechanism. Interestingly, the expression of one of these pili parallels the synthesis of another secreted virulence factor, cholera toxin. This pilus, designated TCP, for toxin-coregulated pilus (48), has been correlated specifically with clinical serotype O1 isolated strains (49) and is the only *V. cholerae* pilus that has thus far been demonstrated to have a role in colonization. This requirement has been demonstrated for colonization of the gut mucosa both of humans (13) and of infant mice (48) used as an experimental cholera model. The coordinate regulation of TCP and cholera toxin, both of which require ToxR for expression, has led to the idea that the corresponding genes are part of a virulence regulon encompassing a number of genes encoding exported factors that are expressed in response to external conditions present in the infected host (29, 30, 34, 48).

The N-terminal amino acid sequence of the TCP major pilin subunit, TcpA, has been determined (48). It displays significant homology with a group of pilus major subunits that belong to a class of pili that has been termed type 4 on the basis of similar morphology and putative function (31) or more recently on the basis of homologies between the major subunits (4, 24, 25). This group has also been termed the *N*-methylphenylalanine (NMePhe) pili on the basis of the modified N-terminal residue of the mature form of the pilin (4, 32). These pili are elaborated by a diverse group of gram-negative bacteria that share the common feature of colonizing mucosal surfaces (10). In the case of *Pseudomonas* species, these pili have been demonstrated to be essential for virulence (32). Their role in colonization has also been implicated for *Neisseria* and *Moraxella* species by correlation between the state of piliation and the ability to adhere to various epithelial cells or colonize their respective hosts (33, 36, 45). Similarly, the TCP pilus of *V. cholerae* has been demonstrated to be required for colonization in both humans and infant mice, using mutants with defined genetic lesions in *tcpA* (13, 48). The mechanism by which type 4 pili mediate colonization has yet to be eludicated to the molecular detail that has been accomplished for some *Escherichia coli* fimbrial types. For example, it is not known whether the colonization-mediating domains lie within the major subunit as has been proposed for K99 fimbriae (16) or are encoded on ancillary adhesin proteins that might occur either at the pilus tip as for PAP (21) or spaced intermittently as well as being at the tip of the fimbrial structure as shown for type 1 fimbriae (1). In the case of type 4 pili, studies on the binding domains are complicated by the lack of identified cellular receptors and the potential role played by these pili in interbacterial interactions involved in microcolony formation, as suggested by their ability to mediate bacterial autoagglutination in culture (48).

There is accumulating evidence that at least part of the colonizing function of type 4 pili is mediated through the major subunit. For example, antibodies directed against proteolytic fragments from the central region of the gonococcal pilus inhibit bacteria expressing the pilus from binding endocytic cells in vitro (35, 39). In recent elegant studies, synthetic peptides corresponding to the carboxy-terminal cystine loop region of the major subunit of the *Pseudomonas* pilus have been shown to compete for binding to an epithelial cell line, attesting to the presence of an adhesive domain within the major pilin subunit (15). These studies also identified a putative cell surface receptor for pilin-mediated binding. In the case of TCP, both polyclonal and monoclonal antibodies directed against the major subunit of the pilus prevent infection by virulent strains in the infant mouse experimental cholera model (46; R. K. Taylor, C. E. Shaw, D. Sun, R. A. Sumrada, and J. A. Rhine, *in* Y. Takeda and R. B. Sack, ed., *Advances in Research on Cholera and*

* Corresponding author.

Related Diarrheas, vol. 8, in press; R. K. Taylor and D. Sun, *in* Y. Takeda and R. B. Sack, ed., *Advances in Research on Cholera and Related Diarrheas*, vol. 9, in press). Only a subset of the monoclonal antibodies that recognize TcpA tertiary structure prevent infection in this model, indicating that they specifically recognize functional domains within TcpA pilin involved in bacterial colonization of the intestine (Taylor and Sun, in press; submitted for publication). Thus, the sequence of *tcpA* and deduced features of the protein will help to define those regions involved in colonization.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *E. coli* MC1061 (2) was the host for cloned Tn*5* IS*50*$_L$::*phoA* (Tn*phoA*) fusions. Strain LE392 (22) was the host strain for the *V. cholerae* cosmid genome library. Strain CC118 (23) was used for manipulations involving *phoA* gene fusions and as donor in triparental matings with *V. cholerae* recipients as previously described (47). Strain JF626 (J. Felton), a JM103 derivative, was the host for recombinant M13 bacteriophage (27). The genome library was constructed from wild-type *V. cholerae* O1 strain O395, which is a virulent classical strain of Ogawa serotype. O395 *tcpA*::Tn*phoA* derivative strains RT104.11 and RT110.21 served as a source of DNA for fusion clones and for genetic complementation with *tcpA*$^+$ clones. Bacteria were cultured under standard conditions (42) except that TCP pilus expression as optimized by growing strain O395 in LB (pH 6.5) at 30°C (48). Antibiotics were used at the following concentrations, in micrograms per milliliter: ampicillin, 100; kanamycin, 40; chloramphenicol, 10; and streptomycin, 100. Isopropyl β-D-thiogalactopyranoside (IPTG) was used at 0.1 mM; 5-bromo-4-chloro-3-indolylphosphate *p*-toluidine salt (XP) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) were each used at 40 μg/ml.

Cloning Tn*phoA* fusion joints. A 0.5-μg sample of total DNA from strain RT104.01 or RT110.21, prepared as described previously (26), was digested with *Xba*I and ligated with 0.1 μg of similarly digested pJM22.1, a pUC derivative carrying a Tc^r determinant flanked by *Xba*I sites (J. Mekalanos). Strain MC1061 (2) was transformed with selection for Ap^r Km^r colonies that appeared blue when XP was incorporated into the selective agar (47).

Genomic library construction. Total DNA from strain O395 was subjected to partial *Sau*3A digestion, accomplished by enzyme dilution as described by Maniatis et al. (22). A portion of each dilution sample was analyzed on a 0.6% agarose gel, and the remainders of samples possessing a majority of fragments in the 25- to 50-kilobase (kb) range were pooled and ligated to CsCl-ethidium bromide gradient-purified cosmid vector pHC79 as described previously (14, 22). The ligated products were packaged into lambda phage particles in vitro, using Gigapack plus (Stratagene), and infected into strain LE392 that had been grown in the presence of 0.4% maltose and 1 mM $MgSO_4$. Then 1,000 resultant Ap^r colonies were inoculated into microtiter dish wells containing LB-ampicillin with 10% glycerol and grown overnight. Samples were replica plated to nitrocellulose filters for growth on ampicillin agar, and the microtiter cultures were stored at −70°C for subsequent retrieval of positive clones. After growth on the filters for 8 h in the presence of ampicillin and overnight in the presence of chloramphenicol, the colonies were lysed and screened by hybridization (22) to nick-translated DNA from two M13 subclones carrying a small restriction fragment adjacent to each end of the cloned RT110.21 *tcpA*::Tn*phoA* fusion.

Southern hybridization of total and plasmid DNA preparations with a *tcpA*-specific probe. To determine the location of *tcpA*, total DNA or CsCl-gradient purified DNA from strain O395 was digested with restriction endonucleases, separated by agarose gel electrophoresis, transferred to nitrocellulose, and hybridized to a ^{32}P-labeled 5-kb *Xba*I probe containing *tcpA* as described previously (22, 43).

Subcloning and DNA sequencing. Appropriate portions of *tcpA* or *tcpA-phoA* fusions were subcloned into M13mp18 and M13mp19 (53), or tg130 and tg131 (18), and plaques from transfected JF626 cells to be screened were detected as white in presence of IPTG and X-Gal or blue in the presence of IPTG and XP in the case of intact *phoA* fusion subclones. Single-stranded DNA from phage carrying the appropriate inserts as determined by restriction enzyme analysis was sequenced by the dideoxyoligonucleotide method, using either a universal *lac* primer or a *phoA* primer to initiate synthesis (27, 37, 47).

Transmission electron microscopy. Bacterial samples were grown in Luria broth (pH 6.5) at 30°C, placed onto 300-mesh Formvar-coated grids, and stained with 0.5% phosphotungstic acid at pH 6.5 to determine the presence of TCP pili. Samples were visualized with a Zeiss transmission electron microscope.

Western immunoblot analysis. Western analysis was performed as described previously (50). Samples were prepared from broth cultures inoculated with fresh colonies and allowed to grow for 8 h. Growth was in LB (pH 6.5) at 30°C for *V. cholerae* and in LB (pH 7.0) at 37°C for *E. coli* carrying cloned genes. Bacteria were pelleted from 0.5 ml of culture, suspended in 0.2 ml of protein sample buffer, boiled for 5 min, and pelleted for 5 min at 15,400 × *g*. A 4-μl amount of supernatant was loaded per lane for sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (20) before protein transfer and immunoblotting. *E. coli* samples were treated the same way except that 1.5 ml of culture was resuspended in 75 μl of protein sample buffer and 50 μl of final supernatant was loaded per lane. After separation by electrophoresis, the proteins were transferred to nitrocellulose and the membrane was incubated with antiserum from a rabbit that had been immunized with the 20.5-kilodalton (kDa) TcpA band extracted from an SDS-polyacrylamide gel of a TCP preparation (48).

Mobilizing cosmid clones into *V. cholerae*. Cosmid clones carrying *tcpA* were mobilized into the *tcpA*-mutant *V. cholerae* strain RT110.21 (48) by triparental matings as previously described (47).

RESULTS

Isolating *tcpA*. The *tcpA* pilin gene was initially identified by the Tn*phoA* insertions RT110.21 and RT104.11, each of which lies within *tcpA* and creates a gene fusion between *tcpA* and *phoA* marked by the adjacent Tn*5*-derived Km^r gene (23). These *tcpA*::Tn*phoA* insertions, which map within a 5-kb *Xba*I restriction fragment (Fig. 1), were cloned into the *Xba*I site of vector pJM22.1 by selecting for the transposon-associated Km^r gene and screening for active alkaline phosphatase by including XP in the agar. Specific DNA fragments subcloned from the resulting plasmids, designated pRT110.21 and pRT104.11, were inserted into M13 vectors, enabling much of the DNA sequence of *tcpA* to be determined. To obtain the intact *tcpA* gene, a *V. cholerae* genomic library was constructed in the cosmid pHC79, which was packaged in vitro and transfected into *E. coli* LE392. Ap^r colonies were selected, and those containing

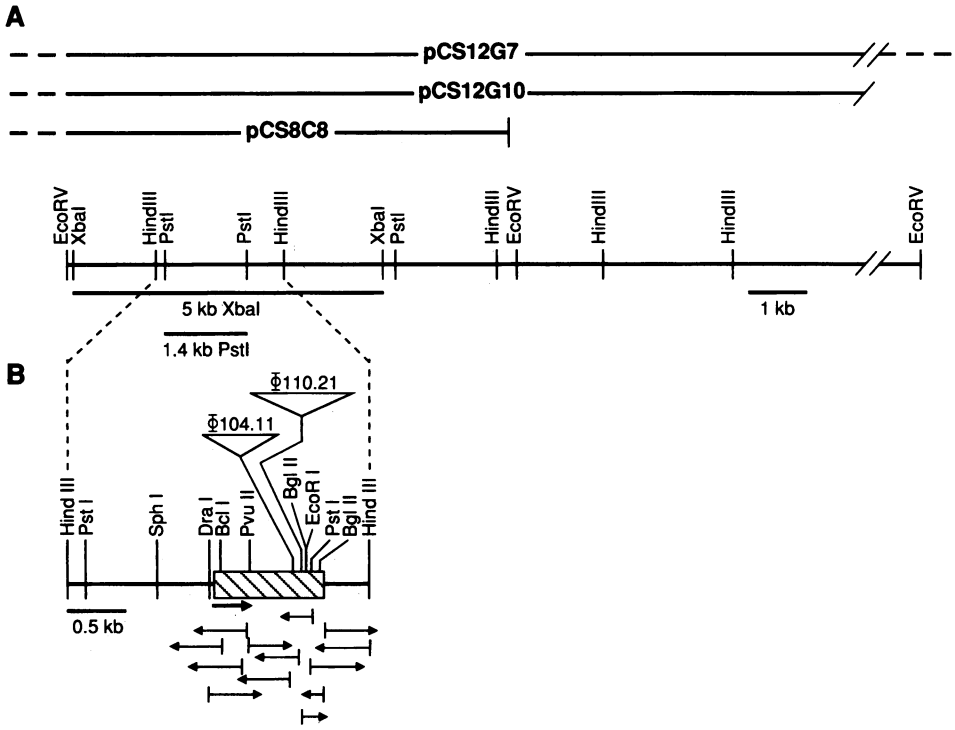

FIG. 1. Partial restriction map of the cosmid clones and sequencing strategy for *tcpA*. (A) The extents of the three cosmid clones are shown or, where not known, are indicated by a dashed line. The locations of the restriction fragments used in the study are noted. (B) The hatched box represents the location of *tcpA* within a 2-kb *Hin*dIII restriction fragment on pRTG7H3. The heavy arrow indicates the direction of transcription. Two active Tn*phoA* fusions used as probes for *tcpA* specific sequences and to verify the reading frame are indicated. The light arrows show the regions and direction of DNA sequencing reactions used to determine the *tcpA* gene sequence.

plasmids which carried *tcpA* sequences were identified by screening the clones by colony hybridization, using a mixture of two *tcpA*-specific DNA probes derived from the pRT110.21 insertion, one representing the 5′ end and the other the 3′ end, and immediately flanking sequences to the gene. Plasmid DNA was isolated from the strains that hybridized positively to the *tcpA* probes, yielding several cosmids which were then characterized by restriction analysis. Cosmids pCS12G7, pCS12G10, and PCS8C8 each contained a 5-kb *Xba*I DNA fragment with an internal 1.4-kb *Pst*I fragment (Fig. 1) that previous Southern hybridization experiments had shown to contain *tcpA* sequences (49).

Expression of TcpA from cloned fragments. The cosmid clones were expected to carry a complete copy of the *tcpA* gene because of its central location within the 5-kb *Xba*I fragment. Western immunoblot analysis using TcpA-specific antiserum and total protein extracts from *E. coli* LE392 carrying each of the cosmids demonstrated that this was indeed the case. Cosmid clones pCS12G7 and pCS12G10 expressed a cross-reactive protein migrating in the same position as the 20.5-kDa TcpA pilin band produced by *V. cholerae* strain O395 (Fig. 2). In the case of *E. coli* harboring cosmid clone pCS8C8, a cross-reactive band that migrated with an apparent molecular mass 2.5 kDa larger than that of native pilin was visualized. This size difference was thought most likely to be due to either an aberrant DNA rearrangement that occurred during cloning, resulting in an altered or fused *tcpA* gene encoding a larger product, or else the expression of a precursor pilin form that was not processed to mature form in *E. coli* LE392 unless additional gene products encoded by clones pCS12G7 or pCS12G10 were present.

The possibility of additional gene products required for TcpA processing was addressed in two ways. First, the internal 2-kb *Hin*dIII fragment on cosmid pCS12G7 was subcloned into plasmid vector pHSS6 (40). The resulting plasmid, pRTG7H3, expressed the larger cross-reactive band, indicating that when *tcpA* was removed from adjacent cloned sequences present on pCS12G7, its product now assumed the larger form. A second method, to rule out that pCS8C8 encoded an altered form of TcpA, was addressed by genetic complementation.

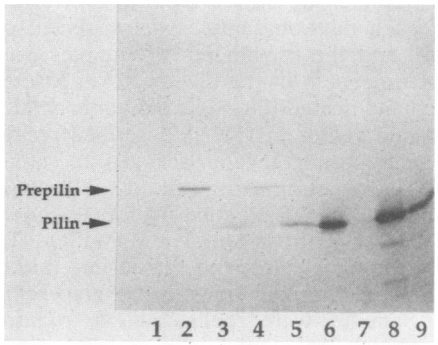

FIG. 2. Western immunoblot analysis showing expression and processing of the cloned *tcpA* gene product. Total protein extracts were separated on 12.5% SDS-polyacrylamide gels, transferred to a nitrocellulose membrane, and probed with rabbit polyclonal anti-TcpA antiserum. Lanes: 1, *E. coli* LE392(pHSS6); 2, LE392(pRTG7H3); 3, LE392(pCS12G7); 4, LE392(pCS8C8); 5, LE392(pCS12G10); 6, *V. cholerae* O395; 7, *V. cholerae* RT110.21; 8, RT110.21(pCS12G7); 9, RT110.21(pCS8C8).

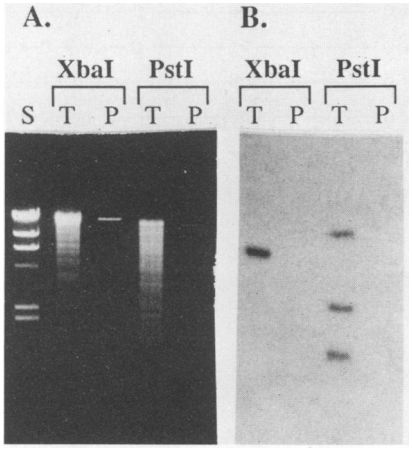

FIG. 3. Southern hybridization analysis to determine the location of *tcpA*. (A) Total (T) or plasmid (P) DNA was prepared from strain O395 and digested with either *Xba*I or *Pst*I, and the resulting fragments were electrophoretically separated on a 0.8% agarose gel. (B) The DNA from the gel was transferred to nitrocellulose and probed with a labeled 5-kb *Xba*I fragment that contains the entire *tcpA* gene and adjacent DNA, as shown in Fig. 1A.

Genetic complementation of the RT110.21 pilin-negative phenotype, using the cosmid clones. Total protein extracts of *V. cholerae* O395 reveal an abundance of TcpA after growth under expressing conditions that could be visualized directly on Coomassie blue-stained SDS-PAGE or by Western blot with anti-TcpA antibody (Fig. 2). The O395 *tcpA*::Tn*phoA* insertion strain RT110.21 exhibited loss of this band. When pCS12G7 was mobilized into strain RT110.21, the cross-reactive pilin band was again apparent on a Western immunoblot (Fig. 2), demonstrating complementation of the pilin-negative phenotype; a similar result was seen with RT110.21 carrying pCS8C8 (Fig. 2). Thus, both cosmids tested were able to complement the pilin defect of strain RT110.21, and the aberrant size of the pCS8C8-encoded protein when expressed in *E. coli* was also complemented to produce the wild-type-size pilin in *V. cholerae*. This result further demonstrates that pCS8C8 does not carry an altered *tcpA* gene and that the cosmid clones pCS12G7 and pCS12G10 carry, in addition to *tcpA*, adjacent genes that encode products necessary for pilin maturation. In the case of each cosmid tested in the complementation analysis, the pilin produced was assembled and visualized on the surface of strain RT110.21 by electron microscopy after negative staining (data not shown).

Localization of *tcpA* to the chromosome. Previous hybridization analysis using a DNA fragment encompassing a short region upstream of the RT110.21 fusion joint allowed us to localize *tcpA* on a 5-kb *Xba*I or 1.4-kb *Pst*I restriction fragment of strain O395 DNA (49). To determine whether these restriction fragments were of chromosomal origin or belonged to any of several cryptic plasmids identified in various *V. cholerae* O1 isolates (3), the 5-kb *Xba*I fragment was used to probe total DNA and CsCl-ethidium bromide gradient-purified plasmid DNA extracted from strain O395. Strain O395 appeared to have a single plasmid species that contained a unique *Xba*I site and migrated at approximately 23 kb (Fig. 3A). As expected on the basis of this size, the plasmid sequences did not hybridize with the 5-kb *Xba*I probe in either the *Xba*I- or *Pst*I-digested lanes (Fig. 3B), whereas the total DNA lanes showed hybridization patterns corresponding to the correct size fragments, thus demonstrating that *tcpA* is chromosomally encoded.

```
TTGAATTGAATAAGTTGGCCTTTTTTTAGTGTCTGATTTATTTTGTGCGTGAATGTTACTCGTGTTT

CTTTCAATGCAAGTGTGTTATTAAAAAAATAAAAAAACACAGCAAAAAATGACATCTGTCAATTGTA

GGTGACTTTGTGTGGTTAAATGTGCGTGTTGCTTACGTTATCTAAAAAAGACCAAGCGACGCATTTC

CTTTAAAGACAGTAAAATGGTGGAGTTACATAAAT ATG CAA TTA TTA AAA CAG CTT TTT
                                    MET Gln Leu Leu Lys Gln Leu Phe

AAG AAG AAA TTT GTA AAA GAA GAA CAC GAT AAG AAA ACC GGT CAA GAG GGT
Lys Lys Lys Phe Val Lys Glu Glu His Asp Lys Lys Thr Gly Gln Glu Gly

ATG ACA TTA CTC GAA GTG ATC ATC GTT CTA GGC ATT ATG GGG GTG GTT TCG
Met Thr Leu Leu Glu Val Ile Ile Val Leu Gly Ile Met Gly Val Val Ser

GCG GGG GTT GTT ACT CTG GCG CAG CGT GCG ATT GAT TCG CAG AAT ATG ACC
Ala Gly Val Val Thr Leu Ala Gln Arg Ala Ile Asp Ser Gln Asn Met Thr

AAG GCC GCG CAA AGT CTC AAT AGT ATC CAA GTT GCA CTG ACA CAG ACA TAC
Lys Ala Ala Gln Ser Leu Asn Ser Ile Gln Val Ala Leu Thr Gln Thr Tyr

CGT GGT CTA GGT AAT TAT CCA GCA ACA GCT GAT GCG ACA GCT GCT AGT AAG
Arg Gly Leu Gly Asn Tyr Pro Ala Thr Ala Asp Ala Thr Ala Ala Ser Lys

CTA ACT TCA GGC TTG GTT AGT TTA GGT AAA ATA TCA TCC GAT GAG GCA AAA
Leu Thr Ser Gly Leu Val Ser Leu Gly Lys Ile Ser Ser Asp Glu Ala Lys

AAC CCA TTC ATT GGT ACA AAT ATG AAT ATT TTT TCA TTT CCG CGT AAT GCA
Asn Pro Phe Ile Gly Thr Asn Met Asn Ile Phe Ser Phe Pro Arg Asn Ala

GCA GCT AAT AAA GCA TTT GCA ATT TCA GTG GAT GGT CTG ACA CAG GCT CAA
Ala Ala Asn Lys Ala Phe Ala Ile Ser Val Asp Gly Leu Thr Gln Ala Gln

TGC AAG ACA CTT ATT ACC AGT GTC GGT GAT ATG TTC CCA TAT ATT GCA ATC
Cys Lys Thr Leu Ile Thr Ser Val Gly Asp Met Phe Pro Tyr Ile Ala Ile

AAA GCT GGT GGC GCA GTA GCA CTT GCA GAT CTA GGT GAT TTT GAG AAT TCT
Lys Ala Gly Gly Ala Val Ala Leu Ala Asp Leu Gly Asp Phe Glu Asn Ser

GCA GCA GCG GCT GAG ACA GGC GTT GGT GTG ATC AAA TCT ATC GCT CCC GCT
Ala Ala Ala Ala Glu Thr Gly Val Gly Val Ile Lys Ser Ile Ala Pro Ala

AGT AAG AAT TTA GAT CTA ACG AAC ATC ACT CAC GTT GAG AAA TTA TGT AAA
Ser Lys Asn Leu Asp Leu Thr Asn Ile Thr His Val Glu Lys Leu Cys Lys

GGT ACT GCT CCA TTC GGC GTT GCA TTT GGT AAC AGC TAATTCAAATAAGTTTGTT
Gly Thr Ala Pro Phe Gly Val Ala Phe Gly Asn Ser

TAACTTAATCTTAACGTTGCCCATTAAATAATGGGCAACTTATTAAATTCAATGTGGTATCAATATG
```

FIG. 4. The *tcpA* gene and deduced protein sequence. The open reading frame encodes a protein of 224 residues that is processed to the 199-residue protein at the position indicated by the vertical arrow and previously determined by protein sequence of the amino end of the mature protein (48). A putative ribosome-binding site and an inverted repeat that may function in transcriptional termination or down regulation are underlined.

DNA sequence of *tcpA*. The amino-terminal protein sequence of TcpA was determined previously, and these studies revealed amino acid homology between TcpA and the structural subunits of several NMePhe pilins (48). To ascertain the extent of homology between these pilins throughout both the complete gene and the deduced protein sequences, as well as to characterize and define *tcpA* in more detail, overlapping fragments of DNA were subcloned into various M13 vectors and sequenced by the dideoxynucleotide method (37). Both strands of the DNA were sequenced according to the strategy presented in Fig. 1. In addition to the wild-type gene, two cloned *tcpA*::Tn*phoA* insertions, from strains RT104.11 and RT110.21 (48), were used to generate sequence and to confirm the open reading frame. The *tcpA* nucleotide sequence and deduced protein sequence are shown in Fig. 4.

The deduced amino acid sequence is in complete agreement with the previously reported 25 N-terminal residues of the mature pilin protein (48) beginning to the right of the vertical arrow in Fig. 4. In the previous protein sequence determination, the first residue could not be identified because its chromatographic properties indicated it to be a modified amino acid. The DNA sequence reveals a Met codon at this position, implying that the resultant protein contains a modified N-terminal methionine other than formyl-Met, which would have been blocked in the original sequence determination. The calculated size of the deduced

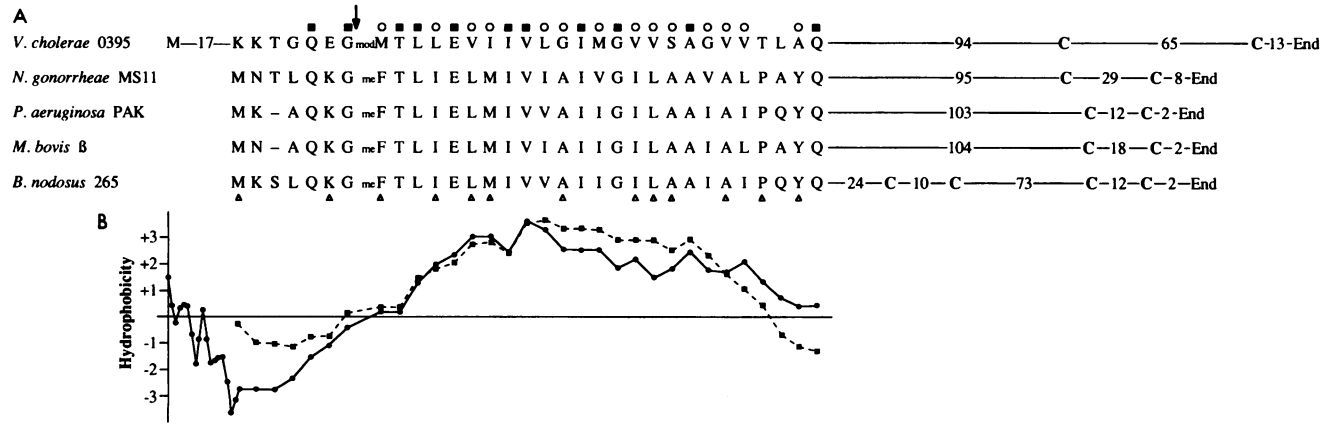

FIG. 5. (A) Comparison of the primary sequence of TcpA to the sequences of type 4 pilins from *Neisseria gonorrheae* MS11 (28), *Pseudomonas aeruginosa* PAK (38), *Moraxella bovis* B (24), and *Bacteroides nodosus* 265 (8). Symbols: ↓, processing site; ■, residues that are identical between TcpA and all of the type 4 pilins; ○, conservative substitutions; △, residues that are identical in all of the type 4 pilins but different from those in TcpA. (B) Kyte and Doolittle hydropathicity analysis (19) showing the similar profiles for TcpA (——) and the averaged values of the type 4 pilins (----). The pairs of Cys residues common to the carboxyl regions of the pilins are also indicated.

mature TcpA protein is 199 residues with a molecular weight of 20,330, in excellent agreement with the apparent molecular weight of 20,500 on SDS-PAGE.

The DNA sequence did not reveal any obvious ribosome-binding site (41) optimally positioned upstream of the coding sequence for mature pilin. In addition, since it is likely that the larger cross-reactive band expressed from plasmid pCS8C8 in *E. coli* represents a TcpA precursor form, the translational start site should be upstream of the codon for the N-terminal residue of the mature pilin in a position to encode a protein of approximately 2,500-Da-larger molecular size. In fact, the sequence reveals an in-frame Met codon with a putative ribosome-binding site 25 codons upstream of the N-terminal residue of TcpA (Fig. 4). Since the increased size of a precursor form of TcpA using this initiator Met would be very close to 2,500 Da, it is likely that translation of a precursor protein begins here, with subsequent processing to the mature 20.33-kDa form. There are no other in-frame Met codons upstream of this position before a nonsense codon is reached.

Predictive features of TcpA structure and comparison to type 4 pilin sequences. Figure 5 summarizes the major features of the protein sequence data derived from the *tcpA* DNA sequence and shows the comparison between TcpA and four NMePhe (type 4) pilin proteins. On the basis of analysis of the open reading frame, identified by the nucleotide sequence of *tcpA* and the molecular weight of the precursor pilin protein, it is likely that TcpA translation initiates 25 amino acids upstream of the N terminus of the native protein, yielding a cleaved leader peptide that is 18 to 19 residues longer than the six- to seven-amino acid leader that has so far been found for NMePhe pilins. As seen by Kyte-Doolittle analysis (19), the hydrophilic nature of the TCP leader peptide is remarkably similar to that of the shorter peptides characteristic of NMePhe pilin proteins. Amino acid homology between the leader peptides exists near the processing site, except for the substitution of a glutamate for lysine at the −2 position and a lysine residue in the TCP pilin sequence where the other pilins possess their initiating methionine. The N-terminal region of the mature TcpA pilin, like that of the NMePhe pilins, is rich in hydrophobic residues. The predicted hydrophobicity profile of this region of TcpA parallels that of the NMePhe pilins and contains a number of identical residues. The conservation of amino acid residues near the cleavage site as well as throughout the amino terminus of the mature protein suggests that the structure of this region is critical for a common function.

Beyond the hydrophobic N-terminal region of type 4 pilins, there is little amino acid homology with the exception of the common pair of cysteine residues toward the C terminus, which is also present in TcpA (Fig. 5) (7). Despite the lack of homology of primary sequences, similarity in overall structure has been reported on the basis of secondary-structure prediction analysis (4, 7). We used the Jameson-Wolf antigenic index program (17, 52) of the University of Wisconsin Genetics Computer Group sequence analysis package (5) to consider several parameters simultaneously in comparing predicted structural features among type 4 pilins and to TcpA. Except for the region between the two cysteine residues, all of the pilins, including TcpA, could be aligned in a common antigenic profile (Fig. 6) that followed a pattern similar to the secondary-structure alignment previously reported for type 4 pilins (7).

DISCUSSION

Two *tcpA*::Tn*phoA* insertions were cloned from strain O395 and used to determine the *tcpA* gene sequence and as probes to isolate the intact gene. The DNA sequence from the fusions and the cloned gene revealed an open reading frame that matched the previously determined TcpA N-terminal protein sequence. This reading frame was additionally confirmed by sequencing across the junctions between *tcpA* and the in-frame *phoA* gene of the Tn*phoA*-generated gene fusions. The open reading frame extended for 597 bases encoding 199 residues, which results in a protein with a predicted molecular weight of 20,330, in good agreement with the apparent molecular weight of 20,500 previously determined by SDS-PAGE for TcpA. The previous amino acid sequence placed an undetermined modified amino acid at the N terminus of the pilin. We originally thought that this would likely be an NMePhe residue, in keeping with the modified N-terminal residue so far characteristic of type 4 pilins. However, the *tcpA* DNA sequence places a methionine codon in this position, suggesting that the N-terminal TcpA residue is likely to be a modified methionine.

The open reading frame continues upstream of the N-ter-

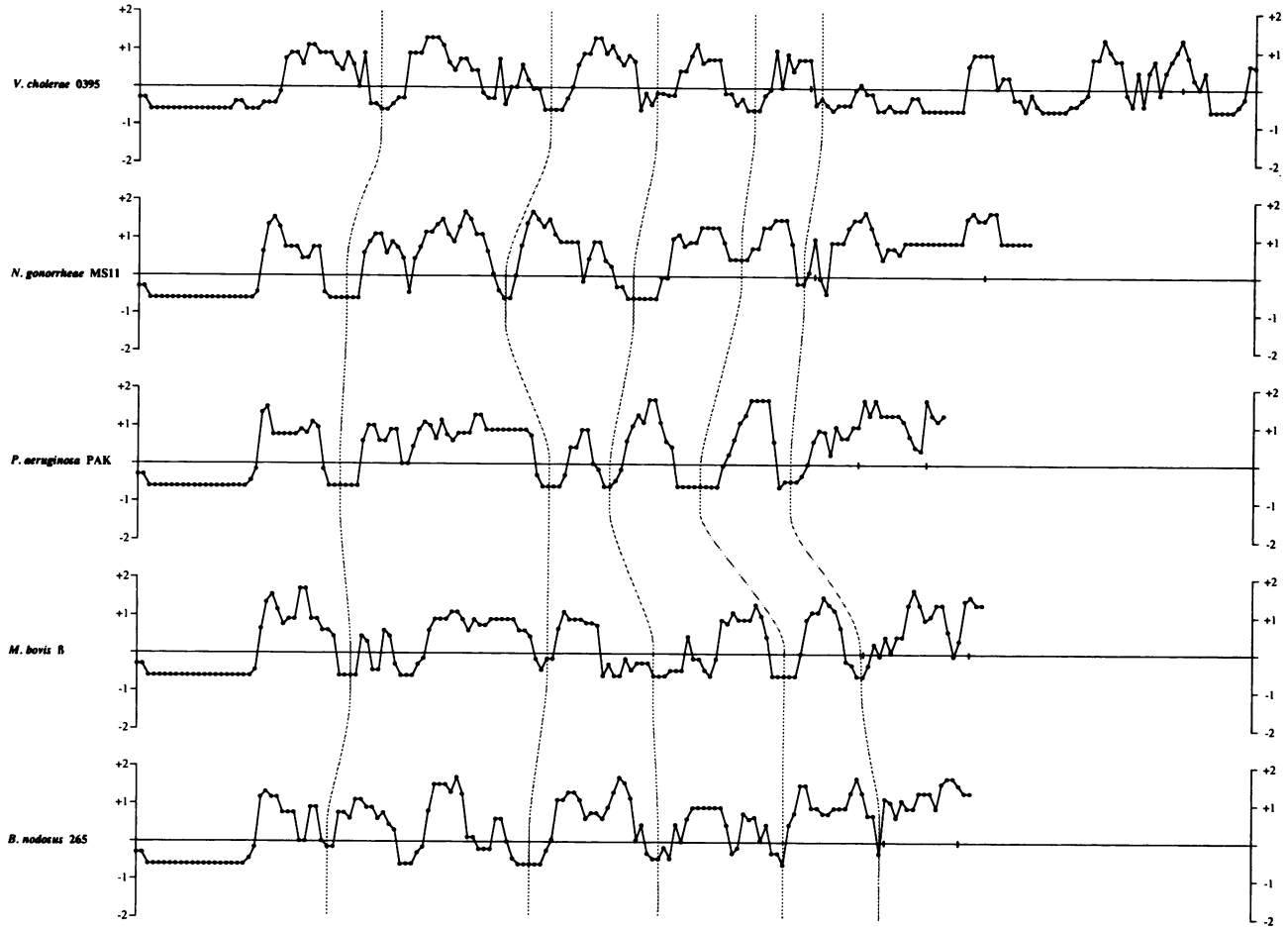

FIG. 6. Comparison of the Jameson-Wolf antigenic indices (17, 52) calculated for the pilins shown in Fig. 4. The vertical broken line indicates the similarly positioned predicted antigenic regions among the pilins.

minal codon of the mature protein for an additional 25 codons, where an ATG codon preceded by a likely ribosome-binding site is found. Translation initiated from this point would yield a precursor protein with a predicted molecular weight 2,500 larger than that of the mature pilin. Results from the expressed protein from various *tcpA* cosmid clones suggests that this is indeed the case. Of the three cosmid clones analyzed for TcpA expression by Western analysis, two expressed a cross-reactive protein of the mature size, but one (from cosmid pCS8C8) expressed a protein with an apparent molecular weight of 23,000, in good agreement with the predicted precursor size. Two results suggest that this larger protein is a precursor form of pilin and not the result of an aberrant event that occurred during cloning. When pCS8C8 was mobilized back into O395 *tcpA*::Tn*phoA* strain RT110.21, a product of the correct molecular mass was expressed. Also, when the *tcpA* gene was subcloned away from the adjacent DNA and expressed in *E. coli*, the resulting protein was of the larger molecular weight. This finding suggests that there is at least one adjacent gene that encodes a product required for processing TcpA from a precursor to mature form and, furthermore, that *E. coli* itself does not encode sufficient information for this event. This function is likely to be part of the overall biogenesis functions encoded by a cluster of adjacent genes, mutations that result in synthesis of a mature form of pilin that is not assembled into surface pili (31, 45). This gene cluster, including *tcpA* and pilus biogenesis genes, is located on the bacterial chromosome and not on the cryptic plasmid present in O395, as determined by the hybridization experiments using the 5-kb *tcpA*-containing *Xba*I restriction fragment as a probe of total and plasmid DNA.

Regions of the predicted primary sequence, especially near the N terminus, show a close relationship between TcpA and the NMePhe, or type 4, pilins (Fig. 5). The most striking feature of the type 4 pilins thus far characterized at the protein sequence level is the modified N-terminal residue of the mature pilin arising from a posttranslational N methylation of the Phe residue encoded at that position in the gene. The protein sequence of TcpA also reveals a modified N-terminal residue, which in this case is likely to be a modified methionine corresponding to a Met codon found in the +1 position that is occupied by a Phe in the NMePhe pilins. Another common feature of type 4 pilin sequences is the conserved hydrophobic domain at the very N terminus of the mature pilin which shares a number of identical amino acids between pilins. This region of the protein is likely to have a dual role, being required both for the secretion of the pilin, as demonstrated for PAK pilin (44), and for mediating subunit-subunit interactions within the pilus structure (Taylor et al., in press). TcpA has a similarly positioned hydrophobic stretch of amino acids, retaining some of the identical residues but showing many more conservative changes than

among the previously characterized members of the pilin group.

Previous studies comparing the sequences of type 4 pilins between species or serotypes have not revealed any striking homologies beyond the hydrophobic domain and the occurrence of one or two pairs of cysteine residues near the C terminus. At the level of predicted secondary structures, a similar pattern is seen between all of the type 4 pilins. We ran a similar comparison, using antigenic structure predictions and including TcpA in the analysis. Again, an overall similarity of patterns is seen which can be extended to TcpA (Fig. 6). These pattern similarities, while only predictive, suggest common domains among these pilins that are likely to have analogous roles in pilus biogenesis and function.

The sequence and predicted structural similarities between TcpA and the type 4 pilins thus far characterized raises the question as to whether TcpA should be considered a member of this pilin class. The type 4 class of pilins was first based on morphological and functional considerations, defined as thin (5 to 7 nm), wavy filaments generally of polar distribution that were likely to have a role in bacterial adhesion (31), similar to TCP. Further characterization showed that some of these pili had a retractile property associated with phage adsorption that also conferred upon the cell the phenomenon of twitching motility (12), properties not yet determined for TCP. When subsequent sequence analysis revealed the NMePhe terminal residue on the first type 4 pilins to be characterized, the nomenclature became more specific, classifying them as the NMePhe pili on the basis of this shared modification of the major subunit. Mature TcpA contains a modified residue, albeit a different one, at the same position in the mature protein and shares extensive homology near the processing site even though it has a longer leader peptide. Another feature that these pili have in common, when it has been analyzed, is that at least part of the colonization function resides within the major subunit. This has been postulated for gonococcal pilin (35, 39), *Pseudomonas* pilin (16), and, most recently, TCP (Taylor and Sun, in press; Sun and Taylor, submitted). Thus, our analysis of TcpA pilin indicates that it is at least highly related to the type 4 pilins.

To incorporate our observations into the current literature regarding type 4 (NMePhe) pilins, we propose that the definition of type 4 pilins be extended to include TcpA. Within this type, subtypes would be formed according to amino acid modification to distinguish NMePhe-containing pili from modified methionine-containing pili (TCP), as well as any additional modifications to be discovered. This proposed scheme defines the type 4 classification in a manner closer to its original conception based on overall morphology and probable mechanism of function, rather than being limited or based solely on a single modified residue present in a position that may be redundant among related members.

ACKNOWLEDGMENTS

We thank L. Hatmaker for electron microscopy, B. Blalock for the Southern hybridization analysis presented in Fig. 2, D. Schifferli for many helpful discussions and critical reading of the manuscript, and D. Sun for suggestions regarding the manuscript. We also acknowledge the insightful discussions with the late E. H. Beachey during the course of these studies.

This work was supported by Public Health Service grant AI25096 to R.K.T. from the National Institutes of Health. Sequence analysis was performed by using the University of Wisconsin Genetics Computer Group programs supported through the Molecular Resource Center at the University of Tennessee, Memphis.

ADDENDUM

The *tcpA* sequence presented here has two differences at the nucleotide level from that recently reported for *tcpA* isolated from another classical biotype strain (9). Neither change alters the corresponding residue in the pilin, suggesting a strong conservation of the primary structure between strains within the same biotype.

LITERATURE CITED

1. **Abraham, S. N., J. D. Goguen, D. Sun, P. Klemm, and E. H. Beachey.** 1987. Identification of two ancillary subunits of *Escherichia coli* type 1 fimbriae using antibodies against synthetic oligopeptides of *fim* gene products. J. Bacteriol. **169**:5530–5536.
2. **Casadaban, M. J., A. Martinez-Arias, S. K. Shapira, and J. Chou.** 1983. β-Galactosidase gene fusions for analyzing gene expression in *Escherichia coli* and yeast. Methods Enzymol. **100**:293–308.
3. **Cook, W. L., K. Wachsmuth, S. R. Johnson, K. A. Birkness, and A. R. Samadi.** 1984. Persistence of plasmids, cholera toxin genes, and prophage DNA in classical *Vibrio cholerae* O1. Infect. Immun. **4**:222–226.
4. **Dalrymple, B., and J. S. Mattick.** 1987. An analysis of the organization and evolution of type 4 fimbrial (MePhe) subunit proteins. J. Mol. Evol. **25**:261–269.
5. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
6. **Ehara, M., M. Ishibashi, Y. Ichinose, M. Iwanaga, S. Shimodori, and T. Naito.** 1987. Purification and partial characterization of *Vibrio cholerae* O1 fimbriae. Vaccine **5**:283–288.
7. **Elleman, T. C.** 1988. Pilins of *Bacteroides nodosus*: molecular basis of serotypic variation and relationships to other bacterial pilins. Microbiol. Rev. **52**:233–247.
8. **Elleman, T. C., P. A. Hoyne, N. M. McKern, and D. J. Stewart.** 1986. Nucleotide sequence of the gene encoding the two subunit pilin of *Bacteroides nodosus* 265. J. Bacteriol. **167**:243–250.
9. **Faast, R., M. A. Ogierman, U. H. Stroeher, and P. A. Manning.** 1989. Nucleotide sequence of the structural gene, *tcpA*, for a major pilin subunit of *Vibrio cholerae*. Gene **85**:227–331.
10. **Finlay, B. B., and S. Falkow.** 1989. Common themes in microbial pathogenicity. Microbiol. Rev. **53**:210–230.
11. **Hall, R. H., P. A. Vial, J. B. Kaper, J. J. Mekalanos, and M. M. Levine.** 1988. Morphological studies on fimbriae expressed by *Vibrio cholerae* O1. Microb. Pathog. **4**:257–265.
12. **Henrichsen, J.** 1983. Twitching motility. Annu. Rev. Microbiol. **37**:81–93.
13. **Herrington, D. A., R. H. Hall, G. Losonsky, J. J. Mekalanos, R. K. Taylor, and M. M. Levine.** 1988. Toxin, toxin-coregulated pili and the *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. J. Exp. Med. **168**:1487–1492.
14. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11**:291–298.
15. **Irvin, R. T., P. Doig, K. K. Lee, P. A. Sastry, W. Paranchych, T. Todd, and R. S. Hodges.** 1989. Characterization of the *Pseudomonas aeruginosa* pilus adhesin: confirmation that the pilin structural protein subunit contains a human epithelial cell-binding domain. Infect. Immun. **57**:3720–3726.
16. **Jacobs, A. C., B. H. Simons, and F. K. DeGraaf.** 1987. The role of lysine-132 and arginine-136 in the receptor-binding domain of the K99 fibrillar subunit. EMBO J. **6**:1805–1808.
17. **Jameson, B. A., and H. Wolf.** 1988. The antigenic index—a novel algorithm for predicting antigenic determinants. CABIOS **4**:181–186.
18. **Kieny, M. P., R. Lathe, and J. P. Lecocq.** 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. Gene **26**:91–99.
19. **Kyte, J., and R. F. Doolittle.** 1982. A method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
20. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.

21. **Lindberg, F., B. Lund, L. Johansson, and S. Normark.** 1987. Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. Nature (London) **328**:84–87.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Manoil, C., and J. Beckwith.** 1985. Tn*phoA*: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA **82**:8129–8133.
24. **Marrs, C. F., G. Schoolnik, J. M. Koomey, J. Hardy, J. Rothbard, and S. Falkow.** 1985. Cloning and sequencing of a *Moraxella bovis* pilin gene. J. Bacteriol. **163**:132–139.
25. **Mattick, J. S., M. M. Bills, B. J. Anderson, B. Dalrymple, M. R. Mott, and J. R. Egerton.** 1987. Morphogenetic expression of *Bacteroides nodosus* fimbriae in *Pseudomonas aeruginosa*. J. Bacteriol. **169**:33–41.
26. **Mekalanos, J. J., D. J. Swartz, D. N. Pearson, N. Harford, G. Francoise, and M. de Wilde.** 1983. Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development. Nature (London) **306**:551–557.
27. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
28. **Meyer, T. F., E. Billyard, R. Haas, S. Storzbach, and M. So.** 1984. Pilus genes of *Neisseria gonorrheae*: chromosomal organization and DNA sequence. Proc. Natl. Acad. Sci. USA **81**:6110–6114.
29. **Miller, V. L., and J. J. Mekalanos.** 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J. Bacteriol. **170**:2575–2583.
30. **Miller, V. L., R. K. Taylor, and J. J. Mekalanos.** 1987. Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. Cell **48**:271–279.
31. **Ottow, J. C. G.** 1975. Ecology, physiology, and genetics of fimbriae and pili. Annu. Rev. Microbiol. **29**:79–108.
32. **Paranchych, W., and L. S. Frost.** 1988. The physiology and biochemistry of pili. Adv. Microb. Physiol. **29**:53–114.
33. **Pedersen, K. B., L. O. Froholm, and K. Bovre.** 1972. Fimbriation and colony type of *Moraxella bovis* in relation to conjunctival colonization and development of keratoconjunctivitis in cattle. Acta Pathol. Microbiol. Scand. Sect. B **80**:911–918.
34. **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.
35. **Rothbard, J. B., R. Fernandez, L. Wang, N. N. H. Teng, and G. K. Schoolnik.** 1985. Antibodies to peptides corresponding to a conserved sequence of gonococcal pilins block bacterial adhesion. Proc. Natl. Acad. Sci. USA **82**:915–919.
36. **Ruehl, W. W., C. F. Marrs, R. Fernandez, S. Falkow, and G. K. Schoolnik.** 1988. Purification, characterization, and pathogenicity of *Moraxella bovis* pili. J. Exp. Med. **168**:983–1002.
37. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
38. **Sastry, P. A., J. R. Pearlstone, L. B. Smillie, and W. Paranchych.** 1983. Amino acid sequence of pilin isolated from *Pseudomonas aeruginosa* PAK. FEBS Lett. **151**:253–256.
39. **Schoolnik, G. K., R. Fernandez, J. Y. Tai, J. Rothbard, and E. C. Gotschlich.** 1984. Gonococcal pili: primary structure and receptor binding domain. J. Exp. Med. **159**:1351–1370.
40. **Seifert, H. S., M. So, and F. Heffron.** 1986. Shuttle mutagenesis: a method of introducing transposons into transformable organisms, p. 123–134. *In* J. K. Setlow and A. Hollaender (ed.), Genetic engineering, principles and methods, vol. 8. Plenum Publishing Corp., New York.
41. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
42. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Southern, E. M.** 1975. Detection of sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
44. **Strom, M. A., and S. Lory.** 1987. Mapping of export signals of *Pseudomonas aeruginosa* pilin with alkaline phosphatase fusions. J. Bacteriol. **169**:3181–3188.
45. **Swanson, J., and J. M. Koomey.** 1989. Mechanisms for variation of pili and outer membrane protein II in *Neisseria gonorrhoeae*, p. 743–761. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
46. **Sun, D., J. J. Mekalanos, and R. K. Taylor.** 1990. Antibodies directed against the toxin-coregulated pilus from *Vibrio cholerae* provide protection in the infant mouse experimental cholera model. J. Infect. Dis. **161**:1231–1236.
47. **Taylor, R. K., C. Manoil, and J. J. Mekalanos.** 1989. Broad-host-range vectors for delivery of Tn*phoA*: use in genetic analysis of secreted virulence determinants of *Vibrio cholerae*. J. Bacteriol. **171**:1870–1878.
48. **Taylor, R. K., V. L. Miller, D. B. Furlong, and J. J. Mekalanos.** 1987. The use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc. Natl. Acad. Sci. USA **84**:2833–2837.
49. **Taylor, R., C. Shaw, K. Peterson, P. Spears, and J. Mekalanos.** 1988. Safe, live Vibrio cholerae vaccines? Vaccine **6**:151–154.
50. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
51. **Tweedy, J. M., R. W. A. Park, and W. Hodgkiss.** 1968. Evidence for the presence of fimbriae (pili) on *Vibrio* species. J. Gen. Microbiol. **51**:235–244.
52. **Wolf, H., S. Modrow, M. Motz, B. A. Jameson, G. Hermann, and B. Fortsch.** 1988. An integrated family of amino-acid sequence-analysis programs. CABIOS **4**:187–191.
53. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.