

Identification of a Glycoprotein Produced by Enterotoxigenic *Escherichia coli*

CHRISTOPH LINDENTHAL AND ERIC A. ELSINGHORST*

Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66045-2106

Received 8 March 1999/Returned for modification 28 April 1999/Accepted 26 May 1999

Enterotoxigenic *Escherichia coli* (ETEC) strain H10407 is capable of invading epithelial cell lines derived from the human ileocecum and colon in vitro. Two separate chromosomally encoded invasion loci (*tia* and *tib*) have been cloned from this strain. These loci direct nonadherent and noninvasive laboratory strains of *E. coli* to adhere to and invade cultured human intestinal epithelial cells. The *tib* locus directs the synthesis of TibA, a 104-kDa outer membrane protein that is directly correlated with the adherence and invasion phenotypes. TibA is synthesized as a 100-kDa precursor (preTibA) that must be modified for biological activity. Outer membranes of recombinant *E. coli* expressing TibA or preTibA were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and blotted to nitrocellulose. The presence of glycoproteins was detected by oxidation of carbohydrates with periodate and labeling with hydrazide-conjugated digoxigenin. Only TibA could be detected as a glycoprotein. Complementation experiments with *tib* deletion mutants of ETEC strain H10407 demonstrate that the TibA glycoprotein is expressed in H10407, that the entire *tib* locus is required for TibA synthesis, and that TibA is the only glycoprotein produced by H10407. Protease treatment of intact H10407 cells removes the carbohydrates on TibA, suggesting that they are surface exposed. TibA shows homology with AIDA-I from diffuse-adhering *E. coli* and with pertactin precursor from *Bordetella pertussis*. Both pertactin and AIDA-I are members of the autotransporter family of outer membrane proteins and are afimbrial adhesins that play an important role in the virulence of these organisms. Analysis of the predicted TibA amino acid sequence indicates that TibA is also an autotransporter. Analysis of the *tib* locus DNA sequence revealed an open reading frame with similarity to RfaQ, a glycosyltransferase. The product of this *tib* locus open reading frame is proposed to be responsible for TibA modification. These results suggest that TibA glycoprotein acts as an adhesin that may participate in the disease process.

Diarrheal disease was responsible for the death of 2.5 million people in 1996, making it the third leading cause of death by infectious diseases worldwide (47). Most cases of diarrhea occur in developing countries, where an estimated 700,000 children under 5 years of age die each year as a result of diarrhea caused by infection with enterotoxigenic *Escherichia coli* (ETEC) (24). ETEC is also responsible for most cases of travelers' diarrhea, the most common disease among travelers from industrialized countries (3).

ETEC infection is acquired by the ingestion of contaminated food or drink. Bacteria initially colonize the proximal small intestine and produce at least one member of two defined groups of enterotoxins: heat-stable toxin (ST) and heat-labile toxin (LT) (28, 36). Initial colonization of the intestinal mucosa is a critical step in the disease process and is mediated by fimbrial colonization factor antigens (CFAs) (18). The action of enterotoxins on the intestine is well understood and is believed to be the major cause of water loss associated with ETEC infection (28). However, human and animal studies performed with ETEC strains that have lost the ability to produce ST or LT enterotoxins indicate that enterotoxins may not be exclusively required for diarrhea (27). This result suggests the presence of previously uncharacterized enterotoxins or other virulence factors (28, 45).

Although there is currently no direct evidence that ETEC strains invade human intestinal cells in vivo, intestinal biopsies taken from ETEC-infected piglets revealed intracellular bac-

teria (30). These findings agree with data showing that ETEC strain H10407 is capable of invading epithelial cell lines derived from the human ileocecum and colon (14). Two separate chromosomally encoded invasion loci (*tia* and *tib*) have been cloned from the classical ETEC strain H10407 (14). These loci direct nonadherent and noninvasive laboratory strains of *E. coli* to adhere to and invade cultured human intestinal epithelial cells. ETEC strains comprising a variety of serotypes, CFA types, and enterotoxin phenotypes were screened for *tia* and *tib* sequences by hybridization (15, 20). Of these strains, 63% hybridized with *tia* or *tib* probes, 41% hybridized with a *tia* probe, and 30% hybridized with a *tib* probe. Interestingly, all *tib*-hybridizing strains expressed CFA type CFA/I. While the genes coding for CFA/I production are located on a plasmid, the *tib* locus is chromosomally encoded (14, 18).

The *tib* locus directs the synthesis of TibA, a 104-kDa outer membrane protein (15). Subcloning and transposon mutagenesis experiments have shown that the adherence and invasion phenotypes of the *tib* locus are directly correlated with the presence of TibA in the outer membranes of *tib*-expressing recombinants. The TibA protein is the product of the *tibA* gene. Plasmids containing the *tibA* gene under the transcriptional control of an exogenous promoter direct the synthesis of a 100-kDa outer membrane TibA precursor (15), which we refer to as preTibA. The preTibA protein does not direct epithelial cell adherence and invasion by recombinant laboratory strains of *E. coli*. When a separate compatible plasmid containing the *tib* locus sequence upstream of the *tibA* gene is introduced into these strains, TibA production is restored (15). This result suggests that preTibA is modified to form TibA. Here we report that this modification is glycosylation. We also demonstrate that the TibA glycoprotein is produced by the

* Corresponding author. Mailing address: University of Kansas, Department of Molecular Biosciences, 7049 Haworth Hall, Lawrence, KS 66045-2106. Phone: (785) 864-4299. Fax: (785) 864-5294. E-mail: elsingh@ukans.edu.

TABLE 1. Plasmids used in this study

Plasmid	Relevant description	Reference
pACYC184	Cloning vector, p15A replication origin	7
pET109	<i>tib</i> locus cloned into pHG165	15
pET139	<i>tibA</i> gene (in pHG165) under transcriptional control of an endogenous promoter	15
pET140	<i>tibA</i> gene under transcriptional control of the <i>lac</i> promoter found in pHG165	15
pET146	<i>tibB</i> , <i>tibC</i> , and <i>tibD</i> ORFs cloned into pACYC184	15
pHC79	Cosmid cloning vector	23
pHG165	pBR322 copy number derivative of pUC8; ColE1 replication origin	38

wild-type ETEC parent strain, H10407, where it is present in the outer membrane, and that the carbohydrate residues are surface exposed. Additionally, TibA appears to belong to the autotransporter family of outer membrane proteins and shows homology with known bacterial adhesins.

MATERIALS AND METHODS

Bacterial strains and plasmids. ETEC H10407 (18) (serotype 078:H11; CFA/I) was the parent strain from which the *tib* locus was cloned. TIB3 is a *tib* deletion mutant of H10407 (15). *E. coli* HB101 (4) (*hdsS20 recA13 rpsL20*) was used as a noninvasive recipient of *tib* locus-containing plasmids. Organisms were grown in Luria broth (10 g of tryptone, 5 g of yeast extract, 5 g of NaCl [pH 7.6]) at 37°C and 200 rpm. Where indicated, antibiotics were added to growth media to the following final concentrations: ampicillin, 100 µg/ml; streptomycin, 100 µg/ml; and chloramphenicol, 20 µg/ml. The plasmids used are listed in Table 1.

Membrane fractionation. Outer membrane fractions were isolated as previously described (15, 37). Luria broth cultures (500 ml each) were grown with shaking at 37°C to late log phase, harvested by centrifugation, and then lysed by two passages through a French press. Inner and outer membranes were separated by sucrose density ultracentrifugation.

Electrophoresis of proteins. Electrophoresis of whole-cell extracts or membrane fractions was performed under denaturing conditions by the method of Laemmli (26). Samples were prepared for electrophoresis by being heated in treatment buffer at 95°C for 10 min. Gels were run for 12 to 14 h at 70 V at room temperature. Gels were stained for proteins with Coomassie blue. The protein concentration of membrane fractions was determined by the Bradford method (5) with a kit from Bio-Rad. For the analysis of outer membranes, 10 µg of total protein was loaded per well. For the analysis of whole-cell extracts, 2-ml Luria broth cultures were grown to late log phase by shaking at 37°C, and then 1 ml was harvested by centrifugation and lysed in 200 µl of treatment buffer. Ten-microliter aliquots were loaded on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels. For glycoprotein staining, freshly prepared cellular fractions were treated for electrophoresis as described above. After electrophoresis, separated proteins were transferred to nitrocellulose filters at 60 V for 60 min in standard buffer (42), followed by Ponceau S staining. Filters were blocked with Tris-buffered saline solution (10 mM Tris hydrochloride [pH 7.5], 150 mM NaCl) containing 2% (wt/vol) casein.

Detection of glycoproteins. Glycoproteins were detected on nitrocellulose filters by using method B of the Roche Molecular Biochemicals digoxigenin (DIG) glycan detection kit according to the manufacturer's recommendations. Briefly, proteins were separated by SDS-PAGE and blotted to nitrocellulose as described above. Subsequently, membranes were washed in phosphate-buffered saline (50 mM potassium phosphate, 150 mM NaCl [pH 6.5]), and carbohydrates were oxidized with sodium metaperiodate. Oxidized carbohydrates were labeled with DIG-conjugated hydrazide, and nitrocellulose membranes were stained with Ponceau S to confirm that equal amounts of proteins had been loaded in each well and that electroblotting had occurred evenly. DIG-labeled proteins were visualized with alkaline phosphatase-conjugated anti-DIG antibodies followed by a color reaction with nitroblue tetrazolium and X-phosphate.

Complementation analysis. The *tib* locus deletion mutation found in TIB3 was complemented by transformation with plasmids pET109, pET139, or pET139 and pET146 by electroporation. For electroporation, late-log-phase-grown bacteria were diluted 1:100 in Luria broth and grown to an optical density at 600 nm of 0.5 with shaking at 37°C. Cultures were chilled and then harvested by centrifugation at 7,700 × g for 10 min. Cells were washed twice and resuspended in cold 10% glycerol in 1/100 of the starting volume. Ten micrograms of gel-purified plasmid DNA in a total volume of 10 µl was added to 100 µl of prepared bacteria and incubated for 1 min on ice. Electroporation was performed in prechilled 0.1-cm Gene Pulser/*E. coli* Pulser cuvettes (Bio-Rad) at 2.5 kV, 25 µF, 600 Ω. After addition of 0.9 ml of SOC medium (Bacto-tryptone, 20 g/liter; Bacto-yeast extract, 5 g/liter; NaCl, 0.5 g/liter; KCl, 2.5 mM; MgCl₂, 10 mM; glucose, 20 mM [pH 7.0]), bacteria were incubated for 60 min with shaking at 37°C and plated on tryptic soy agar plates with appropriate antibiotics. Clones were screened for the presence of the appropriate plasmids by alkaline lysis miniprep and the production of glycoproteins.

DIG labeling of surface carbohydrates and proteolytic digestion of intact bacterial cells. Glycoproteins on the surface of intact bacterial cells were labeled with DIG as described in method A of the DIG glycan detection kit (Roche Molecular Biochemicals), with some modifications. Briefly, TIB3(pET109) was grown in 2 ml of Luria broth to late log phase with shaking at 37°C. The pellet was washed three times with 10 mM HEPES (pH 7.4) containing 100 µg of chloramphenicol per ml (HEPES/chloramphenicol). The pellet was resuspended in 400 µl of 0.1 M sodium acetate (pH 5.5), and sodium metaperiodate was added to a final concentration of 0.78 mM. The mixture was incubated for 20 min at room temperature in the dark. After incubation, 0.39 mg of sodium disulfite per ml was added, and this mixture was incubated for 5 min at room temperature to destroy any excess periodate. The pellet was washed twice with 10 mM HEPES/chloramphenicol and resuspended in 400 µl of 0.1 M sodium acetate, and DIG-hydrazide was added. After 1 h, the pellet was washed twice in HEPES/chloramphenicol and resuspended in 0.1 M Tris (pH 8) plus 100 µg of chloramphenicol per ml. Resuspended bacteria were divided into 200-µl aliquots for protease treatment. Each tube was incubated with different concentrations of trypsin or proteinase K. Four tubes containing bacteria were incubated with each protease at 0, 10, 50, or 100 µg/ml for 30 min at 37°C. The proteolytic treatment was stopped by the addition of 100 µg of soybean trypsin inhibitor per ml or 1 mM phenylmethylsulfonyl fluoride to all four tubes. Periplasmic proteins of protease-treated cells, as well as those from nontreated controls, were released by osmotic shock as described previously (31) by incubation in 20% sucrose followed by cold water shock and centrifugation to pellet the bacteria. Supernatants containing osmotic shock periplasmic proteins were analyzed by SDS-PAGE (10% polyacrylamide) and stained with Coomassie blue. Bacterial pellets were resuspended in 200 µl of treatment buffer and analyzed for the presence of glycoproteins as described above.

DNA sequence analysis. Plasmid pET109 was used as the template for DNA sequencing of the *tibA* and *tibC* genes. Both strands of the *tibA* and *tibC* genes in pET109 were sequenced by the Sanger dideoxy chain-termination procedure (35) with thermal cycle-sequencing protocols and fluorescently labeled dideoxynucleotides (Thermo Sequenase dye terminator cycle sequencing kit; Amersham Life Science). Sequencing reactions were analyzed with an ABI Prism 310 automated DNA sequencer. The resulting DNA sequences were analyzed by using MacVector sequence analysis software (Oxford Molecular Group) to assemble contigs and predict sequencing primers. Oligonucleotide primers were synthesized by Operon Technologies.

Amino-terminal sequencing of proteins. Proteins contained in outer membrane fractions purified from *E. coli* HB101 containing either pET109 (TibA expressing) or pET140 (preTibA expressing) were separated by SDS-PAGE (7.5% polyacrylamide separation gel). Proteins corresponding to TibA or preTibA were excised from the gel and eluted passively. Eluted proteins were sequenced by the Edman degradation method (13) with a Hewlett-Packard G1000-A protein sequencer.

Amino acid analysis. Computer analyses were performed by using BLAST software for homology search and MacVector software (Oxford Molecular Group) for surface exposure and secondary structure prediction. Alignments were performed with Clustalx software, and locations of amphipathic β-sheets and transmembrane domains were determined with TopPred II 1.3 software (10).

Nucleotide sequence accession number. The nucleotide sequences of the *tibA* and *tibC* genes have been submitted to GenBank under accession no. AF109215 and AF131891, respectively.

RESULTS

The TibA protein is glycosylated. The 104-kDa TibA protein is synthesized as a 100-kDa precursor (15), which we refer to as preTibA. To determine if glycosylation is involved in the synthesis of TibA, we purified outer membranes from *E. coli* HB101 bearing one of three different plasmids (Fig. 1A): pET109, containing the entire *tib* locus; pET140, a plasmid with the *tibA* gene under the transcriptional control of the *lac* promoter; and pHC79, the cosmid vector used for the initial

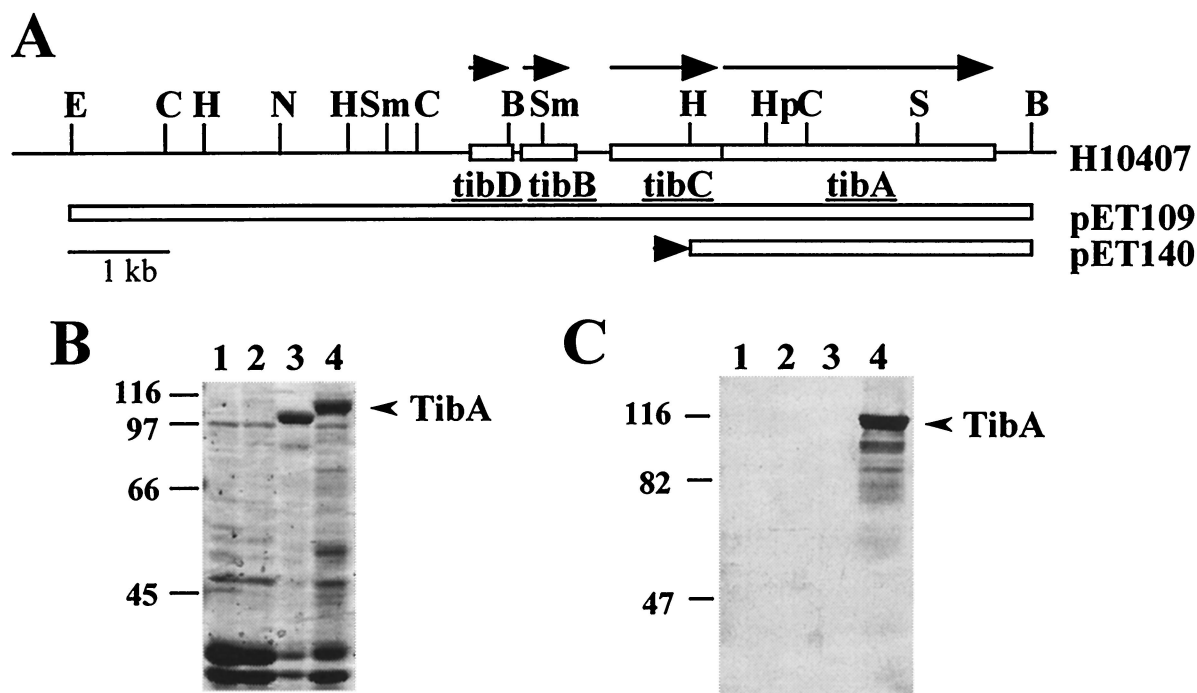


FIG. 1. Detection of glycoproteins in outer membranes of recombinant *E. coli* HB101. (A) Restriction endonuclease map of the *tib* locus as found in the H10407 genome or in the indicated plasmids. The direction of *tib* gene transcription is indicated by arrows above the H10407 map. The extent of the *tib* locus contained by each plasmid is indicated by an open box. The black arrowhead to the left of the pET140 map indicates the direction of transcription from an exogenous promoter found in the plasmid vector. B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hind*III; Hp, *Hpa*I; N, *Nru*I; S, *Sal*I; Sm, *Sma*I. (B) Coomassie blue-stained SDS-PAGE (7.5% polyacrylamide) of outer membranes purified from the following strains (by lane): 1, HB101; 2, HB101(pHC79); 3, HB101(pET140); 4, HB101(pET109). Plasmid pET140 expresses the 100-kDa preTibA protein, whereas pET109 expresses the 104-kDa TibA protein. (C) Samples identical to those shown in panel B were transferred to nitrocellulose and then stained for glycoprotein as described in Materials and Methods. The migration of molecular mass standards is shown to the left of panels B and C. The mobility of TibA is shown by an arrow to the right of panels B and C.

cloning of the *tib* locus. Proteins were separated by SDS-PAGE and then transferred to nitrocellulose. Carbohydrates were detected by periodate oxidation of proteins immobilized on nitrocellulose followed by labeling with DIG-conjugated hydrazide. DIG-labeled proteins were visualized by alkaline phosphatase-conjugated anti-DIG antibodies (Roche Molecular Biochemicals). Only outer membranes containing TibA stained positive for glycoprotein (Fig. 1C). To verify that this result was not an artifact of the system, experiments were performed in which steps of the labeling and detection reactions were omitted. Under these conditions, no glycoproteins were detected (data not shown). TibA stains as a glycoprotein even when purified TibA is treated with strong denaturing agents, such as 6 M guanidine hydrochloride (data not shown), indicating that the carbohydrates are covalently attached. Glycoproteins of lower molecular mass observed in lane 4 of Fig. 1C are most likely breakdown products of mature TibA, because they also can be found when purified TibA is analyzed by SDS-PAGE (data not shown) and are absent in a *tib* deletion mutant of ETEC strain H10407 (see below).

The *tib* locus is required for the synthesis of TibA by H10407. The *tib* locus was originally cloned from H10407, a wild-type ETEC strain. To confirm that H10407 produces glycoproteins and to link them to the *tib* locus, whole-cell lysates of H10407, TIB3 (a *tib* deletion mutant of H10407), and TIB3(pET109) were analyzed for the presence of glycoproteins (Fig. 2B). A 104-kDa glycoprotein is present in H10407 (lane 1) and TIB3(pET109) (lane 3), but not in TIB3 (lane 2). This result shows that TibA glycosylation is not an artifact of *tib* locus expression in HB101 and that the *tib* locus is directly

correlated with the presence of a glycosylated protein in H10407.

The deletion in the TIB3 mutant removes 70% of the *tibA* gene but also removes three other open reading frames (ORFs) found within the *tib* locus (Fig. 2A). To confirm that the 104-kDa glycoprotein in H10407 is actually TibA, complementation studies were performed. TIB3 was transformed with pET146, pET139, or both plasmids. Plasmid pET146 contains the three ORFs upstream of *tibA*, but not the *tibA* gene itself. Plasmid pET139 contains the *tibA* gene under transcriptional control of an endogenous promoter. Whole-cell extracts were prepared from these strains and analyzed for glycoprotein content. The 104-kDa glycoprotein appears only when all *tib* locus ORFs are present. Furthermore, the 104-kDa glycoprotein is not produced in the strain lacking only the *tibA* gene [TIB3(pET146)] (Fig. 2B, lane 4). These results show that the 104-kDa glycoprotein seen in H10407 is indeed TibA and that TibA is the only glycoprotein expressed by H10407 grown under these conditions. When complementation is performed with plasmid pET109, TibA expression is much higher than in wild-type ETEC (Fig. 2B, lane 3). This result is probably due to the copy number of pET109. In complementation with pET139 and pET146, the observed TibA expression drops to almost wild-type levels (Fig. 2B, lanes 5 and 6). This decrease is likely due to a combination of factors. First, maintenance of multiple plasmids is likely to decrease the copy number of these plasmids. Second, complementation and DNA sequence data suggest that *tibA* transcription is directed by two promoters, only one of which is present in plasmid pET139 (data not shown).

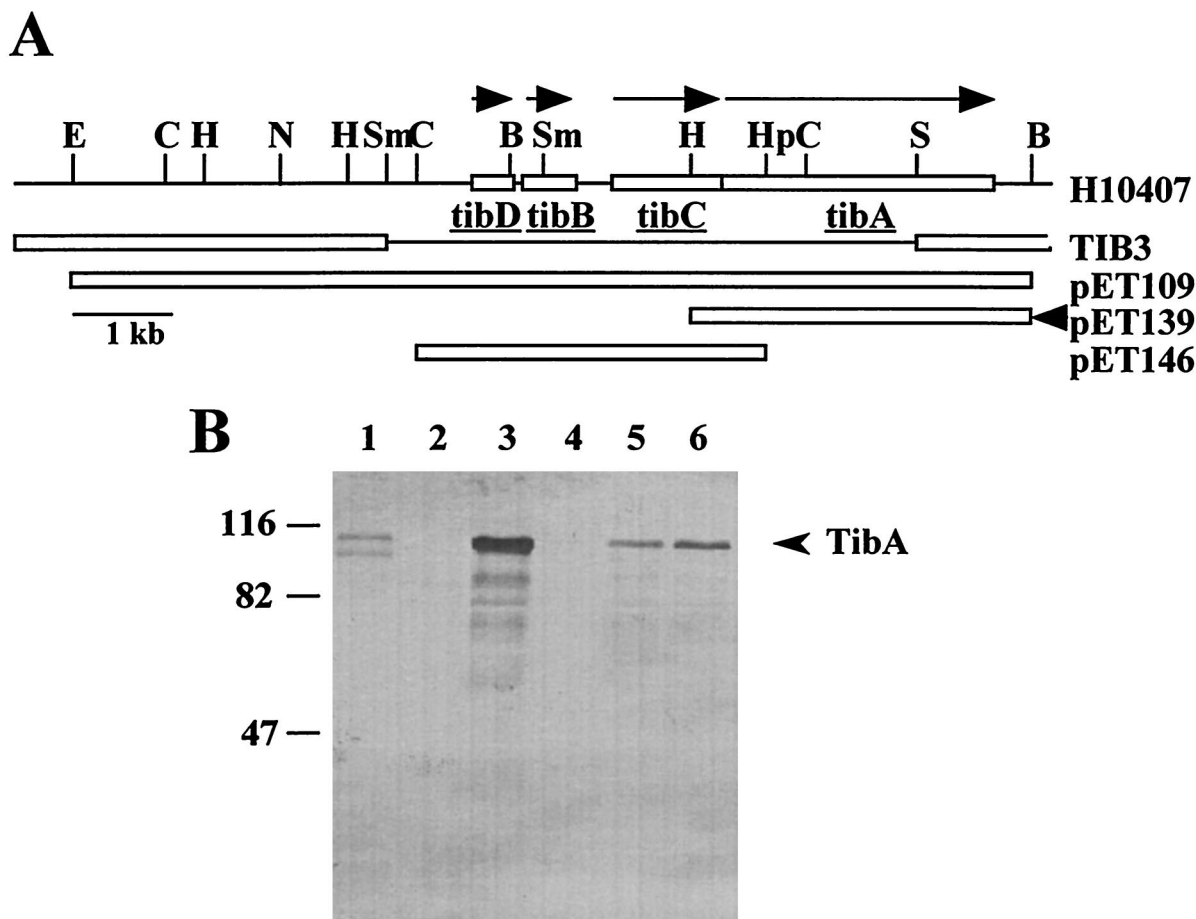


FIG. 2. Complementation analysis of an H10407 *tib* deletion mutant. (A) Restriction endonuclease map of the *tib* locus as found in H10407, TIB3 (an H10407 *tib* deletion mutant), or the indicated plasmids. The direction of *tib* gene transcription is indicated by arrows above the H10407 map. The *tib* locus sequences deleted in TIB3 are indicated by a thin black line in the TIB3 map. The extent of the *tib* locus contained by each plasmid is indicated by an open box. The black arrowhead to the right of the pET139 map indicates the direction of transcription from an exogenous promoter found in the plasmid vector. Restriction enzymes are as indicated in Fig. 1. (B) SDS-PAGE (7.5% polyacrylamide) of whole-cell lysates transferred to nitrocellulose and then stained for glycoprotein as described in Materials and Methods. Lanes: 1, H10407; 2, TIB3; 3, TIB3(pET109); 4, TIB3(pET146); 5 and 6, TIB3(pET139) and TIB3(pET146), respectively. Previously, it had been shown that in the absence of the sequences contained on plasmid pET146, plasmid pET139 did not direct the production of either TibA or preTibA (15). The migration of molecular mass standards (kilodaltons) is shown to the left of panel B. The mobility of TibA is shown by an arrow to the right of panel B.

TibA carbohydrates are surface exposed. As previously reported, deletion of the *tib* locus results in a nonadherent and noninvasive phenotype in H10407. Additionally, recombinant *E. coli* HB101 cells expressing preTibA do not adhere to or invade epithelial cells (15). Therefore, we were interested in determining whether the carbohydrates are exposed on the bacterial surface as we predict them to be involved in binding to epithelial cells. Intact bacteria were oxidized with periodate and subsequently labeled with DIG. The DIG-labeled bacteria were then treated with different concentrations of trypsin or proteinase K, and glycoprotein staining was performed with whole-cell lysates. As shown in Fig. 3, treatment with trypsin does not entirely remove carbohydrates; instead, it cleaves TibA, resulting in a reduced molecular mass of approximately 65 kDa. However, treatment with as little as 10 µg of proteinase K per ml completely eliminates carbohydrate-containing polypeptides. To confirm that the proteases do not penetrate the outer membrane, periplasmic protein profiles were analyzed and compared to those in untreated bacteria for the presence of β-lactamase. This analysis showed that β-lactamase and the other periplasmic proteins are intact after pro-

tease treatment (data not shown). These results show that the carbohydrates attached to TibA are surface exposed.

DNA sequence of the *tibA* gene. Analysis of the pET109 DNA sequence revealed a 2,970-bp ORF coding for a 989-amino-acid residue protein with a predicted molecular mass of 101.1 kDa. To confirm that this ORF was the *tibA* gene, the amino-terminal sequences of the preTibA and TibA proteins were determined. These amino-terminal sequences were identical (AAYDNQTIGRGETSK) and corresponded to residues 55 through 69 of the protein product predicted for the ORF. This result confirmed that this ORF was indeed the *tibA* gene and indicated that preTibA is synthesized as a precursor (pro-preTibA) in which the first 54 amino acids are cleaved during translocation to the outer membrane. The predicted molecular mass of the cleaved *tibA* gene product was 95.2 kDa. This mass corresponded well to the observed mass of preTibA as determined by SDS-PAGE (100 kDa). The observed mass of TibA as determined by SDS-PAGE (104 kDa) represents the mass of the preTibA protein plus the mass of the carbohydrates that are attached to the peptide backbone, including any effect on mobility due to charges on those carbohydrates.

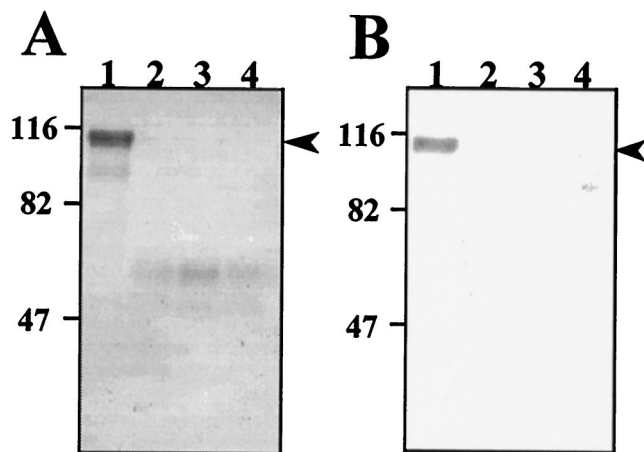


FIG. 3. Protease treatment of intact bacteria. Surface carbohydrates on strain TIB3(pET109) were labeled with DIG after periodate oxidation. Intact bacteria were then incubated with various amounts of trypsin (A) or proteinase K (B). Whole-cell lysates were prepared and then separated by SDS-PAGE (7.5% polyacrylamide). After transfer to nitrocellulose, DIG-labeled glycoproteins were detected by anti-DIG antibodies. Lanes: 1, TIB3(pET109) with 0 μg of protease per ml; 2, TIB3(pET109) with 10 μg of protease per ml; 3, TIB3(pET109) with 50 μg of protease per ml; 4, TIB3(pET109) with 100 μg of protease per ml. The migration of molecular mass standards is shown to the left of each panel. The mobility of TibA is shown by an arrow. In panel A, lane 3 appears to stain more intensely than lanes 2 and 4. In replicates of this experiment, the intensities of staining were identical at each trypsin concentration.

Protein sequence analysis. Analysis of the predicted pro-preTibA amino acid sequence revealed two regions of repetitive amino acid sequences. The first region is located near the amino terminus of the cleaved protein and consists of 12 repeats of a 19-amino-acid sequence. Comparison of those repeats revealed a consensus pattern (Table 2). The second repetitive region is located in the carboxyl half of the protein and consists of eight repeats of a 5-amino-acid sequence with a strong consensus pattern (Table 3). A BLAST search of the National Center for Biotechnology Information databases showed homology between pro-preTibA and two other known adhesins: AIDA-I (26% identity, 44% similarity, 5% gap over 559 amino acids at the N-terminal end starting behind the signal sequence) and pertactin (24% identity, 37% similarity,

TABLE 2. Amino acid sequence of the TibA repeat region 1^a

Amino acid sequence	Residues
TTINSGGKQYVSSGGSATS	82-100
TTINIGGVQHVSSGGSATS	101-119
STINSGGHQHVSSGGSATN	120-138
TTVNNGGRQTVSSGGSAMG	139-157
TTINSGGDQYVSSGGSATS	158-176
ASVTSGARQFVSSGGIIVKA	177-195
TSVNSGGRQYVRDGGSATD	196-214
TVLNNTGRQFVSSGGSAAK	215-233
TTINSGGG MYLYGGSATG	234-251
TSIYNGGRQYVSSGGSATN	252-270
TTVYSSGGRQHVYIDGNVTE	271-289
TTITSGGMLQVEAGGSASK	290-308
TTINSGG·Q·VSSGGSAT·	Consensus

^a The consensus sequence shown on the bottom line represents the amino acids most frequently encountered (at least 7 out of 12 times) at the corresponding location within the 19-amino-acid repetitive sequence. The amino acid residue numbering is based on the first codon of the *tibA* ORF representing residue 1.

TABLE 3. Amino acid sequence of the TibA repeat region 2^a

Amino acid sequence	Residues
PDADN	639-643
PDAGN	644-648
PDAGN	649-653
PDGAN	654-658
PDAGN	569-663
PDAGK	664-668
PGT GK	669-673
PDAGT	674-678
PDAGN	Consensus

^a The consensus sequence shown on the bottom line represents the amino acids most frequently encountered (at least 5 out of 8 times) at the corresponding location within the 5-amino-acid repetitive sequence. The amino acid residue numbering is based on the first codon of the *tibA* ORF representing residue 1.

7% gap over 768 amino acids from the C-terminal end). AIDA-I is an afimbrial adhesin that participates in the diffuse adherence pattern that is characteristic of diarrheagenic diffuse adhering *E. coli* (1, 2). Pertactin is an adhesin that is involved in cell attachment and virulence of *Bordetella pertussis* (8). The highest similarity was found between the C-terminal domains of the pertactin precursor (also known as P.93) and TibA (33% identity, 54% similarity, and 1% gap over 200 residues) (Fig. 4). Pertactin and AIDA-I are members of a rapidly growing class of outer membrane proteins, referred to as autotransporters, that are involved in the virulence of gram-negative bacteria. A common feature of these proteins is the presence of amphipathic sheets within the carboxyl domain that insert in the outer membrane to form a β-barrel that translocates the amino-terminal domain across the outer membrane (25). Because of the strong similarity between the TibA and pertactin C termini, we anticipated such a structure for TibA. Therefore, we used TopPred software (10, 22) to analyze the TibA C-terminal amino acid sequence for the presence of amphipathic helices and sheets. TopPred predicted 10 certain and 4 putative transmembrane amphipathic sheets within the last 311 amino acids of TibA (Fig. 5). To strengthen this prediction, the C-terminal TibA amino acid sequence was analyzed for surface probability by a variation of the method developed by Emimi et al. (16). By this method, TibA residues predicted to be at the surface are always located between transmembrane-spanning sheets, and never within them (Fig. 5). Additionally, the Chou-

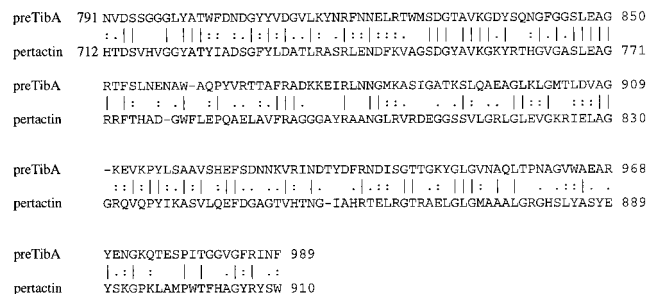


FIG. 4. Alignment of TibA and pertactin precursor carboxy termini. Alignment of the C terminus of TibA (residues 791 to 989) and pertactin precursor (residues 712 to 910) was performed by using Clustalx software. A vertical line indicates positions with identical residues. A colon indicates that one of the following "strong" groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, or FYW. A period indicates that one of the following "weaker" groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEOHRK, FVLIM, or HFY.

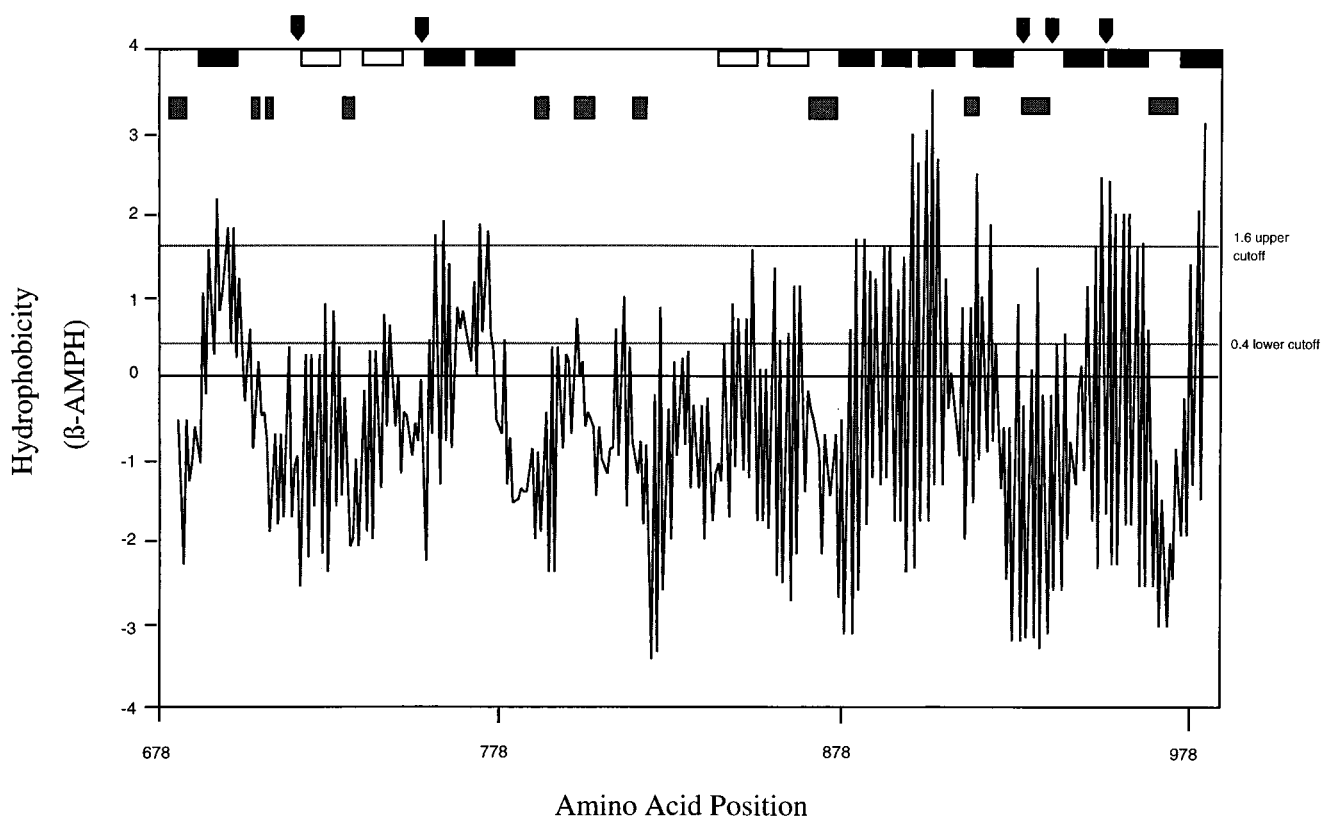


FIG. 5. Prediction of amphipathic β -sheets (β -AMPH). TibA C terminus from residues 678 to 989 was analyzed for the presence of amphipathic β -sheets (full window 10) by using TopPret software and for surface-exposed amino acids and β -turns by using MacVector software. Solid bars along the top of the figure indicate membrane-spanning β -sheets predicted as certain (1.6 upper cutoff), and open bars indicate membrane-spanning β -sheets predicted as putative (0.4 cutoff). Shaded bars indicate amino acids with high surface probability, and arrows indicate amino acids that are likely to be in β -turns, as predicted by Chou-Fasman (9).

Fasman method (9) was used to predict turns in the peptide, as indicated by the vertical arrows in Fig. 5. From these results, we can conclude that the C terminus of TibA most likely contains at least 10 to 14 amphipathic β -sheets spanning the membrane and forming a β -barrel which anchors it in the outer membrane.

DNA sequence of the *tibC* gene. Our complementation experiments indicate that DNA sequences upstream of the *tibA* gene are required for TibA glycosylation. This result suggests that the *tib* locus may encode a glycosyltransferase. Immediately upstream of the *tibA* gene, we have identified an ORF that we have designated *tibC*. Translation of the *tibC* ORF yields a 406-residue protein with similarity (30% identity, 44% similarity, 7% gap over 186 residues) to RfaQ. The RfaQ protein is proposed to be a heptosyltransferase III involved in the biosynthesis of the *E. coli* lipopolysaccharide inner core (32). Based on this similarity, we propose that the product of the *tibC* gene is responsible for glycosylation of TibA.

DISCUSSION

Adherence to intestinal cells plays an important role in infection by enteric pathogens. This adherence is often mediated by protein-protein or protein-carbohydrate interactions between bacterial ligands and eukaryotic receptors. Eukaryotic glycoprotein and glycolipid receptors have been identified that are bound by bacterial lectins. Glycosylation of these bacterial adhesins has long been considered uncommon, even though glycosylation of proteins is well characterized in archaeobacteria

and S-layer proteins of eubacteria. Recently, sensitive analytical techniques, similar to those described here, revealed the presence of carbohydrates on several eubacterial proteins involved in adherence, such as the type 4 pili of *Neisseria meningitidis*, *Neisseria gonorrhoeae*, and *Pseudomonas aeruginosa* (39, 43, 44). Other proteins, such as flagellins of *P. aeruginosa* and *Campylobacter coli* and secreted proteins of *Mycobacterium tuberculosis*, have been shown to be glycosylated (6, 11, 12).

Here we describe the identification of a glycoprotein expressed in the outer membrane of ETEC strain H10407. To our knowledge, this report is the first description of glycoprotein production in *E. coli*. The TibA protein is directly associated with epithelial cell adherence and invasion by ETEC strain H10407 and recombinant *E. coli*. Nonglycosylated forms of TibA (i.e., preTibA) do not direct these actions (15). Therefore, we suspect that TibA is an adhesin/invasin and that the carbohydrates on TibA are associated with its adherence and invasion phenotypes. We also have demonstrated that TibA is the only glycoprotein in broth-grown ETEC strain H10407 and that the carbohydrates are expressed on the bacterial surface. TibA is the product of the *tibA* gene found within the *tib* locus. Production of the mature, glycosylated TibA protein requires additional sequences within the *tib* locus. Analysis of the DNA sequence upstream of the *tibA* gene reveals an ORF that we have designated *tibC*. The *tibC* ORF is immediately upstream of the *tibA* gene. Translation of the *tibC* ORF shows a protein with similarity to RfaQ, a heptosyltransferase. Therefore, we

propose that the TibC protein is responsible for TibA glycosylation.

Based on database searches and analysis of the predicted amino acid sequence, we propose TibA to belong to a group of outer membrane proteins involved in the virulence of a variety of gram-negative bacteria and which show similarity to *Neisseria* immunoglobulin A1 (IgA1) proteases (33). In addition to the IgA1 proteases, pertactin and BrkA from *Bordetella* spp., IcsA from *Shigella* spp., and AIDA-I and Tsh from *E. coli* belong to this family of proteins (19, 21, 29, 34, 41). Because of the ability to mediate their own translocation across the outer membrane, they have been classified as autotransporters. All proteins identified so far as autotransporters are proteins of high molecular weight which contain a signal sequence at the N terminus that is cleaved during transport across the inner membrane. Autotransporter secondary structure has been predicted to contain a series of amphipathic β -sheets at the C terminus which are thought to form a β -barrel through which the rest of the polypeptide chain is translocated. Autotransporters include a very low content of cysteines, because the formation of disulfide bonds in the periplasm appears to interfere with the translocation of the N terminus across the outer membrane. In order to facilitate insertion into the outer membrane, autotransporters share a consensus sequence at their C-terminal end, consisting of an aromatic amino acid at the end, followed by a polar or charged amino acid (25). This sequence of alternating (charged/polar and aromatic/nonpolar) amino acids has also been found in other outer membrane proteins, such as OmpF or PhoE (40), and is thought to be an essential feature of all β -barrel-forming proteins and pores in the outer membrane of gram-negative bacteria. Even though the overall homology among autotransporters is low, except for one or two closely related proteins, these similar features identify them as such (25). The preTibA precursor contains 989 amino acids, including 54 N-terminal amino acids which are cleaved during transport and are therefore thought to be a signal sequence and 10 amino acids at its C terminus that are similar to the consensus sequence described above. Additionally, TibA does not contain cysteines, and its C terminus is predicted to form at least 10 β -sheets, which should be able to form a β -barrel. Because of these features and the overall homology to AIDA-I and pertactin, we propose TibA to be a member of the autotransporter family.

When analyzing the TibA N terminus, secondary structure prediction software predicts the presence of about 20 additional amphipathic β -sheets. Interestingly, the X-ray structure of the extracellular domain of pertactin (also known as P.69) revealed that almost the entire polypeptide folded in a 16-stranded β -helix resulting in a long linear domain protruding from the bacterial surface (17). These authors proposed that this tertiary structure could be an important feature of bacterial adhesins. Because of its similarity to pertactin, it is possible that TibA folds in such a manner.

We have found that TibA contains two regions of repetitive amino acid sequences. Although the role of these repetitive sequences has not been investigated, they may be involved in determining the biological activity of TibA. In general, repetitive amino acid sequences are characteristic of proteins with binding activities. A repetitive sequence similar to the first TibA repeat is found in AIDA-I and is thought to be involved in protein-receptor interactions (2). The second repetitive sequence in TibA shows a high percentage (27%) of prolines and is similar to a repetitive proline-rich region in pertactin (17). Proline-rich regions have been identified in other proteins with binding activity, and they have been suggested to mediate weak nonstoichiometric interactions (46). Therefore, the TibA re-

petitive amino acid sequences are likely to be involved in receptor binding. However, glycosylation of TibA is required for its biological activity, because unglycosylated preTibA is not associated with adherence or invasion (15). Therefore, we propose a cooperative interaction between the TibA repetitive regions and the TibA carbohydrates for receptor recognition. For example, the serine- and threonine-rich sequences within the first repetitive region are possible O glycosylation sites that may be required for TibA-receptor interaction. Overall, the similarity of TibA to known adhesins and the direct correlation between TibA expression and the adherence and invasion phenotypes of ETEC strain H10407 suggest that TibA acts as an adhesin/invasin. TibA may be an ETEC virulence factor that may be useful when developing vaccines and/or treatments for ETEC-mediated diarrhea.

ACKNOWLEDGMENTS

This work was supported by a grant to E. Elsinghorst from the University of Kansas General Research Fund. C. Lindenthal was supported by a Fulbright Fellowship.

REFERENCES

1. Benz, I., and M. A. Schmidt. 1989. Cloning and expression of an adhesin (AIDA-I) involved in diffuse adherence of enteropathogenic *Escherichia coli*. *Infect. Immun.* **57**:1506-1511.
2. Benz, I., and M. A. Schmidt. 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic *Escherichia coli* strain 2787 (O126:H27), is synthesized via a precursor molecule. *Mol. Microbiol.* **6**:1539-1546.
3. Black, R. 1993. Epidemiology of diarrhoeal disease: implications for control by vaccines. *Vaccine* **11**:100-106.
4. Boyer, H. W., and D. Rouland-Dussoix. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* **41**:459-472.
5. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248-254.
6. Brimer, C. D., and T. C. Montie. 1998. Cloning and comparison of *flhC* genes and identification of glycosylation in the flagellin of *Pseudomonas aeruginosa* a-type strains. *J. Bacteriol.* **180**:3209-3217.
7. 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.
8. Charles, I. G., G. Dougan, D. Pickard, S. Chatfield, M. Smith, P. Novotny, P. Morrissey, and N. F. Fairweather. 1989. Molecular cloning and characterization of protective outer-membrane protein P.69 from *Bordetella pertussis*. *Proc. Natl. Acad. Sci. USA* **86**:3554-3558.
9. Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Biochemistry* **13**:222-245.
10. Claros, M. G., and G. Heijnen. 1994. TopPred II: an improved software for membrane protein structure predictions. *CABIOS* **10**:685-686.
11. Dobos, K. M., K. Swiderek, K.-H. Khoo, P. J. Brennan, and J. T. Belisle. 1995. Evidence for glycosylation sites on the 45-kilodalton glycoprotein of *Mycobacterium tuberculosis*. *Infect. Immun.* **63**:2846-2853.
12. Doig, P., N. Kinsella, P. Guerry, and T. J. Trust. 1996. Characterization of a post-translational modification of *Campylobacter* flagellin: identification of a sero-specific glycosylation moiety. *Mol. Microbiol.* **19**:379-387.
13. Edman, P. 1956. On the mechanism of phenylisothiocyanate degradation of peptides. *Acta. Chem. Scand.* **10**:761-768.
14. Elsinghorst, E. A., and D. J. Kopecko. 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli*. *Infect. Immun.* **60**:2409-2417.
15. Elsinghorst, E. A., and J. A. Weitz. 1994. Epithelial cell invasion and adherence directed by the enterotoxigenic *Escherichia coli* *tib* locus is associated with a 104-kilodalton outer membrane protein. *Infect. Immun.* **62**:3463-3471.
16. Emini, E., J. V. Hughes, D. S. Perlow, and J. Boger. 1965. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.* **55**:836-839.
17. Emsley, P. E., I. G. Charles, N. F. Fairweather, and N. W. Isaacs. 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature* **381**:90-102.
18. Evans, D. G., R. P. Silver, D. J. Evans, Jr., D. G. Chase, and S. L. Gorbach. 1975. Plasmid-controlled colonization factor associated with virulence in *Escherichia coli* enterotoxigenic for humans. *Infect. Immun.* **12**:656-667.
19. Fernandez, R. C., and A. A. Weiss. 1994. Cloning and sequencing of a *Bordetella pertussis* serum resistance locus. *Infect. Immun.* **62**:4727-4738.

20. Fleckenstein, J. M., D. J. Kopecko, R. L. Warren, and E. A. Elsinghorst. 1996. Molecular characterization of the *tia* invasion locus from enterotoxigenic *Escherichia coli*. *Infect. Immun.* **64**:2256–2265.
21. Goldberg, M. B., O. Bârzu, C. Parsot, and P. J. Sansonetti. 1993. Unipolar localization and ATPase activity of IcsA, a *Shigella flexneri* protein involved in intracellular movement. *J. Bacteriol.* **175**:2189–2196.
22. Heijne, G. 1992. Membrane protein structure prediction. *J. Mol. Biol.* **225**: 487–494.
23. Hohn, B., and J. Collins. 1980. A small cosmid for efficient cloning of large DNA fragments. *Gene* **11**:291–298.
24. Institute of Medicine. 1986. Disease of importance in developing countries, vol. II., p. 159–169. National Academy Press, Washington, D.C.
25. Jose, J., F. Jaehnig, and F. Meyer. 1995. Common structural features of IgA1 protease-like outer membrane protein autotransporters. *Mol. Microbiol.* **18**:378–380.
26. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
27. Levine, M. M., J. B. Kaper, R. E. Black, and M. L. Clements. 1983. New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development. *Microbiol. Rev.* **47**:510–550.
28. Levine, M. M. 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J. Infect. Dis.* **155**:377–389.
29. Li, J., N. F. Fairweather, P. Novotny, G. Dougan, and I. G. Charles. 1992. Cloning, nucleotide sequence and heterologous expression of the protective outer-membrane protein P.68 pertactin from *Bordetella bronchiseptica*. *J. Gen. Microbiol.* **138**:1697–1705.
30. Moxley, R. A., E. M. Berberov, D. H. Francis, J. Xing, M. Moayeri, R. A. Welch, D. R. Baker, and R. G. Barletta. 1998. Pathogenicity of an enterotoxigenic *Escherichia coli* hemolysin (*hlyA*) mutant in gnotobiotic piglets. *Infect. Immun.* **66**:5031–5035.
31. Neu, H. C., and L. A. Heppel. 1965. The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *J. Biol. Chem.* **240**:3685–3692.
32. Parker, C. T., E. Pradel, and C. A. Schnaitman. 1992. Identification and sequences of the lipopolysaccharide core biosynthetic genes *rfaQ*, *rfaP*, and *rfaG* of *Escherichia coli* K-12. *J. Bacteriol.* **174**:930–934.
33. Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer. 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature* **325**:458–462.
34. Provence, D. L., and R. Curtiss III. 1994. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. *Infect. Immun.* **62**:1369–1380.
35. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
36. Schlager, T. A., C. A. Wanke, and R. L. Guerrant. 1990. Net fluid secretion and impaired villous function induced by colonization of the small intestine by nontoxicogenic colonizing *Escherichia coli*. *Infect. Immun.* **58**:1337–1343.
37. Schnaitman, C. A. 1970. Protein composition of the cell wall and cytoplasmic membrane of *Escherichia coli*. *J. Bacteriol.* **104**:890–901.
38. Stewart, G. S. A. B., S. Lubinsky-Mink, C. G. Jackson, A. Cazzel, and J. Kuhn. 1986. pHG165: a pBR322 copy number derivative of pUC8 for cloning and expression. *Plasmid* **15**:172–181.
39. Stimson, E., M. Virji, K. Makepeace, A. Dell, H. R. Morris, G. Payne, J. R. Saunders, M. P. Jennings, S. Barker, and M. Paniko. 1995. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **17**:1201–1214.
40. Struyve, M., M. Moons, and J. Tommassen. 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol.* **218**:141–148.
41. Suhr, M., I. Benz, and M. A. Schmidt. 1996. Processing of the AIDA-I precursor: removal of AIDA-C and evidence for the outer membrane anchoring as a β -barrel structure. *Mol. Microbiol.* **22**:31–42.
42. 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.
43. Virji, M., J. R. Saunders, G. Sims, K. Makepeace, D. Maskell, and D. J. P. Ferguson. 1993. Pilus-facilitated adherence of *Neisseria meningitidis* to human epithelial and endothelial cells: modulation of adherence phenotype occurs concurrently and with changes in primary amino acid sequence and the glycosylation status of pilin. *Mol. Microbiol.* **10**:1013–1028.
44. Virji, M. 1998. Glycosylation of the meningococcus pilus protein. *ASM News* **64**:398–405.
45. Wanke, C. A., and R. L. Guerrant. 1987. Small-bowel colonization alone is a cause of diarrhea. *Infect. Immun.* **55**:1924–1926.
46. Williamson, M. P. 1994. The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**:249–260.
47. World Health Organization. 1997. The world health report 1997: conquering suffering, enriching humanity. World Health Organization, Geneva, Switzerland.