Nucleotide Sequence of the Gene Encoding the Two-Subunit Pilin of Bacteroides nodosus 265

THOMAS C. ELLEMAN,1* PETER A. HOYNE,1 NEIL M. MCKERN,1 AND DAVID J. STEWART2

Divisions of Protein Chemistry1 and Animal Health,2 Commonwealth Scientific and Industrial Research Organisation, Parkville 3052, Australia

Received 21 January 1986/Accepted 3 April 1986

The nucleotide sequence of the gene encoding pilin from Bacteroides nodosus 265 has been determined. The pilin is encoded by a single-copy gene, from which can be predicted a prepilin comprising a single protein chain of M, 16,637. The prepilin sequence differs in several respects from the mature protein sequence. Seven additional N-terminal amino acid residues are present in prepilin, whereas residue 8, phenylalanine, undergoes posttranslational modification to become the N-methylated amino-terminal residue of mature pilin. In addition, further processing occurs through internal cleavage to produce two noncovalently linked subunits characteristic of pilins from serogroup H of B. nodosus, of which strain 265 is a member. The position of cleavage has been identified between alanine residues at positions 72 and 73 of the mature 149-residue pilin protein. The predicted pilin sequence of B. nodosus 265 shows extensive N-terminal amino acid sequence homology with other pilins of the N-methylphenylalanine type. In addition this sequence also shows homology with these N-methylphenylalanine-type pilins in the C-terminal region of the molecule, especially with pilin from Pseudomonas aeruginosa PAK.

Bacteroides nodosus is the causative organism of ovine footrot. Infection is spread among sheep grazing on wet pasture (1, 8, 31). Although antibody titers after infection are insufficient to provide protection against future infections (7, 11), effective immunity against homologous serogroups can be induced by vaccination with either killed whole cells (6, 7) or purified pili—the major host-protective immunogen (13, 38–41).

The pili of B. nodosus are flexible filamentous structures 5 to 6 nm in diameter and up to 10 μm in length (12, 37, 46). They have a mainly polar distribution on the cell and, in common with pili from bacterial species as diverse as Pseudomonas aeruginosa, Neisseria gonorrhoeae, Moraxella nonliquefaciens, and Moraxella bovis, are composed of identical subunits of the protein pilin, which has a molecular weight of ~18,000 (12, 15, 18, 21, 34). Pilin proteins from these bacterial species show extensive homology in their N-terminal amino acid sequences (23), and all begin with the modified amino acid N-methylphenylalanine (NMephe). Gene sequence studies indicate that the mature protein is produced as a prepilin with six or seven additional N-terminal residues (9, 21, 26, 29). These highly homologous pili are referred to as the NMephe type; they differ considerably in structure and sequence from the several types of pili found among members of the Enterobacteriaceae (14, 30, 44, 45).

At least eight serogroups of B. nodosus (designated A through H) have been identified on the basis of immunological cross-reactivity (3). We have reported previously the gene sequence (9) and the amino acid sequence (23) of pilin from B. nodosus 198, a member of serogroup A, subgroup 1. Pilin from strain 198 resemble other pilin of the NMephe type in being composed of a single-subunit protein. We now report the gene sequence and protein sequence of B. nodosus 265, a member of serogroup H, subgroup 1. The pilins from serogroup H, unlike those from other known serogroups of B. nodosus, comprise two covalently linked subunits as evidenced by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis of purified pilus preparations (D. J. Stewart, J. E. Peterson, J. A. Vaughan, B. L. Clark, D. L. Emery, J. B. Caldwell, and A. A. Kortt, Aust. Vet. J., in press).

MATERIALS AND METHODS

Preparation of DNA and pili from B. nodosus. B. nodosus 265 was harvested from modified Eugin broth (3) after anaerobic growth for 2 days at 37°C. DNA and pili were prepared as previously described (10, 42).

Hybridization probes and primers. Hybridization probes were prepared from DNA of B. nodosus 198 cloned in M13mp8. The M13mp8-cloned template (0.3 μg) and 5 ng of synthetic oligonucleotide primer were boiled for 3 min and annealed at 68°C for 20 min in a buffer of 7 mM Tris hydrochloride (pH 7.4)–7 mM MgCl2–1 mM dithiothreitol–50 mM NaCl. Second-strand synthesis was accomplished by incubating the annealed mixture at room temperature for 45 min with DNA polymerase I Klenow fragment (1 U); dGTP, dCTP, and dTTP each at 40 μM; and 10 μCi of (α-32P)dATP (3,000 Ci/mmol). The reaction was terminated by the addition of EDTA to 25 mM, final concentration. Probes were used without prior purification and were heat denatured at 100°C for 3 min immediately before use. Restriction digests of DNA fractionated in agarose gels (1%, wt/vol) were either probed directly (V. J. Kidd, Focus 6:3, 1984; Bethesda Research Laboratories, Inc.) or after the transfer of DNA to nitrocellulose (36). Recombinant cells harboring the pilin gene were identified by probing colonies grown on nitrocellulose membrane (20). Washing gels or membranes with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% SDS for 1 h at 68°C was sufficient to differentiate positive signals from the background.

Identification and cloning of the pilin gene. DNA (10 μg) prepared from B. nodosus 265 was digested overnight at 37°C with HindIII (24 U), and restriction fragments were...
separated by electrophoresis in a submerged horizontal agarose gel (1%, wt/vol). Fragments ranging in size from 3 to 4 kilobase pairs were recovered from the gel by electrophoresis onto NA45 membrane (Schleicher & Schuell Co.), extracted, and then ligated into the positive selection vector pUN121 (27) with a threefold molar excess of fragments over HindIII-cut vector at a total DNA concentration of 25 μg/ml. Clones identified by hybridization as containing the pilin gene, and confirmed as pilin producing by immunoblotting (28), were used in sequence studies.

Synthetic primers and nucleotide sequence determination. Synthetic 14-mers were manually synthesized by a phosphotriester method (5) and were purified by high-pressure liquid chromatography (HPLC) (9). DNA sequences were determined by the dideoxy chain-termination method (33) with M13mp8- or M13mp9-cloned DNA (25) as templates or with alkaline denatured and neutralized double-stranded DNA (2) as templates (Fig. 1).

Peptide isolation and sequence determination. Before enzymatic digestion pili were reduced and carboxymethylated (23). Enzymatic digestions with trypsin and chymotrypsin were performed overnight at 37°C in 0.05 M ammonium bicarbonate at an enzyme/substrate ratio of 1:50. Peptic digestions were performed in 0.1% (vol/vol) trifluoroacetic acid. Peptides were purified by HPLC (24), and sequences were determined manually by a modified Edman degradation procedure with phenylthiohydantoin derivatives identified by HPLC (24).

Subunits of pilin from B. nodosus 265 were fractionated by HPLC with a Vydac C4 column, equilibrated with 0.1% trifluoroacetic acid, and eluted with a gradient of acetonitrile in trifluoroacetic acid. Amino acid analysis was performed on 1 to 2 nmol of sample hydrolyzed in vacuo for 22 h or 48 h with 200 μl of constant boiling hydrochloric acid containing 0.01% (wt/vol) thiglycolic acid and 0.01% (wt/vol) phenol.

Sequence comparisons. Sequence homology among pilins was investigated by using the computer program ALIGN (4) modified for use with a gap penalty and both positive and negative amino acid comparison scores (4). All sequence alignments presented are considered highly significant (P << 0.001).

RESULTS

Identification of the pilin gene. A hybridization probe was prepared corresponding to the N-terminal coding sequence of B. nodosus 198. This probe hybridized to a single region (3.2 kilobase pairs ± 5%) of a HindIII digest of DNA from B. nodosus 265 fractionated on a 1% agarose gel. The HindIII fragment identified by the probe was inserted into plasmid vector pUN121 for transformation of Escherichia coli. Recombinant clones containing the pilin gene were detected by the hybridization probe. Immunoblotting of these clones identified a protein species of Mr ~ 17,000, similar in size to pilins from other serogroups of B. nodosus.

After DNA sequence studies, probes were prepared from the identified coding sequence of the pilin gene. Hybridization studies on various restriction digests (HindIII, Clal, PstI, PvuI, HindIII-PstI) of strain 265 DNA fractionated by gel electrophoresis identified only a single fragment in each digest, indicating a single copy of the structural pilin gene in the genome. A single locus for the pilin gene is consistent with the identification of a single locus for the pilin structural gene previously determined in B. nodosus 198 (T. C. Elleman, P. A. Hoyne, N. M. McKern, D. L. Emery, D. J. Stewart, and B. L. Clark, Footrot in Ruminants: Proceedings of a Workshop, Melbourne, Australia 1985, in press).

Predicted amino acid sequence. The nucleotide sequence (Fig. 2) contains a long open reading frame coding for a single-chain amino acid sequence. This sequence is consistent with the amino acid composition of pili from B. nodosus 265 and with the amino acid compositions of the separated subunits (Table 1). No discrepancy was found between the predicted amino acid sequence and the partial (50%) protein sequence of pilin (Table 2). A protein of the approximate size predicted from the gene sequence (Mr, 16,637) was expressed in E. coli harboring the recombinant plasmid, and this protein cross-reacted with antiserum raised against B. nodosus 265 pilin. These protein studies and the presence of a single hybridization locus in the genome indicate that the isolated gene is the pilin structural gene.

A leader sequence of seven additional amino acids precedes the phenylalanine which becomes the N-methylated amino-terminal residue of mature pilin. This predicted leader
FIG. 2. Nucleotide sequence encoding pilin of *B. nodosus* 265. Regions of hyphenated dyad symmetry are underlined. Nucleotides different from *B. nodosus* 198 in the region 5' upstream of the coding sequence are indicated below the sequence.

sequence is identical to the leader sequence determined for *B. nodosus* 198 (9). Similar short leader sequences have been predicted for all pilins of the NMePhe type for which gene sequences have been determined (Fig. 3A).

**Codon usage.** The codon usage of the pilin gene of *B. nodosus* 265 is similar to that previously determined for *B. nodosus* 198 (Table 3). Both strains show a marked preference for U rather than C in the third base position of quartet codons, and both duet and quartet codons show a preferential use of A rather than G in the third base position. A similar codon usage has been found in the highly expressed mRNA of *E. coli* and has been interpreted largely in terms of optimizing codon-anticodon pairing energies (16), such that the codons of strongest binding energies are avoided when protein production is high. Many of the codons which are absent from the highly expressed mRNA of *E. coli* (because
The gene sequence predicts a single-chain prepilin precursor molecule for pilin of *B. nodosus 265*. Thus in addition to the processing of prepilin observed in other NMePhe-type pilis (9, 21, 26, 29), where a small signal peptide is removed and the N-terminal phenylalanine residue is methylated, the pilin of *B. nodosus 265* undergoes further processing by internal chain cleavage to produce two subunits. Amino acid analysis of the two subunits suggests that processing occurs by cleavage of the peptide bond between alanine residues at positions 72 and 73 of the mature pilin sequence (Table 1). This cleavage does not result in the creation of a free amino group, since no N-terminal sequence data can be obtained from the intact C-terminal subunit. Although the high content of valine and isoleucine in the N-terminal subunit limit the accuracy of determination of its composition, the isolation of a chromatographic peptide (Table 1) Thr-Thr-Ser-Asn-Leu-Leu-Ala-Ala confirms the suggested cleavage position.

The pilin sequence of *B. nodosus 265* indicates four half-cystine residues in the molecule. Other pilins of the NMePhe type so far characterized have only two half-cystine residues. The ability to separate the subunits of pilin from *B. nodosus 265* both by HPLC and by SDS-polyacrylamide gels (19) under nonreducing conditions (Fig. 4), demonstrates that no covalent linkages occur between subunits of pilin from *B. nodosus 265*. Since free thiol groups are highly unlikely to occur in an extracellular environment, and intersubunit disulfide bridges are excluded, the half-cystine residues of pilin can only be linked to form disulfide bridges within each subunit. The two half-cystine residues of the C-terminal subunit are in positions characteristically found for the two half-cystine residues in pilins of the NMePhe type from other bacterial species (Fig. 5). These C-terminal half-cystine residues have been identified in pilin from *N. gonorrhoeae* as forming a disulfide bridge (35). In contrast, a bridge formed by the two half-cystine residues in the N-terminal subunit of *B. nodosus 265* (Fig. 5) appears to be unrelated to that formed by the two half-cystine residues of strain 198 (23), since these bridges would occur in a different location based on sequence homology. This bridge in strain 265 is present within the pilin N-terminal region, which shows interspecies homology (Fig. 5), whereas the bridge of strain 198 encompasses the central region, which shows virtually no homology with other NMePhe-type pilis.

### Table 1. Amino acid analyses of pili and pilin subunits from *B. nodosus 265*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Residues per molecule</th>
<th>N-terminal subunit*</th>
<th>C-terminal subunit*</th>
<th>Pili*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>7.5 (8)</td>
<td>10.8 (11)</td>
<td>17.3 (19)</td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>3.0 (3)</td>
<td>1.0 (1)</td>
<td>4.3 (4)</td>
<td></td>
</tr>
<tr>
<td>Asx</td>
<td>4.0 (4)</td>
<td>7.9 (8)</td>
<td>12.5 (12)</td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>2.8 (2)</td>
<td>2.7 (2)</td>
<td>4.1 (4)</td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>4.8 (5)</td>
<td>7.6 (8)</td>
<td>15.9 (13)</td>
<td></td>
</tr>
<tr>
<td>Glx</td>
<td>8.4 (8)</td>
<td>7.4 (7)</td>
<td>15.0 (15)</td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>0.0 (0)</td>
<td>0.9 (1)</td>
<td>1.1 (1)</td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>7.5 (10)</td>
<td>1.9 (2)</td>
<td>7.2 (12)</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>6.3 (7)</td>
<td>4.0 (4)</td>
<td>9.2 (11)</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>2.0 (2)</td>
<td>6.8 (7)</td>
<td>9.0 (9)</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>1.9 (3)</td>
<td>(0)</td>
<td>1.8 (3)</td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>1.1 (1)</td>
<td>2.0 (2)</td>
<td>3.3 (3)</td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>1.2 (1)</td>
<td>3.1 (3)</td>
<td>4.4 (4)</td>
<td></td>
</tr>
<tr>
<td>Ser4</td>
<td>3.2 (4)</td>
<td>4.3 (5)</td>
<td>8.3 (9)</td>
<td></td>
</tr>
<tr>
<td>Thr4</td>
<td>5.3 (6)</td>
<td>11.4 (12)</td>
<td>16.2 (18)</td>
<td></td>
</tr>
<tr>
<td>Thr4</td>
<td>(2)</td>
<td>(0)</td>
<td>1.8 (3)</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>2.0 (2)</td>
<td>1.1 (1)</td>
<td>3.4 (3)</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>2.9 (4)</td>
<td>1.0 (1)</td>
<td>3.9 (5)</td>
<td></td>
</tr>
<tr>
<td>NMePhe</td>
<td>(1)</td>
<td>(0)</td>
<td>(0)</td>
<td></td>
</tr>
</tbody>
</table>

* Amino acid analysis was carried out as described previously (20).

### Table 2. Amino acid sequences determined from pilin of *B. nodosus 265*

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Digest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thr-Thr-Ser-Asn-Leu-Leu-Ala-Ala</td>
<td>Chymotrypsin</td>
</tr>
<tr>
<td>Ser-Gln-Val-Ser-Arg</td>
<td>Trypsin</td>
</tr>
<tr>
<td>Val-Met-Ser-Glu-Thr-Gly-Gin-Met</td>
<td>Trypsin</td>
</tr>
<tr>
<td>Thr-Ala-Ile-Glu-Thr-Cys-Leu-Leu-Asp-Gly-Lys-Glu</td>
<td>Trypsin</td>
</tr>
<tr>
<td>Glu-Gly-Lys-Asp-Cys-Phe-Ile-Gly-Tyr-Thr-Thr-Ser-Asn</td>
<td>Trypsin</td>
</tr>
<tr>
<td>Lys-Leu-Thr-Trp-Thr-Arg</td>
<td>Trypsin</td>
</tr>
<tr>
<td>Ser-Pro-Glu-Ala-Thr-Trp-Ser-Cys-Thr-Thr-Asp-Val-Asp</td>
<td>Trypsin</td>
</tr>
<tr>
<td>Val-Asp-Glu-Lys</td>
<td>Trypsin</td>
</tr>
<tr>
<td>MePhe-Thr-Leu-Ile-Glu-Leu-Met-Ile-Val-Val-Ala-Ile-Glu</td>
<td>Intact protein</td>
</tr>
</tbody>
</table>
pilin-coding sequence is even more extensive than that in the N-terminal coding sequence of the pilin genes. The Shine-Dalgarno sequences, extended ribosome binding sites, and predicted signal peptides of *B. nodosus* 198 and 265 are identical. These features show homology with those of other bacterial species producing pilins of the NMePhe type (Fig. 3A). Regions of hyphenated dyad symmetry occur upstream of the ribosome-binding site. Such symmetrical regions have often been implicated in the control of expression.

Within the limited sequences determined 3′ downstream of the coding sequence in *B. nodosus* 198 and 265, homology is only apparent in the region of 80 nucleotides adjacent to the coding sequence. This region shows potential ribo-independent transcription terminators of comparable free energies in both sequences, but differing considerably in the sequences of the inverted repeats (Fig. 3B and 6). Sequences characteristic of transcription terminators can be identified downstream in all pilin coding sequences of the NMePhe type so far determined.

**DISCUSSION**

Extensive sequence homology is shown among pilins from the diverse bacterial species which produce the NMePhe type of pili (Fig. 5), yet the complexity of the pilin gene systems differs considerably. A single locus for the pilin structural gene of *B. nodosus* 198 and 265 has been identified by hybridization studies. In *P. aeruginosa* in a single gene locus was likewise identified, whereas multiple gene loci are present in *M. bovis* and in *N. gonorrhoeae*. The presence of these multiple gene copies permits antigenic switching in *M. bovis* and extensive antigenic variation in *N. gonorrhoeae* (17).

Little information is available regarding transcription of the pilin genes. Although a promoter has been identified 90

---

TABLE 3. Frequency of codon usage in the predicted coding region of the pilin gene from *B. nodosus* 265a

<table>
<thead>
<tr>
<th>Phe</th>
<th>1 (2)</th>
<th>UUU</th>
<th>Ser</th>
<th>1 (3)</th>
<th>UCU</th>
<th>Tyr</th>
<th>0 (4)</th>
<th>UAU</th>
<th>Cys</th>
<th>1 (2)</th>
<th>UGU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu</td>
<td>7 (8)</td>
<td>UUA</td>
<td>3 (3)</td>
<td>UCA</td>
<td>0 (0)</td>
<td>UCC</td>
<td>3 (3)</td>
<td>UAC</td>
<td>Cys</td>
<td>3 (0)</td>
<td>UGC</td>
</tr>
<tr>
<td>UUG</td>
<td>4 (1)</td>
<td>0 (0)</td>
<td>UCA</td>
<td>1 (0)</td>
<td>UCG</td>
<td>0 (0)</td>
<td>UAG</td>
<td>Arg</td>
<td>1 (1)</td>
<td>UGA</td>
<td></td>
</tr>
<tr>
<td>CUU</td>
<td>3 (2)</td>
<td>Pro</td>
<td>1 (1)</td>
<td>CCU</td>
<td>0 (0)</td>
<td>CCC</td>
<td>1 (0)</td>
<td>CAU</td>
<td>0 (0)</td>
<td>UGU</td>
<td></td>
</tr>
<tr>
<td>CUA</td>
<td>3 (1)</td>
<td>0 (0)</td>
<td>CCC</td>
<td>1 (1)</td>
<td>CCA</td>
<td>6 (4)</td>
<td>CAA</td>
<td>0 (0)</td>
<td>UGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUG</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>CCA</td>
<td>0 (0)</td>
<td>CCG</td>
<td>1 (0)</td>
<td>CAG</td>
<td>0 (0)</td>
<td>UGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCU</td>
<td>6 (6)</td>
<td>AUU</td>
<td>8 (4)</td>
<td>ACU</td>
<td>3 (1)</td>
<td>ACC</td>
<td>4 (3)</td>
<td>AAC</td>
<td>1 (1)</td>
<td>AGU</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>6 (7)</td>
<td>0 (0)</td>
<td>ACC</td>
<td>3 (1)</td>
<td>ACG</td>
<td>2 (5)</td>
<td>AAC</td>
<td>3 (1)</td>
<td>AGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUA</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>ACG</td>
<td>2 (5)</td>
<td>AAC</td>
<td>3 (1)</td>
<td>ACG</td>
<td>3 (1)</td>
<td>AGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>3 (1)</td>
<td>AUG</td>
<td>0 (0)</td>
<td>AGC</td>
<td>2 (5)</td>
<td>ACG</td>
<td>0 (0)</td>
<td>AGG</td>
<td>0 (0)</td>
<td>AGG</td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>4 (7)</td>
<td>GUU</td>
<td>11 (10)</td>
<td>GCU</td>
<td>4 (10)</td>
<td>GCA</td>
<td>9 (7)</td>
<td>GAA</td>
<td>3 (2)</td>
<td>GGA</td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>4 (7)</td>
<td>0 (0)</td>
<td>GCG</td>
<td>0 (0)</td>
<td>AGG</td>
<td>0 (0)</td>
<td>GAG</td>
<td>0 (0)</td>
<td>GGG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Possible initiation codons are excluded. Codons corresponding to minor (or weakly interacting) tRNA species in *E. coli* are underlined. First-position totals: U, 27; C, 20; A, 56; G, 52. Second-position totals: U, 36; C, 46; A, 43; G, 30. Third-position totals: U, 57; C, 34; A, 54; G, 10. The codon frequency in *B. nodosus* 198 is shown in parentheses.
than loop has been both pilins these be determined. This related nonrandom DNA base pairs upstream of the coding sequence in the pilE1 gene of N. gonorrhoeae MS11 (26), it would be unwise to extrapolate from this example to the less complex single-locus systems. We have previously reported the possibility of a similarly located promoter in B. nodosus on the basis of the high A+T content of the upstream region (9). However, in view of the proximity of an open, upstream reading frame in both P. aeruginosa PAK and B. nodosus, the possibility that these pilins may be encoded on polycistronic messages cannot be discounted.

The region upstream of the pilin genes of B. nodosus 198 and 265 shows extremely high conservation of the nucleotide sequence, where only 6 nucleotides differ in 381 nucleotides determined. This high sequence conservation may be partly related to the presence of open reading frames in the first 250 nucleotides of the determined sequences. We previously reported the nonrandom location of the pilin gene in DNA fragments of strain 198 cloned in E. coli (10). The pilin gene was consistently located within 2 kilobases of one end of 7-15-kilobase fragments of DNA from a partial Sau3A digestion. One possible explanation of this observation is a lethal product from a gene upstream of the pilin gene of B. nodosus. It is interesting that the proposed basal protein of pilin (Mr, ~80,000 [22, 28]) would require a 2-kilobase gene for encoding, and this protein may be the presumptive lethal product which is partially seen as the upstream open reading frame.

Sequence homology downstream of the pilin-coding sequences ends 80 nucleotides after the translation terminator in B. nodosus 198 and 265. The identification of potential transcription terminators in this region suggests that the pilin gene in B. nodosus is terminal in any transcript. Typical transcription termination signals are likewise found in P. aeruginosa PAK downstream of the pilin-coding sequence (Fig. 3B).

The extensive homology of amino acid sequences within the hydrophobic N-terminal regions of the NMePhe type of pil has long been recognized (23). Sequence similarity gradually disappears after 50 residues, but reappears 50 or so residues before the C termini. This C-terminal homology is found around a disulfide loop common to at least one strain from each of the bacterial species producing pilins of the NMePhe type (Fig. 5). Homology is most notable between B. nodosus 265 and P. aeruginosa PAK, whereas B. nodosus 198 lacks the C-terminal disulfide loop but still retains sequence homology with other pilins. Comparison of the nucleotide sequences of B. nodosus 265 and 198 suggests that strain 198 is related to strain 265 in the C-terminal region by a frameshift mutation (Fig. 6).

Other than the N- and C-terminal regions of homology, little sequence similarity is found among pilins of the NMePhe type. The central region varies quite considerably between otherwise similar sequences of representatives from B. nodosus serogroups A, B, C, and E, which show overall sequence homology of ~70%. (24; N. M. McKern, unpublished results). This region may contain the major serogroup-specific antigenic determinants of B. nodosus. In N. gonorrhoeae this central region is considerably shorter than in B. nodosus, and the major serotype-specific antigenic determinant per base pair (Fig. 5).
determinants of pili from *N. gonorrhoeae* are located within the large C-terminal disulfide loop. However, some variability occurs in this terminal region, as does a common antigenic determinant (17, 32). The conserved sequence within the common antigenic determinant is \( \gamma \), which shows similarity with the sequence \( \gamma \) found in representatives of *B. nodosus* serogroups A, B, C, and E, where it precedes the highly variable region.

With the exception of members of serogroup H of *B. nodosus* (e.g., strain 265), pili from all other known serogroups of *B. nodosus* are composed of a single chain subunit of \( M_r \sim 17,000 \) to 19,000 (Stewart et al., in press). This present study has shown that the pili of strain 265 (a serogroup H representative) are composed of two smaller subunits encoded by a single gene which predicts a single chain preplin of \( M_r \) 16,637. In recombinant *E. coli* strains harboring the gene, a single protein species of this approximate size has been demonstrated. It is therefore apparent that in *B. nodosus* 265 a single polypeptide chain undergoes further processing to produce two noncovalently linked subunits. The position of the cleavage in the polypeptide chain has been identified as occurring between alanine residues 72 and 73 in the mature pilin sequence. Since an Ala-Ala peptide bond is not normally a chemically labile linkage, the two-chain pilin subunit observed in serogroup H might result from proteolytic cleavage. The fact that a two-chain structure is observed only in pili from serogroup H suggests that a unique proteolytic enzyme may be present only in serogroup H; alternatively, there may be present a protease common to all serogroups, in which case the absence of a free amino group after peptide bond cleavage.

We are currently attempting to express pili from *B. nodosus* on the surface of the less fastidious host, *P. aeruginosa*, with a view to production of a plus-based, single-protein-folate vaccine. The extensive homology within the N-terminal pilin sequences of these bacteria and the similar unusual signal sequences suggested the feasibility of such a strategy. The demonstration that the NMePhe-type pilins show much more extensive similarity in amino acid sequence, not only limited to the N-terminal sequence, reinforces the possibility of successful morphogenetic expression of the NMePhe group of pilins in the easily grown, stable genetic environment of *P. aeruginosa*.

**LITERATURE CITED**


15. Fryholm, L. O., and K. Sletten. 1977. Purification and N-
terminal sequence of a fimbrial protein from Moraxella
Mercier. 1981. Codon catalog usage is a genome strategy
recombination leads to pilus antigenic variation in Neisseria
Neisseria pilin proteins: amino-terminal amino acid sequences
and identification of an unusual amino acid. Biochemistry
17:442–445.
cloning: laboratory manual, p. 326. Cold Spring Harbor Lab-
oratory, Cold Spring Harbor, N.Y.
21. Marrs, C. F., G. Schoolnik, J. M. Koomen, J. Hardy, J.
Rothbard, and S. Falkow. 1985. Cloning and sequencing of a
1984. Isolation and characterization of Bacteroides nodosus
fimbriae: structural subunit and bacterial antigen. J. Bact-
23. McKern, N. M., I. J. O’Donnell, A. S. Inglis, D. J. Stewart, and
B. L. Clark. 1983. Amino acid sequence of pilin from Bacte-
roides nodosus (strain 198), the causative organism of ovine
1985. Primary structure of pilin protein from Bacteroides
nodosus strain 216: comparison with the corresponding protein
Enzymol. 101:20–78.
1984. Pilus genes of Neisseria gonorrhoeae: chromosomal
organization and DNA sequence. Proc. Natl. Acad. Sci. USA
81:6110–6114.
Phillipsen. 1983. An improved positive selection plasmid vector
constructed by oligonucleotide mediated mutagenesis. Nucleic
Acids Res. 11:8019–8030.
identification of pilus antigen and other protein antigens of
Bacteroides nodosus using electrobolt radioimmunoassay after
electrophoretic fractionation of the proteins on sodium dodecyl
and sequencing of the Pseudomonas aeruginosa PAK pilin
Comparison of the nucleotide sequences of the genes encoding
the KSF1A and F7, fimbrial antigens of uropathogenic Esche-
pathogenesis of ovine footrot. II. The pathogenic association of
Fusimodis nodosus and F. necrophorus. J. Comp. Pathol.
of gonococcal pilin: approaches towards a gonococcal vaccine. p.
Vaccines 85: molecular and chemical basis of resistance to
parasitic, bacterial, and viral diseases. Cold Spring Harbor
Laboratory, Cold Spring Harbor, N.Y.
33. Sangen, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequenc-
USA 74:5463–5467.
34. Sastry, P. A., J. R. Pearlstone, L. B. Smillie, and W.
Pararchych. 1983. Amino acid sequence of pilin isolated from
35. Schoolnik, G. K., R. Fernandez, J. Y. Tai, J. Rothbard, and
E. C. Gotschlich. 1984. Gonococcal pilin: primary structure and
transfer of DNA and RNA to nitrocellulose or diazobenzoyl-
37. Stewart, D. J. 1973. An electron microscopic study of
38. Stewart, D. J. 1978. The role of various antigenic fractions of
Bacteroides nodosus in eliciting protection against foot-rot in
39. Stewart, D. J. 1978. Studies on the antigenic structure of
40. Stewart, D. J., B. L. Clark, D. L. Emery, J. E. Peterson, and
K. J. Faihey. 1983. A Bacteroides nodosus immunogen, distinct
from the pilus, which induces cross-protective immunity in
41. Stewart, D. J., B. L. Clark, J. E. Peterson, D. A. Griffiths, and
E. F. Smith. 1982 Importance of pilus-associated antigens in
42. Stewart, D. J., B. L. Clark, J. E. Peterson, D. A. Griffiths, E. F.
of Freund’s adjuvant on the antibody and protective responses of
35:130–137.
43. Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C.
Uhlenbeck, D. M. Crothers, and J. Gralla. 1973. Improved
44. van Die, I., B. van Geffen, W. Hoekstra, and H. Bergmans.
1984. Type 1C fimbiae of a uropathogenic Escherichia coli strain:
cloning and characterization of the genes involved in the
expression of the 1C antigen and nucleotide sequence of the subunit
45. Waalen, K., K. Sletten, L. O. Fryholm, V. Vaisanen, and T. K.
Korhonen. 1983. The N-terminal amino acid sequence of type 1
timbria (pili) of Salmonella typhimurium LT2. FEMS Microbiol.
Lett. 16:149–151.
1973. The fine structure of Fusimodis nodosus with special
references to the location of antigens associated with immu-
47. Williams, J. G., and W. B. Grazier. 1971. Limitations of the
detergent-polyacrylamide gel electrophoresis method for mole-
cular weight determination of proteins. J. Chromatog.
57:121–125.