Nucleotide Sequence of the Pilin Gene of Bacteroides nodosus 340 (Serogroup D) and Implications for the Relatedness of Serogroups

By Kenneth G. Finney, Thomas C. Elleman and David J. Stewart

Commonwealth Scientific and Industrial Research Organization, Parkville 3052, Australia

(Received 2 October 1987)

The gene encoding pilin of Bacteroides nodosus 340 has been isolated and the nucleotide sequence determined. The gene is present as a single copy within the B. nodosus genome and a protein of $M_r$ 16683 can be predicted from the proposed coding region. A comparison of the predicted amino acid sequence with pilin from other strains of B. nodosus indicated that the protein of strain 340 (serogroup D) has a high degree of similarity with pilin of strain 265 (serogroup H). The degree of similarity between pilins from these strains and from other B. nodosus serogroups is no greater than that between B. nodosus pilins and the homologous proteins of several different bacterial species. These findings suggest that serogroups D and H may form a subset of B. nodosus serogroups.

INTRODUCTION

Bacteroides nodosus is the infectious agent of ovine footrot, a contagious disease which has a severe effect on the profitability of sheep production. Although immunity is not a consequence of natural infection, vaccination with killed whole cells of B. nodosus elicits a prophylactic response (Egerton & Burrel, 1970; Egerton & Roberts, 1971) and cell-surface appendages referred to as pili (Brinton, 1959), or fimbriae (Duguid et al., 1955), have been identified as the major host-protective immunogen (Every & Skerman, 1982; Stewart, 1978a, b; Stewart et al., 1982, 1983a).

The pili of B. nodosus are flexible filamentous structures 5–6 nm in diameter and up to 10 µm in length (Every, 1979; Stewart, 1973; Walker et al., 1973). They are composed solely of the protein pilin, which has an $M_r$ of 16500 to 19000 (Anderson et al., 1986; Stewart et al., 1986). Many bacterial species are piliated and pilins from B. nodosus, Moraxella bovis, M. nonliquefaciens, Neisseria gonorrhoeae, N. meningitidis and Pseudomonas aeruginosa display extensive N-terminal amino acid sequence homology (Elleman et al., 1986b; Marrs et al., 1985; McKern et al., 1983; Olafson et al., 1985). The similarity of these pilins, which have been designated N-methylphenylalanyl (N-mPhe) pilins (Elleman et al., 1986b; Marrs et al., 1985), has been further demonstrated by the expression of B. nodosus pili in P. aeruginosa as morphologically correct structures (Elleman et al., 1986c; Mattick et al., 1987).

Despite the conservation of sequence at the N-terminus, pilins from different strains of B. nodosus are highly variable beyond this region (McKern et al., 1986), giving rise to serological variation. On the basis of cross-reactivity in agglutination reactions involving the pilus, field isolates of the organism have been divided into a number of serogroups, with some groups further divided into subtypes (Claxton et al., 1983). Nine serogroups, designated A through I, are presently recognized (Claxton, 1986) and protection against footrot following vaccination is restricted to the serogroup from which the vaccine was prepared.
An analysis of the differences between the serogroups of *B. nodosus* will require pilin sequence data for at least one representative of each serogroup. The amino acid sequences of pilin from strains 198, 238 and 265, representatives of serogroups A, G and H respectively, have been derived from the nucleotide sequences of the isolated pilin genes (Elleman & Hoyne, 1984; Elleman *et al*., 1984, 1986b; Elleman & von Ahlefeldt, 1987). In addition, the amino acid sequences of pilin from representatives of serogroups B, C and E have been determined (McKern *et al*., 1986; N. M. McKern, personal communication). In this report, we describe the isolation and characterization of the pilin gene from a representative of the D serogroup, strain 340. The predicted amino acid sequence of strain 340 pilin is of special interest as this additional sequence, together with other data, indicates that there are two subsets of *B. nodosus* serogroups.

**RESULTS**

**Identification of the pilin gene**

The probe derived from strain 198 DNA (see Methods) hybridized to a single band of 3.0 kbp in *HindIII*-digested *B. nodosus* 340 DNA, so fragments of about this size were isolated for cloning of the pilin gene from the latter strain. *E. coli* transformants selected with the same probe contained plasmids bearing a single 3 kbp insert at the *HindIII* site (Fig. 1), with the only variation between clones being the orientation of the insert. To verify that strain 340 pilin was encoded by the cloned DNA, protein extracts of these cells were subjected to immunoblotting. A single immunologically reactive protein of *M*, approximately 17000 was detected in each case.
B. nodosus pilin

Fig. 1. Characterization of the pilin gene. A 3 kbp fragment of B. nodosus 340 DNA cloned in pBR322 is shown, with the pilin-coding sequence boxed. Transcription of the pilin gene occurs from left to right. Nucleotide sequences were derived directly from the plasmid or from subfragments cloned in M13 and are indicated by arrows which define the extent and direction of the determined sequence. The filled circles indicate the use of synthesized primers whose sequences were derived from the strain 340 sequence. HindIII and KpnI sites are abbreviated as H and K respectively.

(Fig. 2), a size comparable to that of pilin from strain 340 and similar to the protein expressed in E. coli cells harbouring the pilin gene from other strains of B. nodosus (Elleman et al., 1984, 1986b).

Sequence studies indicated a cloned subfragment of strain 340 DNA suitable as a specific hybridization probe for the pilin gene of this strain. The fragment extended from a KpnI site within the pilin coding region to a HinPI site 173 nucleotides downstream from the termination codon (see below). Various restriction enzyme digests of genomic DNA were analysed with this probe to investigate the possibility that multiple pilin genes might be present within the B. nodosus genome (Fig. 3). Single bands were detected in HindIII, HindIIIIKpnI and HaeII digests with sizes of 3-0, 1-2 and 0-9 kbp respectively, in keeping with the restriction map generated from the pBR322-cloned material (Fig. 1) and the nucleotide sequence (Fig. 4), which revealed that the proposed coding region lay between HaeII sites 894 bp apart. These results indicate that the pilin gene occurs as a single copy at a unique locus within the B. nodosus genome.

Nucleotide sequence and predicted amino acid sequence

The nucleotide sequence (Fig. 4) contains an open reading frame with features of the pilin-coding sequence identified in other strains of B. nodosus (Elleman & Hoyne, 1984; Elleman et al., 1986b). These include a Shine–Dalgarno sequence (Shine & Dalgarno, 1974), 34AGGAG, ahead of a possible initiation codon. Although two adjacent ATG triplets are present, translation would most probably be initiated at the second triplet since a minimum of five bases is generally found between the Shine–Dalgarno sequence and the initiation codon (Stormo et al., 1982). With translation initiated at the second ATG triplet, the reading frame would code for a protein of M, 16683, a value in agreement with that estimated for the immunologically reactive protein expressed by the transformed E. coli cells. Codon usage within the predicted amino acid sequence reflects trends previously noticed with other pilin structural genes (Elleman & Hoyne, 1984; Elleman et al., 1986b). There is a preferential use of U rather than C in the third base position of quartet codons and a preferential use of A rather than G in the third base position of both quartet and duet codons.

Comparison of sequences within the coding region

The predicted amino acid sequence of the coding region of strain 340 was compared with the sequences of the same region from strains 198 and 265 (Fig. 5). Absolute identity occurs over the
Fig. 2. Immunoblot of *E. coli* cells harbouring recombinant plasmid. Lane A, purified *B. nodosus* 340 pili; lanes B and C, *E. coli* proteins from cells harbouring the pilin gene of *B. nodosus* 340 cloned as a *Hind*III fragment in opposite orientations in pBR322. The positions of *M*, standards are indicated to the left of the figure.

Fig. 3. Hybridization analysis. *B. nodosus* 340 DNA (0.5 μg) was digested and fractionated in a submerged agarose gel (1% w/v). After transfer to nitrocellulose membrane, the DNA was probed with a M13mp18-cloned fragment of the pilin gene of *B. nodosus* 340. Lanes A, B and C; *Hind*III, *Hind*III/KpnI, and *Hae*II digests respectively. The positions of size standards (*Eco*RI/*Hind*III-digested phage λ DNA) are indicated to the left of the figure.

first 25 residues of each sequence, a segment which has been suggested as a topogenic signal common to all pilins of the N-mPhe type (Elleman *et al.*, 1986c). Identity of sequence continues for an additional 32 residues between strains 340 and 265, and further extensive similarity throughout the remainder of these sequences results in an overall identity of 70%. In contrast, there is little homology between the strain 198 sequence and that of the other two strains beyond residue 39, and the total identity is not greater than 38% (Fig. 5). This value is within the range of identities found (33–43%) between *B. nodosus* pilins and the N-mPhe-type pilins of other bacterial species (Table 1).

Diagonal matrices (Staden, 1982) were used to compare the nucleotide sequences of the pilin-
**B. nodosus pilin**

5'GCCACGGTAAAAAGCGGCGGATTTCGGTTCTCTTCTGCTAAAGGCTGAT

ATAATAAGGCCGATGCAGATTTATGTTTATGCGCATATCGTTATTTTCGGGAAAAGCGCTGAGACGGGC

GAT

GTGCCATATATTTATCATATTATTCTCTTTTTTGAAACGCGGTTAGTTGCGGGAAAAATATGAGAAGAAAAATGC

AAAAGGCCGCGTGCCAGAAAAATAATTTTTTTTTTTTTTTCATAAAATAAATAATTTGTGATTTGAC

CATAATGAAAGGCCGCTCGCAACTGCTCTCAAAACAGATGATATTTAATGTTGATCATGAGC

GAG

Met Lys Ser Leu Gln Lys Gly Phe Thr Leu Ile Glu Met Ile Val Val Ala Ile

ATG AAA AGT TTA CAA AAA GGT TTT ACC TTA ATC GAA ATG GTT GTA GTC ATT 400

Ile Gly Ile Leu Ala Ala Ala Ile Pro Gln Tyr Gln Asn Tyr Ile Ala Arg Ser

ATC GGT ATC TTA GCT GCA ATC GCT ATT CCA CAA TAC CAA AAC TAC ATC GCT GCT TCA

Gln Val Ser Arg Val Met Ser Glu Thr Gly Gln Met Arg Thr Ala Ile Glu Thr Cys

CAA GGT AGC CGC GCT ATG TCA GAA ACT GGA CAA ATG CGC ACC GCC ACC GAA ACT ATG

Val Leu Asp Gly Lys Glu Ala Asp Lys Cys Phe Ile Gly Trp Thr Gly Ser Asn Leu

TTT TGG GAT GGT

Leu Asp Gly Phe Thr Leu Ala Ala Ala Ile Ala Ala Ala Thr Gly Gin Thr

TTG GAC GGT GAA TTT ACT GCT GGT ACC GAA TCT ACA GCA GCA ACA GGT CAA ACA

Gly Ile Thr Ile Lys Tyr Pro Gln Thr Leu Thr Trp Ser Arg Ser

GGA ATC ACA ATT AAG TAC CCG GCT GCA GCC GAC GAT GAG GGA AAT ATT TGT GCA ACA

Phe Gly Arg Asn Ala Ala Ala Lys Pro Gln Thr Leu Thr Trp Ser Arg Ser

TTT GGT CAA AAT GCT GCA GCC GGG GCT ATT AAA CCT CAA ACA TTG ACA TGG TCA GCT AGC

Lys Gly Thr Trp Thr Cys Ala Thr Thr Val Glu Ala Lys Phe Gin Pro Thr Gly

AAA GAA GGT ACT TGG ACA TGT GCA ACA ACT GGT GAA GCC AAA TTC CAA CCA ACA ACT GGT

Cys Lys Asp Gly Lys

TGT AAA GAT GGT AAA TAGTCTTCCTCGCAGAAACATCAAAGGTTACTTCGGTACACCTTTTTAT

AATTGATCAATACGTTAAAAAAATATTGATTTTTTTCGCGTTTTTCAATAAAAAGGTGATCCGACATGTGCT

GAGCCCTGATCATTTCCGCTTTGTCTTTTCTTTCTGTATGCTGAGCTGCGGC

3'

Fig. 4. Nucleotide sequence containing the pilin gene of *B. nodosus* 340. The predicted amino acid sequence is presented above the nucleotide sequence.

coding regions of *B. nodosus* (data not shown) and revealed homologies corresponding to those seen in the amino acid sequences. Nucleotide sequence similarity was apparent throughout the comparison of strains 340 and 265 while similarity between these two strains and strain 198 was limited to the sequences encoding the N-terminal region (residues 1 to 40) and the last 14 amino acids. Within the region encoding the invariant amino acids (residues 1 to 25), only three silent base changes have occurred between the strains, remarkably few considering the divergence of strains 340 and 265 relative to strain 198.
Fig. 5. Comparison of predicted amino acid sequences of pilin from \textit{B. nodosus} 198, 340 and 265. Sequences were aligned with the aid of the ALIGN program (Dayhoff, 1976); residues in one sequence differing from the other two sequences at corresponding positions are shaded.

Table 1. Percentage identity between predicted amino acid sequences of bacterial pilins

<table>
<thead>
<tr>
<th>Strain</th>
<th>B. nodosus 340</th>
<th>B. nodosus 198</th>
<th>M. bovis EPP63</th>
<th>N. gonorrhoeae MS11</th>
<th>P. aeruginosa PAK</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. nodosus 340</td>
<td>43</td>
<td>36</td>
<td>36</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>B. nodosus 198</td>
<td>36</td>
<td>39</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. bovis EPP63</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N. gonorrhoeae MS11</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of nucleotide sequences flanking the coding region

In an alignment of nucleotide sequences 3’ downstream of the proposed coding regions of the pilin gene from strains 198, 265 and 340, there is substantially greater similarity between strains 340 and 265 than between either of these two strains and strain 198 (Fig. 6). While strains 265 and 340 show 84\% identity over the sequences compared, strain 198 displays less than half the number of identities. Within the downstream sequence of each strain, a region of hyphenated dyad symmetry followed by a string of thymidine residues is present which might constitute a transcription terminator for RNA polymerase (Platt, 1986). The potential stem and loop structures of the terminators from strains 340 and 265 are identical, with free energies ($\Delta G_{25}^\circ$) of $-64$ kJ mol$^{-1}$ compared with $-77$ kJ mol$^{-1}$ for the strain 198 sequence (Elleman et al., 1986b).

Sequences upstream of the pilin-coding regions of strains 198, 265 and 340 (Fig. 7) show more extensive similarities ($>95\%$ identity) than the downstream regions. However, contrary to the relatively greater similarity of strains 265 and 340 in both pilin-coding regions and downstream sequences, the most extensive similarity (99\% identity) in the upstream region is between strains 198 and 265. Very few base changes have occurred in the sequence immediately adjacent to the pilin-coding region and a potential promoter of the pilin gene has been identified in this area (Johnson et al., 1986). Further upstream, an open reading frame is present in the opposite
B. nodosus pilin

Fig. 6. Comparison of nucleotide sequences 3' downstream of the coding region of the pilin gene from B. nodosus 198, 340 and 265. Bases in one sequence differing from the other two sequences at corresponding positions are shaded. The ALIGN program (Dayhoff, 1976) was used to maximize matches between the sequences. Pilin termination codons are underlined and the arrows denote a region of hyphenated dyad symmetry in each sequence, the presumed transcription terminator.

Fig. 7. Comparison of nucleotide sequences from the 5' upstream region of the pilin gene from B. nodosus 198, 340 and 265. The comparison begins at nucleotide 1 of Fig. 4 and ends at the nucleotide (356) immediately preceding the pilin-coding region. For the strain 198 and 265 sequences, only those bases that differ from the strain 340 sequence are shown. Regions of hyphenated dyad symmetry are indicated by arrows.
orientation to that of the pilin coding sequence (Fig. 7) and this extends considerably beyond the
presented data (T. A. A. Dopheide, personal communication). Within the open reading frame,
an ATG triplet (reverse complement of 141CAT) is well situated with respect to a potential
Shine–Dalgarno sequence, AAGGA (reverse complement of 150TCCTT) (Stormo et al., 1982).
Many of the differences between strain 340 and strains 198 and 265 are concentrated in the
extended ribosome-binding site (Stormo et al., 1982), including the first amino acid residues of
this potential coding region. A sequence motif possessing hyphenated dyad symmetry is present
mid-way between the initiation codons of the potential upstream coding region and the pilin-
coding region (Fig. 7).

**DISCUSSION**

Despite an absence of protein sequence data for pilin of *B. nodosus* 340, the hybridization and
immunoblotting results demonstrate that the DNA fragment isolated from this strain contains
the pilin structural gene. In addition, the amino acid sequence predicted from the coding region
of this gene has features characteristic of *B. nodosus* pilins and is especially similar to the protein
from strain 265. This similarity permits several predictions to be made with regard to the mature
pilin of strain 340. Previous comparisons of nucleotide and amino acid sequences from *B.
* nodosus 198 or 265 (Elleman et al., 1986b; McKern et al., 1983) have indicated that pilin is
encoded as a preprotein from which the first seven residues are cleaved post-translationally in *B.
*nodosus* (Elleman & Hoyne, 1984; Elleman et al., 1986a,b). Since the encoded sequences of the
pilin genes of strains 198, 265 and 340 are identical in the N-terminal region (Fig. 5), it is
reasonable to assume that the strain 340 gene also encodes a preprotein which is subsequently
cleaved in the same position as the preprotein of strains 198 and 265. The relative mobilities of
*B. nodosus* pilin and the immunologically reactive protein produced by *E. coli* cells harbouring
the pilin gene (Fig. 2) support this assumption since the former has a slightly greater mobility
than the *E. coli* protein from which the leader sequence is not removed (Elleman & Hoyne,
1984).

The similarity throughout the amino acid sequences of pilin from strains 265 and 340 also
allows a prediction to be made with regard to the disposition of disulphide linkages within the
strain 340 protein. Cysteine residues are present at positions 57, 67, 140 and 153 of the predicted
strain 340 sequence, and by analogy with pilin of strain 265, in which cysteine residues are
similarly located and of known linkage, disulphide linkages may exist between residues 57 and
67 and residues 140 and 153 of strain 340 pilin (Fig. 4). A disulphide loop near the C-terminus
has been recognized as a homologous feature of pilins from several bacterial species (Elleman et
al., 1986b). Although this feature is present in strains 265 and 340, representatives of serogroups
H and D respectively, it is absent from pilins of other *B. nodosus* serogroups, which contain a
single, larger and more centrally located disulphide loop (McKern et al., 1983; N. M. McKern,
personal communication).

Despite the extensive similarity between pilin of strains 340 (serogroup D) and 265 (serogroup
H), further processing by a central cleavage within the polypeptide chain, as found in pilins of
serogroup H, does not occur in pilins of serogroup D. Pilin of strain 265, although encoded as a
single-chain protein of *M*, 16637 (Elleman et al., 1968b), exists in pili largely as two subunits of
*M*, about 8000 (Anderson et al., 1986; Elleman et al., 1986b; Stewart et al., 1986), a presumed
proteolytic cleavage having occurred at an Ala–Ala bond between residues 79 and 80 (Fig. 5).
Sequence differences in the corresponding region of strain 340 pilin might account for the
absence of cleavage if the appropriate protease is present in this strain also. All other serogroups
of *B. nodosus* are like serogroup D in that pilin exists in pili as a single polypeptide chain
(Anderson et al., 1986; Stewart et al., 1986).

Comparison of pilin amino acid sequences from representatives of different *B. nodosus*
serogroups illustrates relationships beyond those recognized on the basis of serology (Claxton et
al., 1983). As noted above, 70% identity exists between strains 265 (serogroup H) and 340
(serogroup D). However, no greater than 38% identity is shown between the protein from either
of these strains and strain 198 pilin (serogroup A). In contrast, approximately 70% identity of
sequence exists between pilins from representatives of serogroups A, B, C and E (Elleman et al.,
An identity of about 70% is also found when the amino acid sequences of pilins from representatives of serogroups F (Dalrymple & Mattick, 1987) and G (Elleman & von Ahlefeldt, 1987) are compared with strain 198 pilin. There are thus two different types of B. nodosus pilins, those of serogroups D and H forming one subset and those of serogroups A, B, C, E, F and G another. This division of serogroups into subsets is further supported by comparison of the nucleotide sequences downstream of the coding region (Fig. 6): a considerably higher degree of similarity exists between strains 265 and 340 than between these sequences and that of strain 198. Conversely, the downstream sequence of strain 198 is virtually identical to strain 238 of serogroup G in this region (Elleman & von Ahlefeldt, 1987).

The comparison of pilins from B. nodosus with N-mPhe-type pilins of other bacterial species shows that the similarity between different species is about the same as between the subsets of B. nodosus (Table 1). Clearly, generation of the subsets has occurred through more extensive nucleotide rearrangement than the changes responsible for the diversity of serogroups within a subset. A single pilin gene is present in the genome of B. nodosus, so the opportunity for homologous recombination does not exist as it does in species carrying multiple pilin gene sequences such as N. gonorrhoeae (Hagblom et al., 1985). However, recombination between non-homologous sequences or foreign DNA may have given rise to sequence variation in the ancestral pilin genes of B. nodosus which resulted in formation of the subsets.

REFERENCES


