Primary Structure of Pilin Protein from Bacteroides nodosus Strain 216: Comparison with the Corresponding Protein from Strain 198

By NEIL M. MCKERN, IAN J. O'DONNELL, DAVID J. STEWART AND BRUCE L. CLARK

CSIRO, Divisions of Protein Chemistry and Animal Health, Parkville 3052, Victoria, Australia

(Received 9 May 1984; revised 13 August 1984)

The amino acid sequence of pilin protein from Bacteroides nodosus strain 216 was determined. The protein had a calculated molecular weight of 15962 and contained the same number of amino acid residues (151) as the pilin from the previously sequenced strain 198. The sequence of the first 44 residues was common to both strains, including the unusual amino-terminal amino acid, N-methylphenylalanine. Of the remaining 107 residues, 37% of them differed between the two strains. Comparison of hydrophilicity profiles constructed from the sequence data indicated that a conserved region around residues 71–72 was probably the site of an antigenic determinant.

INTRODUCTION

The principal organism responsible for producing footrot in sheep is Bacteroides nodosus. Pili from B. nodosus were used to produce vaccines in sheep which were highly effective against footrot induced by virulent strains from homologous serogroups of B. nodosus (Stewart, 1978a, b; Every & Skerman, 1982; Stewart et al., 1982, 1983). Strains of B. nodosus in Australia have been classified into at least eight distinct serogroups, based on K-agglutination tests (Claxton et al., 1983). An understanding of the molecular basis of antigenic variation between pili from different serogroups of B. nodosus may provide information necessary for the production of a synthetic peptide vaccine that is host-protective against all known strains of footrot. The amino acid sequence of B. nodosus strain 198 (serogroup A) was reported by McKern et al. (1983). In the present communication, the complete amino acid sequence of strain 216 (serogroup E) of B. nodosus is reported and compared with that of strain 198.

METHODS

Preparation of purified pilin. Pili from strain 216 (CSIRO, Animal Health Culture Collection) were isolated and purified essentially as described earlier for strain 198 (McKern et al., 1983; O'Donnell et al., 1983). Pilin subunits were reduced and converted to the S-carboxymethyl (SCM) derivative with [¹⁴C]iodoacetic acid (O'Donnell, 1973).

Enzymic digestion and chemical cleavage of pilin protein. Samples of 20 to 100 nmol pilin protein were digested in 0.05 M-ammonium bicarbonate for various times at 37 °C with one of the following enzymes at a concentration of 0.02 mg ml⁻¹: chymotrypsin (Worthington Biochemical Corp.) for 4 h; Staphylococcus aureus V8 protease (Miles Laboratories, Slough, UK) for 16 h; Armillaria mellea protease (a generous gift from the Pharmaceuticals Division, ICI, UK) for 4 h. Subdigestion of some peptides was carried out with TPCK-trypsin (Worthington Biochemical Corp.) for 2 h or with thermolysin (Calbiochem) for 2 h. Chemical cleavage after aspartic acid residues was achieved by incubation of protein at 110 °C under vacuum for 2 h in 2% (v/v) formic acid (Inglis, 1983).

Abbreviations: PTH, phenylthiohydantoin; SCM, S-carboxymethyl.

0001-1932 © 1985 SGM
Following each of the above treatments, solutions were lyophilized and redissolved in 100 to 200 μl 0·1% (v/v) trifluoroacetic acid (TFA). Insoluble material was removed by centrifugation and stored at −20°C for later analysis.

**High performance liquid chromatography (HPLC).** Soluble peptides from each digest and cleavage were dissolved in 0·1% (v/v) TFA and loaded onto a Waters μBondapak C₈ reverse-phase column, using a Waters HPLC system (Milford, Mass., USA). The elution of bound peptides with acetonitrile was monitored at 214 nm and 280 nm. Fractions were dried under vacuum at 50°C and then redissolved in 100 μl 50% (v/v) pyridine. Tubes containing ¹⁴C-labelled peptides were identified by counting radioactivity in 1 μl samples on the ¹⁴C channel of a Packard Tricarb liquid scintillation spectrometer.

**Sequence analysis.** Peptides were manually sequenced by a modified Edman procedure (Peterson et al., 1972) as described before (McKern et al., 1983), except that the volume of TFA in the cleavage step was reduced to 50 μl and the thiazolinone was converted to the phenylthiohydantoin (PTH) derivative by heating at 65°C for 2 min in 50 μl 25% (v/v) TFA.

PTH-amino acid residues were identified by separation on a Zorbax ODS column (Dupont) using a Hewlett Packard 1084A liquid chromatograph.

**RESULTS**

**Isolation of peptides**

Peptides purified by HPLC using a reversed-phase hydrophobic column were generally sufficiently pure to allow direct sequence analysis, except for some that did not bind to the column. In order to promote interaction with the column packing material, these peptides were reacted with phenylisothiocyanate and the resultant derivatives were re-chromatographed as before on the reverse-phase column. In this way, the previously nonbinding tri-peptide Lys-Ser-Gln was rendered sufficiently hydrophobic to bind to the column and was thereby separated from other components including free amino acids.

In some separations of peptides on HPLC, adjacent peaks often contained peptides with identical N-terminal sequences. This phenomenon, also observed by others (Vensel et al., 1983), was presumably due to incomplete enzyme cleavage at some sites in the protein, generating peptides of varying length but with the same N-terminus.

**Amino acid sequence of strain 216 pilin**

Pilin from strain 216 of *B. nodosus* contained 151 amino acid residues with a calculated molecular weight of 15962 in the unreduced form. The N-terminal residue was identified (by the chromatographic behaviour of its dansyl derivative on polyamide sheet) as the uncommon amino acid N-methylphenylalanine.

Sequence analysis of the peptides outlined in Fig. 1 was used to determine the unequivocal structure of the protein as follows: the sequence of the insoluble chymotryptic peptide C1 was identical to that of the first 19 residues of strain 198 (McKern et al., 1983), thus locating it at the amino terminus. Similarly, the exact correspondence of the sequences of peptides C2 and C3 with residues 20 to 37 of strain 198 positioned these peptides and enabled the sequence of strain 216 to be extended to residue 37. Peptide C4 was identified as starting at position 38 of the strain 216 sequence because of the identity of the first seven residues with residues 38-44 of strain 198. The overlap of peptide AMP2 continued the sequence to residue 62, further extensions to residue 95 being possible because of overlap of the subdigest peptides TH3/A3, C4/T3, TH3/A4 and the peptides AMP3, C5 and SA3. By analogy with the sequence of pilin from strain 198, the peptides SA3/A4, C5/T2 and C6 were positioned at residues 96 to 117. The overlapping peptides AMP4 and SA5 continued the sequence to residue 136. The C-terminal sequence of the molecule was ascertained by alignment of peptides C9 and C10 with the corresponding sequence in strain 198.

Evidence that the final residue in peptide C10 was the C-terminal residue came from the finding that peptides AMP6, SA6 and TH11 also terminated at this residue, as shown by both amino acid and sequence analysis. These results were confirmed by the finding that, upon digestion of the protein with carboxypeptidase Y, glutamine was the first amino acid released (data not shown).
Fig. 1. Primary structure of pilin from *B. nodosus* strain 216. Those peptides necessary for the elucidation of the sequence are shown. Peptides not sequenced to completion are indicated by arrows. C, chymotryptic peptides; SA, *Staphylococcus aureus* V8 protease peptides; AMP, *Armillaria mellea* protease peptides; T, tryptic peptides; TH, thermolytic peptides; A, peptides from dilute acid cleavage. The amino acid sequence of pilin from *B. nodosus* strain 198 is also shown for comparison. Residues that differ between the two strains are enclosed in boxes.
Fig. 2. Profiles of the hydrophilicity index (Hopp & Woods, 1981) of pilin from B. nodosus strain 216 (dashed line) and strain 198 (solid line).

Table 1. Amino acid composition of pilin from B. nodosus strain 216

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No. of residues per pilin molecule</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analysis*</td>
</tr>
<tr>
<td>CMC†</td>
<td>1.64</td>
</tr>
<tr>
<td>Asx</td>
<td>19.7</td>
</tr>
<tr>
<td>Thr</td>
<td>9.6</td>
</tr>
<tr>
<td>Ser</td>
<td>6.2</td>
</tr>
<tr>
<td>Glx</td>
<td>16.3</td>
</tr>
<tr>
<td>Pro</td>
<td>3.9</td>
</tr>
<tr>
<td>Gly</td>
<td>16.5</td>
</tr>
<tr>
<td>Ala</td>
<td>20.9</td>
</tr>
<tr>
<td>Val</td>
<td>6.4</td>
</tr>
<tr>
<td>Met</td>
<td>0.5</td>
</tr>
<tr>
<td>Ile</td>
<td>10.3</td>
</tr>
<tr>
<td>Leu</td>
<td>12.8</td>
</tr>
<tr>
<td>Tyr</td>
<td>6.4</td>
</tr>
<tr>
<td>Phe</td>
<td>2.7</td>
</tr>
<tr>
<td>Lys</td>
<td>13.1</td>
</tr>
<tr>
<td>His</td>
<td>1.3</td>
</tr>
<tr>
<td>Arg</td>
<td>2.7</td>
</tr>
<tr>
<td>MePhe†</td>
<td>ND</td>
</tr>
<tr>
<td>Total</td>
<td>150.9</td>
</tr>
</tbody>
</table>

* Samples were hydrolysed for 18 h at 110 °C under vacuum in constant-boiling HCl containing 0.01% phenol. ND, not determined.
† CMC, carboxymethylcysteine; MePhe, N-methylphenylalanine.

The amino acid analysis of intact SCM-pilin from strain 216 was in agreement with the composition calculated from the sequence data (Table 1), after making allowance for the juxtaposition of several valine and isoleucine residues in the sequence, which can give rise to low yields of these amino acids after acid hydrolysis (Tsugita & Scheffler, 1982).

The presence of Asn-Gly in the sequence at positions 53–54, 95–96 and 128–129 made it difficult to sequence past these residues, presumably due to cyclization which occurs under the acidic cleavage conditions during sequencing (Bornstein & Balian, 1977). However, by shortening the acid cleavage step in peptides containing this bond to approximately 2 min instead of the usual 6 min, sufficient sequence data were obtained to enable residues past the Asn-Gly bond to be unequivocally assigned.
Comparison of hydrophilicity profiles (Hopp & Woods, 1981) of pilin from strains 216 and 198 revealed a similarity of patterns (Fig. 2). Of particular interest was the conservation of the largest peak with its apex at residues 71–72, since the data of Hopp & Woods (1981) suggested that the point of highest local hydrophilicity was associated with an antigenic determinant of the protein. On this basis it is likely that the largely conserved region flanking and including residues 71–72 is both an antigenic site and (as a corollary) a surface-located region of the whole pilus. It is of interest to note that one of the antigenic regions identified in P. aeruginosa PAK pilin spans residues 71–72 in that molecule (Watts et al., 1983).

Three regions of the hydrophilicity profile of pilin from strain 216 differed substantially from the equivalent region of pilin from strain 198 (Fig. 2). These differences (residues 53–58, 85–94 and 132–140) reflected regions of the sequence where there was little identity between the two proteins (Fig. 1), with only two out of six, two out of ten and two out of nine residues, respectively, conserved.

**DISCUSSION**

The N-terminal 44 residues in the pilin from strain 216 of B. nodosus were identical to those from strain 198 (Fig. 1) and this sequence (which was extremely hydrophobic for the first 24 residues) was homologous with the N-terminal sequence of pilin from Pseudomonas aeruginosa (Sastry et al., 1983), Moraxella nonliquefaciens (Froholm & Sletten, 1977) and Neisseria meningitidis (Hermodson et al., 1978). In addition, the N-terminal residue of each of these proteins was the uncommon amino acid N-methylphenylalanine. The conservation of these characteristics between different families of bacteria suggested a special role for the N-terminal region of the subunit pilin protein. This function is likely to be related to the assembly and association of subunits to form the pilus structure, via hydrophobic interactions. Further comparison of the sequence of pilin from strain 216 with strain 198 (Fig. 1) revealed that, in contrast to the absolute identity of the first 44 residues, 37% of the remaining residues differed between the strains [having first made corrections to the published sequence of pilin from strain 198 (McKern et al., 1983) at residues 73, 95 and 128]. Significant sequence differences were expected, since these strains fell into different serogroups as classified by Claxton et al. (1983) – serogroup A for 198 and serogroup E for 216.

It was apparent from the sequence data that there was a deletion of two residues from the sequence of pilin from strain 216 around residue 132 but an addition of two residues at the C-terminus, when compared with the sequence from strain 198. Although it is possible that there is some structural advantage in this conservation between strains of the number of residues per molecule, the fact that pilin from Pseudomonas aeruginosa PAK (Sastry et al., 1983) contains only 144 residues per molecule suggests that 151 residues per pilin chain are not essential to form functional pili of this type.

Absorption with homologous or heterologous pili of ovine antisera to pili of strain 198 or 216 revealed 8–25% cross-reaction between pili from the different strains, as measured by ELISA (D. L. Emery, personal communication). The presence of antigenic specificities unique to each pilus serotype is also inferred from bacterial agglutination tests, where the cross-reactivity is less than 1% of the homologous reaction.

The present work has enabled the regions of conservation and variability between the sequence of pilin from two serogroups of B. nodosus to be discerned. It remains to be determined whether the regions of variability between these two sequences accurately reflect the positions of antigenic determinants on the assembled pilus structure. Further sequence studies, however, on strains from the six other known serogroups of B. nodosus may well provide a comprehensive understanding of the regions of these molecules that are antigenically important.

The authors are indebted to Miss Sue Arendshorst, Mr David Jones and Miss Jill Vaughan for expert technical assistance and to Dr Charles Roxburgh and Mr Neil Cook for amino acid and PTH-amino acid analyses.
N. M. McKern AND OTHERS

REFERENCES


