Molecular analysis of a *Leptospira borgpetersenii* gene encoding an endoflagellar subunit protein

MARINA MITCHISON, JULIAN I. ROOD, SOLLY FAINÉ and BEN ADLER*

Department of Microbiology, Monash University, Clayton, Victoria 3168, Australia

(Received 21 January 1991; revised 15 March 1991; accepted 26 March 1991)

A flagellin gene, flaB, from *Leptospira borgpetersenii* (formerly *L. interrogans*) serovar hardjo was cloned and expressed in *Escherichia coli*. Expression of the 32 kDa FlaB protein was dependent upon the lacZ promoter from pUC18. Nucleotide sequence data showed an open reading frame encoding 283 amino acid residues, corresponding to a protein of molecular mass 31.3 kDa. The G+C content of the flaB gene was 54.7 mol%. Comparison of the deduced FlaB amino acid sequence with flagellins from other bacteria revealed a high level of identity with the *Treponema pallidum* FlaB proteins.

Methods

**Bacterial strains and plasmids.** *L. borgpetersenii* serovar hardjo strain L171 (Farrelly *et al.*, 1987) was cultured in Tween 80-bovine albumin (EMJH) medium with added pyruvate (Johnson *et al.*, 1973). All *Escherichia coli* recombinants were derivatives of strain DH5α (Bethesda Research Laboratories) and were cultured in 2×YT medium (Vieira & Messing, 1982), supplemented with ampicillin (100 μg ml⁻¹) as required. All recombinant plasmids were derivatives of the cloning vector pUC18 (Yanisch-Perron *et al.*, 1985).

**Preparation and analysis of DNA.** Whole-cell leptospiral DNA was prepared from serovar hardjo strain L171 as described previously (Marshall *et al.*, 1984; Doherty *et al.*, 1989). *E. coli* plasmid DNA was isolated by the alkaline lysis method (Birnboim & Doly, 1979). Ligation, transformation and agarose gel electrophoresis methods were as described by Maniatis *et al.* (1982). Restriction endonucleases, T4 ligase and calf intestinal alkaline phosphatase (molecular biology grade) were from either Promega or Boehringer-Mannheim.

**Preparation of *L. borgpetersenii* gene bank.** *L. borgpetersenii* DNA (20 μg) was partially digested with Sau3A, excised from a 1% low-melting-point agarose gel and ligated overnight at 15 °C, within the gel matrix, with 1 μg of *BamHI*-digested, dephosphorylated pUC18 (Yanisch-Perron *et al.*, 1985). The ligation mixture was used to transform competent *E. coli* DH5α cells to ampicillin resistance. Transformants were selected on 2×YT-ampicillin (100 μg ml⁻¹) plates containing X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside). Colourless recombinant clones were screened by colony immunoassay as described previously (Doherty *et al.*, 1989).

**Antiserum production.** All antisera were produced in New Zealand White rabbits. Antiserum used in the colony immunoassay, SB288, was kindly provided by S. Ballard (Department of Microbiology, Monash University) and contained antibodies against a large range of leptospiral serovar hardjo proteins as assessed by Western blotting against leptospiral whole-cell lysate. It was absorbed with lysed *E. coli* DH5α cells prior to use. The R390 antiserum used to detect the flagellar subunits in Western blotting was prepared against isolated, SDS-PAGE-denatured flagellum proteins consisting primarily of the genus *Treponema pallidum* FlaB proteins.
cross-reactive 33–34 kDa doublet (Kelson et al., 1988), from *L. interrogans* serovar pomona (Boulard & Lecroisey, 1982).

**Protein electrophoresis and immunoblotting.** Leptospiral and *E. coli* components were separated by electrophoresis on discontinuous 10–15% (w/v) SDS-polyacrylamide gels (Laemmli, 1970). Leptospiral sonicates were prepared as described previously (Adler et al., 1980). *E. coli* preparations were prepared in sample buffer containing SDS (1.5% w/v), bromophenol blue (0.002% w/v) and glycerol (4% v/v) immersed in boiling water for 10 min and then centrifuged at 11600 × *g* for 5 min prior to loading. Flagellar protein was prepared by boiling in sample buffer for 5 min. Electrophoresis conditions, Western blotting and immunostaining procedures have been described previously (Chapman et al., 1987), as has the method for eluting antibodies from nitrocellulose (Beall & Mitchell, 1986).

**Isolation of flagella.** Flagella isolated from approximately 5 × 10⁹ cells of strain L171 were separated as described by Chang & Faine (1970) and were further purified by caesium chloride density-gradient centrifugation (Kelson et al., 1988).

**DNA hybridization.** Southern blots were performed as described by Southern (1975). Randomly primed DNA was labelled with [³²P]dATP (Amersham) using an oligolabelling kit (Bresatec) following the supplied protocol. Prehybridization and hybridization solutions contained 50% (v/v) formamide, 1% (w/v) SDS, 0.5% (w/v) skim milk powder, 4 mM-EDTA, 32 mM-NaOH, 40 mM-NaH₂PO₄, 0.72 mM-NaCl and 0.3 mg salmon sperm DNA ml⁻¹. Following hybridization overnight, blots were washed under high-stringency conditions prior to autoradiography at ~70 °C using intensifying screens.

**DNA sequencing and computer analysis.** DNA sequencing was performed using [α-³²P]dATP (Amersham) and a T7 sequencing kit (Pharmacia) by following the manufacturer’s protocol for double-stranded sequencing. DNA to be sequenced was purified on a caesium chloride density gradient or by using Gene-clean (Bio 101). Unequivalent sequence was obtained on both DNA strands. Nucleotide and amino acid sequence data were analysed using MELDBBYS, a suite of programs developed by the Walter and Eliza Hall Institute, the Ludwig Institute for Cancer Research and the Howard Florey Institute, Melbourne, Australia. The Treealign program, which performs phylogenetic alignment of homologous sequences (Hein, 1989a, b), was used to compare the FlaB amino acid sequence with other flagellin proteins.

**Results**

**Cloning and definition of the flaB coding region**

Recombinant clones containing chromosomal DNA isolated from serovar *hardjo* strain L171 were initially screened by colony immunoassay using SB288 anti-*hardjo* antiserum. Positive clones were further analysed by Western blotting. One clone, carrying the recombinant plasmid pLBA21, reacted weakly in colony immunoassays and two weak bands, 32 and 28 kDa in size, were detected by Western blotting. To determine whether the cloned products were flagellar antigens they were further analysed by Western blotting using R390 antiserum raised against denatured flagellar components. The antiserum reacted strongly with the 32 and 28 kDa bands (Fig. 1a). Antiserum raised against undenatured whole flagella did not react with the cloned products by Western blotting (data not shown).

The plasmid pLBA21 (Fig. 2) contained a 3.9 kb *Sau3A* leptospiral DNA fragment inserted into the *Bam*HI site of pUC18. By subcloning and subsequent testing for antigen expression using Western blotting with R390 antiserum, the gene encoding the 32 kDa protein was localized to a central 1.2 kb *Nru*I fragment (Fig. 2), which also contained an internal *Nru*I site. In accordance with the terminology used for the cloned *T. pallidum* genes (Pallesen & Hindersson, 1989; Champion et al., 1990), the gene encoding the 32 kDa protein was designated *flaB*. There was no expression of the 32 or 28 kDa proteins when the entire 3.9 kb region encoded by pLBA21, or the 1.2 kb *Nru*I fragment from pLBA38, was inserted in the opposite orientation relative to the lacZ promoter (Fig. 2). Expression of the cloned products encoded by pLBA21 therefore appeared to be under the control of the lacZ promoter in pUC18, indicating a direction of transcription from left to right in Fig. 2. Interestingly, the strain harbouring the *Pst*I subclone pLBA31 produced an 18 kDa protein identified by R390 antiserum. To confirm the origin of the insert, chromosomal DNA isolated from strain L171 was digested with *Eco*RV and *Cla*I and probed by Southern blotting using the 1.2 kb *Nru*I fragment from pLBA38. The probe hybridized to a single 3.4 kb chromosomal band which was of identical size to the fragment obtained after *Eco*RV/*Cla*I digestion of pLBA21 (data not shown). To determine whether any other regions of pLBA21 had *flaB*-related sequences, a 0.5 kb *Pst*I/SalI fragment contained wholly within the *flaB* gene was used to probe pLBA21 digested with *Nru*I. The probe hybridized to three bands, of 1.2, 0.7 and 0.5 kb, which corresponded to the partial *Nru*I digestion of the central 1.2 kb *Nru*I region encoded by pLBA38 (Fig. 2).

**Serological relatedness of the cloned antigens**

A cell lysate of *E. coli* containing pLBA21 was separated by SDS-PAGE, and following Western transfer and overnight incubation with anti-flagellar serum (R390), antibodies were eluted from the 32 and 28 kDa protein bands. These antibodies were reacted by Western blotting with serovar *hardjo* L171 sonicate, isolated flagella and a cell preparation of the FlaB clone, harbouring pLBA21 (Fig. 1b, c). Antibodies eluted from the 32 kDa band reacted with the 28 kDa band and also with the 33 kDa band present in both the leptospiral sonicate and the purified flagella preparation and corresponding to the lower doublet band described by Kelson et al. (1988). Antibodies eluted from the 28 kDa band showed a similar reactivity pattern, and reacted with the 32 kDa band.
Fig. 1. Western blot analysis of the cloned flagellin proteins immunostained with anti-flagellar serum R390 (a), antibodies eluted from the 32 kDa FlaB band (b), or antibodies eluted from the 28 kDa band (c). Lanes: A, E and H, L171 whole-cell sonicate; B, F and I, purified L171 flagella; D, G and J, cell extracts prepared from E. coli DH5α cells containing pLBA21; C, control strain of E. coli DHSa containing pUC18. The 32 kDa FlaB protein and the 28 kDa protein (p28) are indicated by arrows on the right. Minor bands in lanes G and J are due to co-eluted E. coli antibodies and are also seen in the E. coli control (lane C). The numbers on the left indicate the positions of the standard molecular mass markers.

**Nucleotide sequence of the cloned flaB gene and the deduced amino acid sequence**

DNA sequence analysis of a series of overlapping subclones was used to obtain the unequivocal nucleotide sequence of the 1.2 kb NruI fragment from pLBA21, on both DNA strands (Fig. 3). An open reading frame of 849 nucleotides, corresponding to 283 amino acids, was identified. The calculated molecular mass of the encoded protein was 31.3 kDa. A consensus E. coli Shine–Dalgarno sequence was present 7 base pairs upstream from the initiation codon. An inverted repeat sequence \( \Delta G = -22.8 \text{ kcal mol}^{-1} \) \((-95.4 \text{ kJ mol}^{-1})\) which may function as a rho-independent transcriptional terminator was located 37 bases downstream of the stop codon (Tinoco et al., 1973). There were no identifiable promoters upstream of flaB within the sequenced region. The G + C content of flaB was 54-7 mol\%, in contrast to the reported 39 mol\% G + C content of serovar hardjo genomic DNA (LeFebvre et al., 1987). The encoded protein contained 31 (10-9\%) acidic and 33 (11-7\%) basic amino acids. Eighty-nine (31\%) of the amino acids were hydrophobic, but there were no significant hydrophobic domains within the protein. There was a greater similarity of codon usage between the flaB gene and E. coli genes (Ikemura & Ozeki, 1982) than with codon usage observed with the sphingomyelinase C gene from serovar hardjo (Segers et al., 1990), in particular for the amino acids Gln, His, Phe, Tyr and Val. Interestingly, CCG was the only codon used for the amino acid Pro.

**Comparative amino acid sequence analysis of the flagellin proteins**

The predicted amino acid sequence of the serovar hardjo L171 FlaB protein had significant sequence similarity with flagellar proteins from other bacteria (Fig. 4). The aligned sequences do not include Salmonella typhimurium, Campylobacter coli, Rhizobium meliloti, Serratia marcescens or E. coli flagellins due to the greater length of these amino acid sequences. However, these proteins are included in a dendrogram derived from an alignment of the amino acid sequences (Fig. 5). The alignment data revealed a close relationship between the
**Fig. 2.** Restriction maps of pLBA21 and derived subclones. The location of the flaB gene is shown. Arrows for each plasmid indicate the direction of transcription from the lac promoter. Proteins were detected with anti-flagellar serum R390 by Western blotting analysis of the recombinant strains harbouring the plasmids. A dash indicates that no proteins were detected.

**Discussion**

The flagellin proteins from a diverse range of bacteria have previously been cloned, including those from *B. burgdorferi* and *T. pallidum* which, like leptospires, are members of the order *Spirochaetales*. Leptospiral flagellar components have been identified in SDS-PAGE gels as four major protein bands of 31 kDa, 37 kDa and a 33–34 kDa doublet (Kelson et al., 1988). The doublet observed in profiles of isolated leptospiral flagella has previously been shown to be conserved among the different genera of the family *Leptospiraceae* (Kelson et al., 1988). A similar flagellar doublet profile was observed in flagella isolated from *T. pallidum* and shown to consist of two subunit proteins, FlaB1 and FlaB2. Here we report the cloning and expression in *E. coli* of a flagellin gene from *L. borgpetersenii*. The recombinant clone carrying pLBA21 produced two proteins, both of which were expressed under the control of the lacZ promoter. These consisted of the 31 kDa FlaB protein and a 28 kDa protein designated p28. Antisera raised against denatured flagellar proteins from an SDS-PAGE gel reacted with the cloned proteins. However, antisera raised against non-denatured leptospiral flagella did not react with the FlaB or p28 proteins, suggesting that the epitopes recognized by these antibodies were only exposed by the denaturation of flagellin. The gene encoding FlaB was localized to a central NruI region of pLBA21. A PstI site was present six nucleotides upstream of the flaB stop codon and yet the PstI subclone pLBA31 produced an 18 kDa protein that was much
Fig. 3. Nucleotide sequence of the 1207 bp \textit{NruI} fragment of pLBA38 and deduced amino acid sequence. An \textit{E. coli} consensus Shine-Dalgarno sequence is underlined at position 136. A second potential ribosome-binding sequence is also underlined at position 213 followed by a putative internal initiation codon at position 225. The translation stop codon is indicated by the asterisk. A potential transcription termination signal sequence downstream of \textit{flaB} is double underlined.
<table>
<thead>
<tr>
<th>Species</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. borgerdorferi</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>T. pallidum FlaB1</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>T. pallidum FlaB2</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>T. pallidum FlaB3</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>Bor. burgdorferi</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>Bac. subtilis</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>R. cecilola</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
<tr>
<td>C. crescentus</td>
<td>MI--INHLSAVNESSSFLAVKTLKKASSRIRNNGDASGLVSERKALNINEFDDGHTGS</td>
</tr>
</tbody>
</table>

Fig. 4. Alignment of the predicted amino acid sequences of the flagellin proteins from _Leptospira borgerdorferi_ (this work), _Treponema pallidum_ FlaB1 and FlaB1 (Champion et al., 1990), _T. pallidum_ FlaB2 (Pallesen & Hinderos, 1989), _Borrelia burgdorferi_ (Wallich et al., 1990), _Bacillus subtilis_ (LaVallie & Stahl, 1989), _Roseburia cecilola_ (Martin & Savage, 1988) and _Caulobacter crescentus_ (Gill & Agabian, 1983). The gene products were aligned using the Treealign program. Identical amino acids are boxed.
smaller than the predicted size of the truncated product (31 kDa). It is possible that the truncated product is less stable than the complete protein and is degraded to produce the 18 kDa protein seen in Western blots.

The FlaB protein and the p28 protein were antigenically related, as shown from the results of experiments in which specific antibodies eluted from each of the cloned proteins reacted with each other and with only the 33 kDa flagellar protein in leptospires. Interestingly, antibodies eluted from either FlaB or p28 did not react with the larger of the two leptospiral flagellar doublet proteins, indicating that these proteins were antigenically unrelated. This was in contrast to the findings in T. pallidum, where the two components of the doublet were both flagellin subunit proteins (FlaB1 and FlaB2) and were antigenically related. Further studies are required to determine whether the larger leptospiral protein constitutes a second flagellin subunit, which is antigenically distinct from FlaB.

The location of the gene encoding p28 is uncertain. Within the flaB gene coding region there is a potential in-frame internal initiation (nucleotide 225) codon preceded by a weak ribosome-binding site (Fig. 3). The calculated molecular mass of a protein initiating at this site is 28.3 kDa. The existence of a second gene encoding p28 is unlikely as none of the subclones expressed p28 (Fig. 2). Moreover, in Southern blots only the NruI fragments encoded by pLBA38 hybridized with the flaB gene. An alternative explanation for our data is that p28 is a breakdown product of FlaB, especially since additional bands of lower molecular mass were occasionally observed in profiles of strains carrying pLBA21.

The apparent molecular mass of the cloned FlaB protein in SDS-PAGE gels was approximately 32 kDa, close to that calculated from the deduced amino acid sequence (31.3 kDa). The size of the cloned polypeptide was smaller than the 33 kDa flagellar protein in leptospires. The precise reason for this difference is unclear. One possibility may be post-translational modification of the protein in leptospires, such as the addition of a lipid moiety. However, the protein contains neither cysteine residues nor any identifiable putative signal peptide sequences which are necessary for the addition of lipids (Yamaguchi et al., 1988). Although rarely reported in bacteria, the possibility nevertheless exists of the addition of sugar residues to the protein (Lechner, 1989).

Indeed the flagellins of Halobacterium halobium are glycosylated and the amino acid sequence which surrounds N-glycosidic linkages is commonly of the formula Asn-X-Thr(Ser) (Wieland et al., 1985). This sequence is present in the leptospiral FlaB protein six amino acids from the N-terminus. Previous comparisons of flagellins from different genera of bacteria have revealed extensive homology in the N-terminal and C-terminal regions and these regions in leptospiral flagellin are also highly conserved (Fig. 4). The amino acid sequence alignment also showed a high degree of homology between FlaB proteins from L. borgpetersenii strain L171 and T. pallidum. The G+C content of the leptospiral flaB gene is significantly higher than the reported 39 mol% G+C content of the leptospiral genome (LeFebvre et al., 1987), but remarkably similar to the G+C content of the T. pallidum genome (Miao & Fieldsteel, 1978) and the T. pallidum flaB genes (Pallesen & Hindersson, 1989; Champion et al., 1990). Based on these results, we suggest that the FlaB protein from L. borgpetersenii and the FlaB proteins of T. pallidum may have a common ancestral gene. This hypothesis is consistent with the close phylogenetic relationship between the FlaB proteins from the two species (Fig. 5). The mechanism by which the gene may have been introduced into L. borgpetersenii is not known since genetic exchange has not been demonstrated with these organisms.
This work was supported by a project grant from the National Health and Medical Research Council, Canberra, Australia. The authors wish to thank Trudi Bannam, Susan Ballard, Margaret Katz and Pauline Howarth for their advice and technical assistance.

References


HEIN, J. (1989a). A new method that simultaneously aligns and reconstructs ancestral sequences for any number of homologous sequences, when the phylogeny is given. Molecular Biology and Evolution 6, 649-668.


