Molecular analysis of a gene encoding a serum-resistance-associated 76 kDa surface antigen of *Haemophilus somnus*

S. P. Cole,1,2* D. G. Guiney2 and L. B. Corbeil1

Departments of Pathology1 and Medicine2, 8416, University of California San Diego Medical Center, 200 W. Arbor Dr., San Diego, California 92103-8416, USA

(Received 9 February 1993; revised 17 May 1993; accepted 25 May 1993)

*Haemophilus somnus* is a Gram-negative bacterial bovine pathogen which can cause disease or be carried asymptomatically. We previously showed that four serum-sensitive isolates from asymptomatic carriers lacked a 13.4 kb sequence of chromosomal DNA that was present in two virulent serum-resistant strains. We have since sequenced 5 kb of the 13.4 kb fragment from a serum-resistant strain, which contained an open reading frame (ORF) of at least 4.5 kb. From Western blot analysis, the ORF was shown to encode a 76 kDa protein (p76) that co-migrated with a 76 kDa *H. somnus* surface protein. Both the recombinant and natural p76 reacted with convalescent-phase serum from a cow in an experimental *H. somnus* abortion study. The translational start site for p76 was identified by deletion analysis of subclones of the 5 kb cloned sequence. The 4.5 kb ORF contained 1-2 kb tandem direct repeats (DRs), with 65% identity between the two repeats at the protein level. The 5' DR (DR1) included the start site for the 76 kDa protein, and DR2 had a flanking inverted repeat, suggestive of an insertion-sequence-like element.

Introduction

The bovine pathogen *Haemophilus somnus* is the causative agent of a wide range of both septicaemic and nonsepticaemic syndromes including pneumonia, infertility and the sequelae of septicaemia: abortion, meningoencephalitis, arthritis and myocarditis (for reviews, see Harris & Janzen, 1989; Humphrey & Stephens, 1983; Kwiecien & Little, 1991). The infection may be sexually transmitted also (Kwiecien & Little, 1991), since a genital asymptomatic carrier state is common. Most normal bulls are preputial carriers and some normal cows are vaginal carriers (Kwiecien & Little, 1991). Others have shown that genital isolates from asymptomatic carriers are avirulent or less virulent than isolates from diseased animals (Groom et al., 1988; Humphrey, 1982; Humphrey & Stephens, 1983; Inzana & Todd, 1992). Our own studies (Corbeil et al., 1985) demonstrated that several preputial isolates from asymptomatic carriers were serum-sensitive, whereas all isolates that could be causally associated with disease were serum-resistant. The resistance to killing by fresh bovine serum was associated with the presence of 270 and 41 kDa outer-membrane proteins (OMPs) (Widders et al., 1989) and/or lipooligosaccharide with multiple bands on silver-stained gels (Inzana & Todd, 1992). The 270 and 41 kDa OMPs (p270 and p41) were both immunoglobulin-binding proteins (IgBPs; Yarnall et al., 1988a, b).

A cosmid library from a virulent, serum-resistant strain produced a clone that expressed two additional but related serum-resistance-associated proteins at 76 and 120 kDa (p76 and p120) that reacted with antiserum to the 270 kDa IgBP (Corbeil et al., 1988; Cole et al., 1992). The proteins were visualized by Western blot, reacting with protective convalescent serum or antiserum to the 270 kDa IgBP. Later, the genes for these two proteins were localized to a 13.4 kb sequence (Cole et al., 1992). Southern blots showed that two virulent, serum-resistant isolates contained this sequence, whereas four serum-sensitive strains from asymptomatic carriers lacked this segment of DNA (Cole et al., 1992). Additional Southern blots with probes constructed from fragments of each of these two genes showed no cross-hybridization, suggesting that p76 and p120 had no significant DNA homology. Because both proteins reacted with polyclonal antibody to p270, the lack of

**Abbreviation:** DR, direct repeat.

* The GenBank accession number for the sequence of the entire 4.5 kb ORF reported in this paper is L10282.
homology between the two indicated that p270 may be a polymer of p76 and p120.

In this paper, we present the sequence of the gene for the 76 kDa protein. By analysis of Western blots of proteins expressed by subclones, we determined the probable start codon of the p76 gene. We also report the existence of a large insertion-sequence-like direct repeat within this gene and upstream regions of H. somnus chromosomal DNA.

Methods

Bacterial strains and plasmids. H. somnus isolates were grown in brain heart infusion medium (Difco) supplemented with 0.1% Trizma base and 0.001% thiamin monophosphate (BHITT), as described previously (Inzana & Corbeil, 1987). Solid medium was BHITT agar plus 5-10% (v/v) calf blood, as described (Cole et al., 1992). Serum-resistant strain 2336, from which all the subclones were derived, was isolated from a pneumatic calf and has been used to reproduce H. somnus pneumonia (Gogolewski et al., 1987). The subclones in this paper were derived from cosmid pHs1 (Corbeil et al., 1991), containing a 35 kb insert of H. somnus DNA in E. coli DH1. Transformants in E. coli DH5α were grown in LB medium (Sambrook et al., 1989) supplemented with ampicillin at 100 μg ml⁻¹.

Recombinant DNA techniques and subclone construction. DNA isolation for construction of plasmids and subclone analysis was by the boiling method (Holmes & Quigley, 1986). All other DNA manipulations were carried out using standard protocols (Sambrook et al., 1989). Plasmids pHs139 and a derivative, pHs138, were constructed as previously described (Cole et al., 1992). Briefly, pHs139 consisted of a 5 kb Xbal–PvuII insert from pHs1 in pUC19, while the derivative pHs138 contained a 3 kb Hpal–PvuII insert cloned in pUC18. Deletion derivatives of pHs138 were made using appropriate restriction enzymes, followed by religation of the plasmid. Specifically, pHs176 was constructed by cloning the PvuI–PstI fragment from pHs138 into the Smal–PstI site of pUC18. This resulted in the p76 gene lying just downstream of the pUC18 lacZ start site. Plasmid pHs182 was made by treating pHs176 with EcoRI, filling in the ends with the large fragment (Klenow) of DNA polymerase, then religating the plasmid. This treatment placed the p76 gene out of frame with respect to the lacZ ATG start site. Finally, plasmid pHs171 was prepared by digestion of pHs138 with BstXI and EcoRI, followed by religation. All of these plasmids were tested by restriction analysis and gave the appropriate sized fragments.

SDS-PAGE and Western blots. Western blots were from 8% SDS-polyacrylamide gels of whole cell lysates. The cells were prepared for loading in sample buffer as previously described (Cole et al., 1992). Briefly, cultures were adjusted to an optical density of 0.4 at 600 nm, and then a 1.5 ml vol. was pelleted and resuspended in 100 μl of sample buffer prior to loading 3 μl in the wells. After electrophoresis, proteins were electrotransferred to nitrocellulose (Schleicher and Schnell) as previously described (Cole et al., 1992). Filter blots were made from the gel, and reacted with a 1:2000 dilution of serum P3, as before (Corbeil et al., 1988; Cole et al., 1992). Serum P3 was convalescent phase serum taken from a cow in an experimental abortion study using a serum-resistant strain of H. somnus (Widders et al., 1986). This serum reacts with the 270 kDa, 120 and 76 kDa antigens as well as other H. somnus antigens not expressed in the E. coli recombinants reported in this study (Corbeil et al., 1988; Cole et al., 1992). Specific antisem to p270 was raised in rabbits by inoculation of the gel-purified antigen as previously described (Yarnall et al., 1988b). The antisem reacts only with H. somnus antigens p270, p120 and p76, even though there is no significant DNA homology between p76 and p120 (Cole et al., 1992). This suggested that p76 and p120 may be subunits of p270 (Cole et al., 1992). Gamma-Bind G-horse radish peroxidase conjugate (Genex Corp.) at a 1:2000 dilution was used to detect antigen-antibody reactions, with the addition of 4-chloro-1-naphthol and hydrogen peroxide.

Sequencing strategy. Double-stranded sequencing was carried out using dideoxynucleotide termination reactions with T7 DNA polymerase in a multiwell microtitre plate system (Amersham). To resolve some ambiguous sequences, Taq polymerase (Promega) was used at higher temperatures. Clones and subclones were sequenced using both forward and reverse universal pUC primers and overlapping areas with synthetic primers (from Molecular Biology Core Facility, University of California, San Diego, CA).

Data analysis. DNA was analysed using DNA Strider (shareware) and MacVector software (International Biotechnologies Inc.).

Results and Discussion

Cosmid pHs1 contained a 35 kb insert from H. somnus that expressed a 76 kDa antigen and low levels of a group of H. somnus antigens centred at 120 kDa on SDS-PAGE gels when expressed in E. coli (Corbeil et al., 1988). A 12 kb section of the cosmid insert was subcloned into the vector pUC19 (creating pHs134). This plasmid expressed high levels of the 120 kDa group of proteins, as well as the p76 and lower molecular mass proteins (Cole et al., 1992). Further subcloning of this insert into pHs139 and pHs140 on pUC-based vectors resulted in the expression of 120 kDa proteins from pHs140, while pHs139 produced large amounts of p76, as well as lower molecular mass proteins (Cole et al., 1992). Expression was determined by Western blots using cell lysates of the subclones reacted with convalescent phase serum or polyclonal antisem to p270, which reacts with both p76 and p120 (Cole et al., 1992). The 5 kb Xbal–PvuII insert in plasmid pHs139 was subcloned in pUC-based vectors in either orientation with respect to the lacZ promoter to give pHs138 and pHs137. Both of these plasmids contained 3 kb Hpal–PvuII inserts. For pHs138, expression from the lacZ promoter resulted in large quantities of a 76 kDa protein, as well as several lower molecular mass antigens, all of which reacted with antisum to the 270 kDa IgBP. Plasmid pHs137 produced minimal amounts of the same proteins. The same bands were observed with the larger pHs139, although the 5 kb insert was in the opposite orientation with respect to the lacZ promoter from pHs138. This indicates that expression in E. coli of pHs139 was from a promoter originating from H. somnus, but that high-level expression in pHs138 was from the vector lacZ promoter.

DNA sequence analysis of the 5 kb Xbal–PvuII insert from pHs139 revealed a single 4.5 kb ORF, with no other ORF present in any other frame. The ORF may
2701
TTA GAA GAT GCA AAT GCC ATT CTG AAT TTA GAA GCT AAG AAT GTA AAA
2761
TTA GAA GCA AGA GAG ATT TGA AAG AAA TCA ATT CCT GAA GCG ACA GTT AGA CAA ATG TCT
DEAREISXKSIPEATVKQMS
2821
CTG GAG GAT ATT TCG AAG AAG TTA GAT GCC GCT AAG AAT GTA GAA ACT GTT ATT
HLPBFDDILTEGAKKVESRI
2881
GAAT AAG GCA ATC ACA TTT GCC CCT TCT GTT GAG GAG TTT TCA GAA ATT CAA GAT TTG GTG
NKAITFRPSVEEFSEIQDLV
2941
AAA ACC TTA CGG AAA ACA AAG GTT ATG ATA GAG GAT TCT TCA ACA AAA ACA AAT GAA ATC ACA
KTLFKTKVIDLSTKNEIT
3001
GAA GCT TTA CCT GCA TTT GCA AAG ACC ATT GAA CCT ACA CCG GAG TTG AAA GAA CAG TGG
EALAAATSTKIQRTPELKEQL
3061
GAC ACA CCA ATG GAG GAT TCC TTA CAA AAC AGT CAA GCC AAA CCT TTT ACA GTG CAG ATG
KTAIEDFLQNSQGKPLTVQM
3121
ATC GAG AAT CCT AAT CAC GGA TTA COT CCG GAT GAG GCA GAA GGT GTT TTA CTT TAT AAA
IENLNHGLRPDEGEGRLLYK
3181
AAA GAG AAT TTA ACC AAA GAA AAT GCG GTA TTT TCT AGT CCC GAA GGG GCA AAA ATT CAA
KENLTKEAVFSSPEAAKIQ
3241
TTA CGG GAA ACC GTT GAT TTT ATC AAT CGA GGG AAA AAT GAA GGG ATT GAG CCG AGT GTG
LAETVDFINRANKNEGIEPSV
3301
GTG GCC GCA TTA GGT TAT CAG CGA TTT ATT GCT GAT TAT CAC CCA TTT GCA GAA GGT AAT GGA
VGALVYQRLAIAYHPFAEGNG
3361
GCC ATG GCC AGA GTC ATA GTA AAT AAA ATT TTA CTT GAT GCA GGT TAT CCG GCA TTT ACC
RMARVIVNKILLDAGYPAPFT
3421
AAA TTT AGT GAT GAG TTT GCA CGG CAG ATT ATT CCT GAA ACG AAA GCA TCA ACT AAA TCC
KFSDEFEPQIPQTKASSTKS
3481
GCA AGC AGC AGT GAA Gtg GTA Gtt GAG TTT TTA AAA GAG TGT GCA AAA AAA GGA ACC AAG
ATSSSEVVVVEFLKEALKKGSK
3541
GAA GAT AAC GAG CAC ATT TTA GAA AAA ACT GAC CGC ACT TCT AGC GAG TTG ACA GAA AGT
EDNENEQNLLEKTDRSTDLTSES
3601
GCG GTA GAA ATT TCG GTT GCT TTT AGT TCC GGA ACA GTG AGA TCT GCC ACA GTT TCT GAA
AVENSAALSSGTVRSSATVS
3661
ACA GTT ACT GAA AGC GAA CAG GCA AAA GCG AAA CCA GTT AGT GAT TTG GTG AGC AGT AAA
TVTETEQAKAKPKVDLVS
extend further upstream from the sequence reported in this paper, since no stop codon was encountered. The GC content of the entire 4.5 kb ORF is 39.3%, which is consistent with the published sequence of the *H. somnus* ribosomal RNA gene and flanking two ORFs (38.7%) (Theisen & Potter, 1992), the 40 kDa *H. somnus* lipoprotein (36.2%) (Theisen *et al.*, 1992), and the previously determined value (37.3% ± 0.2) for total *H. somnus* DNA (Bailie *et al.*, 1973). The sequence for the p76 coding region is presented in Fig. 1. The derived amino acid sequence of p76 was analysed for hydrophilicity using the algorithm of Kyte & Doolittle (1982)
with a window size of seven residues (data not shown). The peptide was predicted to be highly hydrophilic, with no long hydrophobic regions characteristic of a membrane-spanning protein, and did not contain an obvious signal sequence (Von Heijne, 1983). Our earlier studies showed the H. somnus 76 kDa surface protein to be shed in the soluble fraction of culture supernatant (S. A. Kania, M. Yarnall & L. B. Corbeil, unpublished data). These observations suggest that p76 may be a peripheral membrane protein or a surface protein rather than an integral membrane protein.

A striking feature of the large ORF is the presence of two large, approximately 1-2 kb direct repeats (DRs – Fig. 2), one of which (DR2) is flanked by an imperfect inverted repeat of 11 bp, with only one mismatch (Fig. 1). Although the start codon of the p76 gene is in the middle of DR1 (see evidence below), the entire DR1 was translated in Fig. 3 in order to show the alignment of DR1 with DR2. The most homologous portion of the DRs is the centre, while the ends of the DRs tend to be more divergent (Fig. 3). Also, the two DR sequences are much more homologous at the protein level (64% identity over 404 amino acids), than at the DNA level (31% identity).

Because of the inverted repeats flanking DR2 and the size similarity to other insertion sequences, it was of interest to explore the possibility that the DRs represent a foreign inserted sequence. Therefore, codon usage of the translated region of DR2 was compared (data not shown) to that of the C-terminus of the p76 ORF as well

Fig. 2. The 4.5 kb ORF of pHS139, showing the two DRs and the p76 gene (arrow). The translational start site for the p76 gene lies between the HpaI and PvuII restriction sites (see Fig. 4). The three repeated methionines within each DR are indicated by vertical bars above the horizontal line representing the sequence.

Fig. 3. Alignment of the DNA and translated protein sequences of the two DRs. The portion of DR1 upstream of the translational start site for p76 is indicated in italics. The coding sequence for p76 is underlined. Vertical lines represent identities at the protein level.
as four other published *H. somnus* ORFs (Theisen & Potter, 1992; Theisen et al., 1992). For five amino acids [leucine (UUG), valine, threonine, alanine and glycine] guanine was preferred more as the third base of the codon for DR2 as compared to the other four *H. somnus* ORFs. There was a concomitant decrease in the use of adenine as the third base for these amino acids, with the exception of glycine. The most notable differences between DR2 and the C-terminus of the p76 ORF were with the codon usage of leucine, isoleucine, proline and tyrosine. For the rest of the 20 main amino acids, there was no significant difference between DR2 and these other *H. somnus* sequences. Since DR2 varied in its codon usage from other *H. somnus* sequences, DR2 codon usage was further compared to that of the closely related bovine pathogen *Pasteurella haemolytica* (Lo, 1992), as well as to *E. coli* (Sharp et al., 1988), host of many other insertion sequences. In both cases, the codon usage differed significantly from DR2 (data not shown); thus it is improbable that the DRs originated from either of these other organisms. Nevertheless, since the sequence of DR1 differs substantially from DR2, it appears likely that if insertion and duplication events occurred, they happened long ago.

There have been numerous instances of insertion elements affecting the expression of genes involved in virulence, including IS1 inserts in the *virF* gene (an invasion-positive regulator) of *Shigella flexneri* 2a (Mills et al., 1992), a mosaic of multiple insertion-sequence-like elements upstream of the *Pseudomonas aeruginosa* exotoxin A gene (Pritchard & Vasil, 1990), and three tandem insertion-sequence-like elements upstream of variable antigen genes in *Borrelia hermsii* (Barbour et al., 1991). Thus precedence exists for these insertion-sequence-like DRs of *H. somnus* to play a role in the expression of the p76 gene.

No ATG codon in the sequence was an obvious translational start site, providing an appropriate length of ORF to give a theoretical 76 kDa protein. Thus, an attempt was made to find the start site for this gene by deletion analysis. Subclones of pHSl38 were constructed (Fig. 4) and their cell lysates analysed for expression by Western blotting. The blots were incubated with P3 bovine convalescent-phase serum, which previously was shown to react with the 270, 120 and 76 kDa surface proteins of *H. somnus* (Cole et al., 1992; Corbeil et al., 1992). The Western blots, also shown in Fig. 4, indicate that some inserts encoded proteins that react with P3 convalescent serum. The control cell lysate (*E. coli* DH5α with pUC18 alone) did not react with the antiserum
under these conditions. In addition, a subclone (pHS145, with a 1 kb insert containing 0.5 kb of the 3' end of the ORF) encodes enough of the sequence to express a 23 kDa protein that reacts with P3 bovine convalescent antisera (data not shown). This suggests that at least one epitope on p76 that reacts with P3 antibody is within 167 amino acids from the C-terminus of the protein.

DR1 overlaps the putative translational initiation region of p76, since pHS138 expresses p76 and DR1 spans the HpaI site in pHS139 that was used to derive pHS138 (Figs 2 and 4). Also, each DR contains three ATG codons, which are repeated faithfully. All three of these ATG codons (labelled pHS138, contains all three of the antigens (weaker bands on the Western blot). Subclone pHS182 is missing codon A, but has B and C in Figs 1 and 4 were candidates for start codons for synthesis of these ATG codons (labelled pHS138, contains all three of the antigens (weaker bands on the Western blot). Subclone pHS182 is missing codon A, but has B and C, and also is not in frame for a fusion with LacZ. However, this subclone does not express any antigens detected by the P3 antiserum. pHS176 contains the same insert as pHS182, but is in the proper frame with the lacZ start site to make a translational fusion protein. This clone expresses p76, plus some of the lower molecular mass antigens appearing as weaker bands on the Western blot. The insert which contains only start codon C, in pHS171, is not in frame to make a lacZ translational fusion protein, but does encode some of the same antigens below 76 kDa as pHS138. Plasmids pHS182, 176 and 171 were all deletion derivatives of pHS138, and thus all contained the lacZ promoter in the proper orientation to the H. somnus ORF. Since pHS138 expressed p76, and pHS182 does not, the start codon for p76 must lie between the HpaI site (contained in pHS138, but not 182) and the PvuI site (present in both plasmids). Since no other known start codons (TTG or GTG) are present within these two restriction sites, the start codon for p76 must be the ATG codon A. A possible ribosome binding site (AAAGAG) located 12 bp from the translational start may be functional in pHS139. This site occurs just upstream of the HpaI restriction site and would thus be deleted in pHS138, explaining the orientation-specific effect of the insert in this clone.

Since translation from codon A could theoretically produce a protein of 87 kDa, the protein must either be processed to form p76 or the protein runs anomalously on SDS-PAGE gels. An example of the former situation was found for the 120 kDa Rickettsia rickettsii OMP (Gilmore et al., 1989, 1991), which contains enough DNA in the ORF to encode a deduced 168 kDa protein. Post-translational processing of the C-terminus apparently gives rise to the 120 kDa product, since a 32 kDa peptide with the same sequence as the 3' end of the large ORF was found (Gilmore et al., 1991). Interestingly, the 120 kDa ORF of R. rickettsii is also similar to H. somnus p76 since both contain only two cysteines (Gilmore et al., 1989), providing minimal intramolecular disulphide bonding. The presence of only two cysteines in the H. somnus ORF explains the fact that no difference was observed (Cole et al., 1992) in protein migration in SDS-PAGE gels run with and without urea, or with and without β-mercaptoethanol in the loading buffer. Those experiments, along with inclusion of protease inhibitors in the samples (Cole et al., 1992), suggested that the multitude of bands observed on SDS-PAGE gels for recombinants containing the 76 kDa ORF was not due to artifacts such as monomer polymerization or in vitro proteolytic degradation of the H. somnus proteins.

The expression of lower molecular mass antigens at 66, 52 and 55 kDa by pHS171 which has codon C but not A or B, could be due to use of additional start codons. Codon C as a start site could produce a protein of theoretical molecular mass 52 kDa. The other two antigens may migrate anomalously on SDS-PAGE gels, or be modification (66 kDa) or degradation (50 kDa) products of the 52 kDa protein. These proteins were also expressed by pHS138. Expression from this alternative start site (codon C) within the 76 kDa ORF of lower molecular mass proteins could explain the multiple banding patterns on Western blots observed previously (Cole et al., 1992). The use of more than one translational initiation site within an ORF has been described for numerous genes, including the McrB restriction system of E. coli (Ross et al., 1989), as well as several phage (Dunn & Studier, 1983; Eisenberg & Finer, 1980; Fulford & Model, 1988) and transposon (Isberg et al., 1982; Johnson & Reznikoff, 1984) genes. Expression of a Mycoplasma hyorhinis gene in E. coli resulted in the translation of related polypeptides of 110, 100, 65 and 55 kDa from an ORF of just over 3 kb (Notarnicola et al., 1990). Tryptic peptide analysis revealed that these proteins all shared a common C-terminus, suggesting that all proteins were translated from a single message. A similar mechanism may occur for H. somnus p76 and lower molecular mass proteins.

In conclusion, the DRs and p76 DNA coding sequence lie within a 13.4 kb segment that is present in two virulent serum resistant strains, and absent in all four of the serum sensitive strains tested (Cole et al., 1992). The 13.4 kb segment contains an ORF of at least 4.5 kb, with multiple start sites functional in E. coli and two 1.2 kb DRs. The coding sequence for p76 starts in DR1 and continues beyond DR2, which has flanking inverted repeats similar to insertion elements. The duplicated DRs are thus essential to the expression of a hydrophilic
76 kDa surface protein which is associated with serum resistance and virulence.

This work was supported, in part, by grant 89-37265-4648 from the United States Department of Agriculture.

References


