INFECTION AND IMMUNITY, Apr. 1995, p. 1440–1445 Vol. 63, No. 4
0019-9567/95/$04.00+0

Cloning and Sequence Analysis of a Protective M-like Protein Gene from *Streptococcus equi* subsp. *zooepidemicus*

JOHN F. TIMONEY,[1]* JOHN WALKER,[1] MEI ZHOU,[2] AND JIABING DING[2]

Gluck Equine Research Center, University of Kentucky, Lexington, Kentucky 40546-0099,[1] and Department of Microbiology, Immunology and Parasitology, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853[2]

Received 14 September 1994/Returned for modification 19 December 1994/Accepted 31 January 1995

***Streptococcus equi* subsp. *zooepidemicus*, a Lancefield group C streptococcus, is a frequently isolated opportunist pathogen from a variety of animal hosts, including the horse. Previous studies have indicated that equine strains carry antigens with characteristics of the antiphagocytic M proteins on the Lancefield groups A and G streptococci. We have cloned a protective M-like protein gene (SzPW60) of an equine strain of *S. equi* subsp. *zooepidemicus* W60 and determined its sequence. This gene encodes a protein with a molecular weight of 40,123 which protects mice against subsp. *zooepidemicus* but not subsp. *equi*, stimulates antibodies which opsonize subsp. *zooepidemicus* but not *equi*, and reacts with antiserum to the protein of the parent strain. The predicted amino acid structure shows significant homology with the carboxy termini of groups A and G M proteins but no other homology. The M-like protein, although showing an extensive region of alpha helix, lacks the A, B, and C repeats found in group A M proteins and has a shorter signal sequence. A proline-rich region upstream from the LPSTGE motif contains 20 repeats of the tetrapeptide PEPK. The presence of this repeat region may account for the slow migration of the M-like protein in sodium dodecyl sulfate-polyacrylamide gel electrophoresis.**

Streptococcus equi subsp. *zooepidemicus*, a Lancefield group C streptococcus, is the most frequently isolated opportunist pathogen of horses. A normal mucosal commensal, it causes purulent respiratory infections of weanling and yearling horses and uterine infections in elderly mares. It also causes disease in a variety of other animal hosts.

The existence of mouse protective antigens on bovine and murine strains was recognized more than 50 years ago (26), and the specificity of these antigens was demonstrated in experiments that showed that mice immunized with a strain of subsp. *equi* were not protected against challenge by two equine strains of subsp. *zooepidemicus* (1). Unlike the closely related subspecies *equi*, which is antigenically conserved (1, 11, 21), equine strains of subsp. *zooepidemicus* exhibit great antigenic variation of an acid- and heat-resistant trypsin-sensitive protein (21). On the basis of this information and the observation that acid extracts of a few equine strains were precipitated by either group A M type 36 or 38 antiserum, Moore and Bryans (21) suggested that subsp. *zooepidemicus* possessed antigens analogous to the M proteins of group A streptococci. Subsequent studies showed that an electropurified mutanolysin-extracted 58-kDa protein was effective in stimulating opsonic responses in guinea pigs to some equine strains of subsp. *zooepidemicus* (28). Both equine convalescent serum and polyvalent antiserum to the 58-kDa protein reacted with bands of various molecular masses in all equine strains of subsp. *zooepidemicus* tested (28).

Almost all horses carry a variety of antigenic types of subsp. *zooepidemicus* in their tonsils and on their nasopharyngeal and genital mucosal surfaces as mucosal commensals (15). The mechanism by which they evade clearance by the mucosal immune responses is unknown, but it is tempting to speculate that antigenic variation plays a role since it is known that the antigenically conserved subsp. *equi*, the cause of equine strangles, is very quickly cleared from the nasopharynx during convalescence in association with local production of immunoglobulin G and immunoglobulin A antibody to M protein (10).

As a necessary preliminary step toward understanding the role of variation in the protective M-like proteins in the interaction of group C streptococci with the equine immune system, we cloned and sequenced for the first time a protective M-like protein gene of a representative equine strain of *S. equi* subsp. *zooepidemicus*. We describe the immunologic, protective, and opsonogenic features of the protein and present an analysis of the sequence and compare it with the general structure previously established for group A streptococcal M genes (7).

MATERIALS AND METHODS

Bacteria, plasmids, and bacteriophages. *S. equi* subsp. *zooepidemicus* W60 was isolated from a nasal discharge of a horse in New York in 1976 and stored at −70°C in Todd-Hewitt broth (THB). Lambda gt11 was obtained from Promega Corporation, Madison, Wis. Plasmids used for subcloning and sequencing were pHG165 (27) and pUC 118 (33). *Escherichia coli* hosts were Y1090 for λgt11 and TBI for plasmids.

Media. *E. coli* was grown overnight in Luria-Bertani medium at 37°C with ampicillin added (50 mg/ml) when appropriate. Streptococci were propagated overnight in THB with 0.2% yeast extract at 37°C.

Isolation of chromosomal, plasmid, and bacteriophage DNA. Streptococcal DNA was extracted by suspending the cells from an overnight culture of 500 ml of THB in 25 ml of sucrose buffer (6.7% sucrose, 50 mM Tris-HCl (pH 8.0), 1 mM EDTA), and this was followed by incubation at 37°C for 10 min. A total of 6.25 ml of lysozyme solution (10 mg of lysozyme per ml of H_2O) and mutanolysin (500 U) were added and incubated at 37°C for 5 min. The cells were further lysed by the addition of 3.12 ml of lysing buffer (0.25 M EDTA, 50 mM Tris-HCl [pH 8.0]) and 2.25 ml of 20% sodium dodecyl sulfate (SDS) and then by swirling and incubation at 37°C for 10 min. The tube was then vortexed at the highest speed setting for 30 s, 0.5 ml of RNase solution (10 mg/ml) was added, and the tube was incubated at 37°C for 45 min. The proteins were removed with 1 volume of phenol saturated with 3% NaCl and then by extraction with 1 volume of chloroform-isoamyl alcohol (24:1). The DNA was precipitated by the addition of 1 volume of isopropanol and harvested by centrifugation at 11,600 × *g* for 30 minutes at 4°C. Plasmid DNA was prepared by rapid boiling (14) or by rapid

* Corresponding author. Mailing address: Gluck Equine Research Center, University of Kentucky, Lexington, KY 40546-0099. Phone: (606) 257-4172.

alkaline lysis (3) when larger quantities were required. Bacteriophage DNA was isolated by the rapid method of Tsonic and Manes (30).

Agarose gel electrophoresis. Electrophoresis of plasmid DNA samples and restriction endonuclease digests was performed in 0.7% agarose in TBE buffer (0.04 M Tris-acetate, 0.002 M EDTA, 0.1 mg of ethidium bromide per ml). Restriction endonuclease digests of streptococcal genomic DNA were electrophoresed for 18 h at 4°C and 25 V on 1% agar gels.

Preparation and purification of M-like proteins. Native M-like protein was extracted with mutanolysin from streptococcal cells harvested from overnight cultures (50 ml of THB) that were incubated at 37°C (11).

An acid extract of *S. equi* subsp. *zooepidemicus* W60, grown at 37°C in 6 liters of THB with continuous shaking, was prepared by the Lancefield method with a slight modification (19). Washed cells were extracted for 10 min at 95°C at pH 2.5 (HCl) and then neutralized with 1.0 M NaOH, and the cells were pelleted by centrifugation at 10,000 × *g* for 10 min. The supernatants from this and from a second similar extract were pooled and concentrated by vacuum dialysis against 10 mM phosphate-buffered saline (PBS) containing 1:10,000 Merthiolate.

M-like protein in acid extracts and *E. coli* lysates containing pHG165 was partially purified by adsorption to hydroxyapatite in 10 mM PBS (pH 7.2) and then by elution in 0.2 M Sorensen phosphate buffer (pH 9.0). Excess phosphate was removed by vacuum dialysis against 10 mM PBS.

Partially purified acid-extracted M-like protein was subjected to SDS electrophoresis on 3-mm-thick agarose (Pro Sieve; FMC BioProducts, Rockland, Maine) according to the manufacturer's protocol. Laemmli loading buffer without mercaptoethanol and standard running buffer were used for each separation. The separations were performed overnight at 4°C and 20 mA. Remazol-stained cytochrome *c* was included in the sample to monitor the progress of the run, after which the protein was electrophoretically transferred from the agarose gel to Immobilon P. Protein bands were visualized with 20% methanol and excised. A narrow strip from the edge of the Immobilon was immunoblotted with antiserum to determine the position of the M-like protein. The section of Immobilon containing the M-like protein was washed once in methanol and five times in double-distilled water and submitted for N-terminal amino acid microsequencing to the Biotechnology Laboratory, Cornell University.

Immunoblotting. The proteins in streptococcal extracts of *E. coli* lysates were separated on SDS–10% polyacrylamide gel electrophoresis (PAGE) gels and electrophoretically transferred to nitrocellulose. The blots were developed in antiserum diluted 1:200 in PBS and then in peroxidase-conjugated protein G (1:4,000). Reactive bands were visualized with 4-chloro-1-naphthol (0.5 mg/ml) as the substrate.

Antisera. Guinea pig antiserum to the electropurified protective M-like protein of *S. equi* subsp. *zooepidemicus* was prepared as described elsewhere (28). Antiserum to the recombinant M-like protein of subsp. *zooepidemicus* W60 was produced in rabbits. Two adult Flemish rabbits were inoculated subcutaneously with 100 μg of recombinant M-like protein partially purified by hydroxyapatite chromatography. The protein was from an *E. coli* lysate containing pHG165 and was mixed with complete Freund's adjuvant (CFA) and PBS (1:2). Booster immunizations consisting of 25-μg doses of recombinant protein purified by preparative electrophoresis in agarose (Gel-Sieve) were performed intravenously on days 15, 51, and 93. The rabbits were exsanguinated on day 102. A control rabbit was similarly immunized with a lysate of *E. coli* TBI containing pHG165 but no streptococcal DNA insert.

Rabbit antiserum to an acid extract of subsp. *zooepidemicus* W60 was prepared by inoculating a rabbit (rabbit 128) subcutaneously with 100 μg of extract protein mixed with CFA. Intravenous booster immunizations with 25 μg of extracted protein were given on days 15 and 28. Serum was harvested on day 35. A pool of mouse antiserum to acid-extracted protein of subsp. *equi* was harvested from mice (29).

Mouse immunization and challenge. Inbred 12-week-old ICR mice were vaccinated by the subcutaneous inoculation of 100 μg of *E. coli* lysate containing the recombinant M-like protein of subsp. *zooepidemicus* W60 in CFA. The M-like protein had been partially purified by hydroxyapatite chromatography (29). The mice received boosters 15 and 41 days later by the subcutaneous inoculation of 50- and 20-μg doses, respectively, of partially purified recombinant protein in 0.15 ml of PBS. A group of 31 control mice was similarly inoculated with protein from *E. coli* TBI in CFA and received sham boosters with inoculations of PBS on days 15 and 41.

A group of 30 vaccinated mice and the 31 control mice were challenged 22 days after the second booster with 10^8 CFU of an 18-h THB culture of *S. equi* subsp. *zooepidemicus* W60 given intraperitoneally in a dose volume of 0.1 ml. Another group of 30 vaccinated mice were challenged with 7×10^5 CFU (approximately two 50% lethal doses [LD_{50}]) of an 18-h THB culture of *S. equi* subsp. *equi* CF32. The mortality in each group was recorded each day for 7 days after the challenge. The significance of the differences in mortality between the vaccinated and control groups was determined by chi-square analysis.

Opsonic assay. Equine neutrophils were separated from freshly collected heparinized horse blood with a discontinuous Percoll gradient (24). Neutrophils from 7 ml of blood were suspended in 3 ml of RPMI medium (Gibco, Grand Island, N.Y.), and 80-μl aliquots (6×10^5 cells) were added in triplicate to the wells of a 24-well cell culture cluster (Costar, Cambridge, Mass.). Each well contained a circular 12-mm-diameter glass coverslip. The cell cluster was incubated for 2 h at 37°C in 5% CO_2, after which the wells were washed once with PBS to remove nonadherent neutrophils.

Test bacteria grown overnight at 37°C in THB were added in 3-μl volumes (10^6 CFU) to 450 μl of RPMI 1640, 10 μl of guinea pig serum as a complement source, and 25 μl of antiserum in each well. The plate was gently shaken for 30 min at 37°C to allow for phagocytosis. The coverslips were then washed once with PBS (pH 7.2), fixed in 10% formalin, and stained with Giemsa. The number of neutrophils with ingested and surface-associated streptococci per 100 cells was estimated for each coverslip and expressed as a percentage. The significance of the differences in the opsonic activities of the control and immune sera was determined by a Student *t* test (unpaired observations) based on the means of three experiments.

Gel filtration. The molecular mass of the recombinant M-like protein expressed in *E. coli* was estimated by gel filtration through Toyopearl HW 55S resin (Toso Haas, Montgomeryville, Pa.). The column was calibrated with bovine serum albumin, ovalbumin, and lysozyme. The fractions were tested for the M-like protein by dot blotting them on cellulose nitrate and developing the blots with a 1:200 dilution of the antiserum to M-like protein, as described above under "Immunoblotting."

Construction and screening of genomic library of *S. equi* subsp. *zooepidemicus* W60. Chromosomal DNA of *S. equi* subsp. *zooepidemicus* W60 (20 μg) was partially digested with *Eco*RI, and the fragments were separated by electrophoresis on a 0.7% agarose gel. DNA fragments 2 to 8 kb in length were excised, and the DNA was electroeluted into dialysis tubing. The fragments were extracted with phenol and chloroform, ethanol precipitated, and treated with Geneclean (Bio 101 Co., San Diego, Calif.). DNA fragments (0.05 pmol) and 0.018 pmol of lambda gt11 DNA were ligated overnight at 16°C. The ligation mixture was incubated at 4°C overnight and then mixed with packaging extract (Promega) according to the manufacturer's instructions. After it was packaged, the phage was plated on Luria-Bertani agar plates with *E. coli* Y1090 being used as the host.

The plaques were transferred to nitrocellulose discs blocked in 3% gelatin in PBS and were probed for 2 h at room temperature with a 1:50 dilution of guinea pig antiserum to the electropurified 58-kDa antigen of *S. equi* subsp. *zooepidemicus* W60. The filters were then washed in 20 mM Tris (pH 7.5)–0.5 M NaCl–0.25% Tween 20 (TTBS) three times at room temperature and then were incubated for 2 h in peroxidase-conjugated antiserum to guinea pig immunoglobulin G. The conjugate was removed by washing the filters in TTBS, and the blot was developed with 4-chloro-1-naphthol. Plaques giving a positive signal were then picked with a sterile Pasteur pipette, suspended in lambda buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM $MgSO_4$, 0.2% gelatin), and rescreened until all plaques gave a positive signal. A representative positive phage recombinant was isolated and designated SzPW60-1a.

Subcloning of DNA fragment encoding M protein of *S. equi* subsp. *zooepidemicus* W60. DNA of SzPW60-1a was prepared by standard methods from phage harvested from a 1-liter Luria-Bertani culture of Y1090. The DNA was digested to completion with *Eco*RI, and the 7.3-kb streptococcal DNA restriction fragment was separated by electrophoresis on 0.7% agarose. The fragment was electroeluted, extracted twice with phenol-chloroform, and precipitated with ethanol. The purified DNA fragment was then digested to completion with *Hin*dIII, and the resulting fragments were subcloned into the low copy vector pHG165 by standard ligation and transformation procedures. A clone with a 2.7-kb streptococcal DNA fragment expressing recombinant M protein was detected by colony immunoblotting. Further subcloning of the *Xba*I or *Xba*I-*Hin*dIII fragment of the 2.7-kb *Hin*dIII fragment in pUC 118 was performed.

Southern blotting. Genomic DNA from *S. equi* subsp. *zooepidemicus* W60 was digested to completion with *Hin*dIII, separated by electrophoresis on a 0.75% agarose gel, and transferred to nitrocellulose. Prehybridization was performed at 37°C for 4 h in a hybridization solution (45% formamide and 5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate]), 5× Denhardt's solution, and 100 mg of salmon sperm DNA per ml. The probe (D1) encoding the entire M-like protein gene was prepared from a 1.1-kb *Xba*I-*Hin*dIII fragment of the larger 2.7-kb fragment and end labeled with [^{32}P]dATP with a commercial kit (Prime-It II; Stratagene). Hybridization with probe DNA, which was denatured by being boiled for 10 min, was performed in the same solution overnight at 37°C. The filters were washed once for 10 min in 2× SSC plus SDS at 37°C and then twice in 0.2× SSC plus SDS at 37°C for 10 min each time. The washed blots were air dried and exposed to X-Omat AR film (Eastman Kodak Co., Rochester, N.Y.).

Nucleotide sequence accession number. Nucleotide sequence determinations for the four subclones (D1, D2, D3, and D4) constituting the entire 2.7-kb *Hin*dIII fragment were performed with single-stranded templates in M13 (31) by means of the Sanger dideoxy chain termination procedure (Sequenase; U.S. Biochemical Corp., Cleveland, Ohio). Universal primers (U.S. Biochemical) as well as oligonucleotides constructed from internal sequences obtained during the study were used as sequencing primers. The nucleotide sequences of both strands of the two subclones (D1 and D4) containing the open reading frame of the M protein gene were determined. The GenBank accession number is UO 4620.

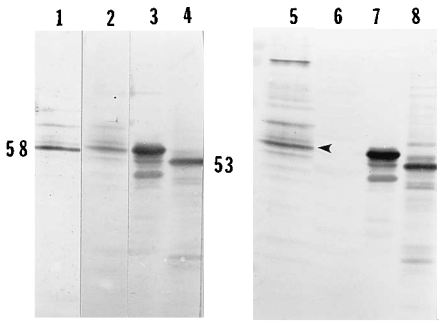

FIG. 1. Immunoblot showing the M-like protein in a mutanolysin extract of *S. equi* subsp. *zooepidemicus* W60 (lanes 1, 2 and 5), in a lysate of *E. coli* TBI containing SzPW60 in pHG165 (lanes 3 and 7), and in an acid extract of strain W60 (lanes 4 and 8). Lane 6 contains a lysate of *E. coli* TBI with pHG145. Lanes 1, 3, and 4 were treated with guinea pig antiserum to the 58-kDa mutanolysin-extracted antigen band of W60, which is indicated by an arrowhead on lane 5. Lane 2 was treated with antiserum to recombinant SzPW60. Lanes 5 to 8 were treated with serum from a horse with peritonitis caused by subsp. *zooepidemicus*. Molecular masses (in kilodaltons) are shown at the left and the center of the figure.

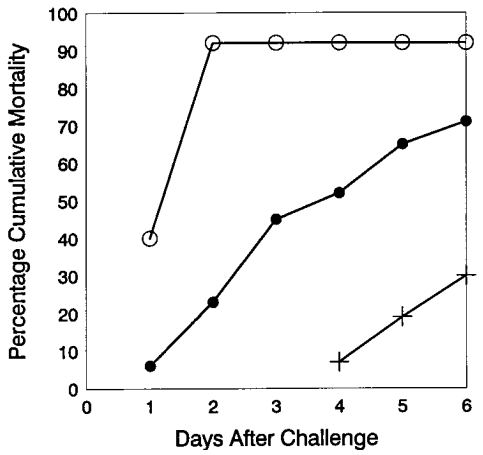

FIG. 3. Cumulative mortality curves for groups of normal control mice (solid circles) and mice vaccinated with the recombinant M-like protein of *S. equi* subsp. *zooepidemicus* W60 and subsequently challenged intraperitoneally with 10^8 CFU of strain W60 (crosses) and 7×10^5 CFU of *S. equi* subsp. *equi* CF32 (open circles). The differences in the mortality of the groups of control and vaccinated mice were significant ($P < 0.01$). *S. equi* subsp. *equi* shows a much higher virulence for mice than subsp. *zooepidemicus*.

RESULTS

The genomic library of *S. equi* subsp. *zooepidemicus* W60 in λgt11 was probed with guinea pig opsonic antiserum against the electropurified 58-kDa antigen in a mutanolysin extract of W60, and four strongly reacting plaques were detected. All four recombinant phage contained similar 7.3-kb inserts, one of which, SzPW60-1a, was chosen for further study. The protein coded for by SzPW60-1a had the same molecular mass (58 kDa) as the native protein used to immunize the guinea pig (Fig. 1). A 2.7-kb *Hin*dIII subfragment of SzPW60-1a subcloned into plasmid pHG165 and transformed into *E. coli* TBI also expressed this protein (Fig. 1), as did a subclone in pUC118 in *E. coli* TBI containing a 1.16-kb *Hin*dIII-*Xba*I fragment (D1) (Fig. 2). One of these clones was designated pSzPW60. Rabbit antiserum against an *E. coli* lysate containing pSzPW60-1 in pHG165 and the guinea pig antiserum to the 58-kDa band electropurified from a mutanolysin extract of W60 showed very similar patterns of reactivity (Fig. 1). These sera as well as opsonic serum from a horse (horse 631) with peritonitis caused by *S. equi* subsp. *zooepidemicus* reacted strongly with a 58-kDa antigen band in the mutanolysin extract. A 53-kDa band in an acid extract of W60 was also strongly reactive with these sera, suggesting that it was a derivative of the 58-kDa native molecule extracted by mutanolysin.

Mouse protective activity of recombinant M-like protein. The recombinant protein stimulated a statistically significant ($P \leq 0.005$) level of protection against subsp. *zooepidemicus* W60 in mice immunized with a lysate of *E. coli* TBI containing pSzPW60-1. Mortality in the challenged, immunized mice was less than half of that observed in the controls, and the mean time to death of immunized mice was about 2 days longer (Fig. 3). Immunized mice were not protected against challenge by two LD_{50}s of *S. equi* subsp. *equi* CF32.

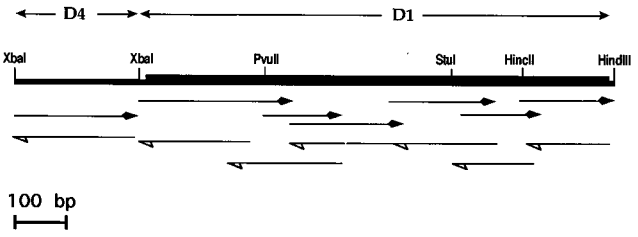

FIG. 2. Partial restriction map of DNA fragments D4 and D1 encoding the M-like protein (SzPW60) from *S. equi* subsp. *zooepidemicus*. The open reading frame for SzPW60 is shown in D1 as a thick line. The promoter and ribosomal binding sites are located in D4, and the sequencing strategy is shown by arrows beneath the figure.

Opsonogenic activity of recombinant 58-kDa protein. Rabbit antiserum to the recombinant 58-kDa protein of subsp. *zooepidemicus* W60 in an *E. coli* lysate and a pool of sera from mice immunized with this protein were opsonic for subsp. *zooepidemicus* W60 but not for subsp. *equi* (Table 1). Conversely, mouse antiserum to an acid extract of subsp. *equi* showed opsonic activity for subsp. *equi* but not for subsp. *zooepidemicus* W60. Rabbit antisera to the acid-extracted protein of W60 and to the recombinant 58-kDa protein showed similar opsonic activities for W60. However, antiserum to the acid-extracted protein also showed some opsonic activity for subsp. *equi*, suggesting the presence of an additional, nonspecific opsonogenic antigen in the acid extract.

Gel filtration. The molecular mass of the recombinant protein expressed by *E. coli* TBI containing pSzPW60-1 was approximately 45 kDa, as determined by filtration through Toyopearl HW 55S size exclusion resin (Toso Haas).

Southern blotting. A ^{32}P-labeled DNA probe, prepared from D1, hybridized with three *Hin*dIII fragments (2.7, 1.8, and 1.4 kb) of chromosomal DNA of subsp. *zooepidemicus* W60 (Fig. 4). The smaller fragments represent portions of one or possibly two other genes with anchor sequence homology to SzPW60 (27a).

Amino acid sequences derived from microsequencing. The N-terminal amino acid sequences of the recombinant 58-kDa protein and acid-extracted 53-kDa protein separated by SDS electrophoresis and transferred to Immobilon were identical (Fig. 5).

DNA sequence analysis of the SzPW60 gene. Following the subcloning of specific *Hin*dIII-*Xba*I fragments, it was determined by immunoblotting that D1 (1,170 bp) alone contained SzPW60. Nucleotide sequence analysis revealed one open reading frame of 1,128 bp with a putative Shine-Dalgarno sequence 6 bases upstream from the start codon. Potential

TABLE 1. Specificity of opsonic activity for equine neutrophils of rabbit and mouse antisera to recombinant M-like protein (SzPW60) of *S. equi* subsp. *zooepidemicus*

Serum	% Neutrophils with ingested or adherent bacteria after a 30-min incubation[a]	
	Subsp. *zooepidemicus* W60	Subsp. *equi*
Rabbit 191 (preimmunization)	13 ± 2.1	3 ± 1.0
Rabbit 191 (postimmunization with recombinant SzPW60 in *E. coli* lysate)	38 ± 6.7	5 ± 1.5
Rabbit EC8 (postimmunization with *E. coli* lysate)	9 ± 2.5	Not done
Rabbit 128 (postimmunization with acid extract of subsp. *zooepidemicus* W60)	47 ± 2.6	9 ± 0.6
Mouse pool A (postimmunization with acid extract of subsp. *equi*)	5 ± 0.6	14 ± 3.6
Mouse pool B (postimmunization with recombinant SzPW60 in *E. coli* lysate)	19 ± 2.0	2 ± 1.2

[a] Means of three experiments ± standard errors of the mean. Sera 128 and 191 (postimmunization) showed significantly higher opsonic activity ($P < 0.01$) than control sera EC8 and 191 (preimmunization) for subsp. *zooepidemicus*. Mouse pool A showed significantly higher ($P < 0.01$) opsonic activity for subsp. *equi* but significantly lower activity ($P < 0.01$) for subsp. *zooepidemicus* than mouse pool B. Statistical analysis was done by the Student *t* test (unpaired observations).

promoter sequences were identified in the adjacent sequence of D4 (Fig. 2). Translation of this open reading frame (Fig. 6) revealed a protein of 376 amino acids with a molecular mass of 40,123 and a pI of 5.13.

The GenBank database was scanned with the Intelligenetics (Mountain View, Calif.) software for sequences homologous to that of SzPW60. No homology was detected, except at the carboxy terminus, where the membrane anchor region showed extensive amino acid homology with group A M protein genes (Fig. 7). This homology included and extended downstream from the LPSTGE consensus sequence.

The N-terminal region of the protein has characteristics commonly found in the signal sequences of gram-positive bacteria (31). The first 12 residues contain seven charged amino acids (five basic and two acidic). The next region is strongly hydrophobic (11 of 14 amino acids) (Fig. 8). A signal cleavage site at amino acid 32 or 33 generates a signal peptide similar in length to that found on other gram-positive secreted proteins (31). A comparison of the signal sequence of SzPW60 with group A or G M protein signal sequences in GenBank revealed no homologies. However, the group C SzPW60 signal sequence of 33 amino acids was similar in length to other streptococcal signal sequences, such as those of streptolysin O, streptokinase, and protein G (33, 26, and 33 residues, respectively), but somewhat shorter than group A or G M protein signal sequences with 41 residues (6, 13, 16, 17, 25). Interestingly, the N-terminal amino acid of the mature recombinant M-like protein of W60 is alanine, whereas methionine is the N-terminal residue of the native acid-extracted protein. This suggests that protease cleavage of signal peptides in *E. coli* and subsp. *zooepidemicus* is slightly different in specificity.

The carboxy-terminal end of the M-like protein shows all the general features well established for surface proteins of gram-positive cocci (8). These features include a small cluster of four charged amino acids at the C terminus followed by a region of predominantly hydrophobic residues (19 of 24) (Fig. 8). Adjoining the hydrophobic domain is the consensus membrane anchor sequence LPSTGE. The next region of 78 amino acids is rich in proline (30 of 78) and lysine (20 of 78) arranged mostly as the tetrapeptide repeat PEPK. This cell wall region is conspicuously lacking in glycine by comparison with group A M protein sequences, which may reflect the fact that *N*-acetylgalactosamine is the main amino sugar of group C cell wall polysaccharide, whereas *N*-acetylglucosamine is the main amino sugar in the group A cell wall (18).

Analysis of the secondary structure of the translated protein showed an extensive region of alpha helix stretching from near the cleavage site of the signal sequence to the proline-rich repeat region in the carboxy terminus (Fig. 8). This region contains a gap in alphahelical potential between residues 130 and 140. The N terminus of the mature protein has a net negative charge.

Molecular mass. The predicted molecular mass of the protein, 40.1 kDa, was much lower than that expected from its migration in SDS-PAGE. On gel filtration, the recombinant M protein eluted in fractions of calibrated molecular mass (45 kDa) much closer to the actual calculated molecular mass of the M-like protein (data not shown).

The principal M-like protein fragment in acid extracts of W60 was smaller than the recombinant protein expressed by *E. coli* or the native protein released by mutanolysin (Fig. 1, lane 5) by about 5 kDa. This suggests that hot acid hydrolysis effectively degrades the cell wall, releasing most of the embedded protein. It is also apparent that the M-like protein of subsp. *zooepidemicus* is more resistant to acid hydrolysis than the protective M-like protein of subsp. *equi* because acid ex-

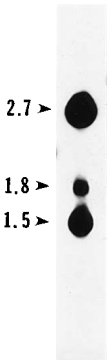

FIG. 4. Southern blot of *Hin*dIII-digested genomic DNA of *S. equi* subsp. *zooepidemicus* W60 following hybridization with a ^{32}P-labeled DNA probe prepared from *Hin*dIII-*Xba*I DNA fragment D1 (Fig. 2) containing SzPW60. The sizes of the fragments are shown in kilobases at the left.

Residue #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
rec M protein	A	D	T	A	D	S	A	T	P	A	V	T	A	T	V	T	D	S	L	D	S
53 k acid extract fragment		D	T	A	D	S	A	T	P	A	V	T	A	T	V	T	D	?	L		

FIG. 5. N-terminal amino acid sequences obtained by direct microsequencing of recombinant M-like protein of subsp. *zooepidemicus* W60 expressed in *E. coli* and of a 53-kDa fragment from an acid extract of *S. equi* subsp. *zooepidemicus* W60. The 53-kDa fragment was recognized by antiserum to the recombinant protein and by serum from a horse with peritonitis caused by subsp. *zooepidemicus*.

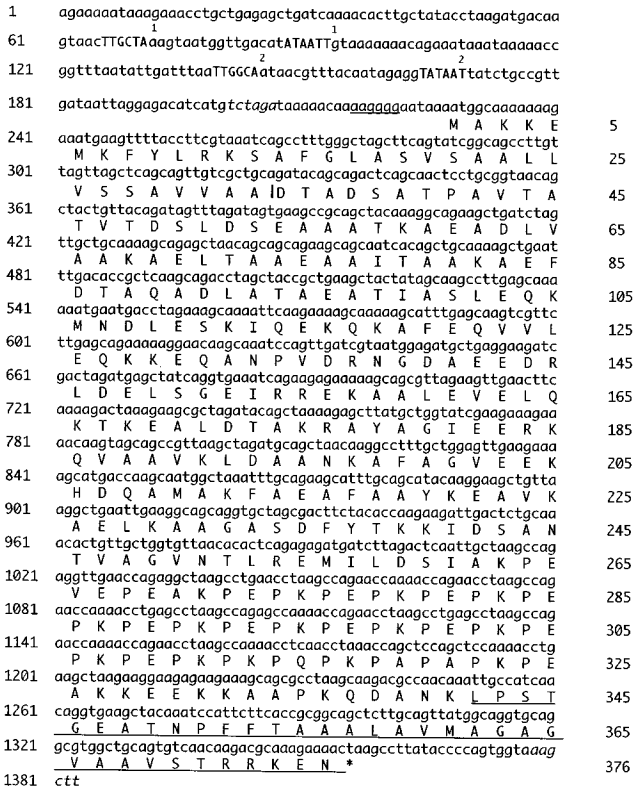

FIG. 6. Nucleotide and deduced amino acid sequence of SzPW60. Base position numbers are shown on the left, and amino acid positions are shown on the right. Putative promoter sites 1 and 2 and the Shine-Dalgarno ribosome binding sequences (underlined) are shown. The signal sequence cleavage site is shown as a heavy vertical bar, and the membrane anchor region showing extensive amino acid homology with group A M genes is underlined. The asterisk denotes the stop codon.

tracts of the latter display a greater number of smaller antigen fragments under the same conditions of time, temperature, and pH (23). This may indicate a higher frequency of acid-labile peptide bonds in the M-like protein of subsp. *equi*.

DISCUSSION

The specificity of the opsonic and mouse protective responses to SzPW60 in combination with its structural features is suggestive of an M-like protein. The most notable difference between SzPW60 and other M proteins so far described (7) is the absence of A and B repeat regions in the N-terminal half of the molecule and C repeats in the carboxy-terminal half.

SzPW60	LPSTGEATNPFFTAAALAVMAGAGVAAVSTRRKEN
Type 2 M-protein (2)	LPSTGETANPFFTAAAATVMVSAG-MLALKRKEEN
Type 5 M-protein (20)	LPSTGETANPFFTAAALTVMATAG-VAVVKRKEEN
IgA-binding (9)	LPSTGETANPFFTAAAATVMVSAG-MLALKRKEEN
Protein H (12)	LPSTGETANPFFTAAALTVMATAGVAAVVKRKEEN
IgG-binding (4)	LPSTGETTNPFFTAAALAVIASAGVFAL-KRKEEN

FIG. 7. Alignment of amino acid sequences (residues 342 to 376) in the membrane anchor region, showing homology of the M-like protein (SzPW60) of *S. equi* subsp. *zooepidemicus* W60 and examples of M and other M-like surface proteins of *Streptococcus pyogenes*. The numbers in parentheses are references.

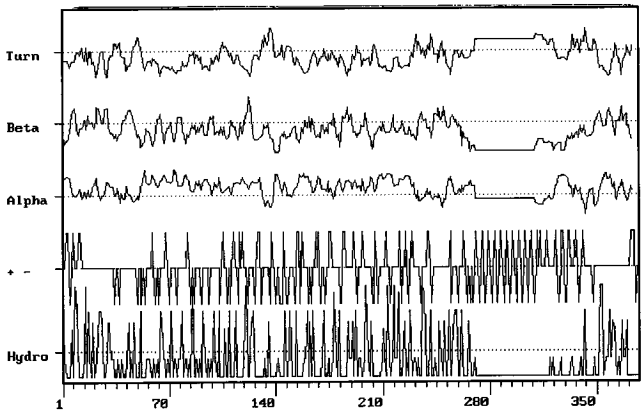

FIG. 8. Predicted plots for the secondary structure of the M-like protein (SzPW60) of *S. equi* subsp. *zooepidemicus* W60. The sequence was analyzed with the Novotny-Auffray algorithm (22) by the Wisconsin Genetics Program. Plots marked Turn, Beta, and Alpha indicate the potential for beta turn-random coil, beta sheet, and alpha helix formation, respectively. The + − plot shows regions of the molecule with net positive (upper) and negative (lower) charges. The hydrophobicity (Hydro) plot shows the highly hydrophobic regions of the protein. The positions of the amino acids are shown on the horizontal axis.

These repeats may be a feature of evolutionarily more recent *emm* genes since it is possible that streptococci such as subsp. *zooepidemicus* that are host adapted to lower animal hosts represent a phylogenetically more ancient population than the group A streptococci that show a high degree of adaptation to primates.

SzPW60 also differs from other M proteins in the shorter length (33 amino acids) of its signal peptide and therefore more closely resembles the signal sequences of streptococcal proteins such as streptolysin O and protein G (6, 15). The discrepancy in the size of its signal sequence suggests that SzPW60 may not be closely related to the group A M proteins and that it has arisen from a precursor protein with a different function but which coincidentally possesses an opsonogenic-protective activity. However, the demonstration of a similar though not identical protein on subsp. *equi* (29a) and the lack of cross protection or cross-reactivity of the opsonic antibody to the proteins from subsp. *equi* and *zooepidemicus* suggest that these group C streptococci express a protein with at least some of the essential functional and structural features of the group A M proteins.

The absence of cross protection between subsp. *zooepidemicus* and *equi*, as can be seen from Table 1 and Fig. 3, is not unexpected, given that mice immunized with subsp. *equi* bacterin are not protected against challenge with subsp. *zooepidemicus* (1). Conversely, although the nasopharyngeal mucosae and tonsils of all horses become colonized early in life by a variety of antigenic variants of subsp. *zooepidemicus* (15) which stimulate mucosal and systemic antibody responses, they are not protected against infection by subsp. *equi*. Recognition of the type specificity of immunity to subsp. *equi* has been a rationale for many years for the inclusion of this organism rather than subsp. *zooepidemicus* in strangles vaccines.

What is the relationship of SzPW60 to the hot-acid-resistant protein antigens described by Moore and Bryans (21)? The 53-kDa antigen in an acid extract of subsp. *zooepidemicus* W60 (Fig. 1, lanes 4 and 8) is the major band recognized by antiserum to acid-extracted protein and by serum 631 from a horse with a chronic invasive infection with subsp. *zooepidemicus*. This antigen has been shown on immunoblots to vary in size among the Moore and Bryans serovars, the M-like protein genes of which have also been shown to vary in their DNA

sequences (32). However, conclusive proof that SzPW60 is the typing antigen will have to await the production of a new set of typing antisera since the original set was discarded some years ago.

The discrepancy between the predicted molecular mass of SzPW60 and that estimated by SDS-PAGE is very substantial and suggests the possibility that a portion of the gene is deleted during cloning. This is highly unlikely, given that the recombinant and native M-like proteins had the same molecular masses as determined by SDS-PAGE and that the cloned DNA insert was the same size as an *Xba*I-*Hin*dIII fragment of W60 genomic DNA that hybridized with a probe of the DNA insert (Fig. 4). Examples of overestimation of the molecular masses of streptococcal proteins with domains rich in proline include the M6 protein (13), a group G albumin-binding protein (25), and protein G (6). Perhaps the most striking example of this phenomenon is the proline-rich acidic procyclic protein of *Trypanosoma brucei*, which has a predicted molecular mass of 15 kDa and yet which migrates at 45 kDa in SDS-PAGE (5). Clearly, small variations in the number of proline-rich repeats could generate significant differences in mobility in SDS-PAGE gels and create the impression of greater size variations than are actually present.

Although they vary in size, the M-like proteins of all equine strains of subsp. *zooepidemicus* tested have shown cross-reactivity with polyvalent antiserum to mutanolysin-extracted M-like protein (28). This suggests a substantial degree of conservation of the immunodominant epitopes. The role of antigenic variation, as shown by the Moore and Bryans serovars (21), in the evasion of mucosal immune responses by subsp. *zooepidemicus* in the horse is at present unclear. The organism is a normal commensal of the nasopharyngeal mucosa of all horses (15), and a number of different M-like protein types may be present in the tonsil at the same time (27a). This situation contrasts with that observed for the antigenically highly conserved subsp. *equi*, which is quickly cleared from the nasopharynx following the onset of mucosal antibody responses (10). Clearly, the variation in epitopes involved in mucosal immune clearance could play a significant role in the differing interactions of subsp. *equi* and subsp. *zooepidemicus* with the equine nasopharyngeal mucosa. Primers based on the sequence of SzPW60 are now being used to amplify the DNA sequences of M-like genes from the Moore and Bryans serovars of subsp. *zooepidemicus*. By this method, the locations and patterns of amino acid sequence variations that might account for antigenic-opsonogenic variation can be identified.

ACKNOWLEDGMENTS

We are indebted to Yung-Fu Chang and Linda Coogle for their help in sequence analysis.

The financial support of the Johnston Memorial Fund and the Keeneland Association is gratefully acknowledged.

REFERENCES

1. **Bazely, P. L.** 1942. Studies with streptococci. 4. Cross immunity to *Streptococcus equi*. Aust. Vet. J. **18:**189–194.
2. **Bessen, D. E., and V. A. Fischetti.** 1992. Nucleotide sequences of two adjacent M or M-like protein genes of group A streptococci: different RNA transcript levels and identification of a unique immunoglobulin A-binding protein. Infect. Immun. **60:**124–135.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:**1513–1523.
4. **Boyle, M. D. P., J. Howlitzky, R. Raeder, and A. Podbielski.** 1994. Analysis of genes encoding two unique type IIa immunoglobulin G-binding proteins expressed by a single group A streptococcal isolate. Infect. Immun. **62:**1336–1347.
5. **Clayton, C. E., and M. R. Mowatt.** 1989. The procyclic acidic repetitive proteins of *Trypanosoma brucei*. J. Biol. Chem. **264:**15088–15093.
6. **Fahnestock, S. R., P. Alexander, J. Nagle, and D. Filpula.** 1986. Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. **167:**870–880.
7. **Fischetti, V. A.** 1989. Streptococcal M protein: molecular design and biological behavior. Clin. Microbiol. Rev. **2:**285–314.
8. **Fischetti, V. A., V. Pancholi, and O. Schneewind.** 1991. Common characteristics of the surface proteins from gram-positive cocci, p. 290–294. *In* G. M. Dunny, P. P. Cleary, and L. L. McKay (ed.), Genetics and molecular biology of streptococci, lactococci, and enterococci. American Society for Microbiology, Washington, D.C.
9. **Frithz, E., L. O. Heden, and G. Lindahl.** 1989. Extensive sequence homology between IgA receptor and M protein in *Streptococcus pyogenes*. Mol. Microbiol. **3:**1111–1119.
10. **Galán, J. E., and J. F. Timoney.** 1985. Mucosal nasopharyngeal immune response of the horse to protein antigens of *Streptococcus equi*. Infect. Immun. **47:**623–628.
11. **Galán, J. E., and J. F. Timoney.** 1988. Immunologic and genetic comparison of *Streptococcus equi* isolates from the United States and Europe. J. Clin. Microbiol. **26:**1142–1146.
12. **Gomi, H., T. Hozumi, S. Hattori, C. Tagawa, F. Kishimoto, and L. Bjorck.** 1990. The gene sequence and some properties of protein H, a novel IgG binding protein. J. Immunol. **144:**4046–4052.
13. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207:**196–203.
14. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114:**193–197.
15. **Kasai, K., R. Nobata, and E. Rya.** 1944. On the incidence of *Streptococcus hemolyticus* in the normal tonsils of horses and the typing of equine tonsillar streptococci. Jpn. J. Vet. Sci. **6:**116–123.
16. **Kehoe, M., and K. Timmis.** 1984. Cloning and expression in *Escherichia coli* of the streptolysin O determinant from *Streptococcus pyogenes*: characterization of the cloned streptolysin O determinant and demonstration of the absence of substantial homology with determinants of other thiol-activated toxins. Infect. Immun. **43:**804–810.
17. **Khandke, K. M., T. Fairwell, and B. N. Manjula.** 1987. Difference in the structural features of streptococcal M proteins from nephritogenic and rheumatogenic serotypes. J. Exp. Med. **166:**151–162.
18. **Kilpper-Bälz, R., and K. H. Schleifer.** 1984. Nucleic acid hybridization and cell wall composition studies of pyogenic streptococci. FEMS Microbiol. Lett. **24:**355–364.
19. **Lancefield, R. C., and G. E. Perlmann.** 1952. Preparation and properties of type-specific M antigen from group A type 1 haemolytic streptococcus. J. Exp. Med. **96:**71–82.
20. **Miller, L., C. Gray, E. Beachey, and M. Kehoe.** 1988. Antigenic variation among group A streptococcal M proteins: nucleotide sequence of the serotype 5 M protein gene and its relationship with genes encoding types 6 and 24 M proteins. J. Biol. Chem. **263:**5668–5675.
21. **Moore, B. O., and J. T. Bryans.** 1969. Antigenic classification of group C animal streptococci. J. Am. Vet. Med. Assoc. **155:**416–420.
22. **Novotny, J., and L. Auffray.** 1984. A program for prediction of protein secondary structure from nucleotide sequence data. Nucleic Acid Res. **12:**243–255.
23. **Piszkiewicz, D., M. Landon, and E. L. Smith.** 1970. Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. Biochem. Biophys. Res. Commun. **40:**1173–1178.
24. **Pycock, J. F., W. E. Allen, and T. H. Morris.** 1987. A rapid procedure for isolation of equine neutrophils. Res. Vet. Sci. **42:**411–412.
25. **Sjobring, U.** 1992. Isolation and molecular characterization of a novel albumin-binding protein from group G streptococci. Infect. Immun. **60:**3601–3608.
26. **Stamp, T. C., and E. B. Henry.** 1937. The immunizing activity of certain chemical fractions isolated from haemolytic streptococci. Lancet **i:**257–259.
27. **Stewart, G. S., A. B. Lubinsky-Mink, C. G. Jackson, A. Cassel, and J. Kuhn.** 1986. pHG165: a pBR322 copy number derivative of pUC8 for cloning and expression. Plasmid **15:**172–181.

27a. **Timoney, J. F.** Unpublished data.

28. **Timoney, J. F., and M. Mukhtar.** 1992. Variability in the M proteins of equine strains of *Streptococcus equi* subsp. *zooepidemicus*, p. 15–20. *In* W. Plowright, P. D. Rossdale, and J. F. Wade (ed.), Equine infectious diseases VI. R&W Publications, Newmarket, England.
29. **Timoney, J. F., and J. Trachman.** 1985. Immunologically reactive proteins of *Streptococcus equi*. Infect. Immun. **48:**29–34.

29a. **Timoney, J. F., and J. Walker.** Unpublished data.

30. **Tsonic, P. A., and T. Manes.** 1988. Rapid phage isolation without the use of enzymes. BioTechniques **6:**950–951.
31. **von Heijne, G., and L. Abrahmsen.** 1989. Species specific variation in signal peptide design. Implications for protein secretion in foreign hosts. FEBS Lett. **244:**439–446.
32. **Walker, J. A., and J. F. Timoney.** 1995. Application of the polymerase chain reaction to the study of the molecular basis of M protein variation in *Streptococcus equi* subsp. *zooepidemicus*, p. 195–200. *In* T. Kumanomido, and W. Plowright (ed.), Equine infectious diseases VII. R&W Publications, Newmarket, England.
33. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.