INFECTION AND IMMUNITY, Aug. 1993, p. 3318–3326
0019-9567/93/083318-09$02.00/0

Vol. 61, No. 8

Repeats in an Extracellular Protein of Weakly Pathogenic Strains of *Streptococcus suis* Type 2 Are Absent in Pathogenic Strains

HILDE E. SMITH,[1]* FRANS H. REEK,[1] URI VECHT,[2] ARNO L. J. GIELKENS,[1] AND MARI A. SMITS[1]

Departments of Molecular Biology[1] and Bacteriology,[2] DLO-Central Veterinary Institute, P.O. Box 65, 8200 AB Lelystad, The Netherlands

Received 28 April 1993/Accepted 10 May 1993

***Streptococcus suis* type 2 strains that are pathogenic for pigs produce a 110-kDa extracellular protein factor (EF). Nonpathogenic and weakly pathogenic strains do not produce EF or produce a protein (EF*) that is immunologically related to EF. To study the pathogenesis of *S. suis* type 2 in pigs and to develop tools and methods for the control of *S. suis* type 2 infections, we cloned and characterized the genes encoding EF and various EF* proteins. Analysis of the deduced amino acid sequences showed that the first 833 amino acids at the N terminus of the EF and EF* proteins were nearly identical. The proteins differed, however, at their C termini. Unlike the 110-kDa EF protein, the EF* proteins contained several repeated units of 76 amino acids. The number and arrangement of the repeats in the EF* proteins varied. The data suggest that the gene encoding EF could have evolved from an *epf** gene by a specific deletion event. The lack of repeated amino acid units in the EF protein may be related to virulence.**

Streptococcus suis type 2 infections are a common cause of septicemia, arthritis, meningitis, and sudden death in young pigs (4, 26) and are also the cause of meningitis in humans (1). During the last few years, *S. suis* type 2 infections have become a major problem in almost all countries with an intensive pig industry. Attempts to control the disease in pigs are hampered by the lack of effective vaccines and sensitive diagnostic tests.

To understand the pathogenesis of the disease and to develop new diagnostic tools, we initially searched for proteins which are associated with pathogenic strains of *S. suis* type 2 (25, 27, 28). In these studies, we identified three phenotypically different *S. suis* type 2 strains. Strains of the first phenotype produced a 136-kDa cell envelope-associated protein, called the muramidase-released protein (MRP), and a 110-kDa extracellular protein factor (EF). Strains of this phenotype (MRP^+ EF^+) were frequently isolated from the organs of diseased pigs and caused severe clinical signs of disease in experimentally infected pigs. Strains of the second phenotype did not produce MRP and EF. Strains of this phenotype (MRP^- EF^-) were frequently isolated from the tonsils of healthy pigs and caused no clinical signs of disease in experimentally infected pigs. Strains of the third phenotype produced MRP and a protein that was immunologically related to EF. This protein was designated EF*. Different strains produced EF* proteins of different sizes (M_r, >110,000). Strains of the last phenotype (MRP^+ EF*) were isolated at low frequencies from the organs of diseased pigs and from the tonsils of healthy pigs. These strains were, however, isolated at a high frequency from human patients. MRP^+ EF* strains caused almost no clinical signs of disease in experimentally infected pigs. These data suggested that MRP and EF are markers for pathogenic strains of *S. suis* type 2. Nonpathogenic or weakly pathogenic strains do not synthesize EF or synthesize EF*.

To get insight into the role of MRP and EF in the pathogenesis of *S. suis* type 2 infections, it would be of benefit to study the molecular structure of MRP and EF. We recently described the cloning and molecular characterization of the gene encoding MRP (23). Here we describe the cloning and molecular characterization of the gene encoding EF. Moreover, we describe the cloning and characterization of the genes encoding five different forms of EF*. The results showed that the *epf* and *epf** genes have extensive regions of homology. A striking feature of all EF* proteins was the presence of long regions of tandem repeats. These repeats were absent in EF. The number of repeats differed between the various EF* types.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Escherichia coli* JM101 (14) and LE392 (16) were used as hosts for recombinant plasmids and bacteriophages. The pathogenic MRP^+ EF^+ strain D282 of *S. suis* type 2 (25) and 30 MRP^+ EF* strains were used in these studies. Sixteen MRP^+ EF* strains were isolated from human patients, five strains were isolated from the tonsils of slaughtered pigs, seven strains were obtained from the organs of diseased pigs, and two strains were of unknown origin (28). The *E. coli* strains were grown in Luria broth (15). Ampicillin was added as needed to a final concentration of 50 μg/ml. *S. suis* strains were grown in Todd-Hewitt broth (Oxoid, Ltd., London, England).

Construction and immunological screening of the DNA library. A DNA library of *S. suis* type 2 strain D282 was constructed in LambdaGEM-11 by the methods recommended by the manufacturer of the cloning vector (Promega, Madison, Wis.). Recombinant bacteriophages were plated on *E. coli* LE392 and incubated for 16 h at 37°C. Nitrocellulose filters (Schleicher and Schuell, Inc., Dassel, Germany) were placed on the plaques, and the agar plates were further incubated for 2 h at 37°C. Recombinants that produced EF were detected by use of a monoclonal antibody (MAb) directed against EF (27). Bound antibodies were visualized with anti-mouse immunoglobulins conjugated with alkaline phosphatase (Zymed Laboratories, Inc., San Francisco, Calif.) as described by Sambrook et al. (19). Selected EF-positive clones were purified by several rounds of single-plaque isolation and immunological screening.

* Corresponding author.

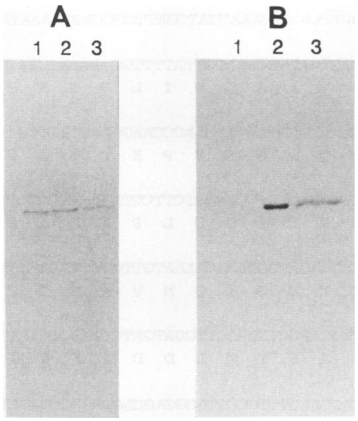

FIG. 1. (A) Western blot analysis of proteins from plaques of two recombinant bacteriophages. Ten plaques of each recombinant bacteriophage were collected from an agar plate. The proteins in the plaques were eluted with 200 μl of Laemmli buffer (13) for 24 h at 4°C. Samples of 5 μl were applied to the gel. Lane 1, positive control, culture supernatant of *S. suis* type 2 strain D282; lane 2, recombinant bacteriophage 1; lane 3, recombinant bacteriophage 2. (B) Western blot analysis of proteins encoded by recombinant plasmid pEF2-19. Lane 1, negative control, JM101(pKUN19); lane 2, culture supernatant of *S. suis* type 2 strain D282; lane 3, JM101(pEF2-19). The blots were screened with the EF-specific MAb.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis and Western immunoblot analysis. Proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis with 4% acrylamide stacking and 6% acrylamide separation gels (13). The separated proteins were transferred to nitrocellulose membranes with a Semi-Dry transfer cell (Bio-Rad Laboratories, Richmond, Calif.). To detect EF and EF* proteins, the blots were incubated with an MAb directed against EF (27). Bound antibodies were visualized as described above.

DNA manipulations and nucleotide sequence analysis. Polymerase chain reaction (PCR)-amplified DNA fragments and restriction fragments were subcloned into plasmid vector pKUN19 (12) by standard molecular biological techniques (19). Progressive unidirectional deletions were made with the Erase-a-Base system from Promega. DNA sequences were determined by the dideoxy chain termination method (20). To exclude errors made by *Taq* DNA polymerase during DNA amplification, we sequenced a mixture of six independently derived PCR clones. DNA and protein sequences were analyzed by the PCGENE (Intelligenetics Corp., Mountain View, Calif.) and the Genetics Computer Group (University of Wisconsin) software packages.

Southern blotting and hybridization. DNA was transferred to GeneScreen Plus membranes (New England Nuclear Corp., Dreieich, Germany) as described by Sambrook et al. (19). DNA probes were labeled with [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) by use of a random-primed labeling kit (Boehringer GmbH, Mannheim, Germany). The blots were hybridized with DNA probes as recommended by the supplier of the GeneScreen Plus membranes. After hybridization, the membranes were washed twice with a solution of 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate [pH 7.0]) for 5 min at room

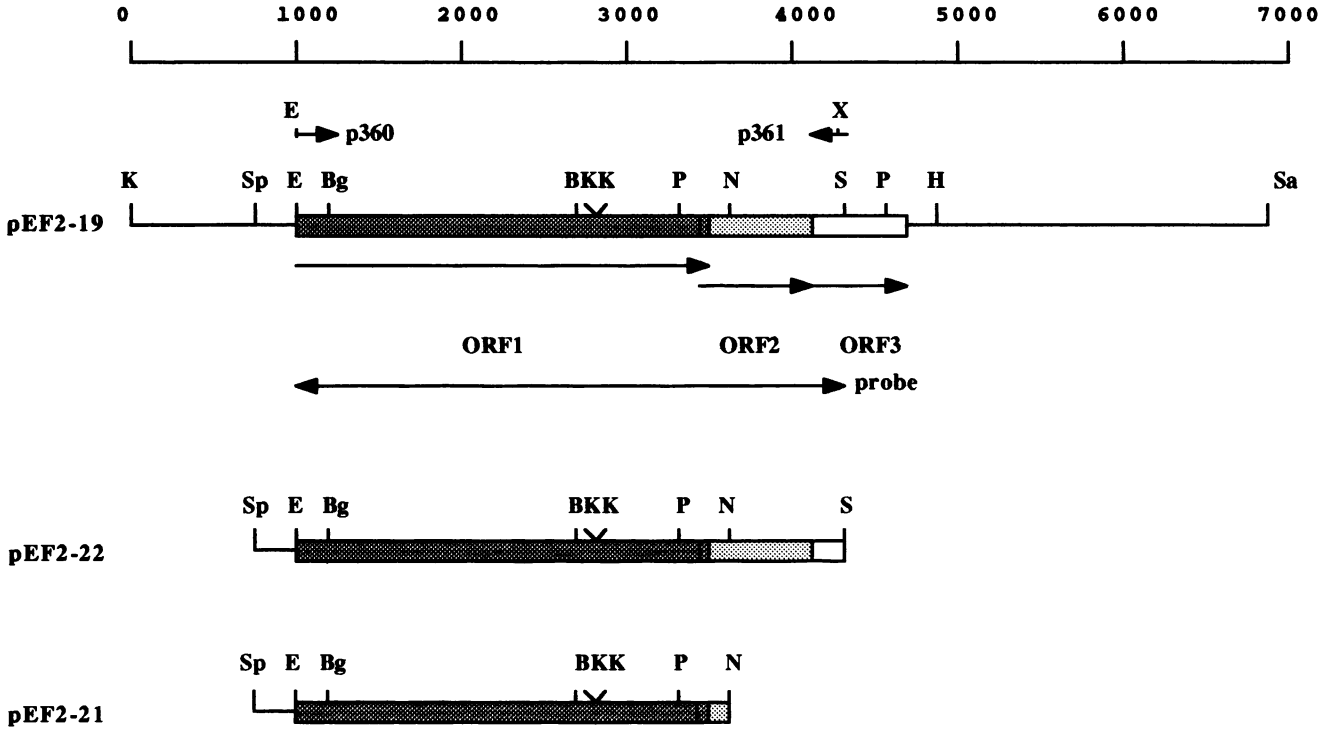

FIG. 2. Restriction maps of *epf*-containing inserts of pEF2-19, pEF2-22, and pEF2-21. ORFs 1, 2 and 3 are boxed. Restriction sites: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RV; H, *Hin*dIII; K, *Kpn*I; N, *Nar*I; P, *Pst*I; S, *Sna*BI; Sa, *Sal*I; Sp, *Spe*I. The positions of oligonucleotides p360 and p361, which include recognition sites for *Eco*RI and *Xba*I, respectively, are indicated by arrows. The *Eco*RV-*Sna*BI fragment was used as a probe in the Southern hybridizations.

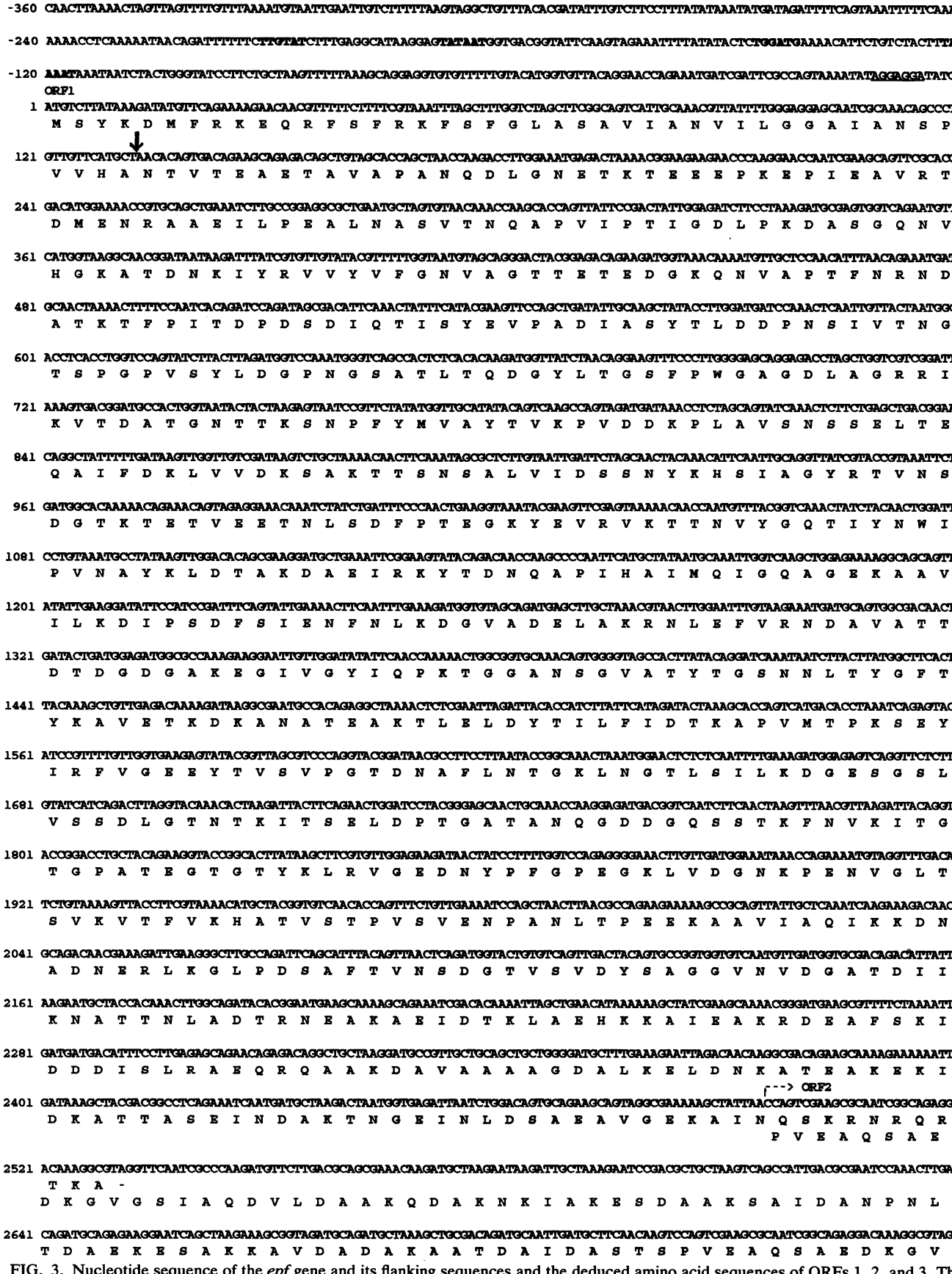

FIG. 3. Nucleotide sequence of the *epf* gene and its flanking sequences and the deduced amino acid sequences of ORFs 1, 2, and 3. The putative ribosome-binding site is underlined. Nucleotides in boldface type indicate the −35 and −10 regions of the putative promoter sequences. The beginnings of ORFs 2 and 3 are indicated. The arrowheads indicate the regions of dyad symmetry. The vertical arrow indicates the potential signal peptidase cleavage site.

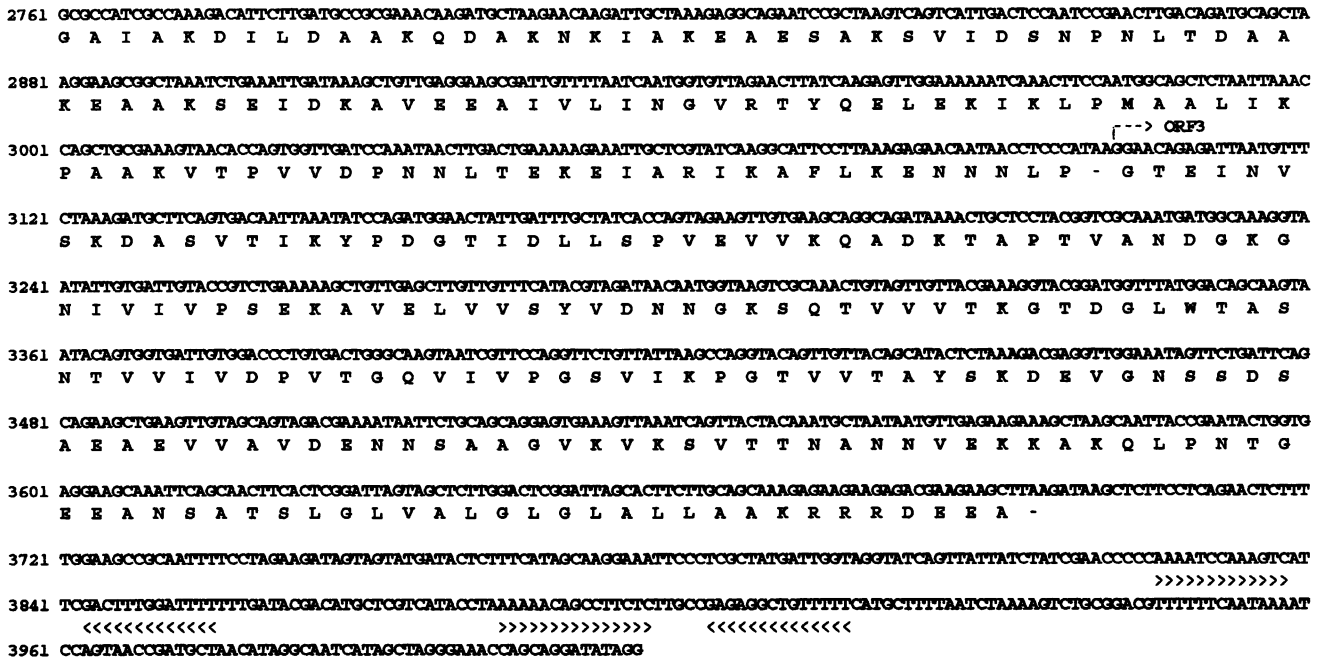

FIG. 3—*Continued.*

temperature and twice with a solution of 0.1× SSC plus 0.5% sodium dodecyl sulfate for 30 min at 65°C.

Amplification and cloning of genomic DNA fragments. To amplify *epf** sequences, genomic DNA from five different MRP^+ EF* strains was used as the template in PCR. The sequences of the PCR oligonucleotide primers were as follows: p360, 5′-ATGTAATTGAATTCTCTTTTTAAGT-3′; and p361, 5′-AAACGTCCGCAGACTTCTAGATTAAAAG C-3′. The positions of these oligonucleotides are indicated in Fig. 2. The underlined sequences indicate the recognition sites for the restriction enzymes *Eco*RI and *Xba*I in p360 and p361, respectively, which were introduced for cloning purposes. Amplified DNA fragments were isolated by agarose gel electrophoresis and extraction from the gel with Gene Clean (Bio101, La Jolla, Calif.). The purified fragments were digested with *Eco*RI and *Xba*I and cloned into plasmid pKUN19 (12).

Nucleotide sequence accession numbers. The *epf* sequence of strain D282 and the *epf** sequence of strain 1890 have been assigned GenBank-EMBL accession numbers X71881 and X71880, respectively.

RESULTS

Cloning of the *epf* gene. We constructed a DNA library of strain D282 of *S. suis* type 2 in the LambdaGEM11 replacement vector. We obtained approximately 5×10^5 recombinants per μg of DNA. Of this library, we tested 2,000 plaques of recombinant phages for the presence of an antigenic determinant of EF. Two plaques reacted positively with an MAb specific for EF. Both recombinants encoded a protein that comigrated with EF from *S. suis* and that reacted with the EF-specific MAb (Fig. 1A). This indicated that both recombinant bacteriophages contained the entire *epf* gene.

From restriction enzyme analyses, we concluded that the two recombinant bacteriophages had a common DNA region of about 13 kb. Parts of this common DNA region were subcloned into plasmid pKUN19, and the proteins expressed by these plasmids were analyzed by Western blotting. A plasmid that contained a 6.8-kb *Kpn*I-*Sal*I fragment (pEF2-19, Fig. 2) coded for EF (Fig. 1B). In contrast, plasmids that contained the 5.8-kb *Eco*RV-*Sal*I or the 5.3-kb *Bgl*II-*Sal*I fragment (Fig. 2) did not express EF (results not shown). These data indicated that the *Eco*RV and *Bgl*II sites are within regions required for the expression of EF and that the gene for EF (*epf*) is located between the *Kpn*I and *Sal*I sites.

Structure of the *epf* gene. We determined the nucleotide sequence of a 4.0-kb *Spe*I-*Hin*dIII fragment containing the entire *epf* gene and the regions flanking it. The sequence (Fig. 3) revealed the presence of three major open reading frames (ORFs). ORF 1 (nucleotides 1 to 2530) coded for a polypeptide of 843 amino acids, ORF 2 (nucleotides 2496 to 3099) coded for a polypeptide of 201 amino acids, and ORF 3 (nucleotides 3102 to 3693) coded for a polypeptide of 197 amino acids. ORF 1 contained a putative ATG initiation codon preceded by a sequence similar to ribosome-binding sites found in other gram-positive bacteria (8). In contrast, ORFs 2 and 3 did not possess putative initiation codons and ribosome-binding sites. The 3′ end of ORF 1 and the 5′ end of ORF 2 overlapped each other, albeit in different frames. ORFs 2 and 3 are separated from each other by a single TAA stop codon. ORF 1 was preceded by potential promoter sequences (Fig. 3). Downstream of ORF 3, we found two regions of extensive dyad symmetry which were followed by stretches of thymidine residues. These DNA motifs may function as rho-independent transcription termination signals (18, 24).

EF protein. From its size, ORF 1 was expected to encode EF. To confirm this, we subcloned two different fragments: an *Spe*I-*Sna*BI fragment that contained ORFs 1 and 2 (pEF2-22, Fig. 2) and an *Spe*I-*Nar*I fragment that contained only ORF 1 (pEF2-21, Fig. 2). Both plasmids coded for a protein that was indistinguishable from the EF secreted by *S. suis* (data not shown). Therefore, we concluded that only ORF 1 is necessary for expression of EF. Because EF is found exclusively in the supernatant of *S. suis* cultures, we

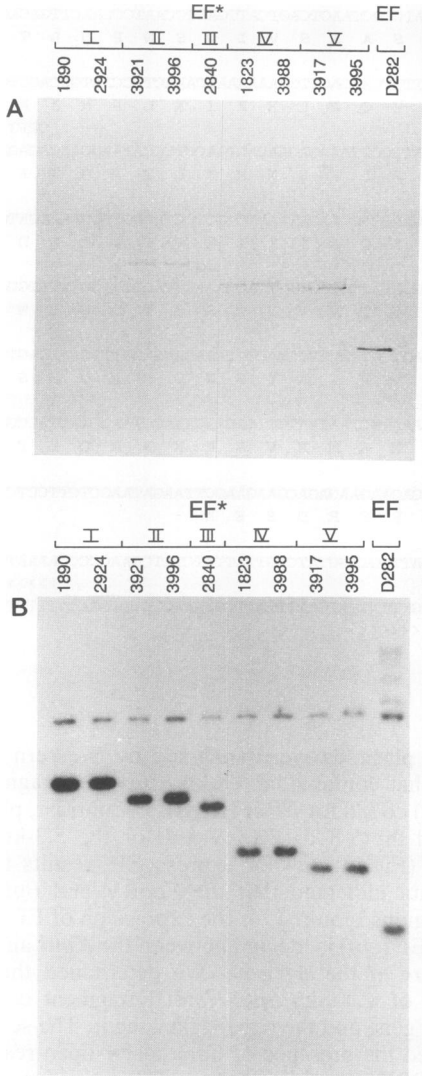

FIG. 4. (A) Western blot of proteins present in 30 μl of the culture supernatants of various *S. suis* type 2 strains of the MRP$^+$ EF* phenotype. Blots were screened with an anti-EF MAb (27). The strain numbers and the classes to which they belong are given above each lane. (B) Southern blot of chromosomal DNA from various *S. suis* type 2 strains of the MRP$^+$ EF* phenotype. Chomosomal DNA was digested with the restriction enzyme *Pst*I. The blot was hybridized with the *Eco*RV-*Sna*BI fragment isolated from plasmid pEF2-19 (Fig. 2). The strain numbers and classes to which they belong are given above each lane.

expected to find a gene encoding a precursor protein with a signal peptide at the N terminus. Indeed, the first 44 residues of EF had the characteristics of a typical signal peptide (Fig. 3) (29). The calculated molecular weight of the mature protein is 85,200. This does not correspond to the molecular weight of EF (110,000) calculated from sodium dodecyl sulfate-polyacrylamide gels.

No significant similarities were found between the deduced amino acid sequence of EF and the protein sequences in the EMBL data base.

Homology between *epf* and *epf genes.** Strains of *S. suis* type 2 that produce MRP but not EF produce EF* proteins of various sizes (27, 28). Among 30 MRP$^+$ EF* strains, we found five different molecular masses for EF*. Class I strains produced an EF* protein of 195 kDa; class II strains produced one of 180 kDa; class III strains produced one of 175 kDa; class IV strains produced one of 160 kDa; and class V strains produced one of 155 kDa (Fig. 4A).

Since the EF and EF* proteins were immunologically related, we examined whether there is homology between the *epf* gene and the genes encoding the EF* proteins. Chromosomal DNA from different MRP$^+$ EF* strains and an MRP$^+$ EF$^+$ strain was digested with the restriction enzyme *Pst*I and hybridized with a ^{32}P-labeled *Eco*RV-*Sna*BI fragment (Fig. 2) containing the entire *epf* gene. All digests contained two *Pst*I fragments that hybridized strongly with the probe (Fig. 4B). The largest fragment was identical in size in all strains. The length of the smallest fragment varied among the strains. This variation correlated well with the variation in the molecular weight of EF and EF*. These data suggest that the genes encoding EF and EF* are homologous. Since the smallest hybridizing fragment of strain D282 covers the 3′ end of the *epf* gene (Fig. 2), these data furthermore suggest that the major differences between the *epf* and *epf** genes are located at their 3′ ends.

Amplification and cloning of *epf genes.** To isolate the different *epf** genes, we amplified the *epf**-containing DNA by PCR with genomic DNA from five MRP$^+$ EF* strains (one representative of each class defined above) as templates. As expected, the lengths of the amplified fragments varied among the strains (results not shown). Again, this variation correlated with the variation in the molecular weight of EF*.

***epf** gene of class I.** We determined the nucleotide sequence of a 6.5-kb *Eco*RI-*Xba*I fragment comprising the entire class I *epf** gene (*epf**I). Analysis of the sequence revealed that the *epf** and *epf* genes are very similar. The first 2,499 nucleotides of the *epf**I and *epf* genes were almost identical (99%). In the *epf**I gene, this region was followed by a 2,433-bp sequence (Fig. 5) of which only 65 bp (positions 4640 to 4704) were found in the *epf* gene. Downstream from this 2,433-bp fragment, the *epf**I gene sequence was again almost identical to the corresponding region of the *epf* gene.

Class I EF* protein. The sequence data (Fig. 6) showed that the C terminus of the class I EF* protein contained 10.5 imperfect repeats of 76 amino acids (denoted R1 to R11, Fig 5 and 6). Repeats R1 through R4 and R6 through R11 were contiguous, but R4 and R5 were separated from each other by 113 amino acids, and R5 and R6 were separated by 22 amino acids (Fig. 5 and 6). The amino acid sequences of R6 through R11 were highly conserved, whereas the sequences of R1 through R5 were more variable (Fig. 7). One particular amino acid sequence, NPNL, was conserved in all repeated units. No significant homology was found between the class I EF* sequence and the protein sequences in the EMBL data base.

EF* and EF. Compared with the *epf**I gene, the *epf* gene lacked two fragments, one of 2,140 bp (*epf**I positions 2500 to 4640) and another of 228 bp (*epf**I positions 4705 to 4933; Fig. 5 and 6). Therefore, R1 through R7, R9, and part of R8 are not present in EF. As a result of the 2,140-bp deletion, the translational reading frame of the downstream region differs in the *epf* and in the *epf**I genes. As a result, EF protein synthesis is aborted within the truncated R8, whereas EF*I protein synthesis continues for 68 amino acids downstream of R11. Consequently, EF lacks all repeated amino acid elements.

***epf** genes of classes II, III, IV, and V.** Southern hybrid-

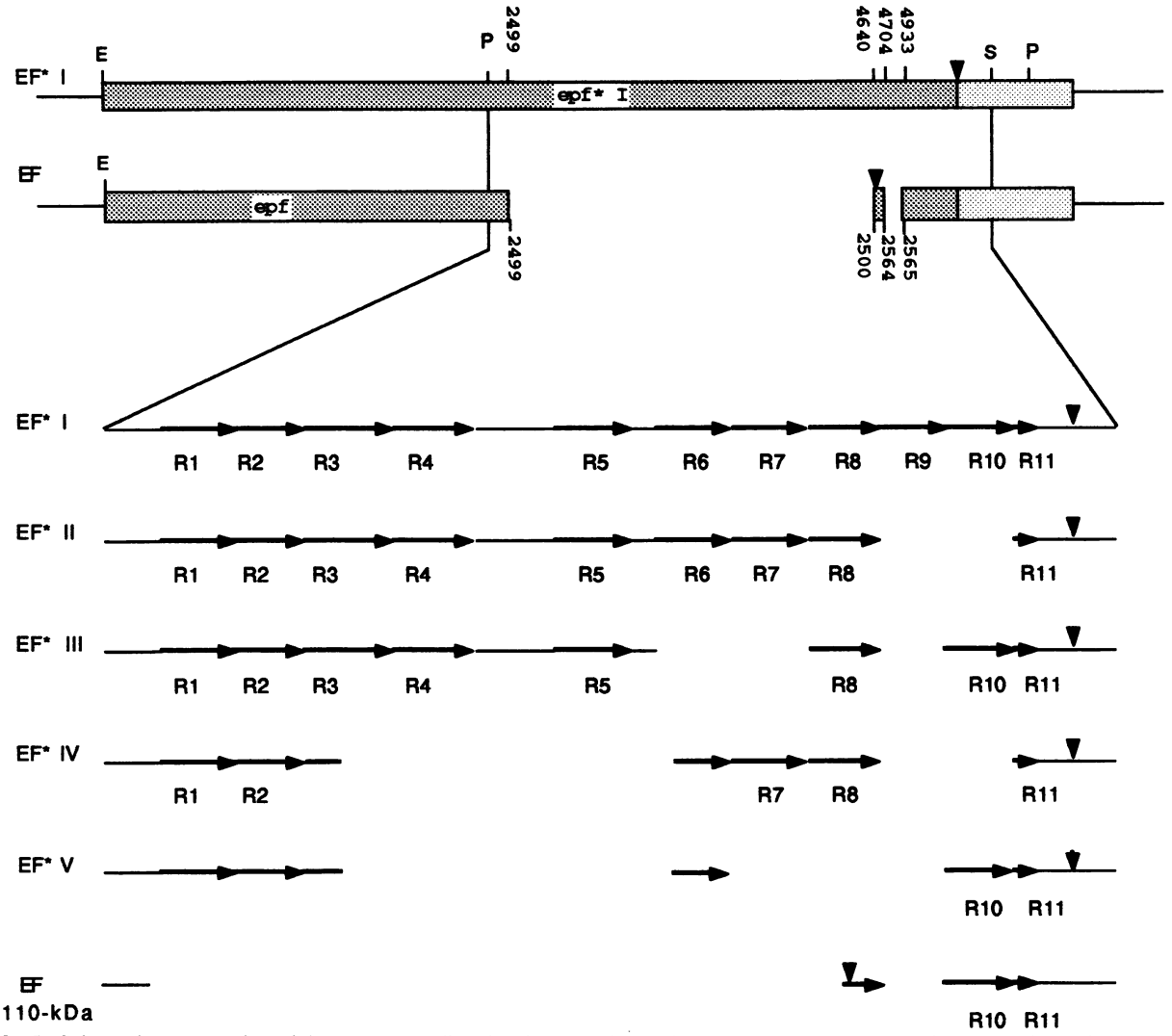

FIG. 5. Schematic presentation of the genes encoding the class I EF* protein and the 110-kDa EF protein. Expansions of the *Pst*I-*Sna*BI regions of the five classes of *epf** genes and of the *epf* gene of the pathogenic strain D282 are presented. The arrows indicate regions encoding the repeated amino acid units R1 through R11. Gaps indicate regions missing in the different strains. E, *Eco*RV; P, *Pst*I; S, *Sna*BI. The vertical arrowheads indicate the positions of in-frame translation stop codons. The numbers refer to the nucleotide sequence positions of the *epf* and *epf**I genes as indicated in Fig. 3 and 6.

ization analyses suggested that the differences between the genes encoding the various EF* proteins were located on *Pst*I fragments at the distal ends of the genes (Fig. 4B). Therefore, we determined the nucleotide sequences of these *Pst*I fragments derived from the class II, III, IV, and V strains. Comparison of these nucleotide sequences with the sequence of the *epf**I gene showed that the various *epf** genes were highly homologous in this region. They differed, however, in the number and arrangement of the sequences encoding the repeated units (Fig. 5). Compared with the *epf**I gene, the *epf**II gene lacked the regions encoding R9 and R10; the *epf**III gene lacked the regions encoding R6, R7, and R9; the *epf**IV gene lacked the regions encoding R9 and R10 and a fragment of 1,032 bp which contained the regions encoding R4 and R5 and parts of R3 and R6; and the *epf**V gene lacked the regions encoding R7, R8, and R9 and the 1,032-bp fragment.

DISCUSSION

We cloned and sequenced the genes coding for the EF and EF* proteins of *S. suis* type 2. It appeared that the sequences were highly homologous, which suggests a close evolutionary relationship between these genes. Since we previously found that EF is associated with pathogenic strains and EF* is associated with weakly pathogenic strains (27, 28), this is an interesting finding.

We located the *epf* gene on a 6.8-kb *Kpn*I-*Sal*I DNA fragment. The established DNA sequence showed three major ORFs. Because ORF 3 and the 3′ end of ORF 2 could be deleted without any effect on the mobility of EF on sodium dodecyl sulfate-polyacrylamide gels, we concluded that EF is encoded by ORF 1. The molecular weight of the expected mature form of the polypeptide encoded by ORF 1 (85,200) differed, however, from the molecular weight of EF

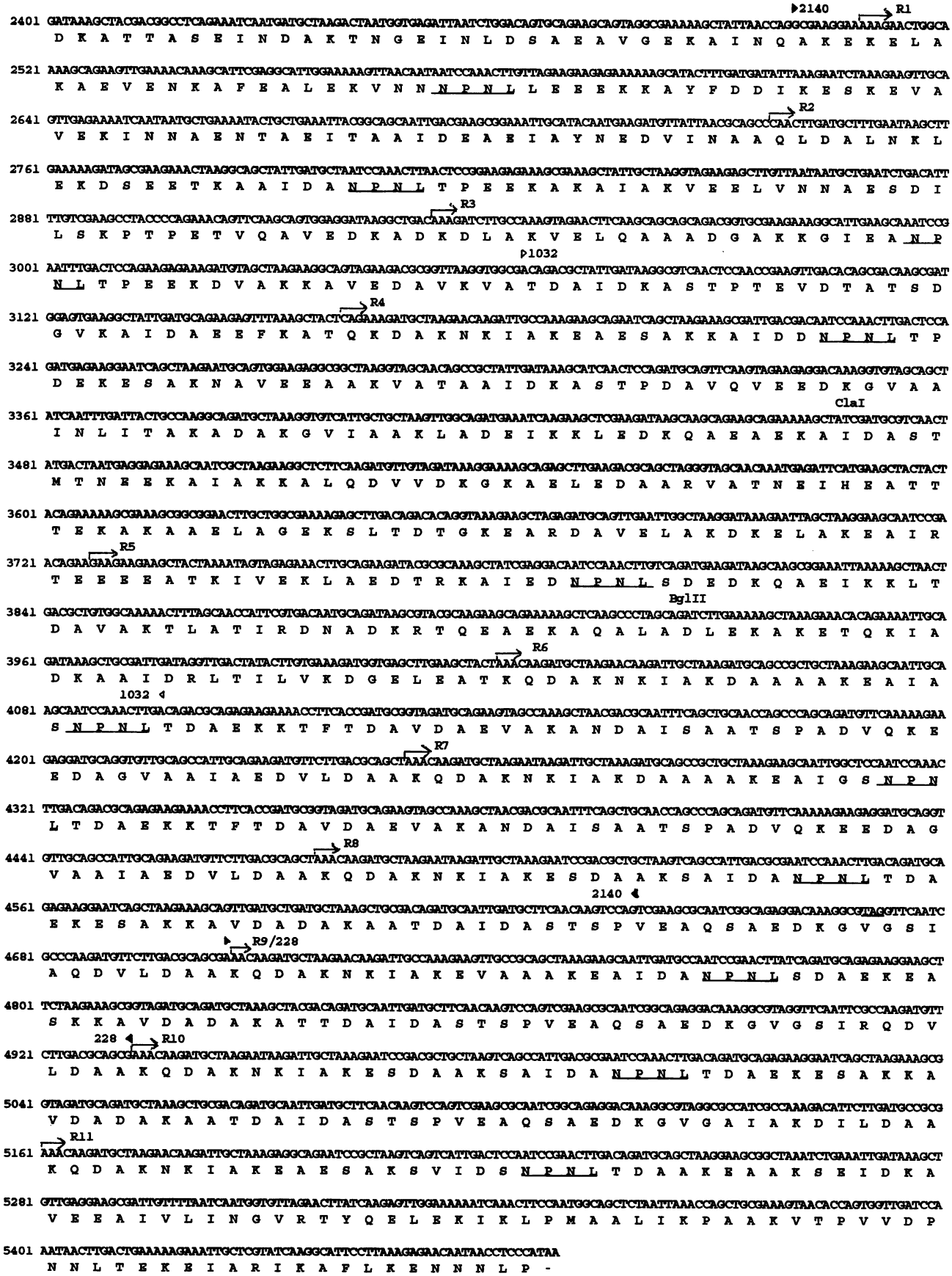

FIG. 6. Nucleotide sequence of the 3′ end of the *epf** gene of strain 1890 and the deduced amino acid sequence of the C terminus of the EF*I protein. The solid arrowheads border the 2,140-bp and the 228-bp regions which are absent in the gene encoding the 110-kDa EF protein. The region between the two open arrowheads is absent in the genes encoding the class IV and class V EF* proteins. Labeled horizontal arrows indicate the starts of repetitive units R1 through R11. The highly conserved amino acid sequence NPNL, which is present in all repeats, is underlined.

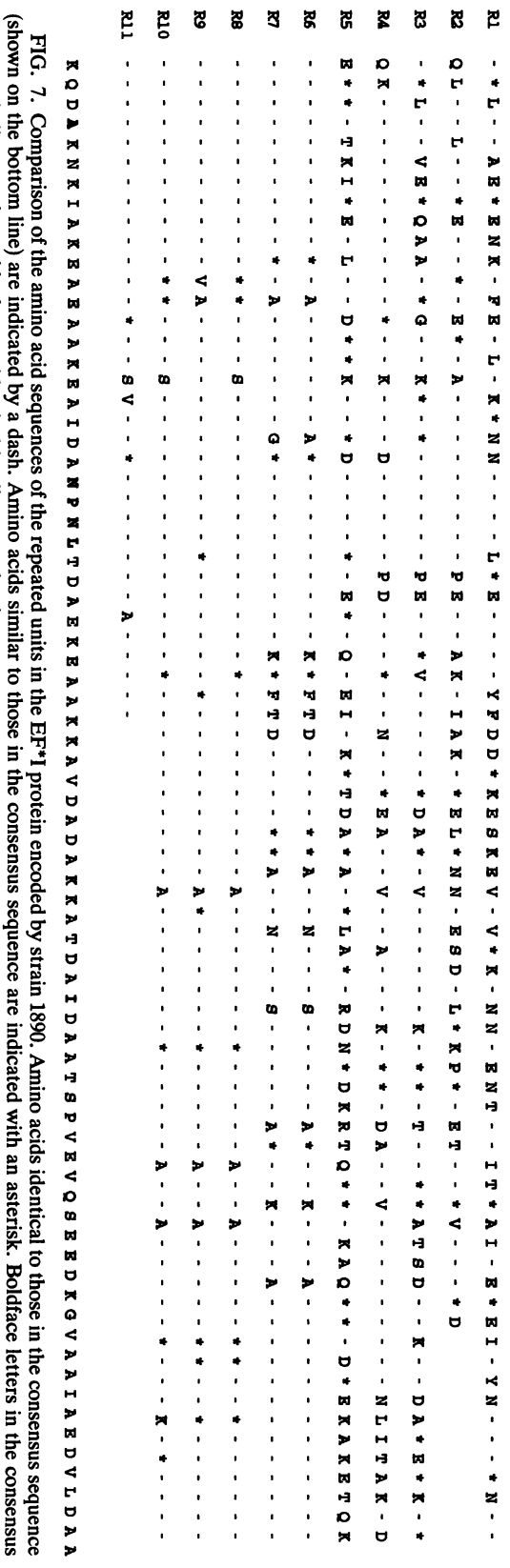

FIG. 7. Comparison of the amino acid sequences of the repeated units in the EF*I protein encoded by strain 1890. Amino acids identical to those in the consensus sequence (shown on the bottom line) are indicated by a dash. Amino acids similar to those in the consensus sequence are indicated with an asterisk. Boldface letters in the consensus sequence indicate amino acids that are identical in all repeated units.

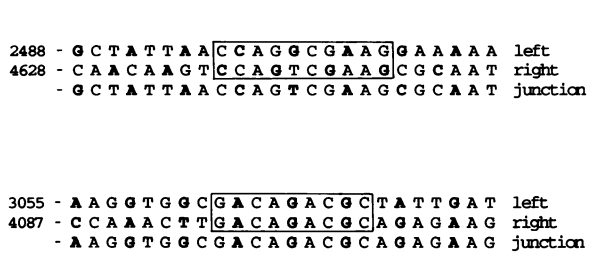

FIG. 8. Nucleotide sequences near the left and right borders of the fragments missing in the *epf* gene (A) and in the *epf** genes of classes IV and V (B). The top and middle sequences represent regions flanking the left and right ends of the deletions, respectively. The bottom sequences show the junctions as found in the *epf* gene (A) and in the *epf** genes of classes IV and V (B). Directly repeated sequences are boxed. Boldfaced nucleotides indicate the first bases of the translational triplets. The numbers refer to the nucleotide positions in the *epf**I gene (Fig. 6).

estimated from sodium dodecyl sulfate-polyacrylamide gels (110,000). It has been observed for other proteins that calculated and estimated molecular weights can differ (3, 7). In particular, highly charged proteins often migrate anomalously in sodium dodecyl sulfate gels (3, 7). The protein encoded by ORF 1 contained 25% charged amino acids. This could explain the apparent higher molecular weight of EF in gels.

Southern analysis indicated that the 5′ ends of the *epf* and *epf** genes were strongly conserved among the strains examined. This was confirmed by sequence analysis of the class I *epf** gene. The 3′ ends of the *epf** genes, however, differed from each other and from the 3′ end of the *epf* gene. We found that the proteins encoded by the *epf** genes are characterized by a variable number of repeated amino acids near the C terminus. In contrast, the EF protein lacks such repeated units. This difference between EF and EF* proteins might be important for the difference in pathogenicity between MRP^+ EF^+ and MRP^+ EF* strains.

Many bacteria produce proteins that contain repeated units of amino acids (6, 9, 22, 30, 31). Amino acid repeats are frequently involved in the binding of a protein to specific ligands (6, 21, 22, 30, 31). Whether the repeats in the EF* proteins likewise affect binding to a specific ligand is unknown. If they do, it is intriguing to speculate that the repeated units in the EF* proteins reduce the virulence of *S. suis* type 2 by binding to a specific ligand.

The EF* proteins varied in the number and arrangement of the repeated units of amino acids. Moreover, the amino acid sequences of some repeats were highly conserved, whereas the sequences of other repeats were more variable. Variation in amino acid repeats has commonly been observed in other bacterial proteins, such as the M protein of *Streptococcus pyogenes* (10, 11, 21), the glucosyltransferases of *Streptococcus downei* and *Streptococcus mutans* (31), and toxin A of *Clostridium difficile* (31). Because of this variation, the M proteins of *Streptococcus pyogenes* showed different antigenic specificities (9–11), and it has been proposed that this is a mechanism by which organisms can escape from the host immune system (9–11). The variation in the EF* proteins might serve a similar function.

The identity of the 5′ ends of the *epf* and *epf** genes indicates that one of these genes originated from the other or that they had a common ancestor. We believe that the variation in these genes resulted from duplication and/or

deletion of part of their sequences. In general, DNA duplications and deletions are believed to result from errors during replication by slipped-mispairing mechanisms or by unequal crossing-over (5). Both mechanisms require regions of DNA homology, which can be very short for slipped mispairing (5). The *epf* gene could have originated from the class I *epf** gene by a specific deletion of 2,140 bp. Analysis of the nucleotide sequences of the regions flanking the left and right borders of the deletion showed direct repeats of 10 bp (one mismatch) (Fig. 8). This short direct repeat could have been involved in the generation of the deletion. The *epf** genes of classes IV and V could have originated from the *epf** gene of class I by a specific deletion of 1,032 bp. The nucleotide sequences flanking the left and right borders of this deletion showed direct repeats of 9 bp (Fig. 8). This 9-bp direct repeat is likely to have been the deletion target site in this case. The variation in the number and arrangement of repeated units in the various *epf** genes could be explained by the duplication or deletion of complete repeats.

If it is true that the *epf* gene originated from a larger *epf** gene and that the absence of repeated units is important for virulence, this would imply that weakly pathogenic *S. suis* strains can be changed into pathogenic strains by one recombination event. If this idea proves true, it will complicate efforts to control *S. suis* infections. The frequency of such events is unknown. Studies with *E. coli* and *Bacillus subtilis* indicate that deletion frequencies resulting from recombination at a 9-bp direct repeat in the chromosome can amount to about 10^{-8} per generation (2, 17).

REFERENCES

1. **Arends, J. P., and H. C. Zanen.** 1988. Meningitis caused by *Streptococcus suis* in humans. Rev. Infect. Dis. **10**:131–137.
2. **Bron, S., S. Holzappel, G. Venema, and B. P. H. Peeters.** 1991. Plasmid deletion formation between short direct repeats in *Bacillus subtilis* is stimulated by single-stranded rolling-circle replication intermediates. Mol. Gen. Genet. **226**:88–96.
3. **Burton, Z., R. R. Burgess, D. Moore, S. Holder, and C. A. Gross.** 1981. The nucleotide sequence of the cloned *rpoD* gene for the RNA polymerase sigma subunit from *E. coli* K-12. Nucleic Acids Res. **9**:2889–2903.
4. **Clifton-Hadley, F. A.** 1983. *Streptococcus suis* type 2 infections. Br. Vet. J. **139**:1–5.
5. **Ehrlich, S. D.** 1989. Illegitimate recombination in bacteria, p. 797–829. *In* D. Berg and M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
6. **Fahnestock, S. R., P. Alexander, J. Nagle, and D. Filpula.** 1987. Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. **167**:870–880.
7. **Gill, D. R., and G. P. C. Salmond.** 1990. The identification of the *Escherichia coli ftsY* gene product: an unusual protein. Mol. Microbiol. **4**:575–583.
8. **Hager, P. W., and J. C. Rabinowitz.** 1985. Translational specificity in *Bacillus subtilis*, p. 1–32. *In* D. A. Dubnau (ed.), The molecular biology of the bacilli. Academic Press, Inc., New York.
9. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. The complete nucleotide sequence of type 6M protein of the group A *Streptococcus*: repetitive structure and membrane anchor. J. Biol. Chem. **261**:1677–1686.
10. **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.
11. **Jones, K. F., S. K. Hollingshead, J. R. Scott, and V. A. Fischetti.** 1988. Spontaneous M6 protein size mutants of group A streptococci display variation in antigenic and opsonogenic epitopes. Proc. Natl. Acad. Sci. USA **85**:8271–8275.
12. **Konings, R. N. H., E. J. M. Verhoeven, and B. P. H. Peeters.** 1987. pKUN vectors for the separate production of both DNA strands of recombinant plasmids. Methods Enzymol. **153**:12–34.
13. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
14. **Messing, J.** 1979. A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. Recombinant DNA technical bulletin. NIH publication no. 79–99, 2, no. 2, p. 43–48. National Institutes of Health, Bethesda, Md.
15. **Miller, J.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Murray, N. E., W. J. Brammer, and K. Murray.** 1977. Lambdoid phages that simplify the recovery of *in vitro* recombinants. Mol. Gen. Genet. **150**:53–58.
17. **Peeters, B. P. H., J. H. de Boer, S. Bron, and G. Venema.** 1988. Structural plasmid instability in *Bacillus subtilis*: effect of direct and indirect repeats. Mol. Gen. Genet. **212**:450–458.
18. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
19. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **Scott, J.** 1990. The M protein of group A *Streptococcus*: evolution and regulation, p. 177–203. *In* B. M. Iglewski and V. L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, Inc., San Diego, Calif.
22. **Signäs, C., G. Raucci, K. Jönsson, P.-E. Lindgren, G. M. Anantharamaiah, M. Höök, and M. Lindberg.** 1989. Nucleotide sequence of the gene for a fibronectin-binding protein from *Staphylococcus aureus*: use of this peptide sequence in the synthesis of biologically active peptides. Proc. Natl. Acad. Sci. USA **16**:699–703.
23. **Smith, H. E., U. Vecht, A. L. J. Gielkens, and M. A. Smits.** 1992. Cloning and nucleotide sequence of the gene encoding the 136-kilodalton surface protein (muramidase-released protein) of *Streptococcus suis* type 2. Infect. Immun. **60**:2361–2367.
24. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Devine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
25. **Vecht, U., J. P. Arends, E. J. van der Molen, and L. A. M. G. van Leengoed.** 1989. Differences in virulence between two strains of *Streptococcus suis* type 2 after experimentally induced infection of newborn germfree pigs. Am. J. Vet. Res. **50**:1037–1043.
26. **Vecht, U., L. A. M. G. van Leengoed, and E. R. M. Verheyen.** 1985. *Streptococcus suis* infections in pigs in The Netherlands (part one). Vet. Q. **7**:315–321.
27. **Vecht, U., H. J. Wisselink, M. L. Jellema, and H. E. Smith.** 1990. Identification of two proteins associated with virulence of *Streptococcus suis* type 2. Infect. Immun. **59**:3156–3162.
28. **Vecht, U., H. J. Wisselink, J. E. van Dijk, and H. E. Smith.** 1991. Virulence of *Streptococcus suis* type 2 strains in newborn germfree pigs depends on phenotype. Infect. Immun. **60**:550–556.
29. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14**:4683–4690.
30. **Welch, R. A.** 1991. Pore-forming cytolysins of Gram-negative bacteria. Mol. Microbiol. **5**:521–528.
31. **Wren, W. B.** 1991. A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeated sequences. Mol. Microbiol. **5**:797–803.