Porcine Lung Surfactant Protein D: Complementary DNA Cloning, Chromosomal Localization, and Tissue Distribution

Martin van Eijk, Henk P. Haagsman, Thomas Skinner, Alan Archibold, Kenneth B. M. Reid and Peter R. Lawson

*J Immunol* 2000; 164:1442-1450; doi: 10.4049/jimmunol.164.3.1442
http://www.jimmunol.org/content/164/3/1442

**References**

This article cites 58 articles, 27 of which you can access for free at:
http://www.jimmunol.org/content/164/3/1442.full#ref-list-1

**Subscription**

Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2000 by The American Association of Immunologists All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Porcine Lung Surfactant Protein D: Complementary DNA Cloning, Chromosomal Localization, and Tissue Distribution

Martin van Eijk,* Henk P. Haagsman,3*† Thomas Skinner,‡ Alan Archibold,‡ Kenneth B. M. Reid,§ and Peter R. Lawson§

Porcine organs and lung surfactant have medically important applications in both xenotransplantation and therapy. We have started to characterize porcine lung surfactant by cloning the cDNA of porcine surfactant protein D (SP-D). SP-D and SP-A are important mediators in innate immune defense for the lung and possibly other mucosal surfaces. Porcine SP-D will also be an important reagent for use in existing porcine animal models for human lung infections. The complete cDNA sequence of porcine SP-D, including the 5' and 3' untranslated regions, was determined from two overlapping bacteriophage clones and by PCR cloning. Three unique features were revealed from the porcine sequence in comparison to SP-D from other previously characterized species, making porcine SP-D an intriguing species addition to the SP-D/collectin family. The collagen region contains an extra cysteine residue, which may have important structural consequences. The other two differences, a potential glycosylation site and an insertion of three amino acids, lie in the loop regions of the carbohydrate recognition domain, close to the carbohydrate binding region and thus may have functional implications. These variations were ruled out as polymorphisms or mutations by confirming the sequence at the genomic level in four different pig breeds. Porcine SP-D was shown to localize primarily to the lung and with less abundance to the duodenum, jejunum, and ileum. The genes for SP-D and SP-A were also shown to colocalize to a region of porcine chromosome 14 that is syntenic with the human and murine collectin loci. The Journal of Immunology, 2000, 164: 1442–1450.

1 This work was supported by a travel grant from the board of trustees of Utrecht University (to M.v.E.).

2 Cloning, Chromosomal Localization, and Tissue Distribution

SP-D, including the 5' and 3' untranslated regions, was determined from two overlapping bacteriophage clones and by PCR cloning. Three unique features were revealed from the porcine sequence in comparison to SP-D from other previously characterized species, making porcine SP-D an intriguing species addition to the SP-D/collectin family. The collagen region contains an extra cysteine residue, which may have important structural consequences. The other two differences, a potential glycosylation site and an insertion of three amino acids, lie in the loop regions of the carbohydrate recognition domain, close to the carbohydrate binding region and thus may have functional implications. These variations were ruled out as polymorphisms or mutations by confirming the sequence at the genomic level in four different pig breeds. Porcine SP-D was shown to localize primarily to the lung and with less abundance to the duodenum, jejunum, and ileum. The genes for SP-D and SP-A were also shown to colocalize to a region of porcine chromosome 14 that is syntenic with the human and murine collectin loci. The Journal of Immunology, 2000, 164: 1442–1450.

S urfactant is an essential constituent of the lungs that coats the alveoli. This barrier material of the alveoli, lung surfactant, is composed of both lipid and protein and has a dual role: firstly, as a true surfactant, lowering the surface tension to prevent the alveolar collapse (1, 2); and secondly, surfactant participates in innate immune defense of the lung and possibly other mucosal surfaces. This function initially became apparent because of structural similarities of the surfactant proteins with other innate immune proteins (3–6) and from transgenic mice (7, 8). At least two of the surfactant proteins participate in lung defense by responding to environmental challenges and by targeting nonself structures, such as micro-organisms and pollen (9–11).

The porcine extracted surfactant is used as a successful therapy for many pulmonary disorders, including respiratory distress syndrome, meconium aspiration syndrome, pneumonia, pulmonary hemorrhage, and lung hypoplasia (12, 13). Mounting evidence suggests that the protein components of surfactant may play a beneficial role in surfactant replacement therapy. Firstly, the use of natural surfactants leads to an improved disease outcome, compared with that obtained with synthetic surfactants (14). Secondly, the enrichment of extracted porcine surfactant with SP-A4 has been shown to prevent inactivation of surfactant function by serum proteins (15). To further understand the surfactant protein components and to initiate research into the protective benefits of surfactant protein in surfactant preparations, the cDNA for porcine SP-D was cloned.

Four proteins are classically thought of as being associated with pulmonary surfactant. Two hydrophobic proteins, SP-B and SP-C, are involved in the surface tension-lowering activity of surfactant, and two hydrophilic proteins, SP-A and SP-D, are involved in nonclonal innate immunity and lipid homeostasis in the lung. SP-A and SP-D are collagenous Cα1-dependent lectins, termed collectins. The other members of the collectin family include mannose-binding lectin (MBL), a serum protein, the only collectin that activates the complement system (16), and two serum proteins found only in bovidae, conglutinin and CL-43 (17).

The basic structural unit of the collectins is a trimer of a polypeptide chain. The polypeptide chain is composed of an N-terminal, cysteine-rich region, a collagen-like region, an α-helical coiled neck region and a carbohydrate recognition domain (CRD). Three polypeptide chains are held together by the collagen-like region; triple helical formation is considered to be initiated by the neck region in the polypeptide chain (18). Each member of the

*Laboratory of Veterinary Biochemistry and Graduate School of Animal Health, Utrecht University, Utrecht, The Netherlands; †Department of the Science of Food of Animal Origin, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; 3The Roslin Institute, Roslin, Midlothian, Scotland; and 4Medical Research Council Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford, United Kingdom

Received for publication June 24, 1999. Accepted for publication November 17, 1999.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 This work was supported by a travel grant from the board of trustees of Utrecht University (to M.v.E.).

2 The nucleotide sequences reported in this paper have been submitted to the GenBank database (accession numbers AF132496 and AF133668).

3 Address correspondence and reprint requests to Prof. Henk P. Haagsman, Laboratory of Veterinary Biochemistry, Utrecht University, P.O. Box 80 176, 3508 TD Utrecht, The Netherlands. E-mail address: h.haagsman@vvdovet.uu.nl

4 Abbreviations used in this paper: SP-A/B/C/D, lung surfactant protein A, B, C, and D; CGN1, bovine conglutinin gene; CRD, carbohydrate recognition domain; HAS, human chromosome; Apg17, Agt11 library clone 17; Apg17, Agt11 library clone 17; MBL, mannose-binding lectin; pBSS, pBluescript; pPIG17d, pig SP-D clone 17, plasmid subclone d; pPIG20j, pig SP-D clone 20, plasmid subclone j; SFTPA, porcine or human SP-A gene; SFTPD, porcine or human SP-D gene; SSC, porcine chromosome; UTR, untranslated region; FISH, fluorescence in situ hybridization; RACE, rapid amplification of cDNA ends.
collectin family has a different degree of higher order oligomerization. CL-43 is found as a monomer of the basic trimeric structural unit, SP-A and MBL are found as hexamers, while SP-D and conglutinin form tetramers. SP-A and MBL resemble C1q in their tertiary structure by having a “kink” in the collagenous stalks, possibly induced by an interruption of the collagen-like, Gly-Xxx-Yyy sequence (4).

The CRD of each collectin has a distinct carbohydrate specificity that influences the targets that can be bound (6). By virtue of the CRD, associated receptors, and additional associated functional components, the collectins are all engaged to some degree in the innate immune system. MBL is the only collectin that has associated serine proteases, like those of C1q, through which it activates the complement system. SP-A and MBL as well as C1q may bind to a common receptor, C1qRβ (19), found on monocytes and macrophages, while SP-D has its own putative receptor on macrophages, pp-340 (20).

Investigations into the role of SP-D in surfactant have come full circle. Initially, work concentrated upon the function of SP-D in lipid homeostasis (21, 22). Later, in vitro evidence built upon a role of SP-D in the innate immune system. SP-D has been shown to bind directly to a vast range of pathogens, including viruses (23), bacteria (24), yeasts (25), and fungi (26), and is mediated via calcium-dependent interactions between the CRD and glycoconjugates expressed on the surface of micro-organisms. This binding can lead to phagocytosis of pathogens (27) and modulation of the immune cell and cytokine response (28), and provides protective resistance to infection (29). Both SP-A and SP-D might also play a role in modulating allergic responses (30). However, recently knockout of the SP-D gene in mice has caused some reappraisal of the function of SP-D, since, unlike the immune-compromised SP-A knockout mice (7), null SP-D mice exhibit a serious surfactant homeostatic imbalance (31, 32).

This paper describes the full-length cDNA cloning of porcine SP-D from a bacteriophage library and by PCR cloning. Sequence analysis reveals that porcine SP-D contains significant sequence differences compared with the other mammalian SP-Ds, which are reminiscent of other members of the collectins. The genes for SP-D and SP-A have also been mapped onto the porcine genome. Finally, the tissue distribution of SP-D has been extensively characterized to clarify the expression at nonpulmonary sites.

Materials and Methods

Construction of the cDNA probe

A recombinant cDNA probe of human SP-D lacking the majority of the collagen-encoding region (containing only the N-terminal region followed by the neck region and the CRD) was used to screen a porcine lung cDNA library. The probe (a gift from N. K. Lim) was generated by digesting the plasmid containing the full-length human SP-D cDNA with NruI and Smal to remove the majority of the collagen-encoding region, then self ligated. This plasmid could be cut with EcoRI and XhoI to release the 0.6-kb recombinant human SP-D probe that was gel purified (QIAquick gel extraction, Qiagen, Hilden, Germany). Radiolabeled DNA was made using a random oligonucleotide priming kit, following the manufacturer’s procedures (Multiprime DNA labeling system, Amersham, Aylesbury, U.K.) using [α-32P]dCTP (3000 Ci/mmol). The radiolabeled DNA was separated from unincorporated nucleotides using a Sephadex G-50 nick column (Pharmacia Biotech, Uppsala, Sweden).

Screening of the bacteriophage cDNA library

A commercially available porcine lung cDNA library, constructed in Agt11 (Clontech, Cambridge BioScience, Cambridge, U.K.), was screened with the α-32P-labeled human SP-D cDNA probe. Twenty-five positive plaques were identified from ~0.9 × 10^6 plaques, which hybridized to the probe under high stringent conditions (0.2× SSC (SSC = 150 mM NaCl and 15 mM Na,C,H,O2·2H2O, pH 7.0) and 0.1% (w/v) SDS at 68°C) and were detected after a 3-day exposure to x-ray film (X-OMAT AR, Kodak, Cambridge, U.K.). These were purified to homogeneity through three rounds of cDNA screening.

Characterization of positive plaques

Before subcloning into pBluescript (pBS), positive plaques identified by hybridization were characterized by PCR screening to eliminate subcloning false positive sequences. The inserts from positive plaques, identified by hybridization, were amplified by PCR using commercially available oligonucleotides to the right and left λ DNA arms (Clontech). A 50-μl standard PCR (25 mM KCl, 1.5 mM MgCl2, 1 U of Taq DNA polymerase, 200 μM of each dNTP, and 0.2 μM of each primer) was performed on a single plaque, picked from a purified bacteriophage stock plate. A standard PCR program run consisted of a 5-min denaturing step at 95°C, followed by usually 25 cycles of 94°C for 45 s, 55°C for 1 min, and a 2-min extension at 72°C. The PCR product was purified (Wizard DNA clean-up system, Promega, Southampton, U.K.) and sequenced by fluorescent dye terminator cycle sequencing using AmpliTaq FS. Each plasmid insert was sequenced on both strands, with independently primed sequence runs, confirming each strand at least three times.

Genomic DNA PCR

To confirm that the three significant differences found in the porcine SP-D cDNA sequence, when compared with those of human, rodent, and bovine SP-D, were not due to mutations or polymorphisms, the regions containing the differences were amplified from genomic DNA of four different breeds of pig by PCR. The genomic DNA was purified from pig sperm following the manufacturer’s recommendations (QIAamp Tissue Kit, Qiagen). Pig semen from the pig breeds, Large White Sireline, Finnish Landrace, Duroc Sireline, and Pietrain, were received as a grateful gift from Dr. H. Fetsma (Varkens KI, Noord-Brabant, CVZ-Group, The Netherlands) to amplify the region of interest, primers were designed from the porcine SP-D sequence and positioned according to the location of intron-exon boundaries in the human and murine SP-D genes. A standard 50-μl PCR was performed to amplify the CRD exon using the primers PD-CRD-F (5′ CAC GAC GCT GAC GAT CAC GAG 3′) and PD-CRD-R (5′ TCA GAA CTC GCA GAT CAC GAG 3′), while two collagen-encoding exons of the porcine SP-D gene, including the intervening intron, were amplified with primers GTG TTC CAG GAG CTG TAG GGC GAC GAG C and CTC TCC TTT GGC TGG TGT GCC. The PCR products were gel purified (QIAquick gel extraction) and sequenced directly on both strands using the amplification primers. 5′ RACE and primer extension

5′ RACE (34) was employed to clone the 5′ region of the porcine cDNA for SP-D. A single-stranded lung DNA template was synthesized using 50 U of monkey leukemia virus reverse transcriptase on porcine total lung RNA using an oligonucleotide specific for the cDNA of porcine SP-D (PD-IR, CCT GGA TCC CCT TCT TCG C), the ssDNAs were tailed with deoxycytidine using TdT (Roche, Lewes, U.K.). A 50-μl standard PCR was performed for 35 cycles on the tailed ssDNA template using an anchor primer (XSCGGSQ, GAG TCG AGT CGA CAT CGA TG3′), and the SP-D antisense primer, PD-IR. After PCR purification (Wizard DNA-cleanup system), a nested PCR was performed using 2.5 μl of the first PCR as a template with an antisense SP-D oligonucleotide (PD-5′PE, GCC TAC ACA TGA CCA GGC GCG AGT CGG TGG 3′) and the anchor oligonucleotide (XSC, GAG TCG AGT CGA CAT CGA TG), the 5′ RACE products were analyzed by agarose gel electrophoresis, gel purified (QIAquick, Qiagen), and then subcloned into pMosBlue using the manufacturer’s procedure (Amersham). A total of 18 independent clones were cycle sequenced using fluorescent dye terminator cycle sequencing with reverse pMosBlue primers.

To determine the 5′ extent of the porcine cDNA, a radiolabeled [α-32P]dCTP oligonucleotide (PD-5′PE) was added to 50 μg of porcine lung RNA in annealing buffer (150 mM KCl, 10 mM Tris-HCl (pH 8.3), and 1 mM EDTA) for 90 min at 65°C, then mixed with an equal volume of extension buffer (30 mM Tris-HCl, 15 mM MgCl2, 8 mM DTT, 1 × NEBuffer 4 (New England Biolabs, Beverly, MA), 2 U/μl TdT, 10 mM DTT, 0.1× DEPC water) and heated to 95°C for 5 min. Then 50 U of Moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI) was added to start the reverse transcription. After 15 min at 37°C and 5 min at 85°C, the reaction mixture was extracted with an equal volume of chloroform, and the RNA was precipitated from the supernatant by adding 2.5 vol of 100% ethanol. The RNA was redissolved in sterile water and used as the cDNA template for the following PCR reactions. A 50-μl PCR was performed using the primers PD-1 (5′ GGT GCG GGG GGA GGG GGG GGG GGG GGG 3′) and PD-2 (5′ GGCG CGG CGG CGG GG 3′), with the initial denaturing step at 95°C for 5 min, followed by 35 cycles of 94°C for 45 s, 55°C for 1 min, and a 2-min extension at 72°C. The PCR product was purified (Wizard DNA clean-up system, Promega, Southampton, U.K.) and sequenced by fluorescent dye terminator cycle sequencing using AmpliTaq FS. Each plasmid insert was sequenced on both strands, with independently primed sequence runs, confirming each strand at least three times.
0.2 mM of four dNTPs, and 20 U RNAsin (Promega) containing 50 U of monkey leukemia virus reverse transcriptase and incubated at 42°C for 60 min. A portion of the extension reaction (5 µl) was separated on a 6% (w/v) denaturing polyacrylamide sequencing gel alongside a radioactive sequencing reaction of pBS as a m.w. marker, then exposed to x-ray film.

Chromosomal localization of SP-A and SP-D

To localize the genes for SP-A and SP-D on the porcine genome, two PCRs were developed to amplify the porcine surfactant genes from a porcine/rodent somatic cell hybrid panel of 27 cell lines (35). The PCRs were optimized for temperature, magnesium concentration, and the number of cycles to specifically amplify the porcine gene only. For SP-A, a standard PCR was chosen with 33 cycles of 94°C for 30 s, 64°C for 30 s, and 72°C for 1 min. For SP-D, the optimized PCR used a buffer of 16 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 67 mM Tris-HCl (pH 8.8), 0.01% (v/v) Tween-20, and 200 µM of each dNTP. The reaction was performed as described above, but with an annealing temperature of 61°C. Both PCRs included 0.2 µM of the relevant surfactant gene-specific oligonucleotides (for SP-A, TCC TCA GTT TCC AGG AGT TCA TGC and AAG GCA CCT AC and TGG CCA GCA GAA GGT CAC) and 50 ng of porcine/rodent somatic cell line genomic DNA or 50 ng of hamster genomic DNA as a negative control template. Part of the PCR product (20 µl) was examined on a 1.2% (w/v) agarose gel, and each lane was scored for the presence or the absence of the expected PCR product (porcine SP-A, 418 bp; porcine SP-D, 257 bp). Statistical analysis of these data was performed using a computer program on the internet (36).

Northern blotting

Total RNA was isolated from different tissues derived from 4- to 6-wk-old male or female pigs using RNAzol-B (Tel-Test, Friendswood, TX) according to the manufacturer’s instructions. Approximately 40 µg of total RNA from each tissue was separated on a 1.2% (w/v) agarose formaldehyde gel, and each lane was scored for the presence or the absence of the expected PCR product (porcine SP-A, 418 bp; porcine SP-D, 257 bp). Part of the PCR product (20 µl) was transferred unto nylon membrane (Nytran-N, Schleicher & Schuell, NY) and fixed by baking at 80°C for 2 h. The probe identified 25 positive plaques from a representative sample of lung cDNAs (0.9 × 10⁶ plaques). To circumvent the

**FIGURE 1.** The arrangement of the clones used to obtain the full-length porcine SP-D cDNA sequence. A, A graphical representation of porcine SP-D, detailing the untranslated regions (sticks) and coding regions (boxes). Shading indicates different regions of the protein: signal peptide (dense horizontal bars), N-terminal, cysteine-rich region (dot shading), collagen-like region (filled bars), neck domain (light horizontal bars), and carbohydrate recognition domain (dense diagonal bars). B, The cDNA clones obtained from subcloning the two λ clones and the 5' RACE clone (filled bars). The dotted bar indicates the region of pPIG17d that is homologous with protein kinase C. C, The genomic PCR products (hatched boxes) used in direct sequence confirmation of the three major variants in porcine SP-D. The genomic PCR covering part of the collagen-encoding region contains an intron; this is indicated by an inserted dotted line.
problem of subcloning cDNAs that were not homologous to human SP-D, the inserts of the positive λ phages were PCR sequenced. This revealed that only two of the clones were potential porcine homologues of human SP-D (Apig17 and Apig20). These clones overlapped and were judged to cover the entire protein-coding region for SP-D. The λ DNA insert was subcloned by EcoRI digestion into pBS, generating the plasmid clones pPIG17d and pPIG20j, and subsequently sequenced.

Analysis of the clones

The inserts of the two plasmid clones, pPIG17d and pPIG20j, were sequenced by primer walking and were found to be 1123 and 606 bp long, respectively. Alignment of the two sequences revealed that they overlapped by 97 bp (Fig. 1).

Sequence analysis of pPIG20j showed that it was homologous over its entire length to the 3′ regions of the cDNA encoding SP-D from other species; it encoded the CRD region and the entire 3′ UTR, as determined by the presence of a polyadenosine tract.

Sequence analysis of pPIG17d showed that the 5′ region of the clone was homologous to the cDNA sequence of human protein kinase C (see Fig. 1, dotted region of the bar), while the 3′ region showed homology to the 5′ region of human SP-D. This chimera is probably a ligation artifact formed from constructing the cDNA library. To confirm that pPIG17d was indeed a chimeric clone and to identify the 5′ sequence of porcine SP-D, 5′ RACE was used to clone the 5′ region from lung RNA, and primer extension was performed to confirm that the RACE clone extended to the 5′ end of the porcine cDNA sequence.

Cloning the 5′ extent of porcine SP-D

The 5′ RACE product of the porcine SP-D was subcloned, and 10 independent clones were sequenced with vector and internal primers. The sequences were aligned and edited, and a consensus was obtained. The longest 5′ extension product sequenced was used to designate the start of transcription (+1). The 5′ RACE product was 168 bp long with 97 bp of sequence, in agreement with the sequence obtained from clone pPIG17d (excluding 31 bp of sequence from the primer, PD-5′). At the point of chimera in the porcine cDNA library clone (pPIG17d), the 5′ RACE product continued and provided an additional 40 bp of novel sequence. A ragged 5′ end was observed, but the majority of clones (7 of 10) extended to position +1, and the remaining three clones extended to position +2. Direct sequencing of the uncloned 5′ RACE product indeed indicated that the majority of the cDNAs extended to the designated start of transcription, as no appreciable drop in sequence signal strength was observed toward the transcription start site. This is also consistent with a strongest signal strength observed in primer extension (results not shown) and the sequence of the 5′ regions of murine, rat, and human SP-D.

Analysis of the sequence

The sequence for porcine SP-D obtained from the overlapping RACE and cDNA clones (Fig. 1B) has the accession number AF132496 and is shown in Fig. 2. The longest open reading frame, running from the first ATG (Fig. 2, bold) predicts a protein sequence of 378 aa, which shows the greatest similarity to bovine SP-D with 78% identity and is shown aligned with the sequences of SP-D from four other species (Fig. 3A). The cDNA has a 43-bp-long 5′ UTR sequence that shows 71% similarity with those of both human and murine SP-D. Curiously, the long 3′ UTRs of porcine and bovine SP-D show a greater similarity with the 3′ UTR of bovine conglutinin, than with the shorter 3′ UTRs of human and rodent SP-D. The 3′ UTR of porcine SP-D contains two potential polyadenylation sites (AATAAAA), 25 and 117 bp upstream of the polyadenosine tract (Fig. 2), a similarity also shared with both bovine SP-D and conglutinin and quite distinct from the single polyadenylation sites seen in the 3′ UTR of rodent and human SP-D.

FIGURE 2. The cDNA sequence of porcine SP-D. The numbering refers to the nucleotide sequence with +1 being the start of transcription, based on the longest RACE product identified and primer extension. Above the nucleotide sequence is the predicted protein sequence, starting at the first methionine, ATG (codon (bold). The protein sequence in italics indicates the predicted leader peptide. The underlined DNA sequence indicates that identified by 5′ RACE. The remaining sequence was obtained from the two λ clones isolated from a porcine lung cDNA library. The two potential polyadenylation signals are double underlined and are 25 and 117 bp upstream from the start of the polyadenosine tract. The arrow indicates an exon-intron-exon junction identified in the gene by PCR sequencing.
FIGURE 3. Alignment of the five mammalian protein sequences of SP-D, including helical wheel diagrams of the neck regions. A. Porcine SP-D (po) aligned with four other mammalian protein sequences for SP-D: human (hu), bovine (bo), mouse (mo), and rat (ra). The three major differences in porcine SP-D are highlighted (bold): an extra cysteine residue, the presence of a potential glycosylation site in the CRD, and a 3-aa insertion in the CRD. Dots indicate residues in common with the porcine sequence. Dashes indicate spaces that are inserted to maximize the identity across the alignment, and underbars show the potential N-linked glycosylation sites. The key residues necessary for coordination of Ca$^{2+}$-ions and hydroxyl groups of oligosaccharides, predicted from MBL (46), are shown with asterisks. A consensus sequence of identity is constructed for the five mammalian SP-D sequences and is shown below the alignment. B. Helical wheel diagrams show the heptad repeating sequence of the neck region in SP-D. Hydrophobic residues in all species to date can be seen at positions 1 and 4 in the heptad repeat, which are implicated in the trimerization of three neck regions. Residues 224–250 are shown for the porcine SP-D neck region.

FIGURE 4. Chromosomal localization of the porcine SP-A and SP-D genes. Solid bars on the left show the regions of porcine chromosome 14 (SSC14) represented in the somatic hybrid cell panel (numbered above). Only those hybrid lines that contain part of SSC14 are shown; asterisks indicate SSC14 regions in hybrid cell line that do not correspond with G band limits (adapted from Yerle et al. (35)). The + or − sign above each bar denotes whether that hybrid tested positive or negative for both SFTPA and SFTPD. All other hybrids tested were negative for both SFTPA and SFTPD. Solid bars on the right of the ideogram represent regions of SSC14 that are homologous to human chromosomes (adapted from Rettenberger et al. (49)).
a heptad repeat of two hydrophobic residues at positions 1 and 4, as shown in the helical wheel diagrams (Fig. 3B).

Alignment of porcine SP-D with the sequences from other species revealed three major differences that are not shared by any of the known SP-D cDNAs (Fig. 3A). First, porcine SP-D has an extra cysteine residue in the collagen region at position 105, through an AGT codon change (serine) in humans and mice to a TGC codon (cysteine) in the pig. The second distinguishing feature is that porcine SP-D has a potential N-linked glycosylation site in the CRD at position 323, resulting from the AAG codon (lysine) in humans and rodents changed to an AAT codon (asparagine) in the pig. Finally, porcine SP-D shows an insertion of 3 aa in the CRD region, a difference to date unreported for any species within any group of the collectins; the size of the CRD region was thought to be constant across different species within each group of the collectins.

**Confirmation of the major sequence variants in porcine SP-D**

The three variants in the porcine SP-D sequence may have important structural and evolutionary consequences. To confirm that these variations are not artifacts, mutations, or polymorphisms, the appropriate areas were sequenced at the genomic level from four different pig breeds (see Fig. 1C). The CRD, which is encoded by a single exon in murine and humans, was amplified from porcine genomic DNA to confirm the novel glycosylation point and the sequence insertion. The PCR products amplified from this genomic DNA region and the plasmid clone gave the same expected size PCR product of 381 bp. For the collagen region variation, the extra cysteine is in the first collagen-encoding exon close to the intron-exon boundary predicted from the human and mouse genes (37, 38). The extra cysteine was examined by amplifying the first two collagen-encoding exons, including the predicted short intervening intron. The PCR yielded a genomic product of ~600 bp compared with the 219-bp product predicted from the cDNA sequence (shown in Fig. 1C).

The PCR products were sequenced on both strands. The genomic sequences from the different pig breeds agreed with that obtained from the cDNA sequence (results not shown). This indicates that these variants are neither polymorphisms nor mutations, but that porcine SP-D contains three major sequence differences in comparison to SP-D from other species.

**Correction of the porcine SP-A CRD sequence**

The only other reported porcine surfactant or collectin DNA sequence is that for porcine SP-A (accession no. L41350). Alignment of this sequence against SP-A from other species, at both the protein and DNA levels, revealed that the reported SP-A sequence appears to contain a frame-shift sequencing error in the CRD region. Two insertions of TG and G appear at positions 498 and 535 in the published cDNA sequence. This was confirmed by PCR sequencing on both strands of the CRD exon from different pig breeds. A consensus was constructed by alignment and submitted to the GenBank database (accession no. AF133668). This will prove important for the recombinant expression of this innate immune molecule and its use in current porcine animal models for human respiratory infections (39, 40).
Chromosomal localization of porcine SP-A and SP-D

For the chromosomal localization of SP-A and SP-D, sequence tag sites were developed that would specifically amplify the porcine genes from a porcine somatic cell hybrid panel constructed on a hamster cell line background (35). For SP-A and SP-D, a 418-bp and a 257-bp product, respectively, were specifically amplified from porcine genomic DNA. The porcine SP-D primers also amplified an additional fragment of about 600 bp from most cell lines. However, this product was ignored because it was present in the negative control hamster genomic DNA. Both primer pairs gave the same pattern of positive signals across the different cell lines; hybrids 12, 20, 22, and 26 all tested positive (Fig. 4). The pattern of these four positives assigns both genes to SSC14 with a correlation coefficient of 0.87 and a chromosomal probability of 98%.

Tissue distribution of porcine SP-D

The expression of SP-D in different porcine tissues was studied by Northern blot hybridization. To eliminate cross-hybridization with other collagen-like containing cDNAs, a radiolabeled probe was used that contained only the CRD sequence of pig SP-D. Analysis of lung and a wide range of nonpulmonary tissues clearly showed a 1.3-kb signal corresponding to SP-D mRNA in the lung. A weaker signal, at the same m.w., was also present in duodenum, jejunum, ileum, and the mucosa of ileum (Fig. 5). No detectable levels of SP-D mRNA were observed in stomach, pylorus, or colon.

Discussion

The full-length cDNA sequence of porcine SP-D has been cloned from two overlapping λ DNA clones containing the entire coding region for the mature polypeptide, and 5′ RACE clones covered the remaining 5′ region (Fig. 1B). All the clones overlapped by at least 97 bp and showed no differences in these areas. The cDNA for porcine SP-D has a 43-bp 5′ UTR, 1134 bp of coding sequence, and a 172-bp 3′ UTR, including two potential polyadenylation sites, 25 and 117 bp upstream of the polyadenosine tail (see Fig. 2).

The coding region contains a predicted hydrophobic leader sequence of 20 aa, followed by a short cysteine rich N-terminal sequence that is believed to be involved in higher order oligomerization of the trimeric SP-D polypeptide chains (41). The 177-aa long collagen-like region is of the same length that of seen in humans and rodents, in contrast to bovine SP-D, for which the collagen-like region is 6 aa shorter. The porcine SP-D gene probably has a similar genomic organization across the collagen-like encoding region as seen in the human and murine genes (37, 38), which are composed of four exons each of 117 bp. Indeed, the position of one of the predicted introns has been confirmed by genomic PCR sequencing (Fig. 2). Despite the overall degree of identity seen between porcine and bovine SP-D (78%) compared with the other mammalian SP-Ds, analysis and alignment of the collagen region would suggest that the bovine SP-D gene has an intron-exon organization that resembles conglutinin rather than the human and murine SP-D genes.

The 28-aa neck region of porcine SP-D conforms well to the heptad repeating motif, with hydrophobic residues at positions 1 and 4 (18) (see Fig. 3B), which is instrumental in trimerization of the collagen region (42).

The CRD of porcine SP-D contains the conserved residues identified in the structure of MBL that are required for the coordination of Ca\(^{2+}\) ions and hydroxyl groups of carbohydrate. The CRD, as judged by PCR sequencing, is encoded by a single exon, like other members of the collectins, despite having a 9-bp insert.

The overall structure and features of the sequence for porcine SP-D are consistent with the sequence of SP-D from different species. However, there are a number of important differences in porcine sequence that set it apart from the other mammalian SP-D cDNA sequences previously characterized (see Fig. 6). Each of the three differences may have important consequences for the structure and functional properties of SP-D, but it remains to be shown whether porcine SP-D has any functional properties different from those found in other species.

These differences include an extra cysteine in the collagen region, which is absent in all other known SP-Ds but is reminiscent of the extra cysteine found in collagen-like regions of bovine conglutinin (43) and human SP-A. The extra cysteine in conglutulin appears to contribute to an elongated central core at the base of the collagen stalks when compared with the other cruciform collectins, SP-D, as judged by electron microscopy (44). For human SP-A, two different genes contribute to its molecular structure; only one of the polypeptide chains has an extra cysteine. It has been proposed that the extra cysteine forms interchain disulfide bonds to either stabilize or form the heteropolymeric structure of SP-A (45). Only one polypeptide chain exists for porcine SP-D; it will be interesting to determine what effect this extra cysteine in the collagen region has on the structure of porcine SP-D.

Chromosomal assignment of two of the porcine surfactant proteins, SP-A and SP-D, was performed by somatic cell hybrid mapping (35). The somatic cell hybrid panel was originally characterized using G banding combined with fluorescence in situ hybridization (FISH) and marker assignments by PCR (35) to define the porcine composition of each hybrid line (Fig. 4). Assignment of markers by PCR occasionally reveals the presence of...
small fragments of porcine chromosomes that had escaped the initial FISH and G-banding analysis. Since one of our results (hybrid 12) suggests this, the chromosomal localization for porcine SFTP A and SFTPD should be limited to an assignment to SSC14 rather than to a specific subchromosomal region. The homologies between the porcine and human genomes have been extensively studied by zoo-FISH (47–49) (see Fig. 4). As the collectin cluster in humans is found on HSA10q22.2–23.1 (50), and HSA10q exhibits conservation of synteny with SSC14q2.3-qter (48), our assignment of porcine SFTP A and SFTPD to SSC14 conforms to our predictions. The identical pattern of positives seen in both panel screens is consistent with the conservation of the collectin cluster in pigs. However, further studies using radiation hybrid mapping or restriction mapping would be required to determine the distance between these two genes in the porcine genome.

Northern blot analysis, using a porcine SP-D CRD probe, showed not only the presence of a 1.3-kb signal in lung, but also a weaker signal in duodenum, jejunum, ileum, and ileal mucosa (Fig. 5). Because no SP-D mRNA could be detected in stomach, pylorus, or colon, it appears that nonpulmonary expression of SP-D in pig is restricted to the small intestinal region of the digestive tract, although the degree of expression is considerably lower than that seen in the lung. Data from several studies on different species have shown that apart from lung, SP-D has been reported to be present in stomach (51, 52), trachea (53), heart, pancreas, small intestine, colon (6), mesentery (54), heart, and kidney (51). SP-A and SP-B have been shown to be associated with secreted surfactant-like particles in the intestinal lumen (55, 56). The detection of SP-D in intestinal tissues and the ileal mucosa supports the potential function of SP-D as a host defense mechanism of mucosal surfaces other than lung. The expression of SP-A, which was reported to be present in the small intestine and colon of the rat (57), furthers the interest in studying the expression of these mediators of local innate immunity at nonpulmonary sites.

This work contributes to the understanding and development of porcine models for human respiratory infectious diseases (39, 40) and the use of porcine lung surfactant as a therapeutic agent in human surfactant disorders. Given the fact that respiratory infections are the most frequent health problems in the pig-breeding industry (58), knowledge of collectin-mediated lung defense may lead to the development of potential markers for disease and the innate immune system for pig breeders. Furthermore, this work contributes to the knowledge of the porcine innate immune self/ nonself recognition system of the lung that will be useful for xenotransplantation research, an area where the importance of innate immunity in contributing toward delayed xenograft rejection is just coming to light (59).

Acknowledgments

We thank Dr. N. Stockhofe-Zurwieden (Institute for Animal Science and Health, ID-Lelystad, Lelystad, The Netherlands) for the preparation of the various porcine tissues.


