The structure of porcine protegrin genes**

Chengquan Zhao, Tomas Ganz, Robert I. Lehrer*

Department of Medicine, UCLA School of Medicine, UCLA Center for the Health Sciences, CHS 37-055, 10833 LeConte Avenue, Los Angeles, CA 90095-1678, USA

Received 22 May 1995

Abstract We cloned the genes of three protegrins, a family of cathelin-associated antimicrobial peptides originally isolated from porcine leukocytes. Each gene comprised 4 exons and 3 introns, wherein Exon I encoded the signal sequence and the first 37 amino acids of cathelin, Exons II and III contained 36 and 24 additional cathelin residues and Exon IV contained the final two cathelin residues followed by the protegrin sequence. This quadripartite gene structure helps explain how structurally diverse antimicrobial peptides can be expressed on common, cathelin-containing precursors. Southern blot probed with an oligonucleotide specific for protegrin genes suggested that several identical or nearly identical protegrin genes were densely clustered in the pig chromosome.

Key words: Protegrin; Gene structure; Antibiotic peptide; PG-5

1. Introduction

Leukocytes are key effector cells in host defense against microbial infections. Their ability to kill bacteria and fungi is mediated, in large part, by antimicrobial peptides and proteins found within their cytoplasmic granules. Although these antibiotic molecules vary in size, structure, and activity, most are amphiphiles with cationic and hydrophobic surface domains. Many of the bovine and porcine antimicrobial peptides are synthesized as larger precursors with cathelin-like [2] pro-sequences that are removed during processing. The family of cathelin-related peptides and proteins includes: bovine indolicidin [3], cyclic dodecapeptide [4], 'Bac5' [5,6], 'Bac7' [7, 8]; rabbit 'P15' [9,10] and 'CAP18' [11,12]; and porcine 'PR39' [13,14], protegrins [15,16], 'C-12' [17,18], 'PMAP-23' [19],'PMAP-36' [20] and 'PMAP-37'[25]. To gain insight into the evolution of cathelin-related antimicrobial peptides, we analyzed the structure of three porcine protegrin genes.

2. Materials and methods

2.1. Genomic DNA purification, and PCR amplification

High molecular genomic DNA was purified from pig white blood cells with the QIAGEN blood DNA kit (QIAGEN, Chatsworth, CA). To amplify protegrin (PG) genes, PCR was performed using genomic DNA as a template.

The sense primer (5'-GTCGGAATTCATGGAGACCCAGAG(A or G)GCCAG-3') corresponded to the 5' regions of PG cDNAs, and provided an *Eco*RI restriction site. The antisense primer (5'-GTCGTCTAGA(C or G)GTTTCACAAGAATTTATTT-3') was complementary to 3' ends of PG cDNAs immediately preceding their poly(A) tails and provided an *Xba*I restriction site. The reaction was carried out in a total volume of 50 μl, which contained 200 ng of purified pig genomic DNA, 25 pmol of each primer, 1 μl of 10 mM dNTP, 5 μl of 10 × PCR buffer (200 mM Tris-HCl, 100 mM $(NH_4)2$, 20 mM $MgSO_4$, 1% Triton X-100, 0.1% BSA), and 2.5 units of cloned Pfu DNA polymerase (Stratagene, La Jolla, CA). Thirty cycles were performed, each with 1 min of denaturation at 94 °C, 1 min of primer annealing at 55°C, 2 min of primer extention at 72°C, and a final extension step at 72°C for 10 min.

2.2. Gene cloning and sequencing

The amplified PCR product was digested with *Eco*RI and *Xba*I, excised from the agarose gel, purified, and ligated into pBluescript KS+ vector (Stratagene, La Jolla, CA) that had been digested with *Eco*RI and *Xba*I and purified. Both strands of DNA were sequenced by the dideoxy method using the Sequenase version 2.0 kit (United States Biochemical, Cleveland, OH), pBluescript universal primers and specific oligomer primers based on PG genomic and cDNA sequences. Computer analysis of the DNA sequences was performed using the PC-Gene program (Intelligenetics, Palo Alto, CA). Segments containing compression artefacts were resequenced with Bst DNA sequencing kit (BIO RAD Laboratories, Melville, NY)

2.3. Screening the porcine genomic library

A porcine liver genomic library in EMBL3 SP6/T7 phages was purchased from Clontech (Palo Alto, CA). *E. coli* strain K803 was used as a host, and DNA from phage plaques was transferred onto nylon membranes (Dupont, Boston, MA). The filters were hybridized with ^{32}P-labeled porcine 691 PG-3 cDNA [16]. The filters were washed several times, finally at 60°C in 0.1 × SSC and 0.1% SDS, and exposed to X-ray film with an intensifying screen at −70° C. Positive clones were subjected to two additional rounds of plaque purification at low density.

2.4. Southern blotting

DNA purified from hybridizing clones was digested with various restriction endonucleases (New England Biolabs, Beverly, MA), fractionated on 0.8% agarose gels, and transferred onto GeneScreen Plus membrane (DuPont, Boston, MA). The hybridization probes were labeled with ^{32}P and included porcine PG-3 cDNA, and 5'-labeled protegrin-specific oligonucleotide complementary to nt 403–429 of PG-1, -2 and -3 cDNAs. For the cDNA probe, the hybridization and washing condition were carried out as for the library screening. For the oligonucleotide probe, the membranes were washed at 42°C in 0.1 × SSC, 0.1% SDS.

2.5. 5' flanking region of the protegrin gene

The PG-positive EMBL3 clone DNA cut with restriction enzyme *Pst*I was probed with an oligonucleotide complementary to nt 49–72 of PG cDNA [16]. A hybridizing 1.2 kb fragment was identified, and inserted into pBluescript KS vector. DNA sequence was determined as described above.

3. Results and discussion

With porcine genomic DNA as template, an upstream primer that corresponded to the 5' end of protegrin cDNAs, and a downstream primer matching the region immediately preceding

*Corresponding author. Fax: (1) (310) 206 8766.
E-mail: rlehrer@medicine.medsch.ucla.edu

**Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries with the following accession numbers: X-84094 (PG-1), X-84095 (PG-3) and X-84096 (PG-5).

SSDI 0014-5793(95)00633-8

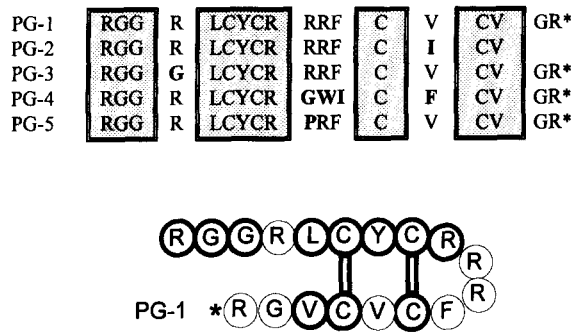

Fig. 1. Primary structures of protegrins. The sequences of PG-1, -2, and -3 were determined from the mature peptides [15] and that of PG-4 was inferred from its cDNA [16]. The sequence of PG-5 was derived from the gene sequence reported here. The C-terminal glycine of PG-5 was assumed to have been converted to an amide in the mature peptide, as reported for PG-1, -2, and -3 [23]. The amino acids of PG 2–5 that differ from those of PG-1 are indicated in boldface. PG-1 is diagrammed below the five sequences, with its disulfide pairing indicated [24] and the residues invariant to all five protegrins heavily outlined.

the poly (A) tail, we generated a PCR product of about 1.85 kb. We confirmed that the band was protegrin-related by hybridization of the PCR product with a protegrin-specific oligonucleotide probe complementary to nt 403–429 of the protegrin cDNA sequences (data not shown). The PCR product was then subcloned into pBluescript vector, and recombinant plasmids were subjected to DNA purification and sequencing. Gene sequences for three different protegrins were identified, two of which encoded the previously described protegrin peptides PG-1 and PG-3 [15,16], and another that coded for a novel protegrin that we named PG-5. The amino acid sequence of PG-5 is shown in Fig. 1, which compares it to the other known protegrin peptide sequences.

Comparison of protegrin cDNAs and genes revealed that the coding regions of protegrin genes consisted of four exons, interrupted by three introns (Figs. 2 and 3). A 1.2 kb *Pst*I fragment covered the 5′ flanking region of 690 bp, exon I, and most of intron I. A TATA-box like consensus sequence and several potential regulatory motifs were identified by examination of the 5′ flanking sequence (Fig. 3 and Table 2). The first exon contained a short 5′ non-coding region and codons for the first 66 amino acids of the protegrin prepro-peptide, including a 29 residue signal peptide and the first 37 cathelin residues. Exons II and III were relatively small, only 108 and 72 bp, respectively, and together contained the next 60 cathelin residues. The final two cathelin residues were on Exon IV, and were followed by the protegrin sequences. The exon-intron splice site sequences are shown in Table 1, and conform to the consensus

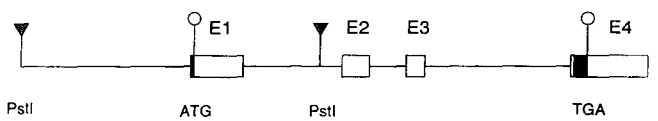

Fig. 2. Organization of protegrin genes. The positions of the start codon ATG and the stop codon TGA are indicated by circles. The *Pst*I fragment used for sequencing the 5′ flanking region is denoted by triangles. Open boxes denote exons (E), the crosshatched box in E1 denotes the 5′ untranslated region and the black box in E4 represents the sequence encoding the mature protegrin peptide. The map is drawn to scale.

Table 1
Exon–intron structure of the PG1 gene

Exon	Size (bp)	5′ splice donor	Intron	Size (bp)	3′ splice acceptor
1	211	AAGGCCgtgagtcg	1	404	ttgaccagGACGAG
2	108	AACGGGgtgaggct	2	152	ccttccagCGGGTG
3	72	AATGAGgtgagtgg	3	595	ggtcacagGTTCAA
4	313				

rule [21]: all introns end on an AG doublet, preceded by a T/C rich stretch of 8–12 bases, while all introns start with GT, followed predominantly by A/G A/G G sequence.

In summary, the highly conserved cathelin region spans exons I–IV, and Exon IV contains the full sequence of the mature protegrin peptide followed by an amidation consensus sequence, a 3′ untranslated region, and the putative polyadenylation site. The three introns range in size from 152 to 595 bp. If the protegrin genes are representative of other cathelin-like genes, the third intron of cathelin-associated peptides will be found to separate all but the last two residues of the highly conserved cathelin region from the variable antimicrobial peptides encoded in Exon IV. Such a layout would favor recombination mechanisms involving association of diverse Exon IV's with the first three exons specifying cathelin containing prepro-regions.

The protein sequences deduced from the genomic DNA are consistent with the amino acids from previously characterized porcine PG-1 and PG-3 cDNAs [16]. PG-1, PG-3, and PG-5 gene sequences are nearly identical. PG-1 and PG-3 genes differ in only five base pairs within the coding sequence. One substitution was located in the second intron, two in the third intron, and the remaining two in exon IV. PG-1 and PG-5 genes differed by seven nucleotide substitutions, two located in third intron, one each in exon I and exon II, and the remaining three in exon IV. One of the two differences in the mature peptide region involved substitution of a Pro in PG-5 for an Arg in PG-1 at residue 10 (Fig. 1). The high degree of homology within the protegrin family suggests these genes have arisen from very recent duplications.

Homology search of protegrin genes against the EMBL/GenBank identified no significantly homologous genes. More specifically, the gene structures and nucleotide sequences of protegrins were very different from those of defensins, which contain

Table 2
Selected consensus sites in the 5′ flanking region

Consensus site	Reference	Strand	Location
NF-IL6	26	−	−632
		+	−126
		+	−108
NF-kB	27	−	−56
Gamma interferon response element	28	+	−419
		+	−345
		−	−336
		−	−207
		−	−92, −91
		+	−58
		−	−34

The consensus sites were identified by the Signal Scan program using David Ghosh's Transcription Factor Database (Prestridge, D.S. (1991) SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. CABIOS 7, 203–206).

```
ctgcagtggcctggagtgtgagacggatagaaagagttggcgtcacagcatagccgacct  -631

ttcctcatacttggttgggaaggggccggggaacaggcatgtcccggggaacgtggggat  -571

ggggcatttttagactgagagagcagtcaagaagagtgcaggagaggagaccagggagct  -511

gaccctatatcataggtagtgtggtgttgactatcaactaggaactggcgggtctgtaag  -451

actctggtttgtgttggtactatcatccctccattacactttaatttagggcttcctagg  -391

tgccaggcagcacctagagctggggacagagtgctgactacactcctggaccctggttcc  -331

tgaggcggccaacctggggacccagagcaggtgtgagacagtgcctgcctctccttctcc  -271

tgaaggacccaaccctcttgagcaaccctgcctgtaacactgcctgtggggcccaggaga  -211

ggccaggacactgtcagggcgatggagaaaccacttcttcaccttgcacaggcacccatc  -151

cttccccactgcccagcaatcccctgaggcaagagccagggctgaggcaagtccagccag  -91

catcctgggaggcagccagggatggggccgctcaggaagtcccagacccaggctgggcat  -31
                                -1*
aaaggagggtcccgcaggctgggaggaggcTCACCTGGGCACCATGGAGACCCAGAGGGC   30
                                                M  E  T  Q  R  A
                                                G5
CAGCCTGTGCCTGGGGCGCTGGTCACTGTGGCTTCTGCTGCTGGCACTCGTGGTGCCCTC   90
  S  L  C  L  G  R  W  S  L  W  L  L  L  L  A  L  V  V  P  S
                                                G5
GGCCAGCGCCCAGGCCCTCAGCTACAGGGAGGCCGTGCTTCGTGCTGTGGATCGCCTCAA  150
  A  S  A  Q  A  L  S  Y  R  E  A  V  L  R  A  V  D  R  L  N
CGAGCAGTCCTCGGAAGCTAATCTCTACCGCCTCCTGGAGCTGGACCAGCCGCCCAAGGC  210
  E  Q  S  S  E  A  N  L  Y  R  L  L  E  L  D  Q  P  P  K  A
Cgtgagtcgggcaggggctcaggaggggctgggggggcgggggctgtccccacccgcccc  270

ggggctccctgtccctcccccctgctcaggctgtcctcctgccaggaaggcacttgtccc  330

tctaaggggaccccctctgccaggaaaccttcccagagctgggtgccctgcccgcgtga  390

gagcttcccgccttagcctctgggctgtgggctcagggccctgcacagcctgtgaggcag  450

gagcgggctctgtcccctcccctgtgcacccagcaccaagcccagggccaggctcccagc  510

aggggctgcagaggctgctgtctaggtgggggcggggaggggggtgacagatccgagggggg  570

aagcctgagcccgagtcccatctccccactttgatccttgaccagGACGAGGACCCGGGC  630
                                                D  E  D  P  G
    A5
ACCCCGAAACCTGTGAGCTTCACGGTGAAGGAGACTGTGTGTCCCAGGCCGACCCGGCAG  690
  T  P  K  P  V  S  F  T  V  K  E  T  V  C  P  R  P  T  R  Q
CCCCCGGAGCTGTGTGACTTCAAGGAGAACGGGGtgaggctgggggctgggggcgctggc  750
  P  P  E  L  C  D  F  K  E  N  G
ggatgcttcccaaggagctgaacaggagagcctgctggggaagatgtccaggccctgggg  810

tgaggctgggagctcatggatggaggaggggggtcccagtttgaccttgagtctccccct  870
  t3
tccagCGGGTGAAACAGTGTGTGGGGACAGTCACCCTGGATCAGATCAAGGACCCGCTCG  930
      R  V  K  Q  C  V  G  T  V  T  L  D  Q  I  K  D  P  L
ACATCACCTGCAATGAGgtgagtggccccttattggtgtcaagttgctaatgggttggtg  990
D  I  T  C  N  E
```

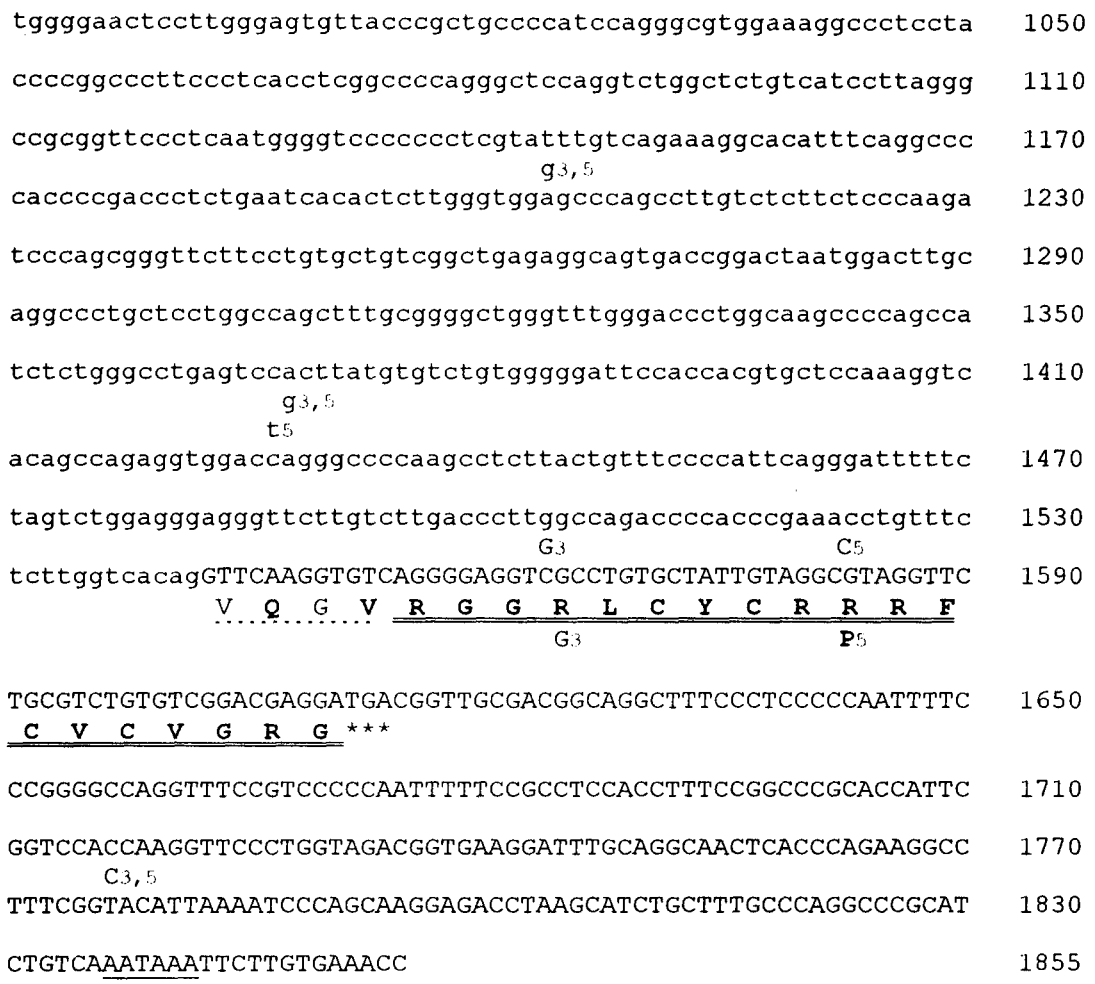

Fig. 3. Nucleotide sequences of protegrin genes. The PG-1 sequence is indicated in the main rows. The other protegrin sequences, shown only where they differ from PG-1, are identified by a number (^{3} = PG-3, ^{5} = PG-5). Nucleotide sequences of exons are indicated by capital letters, and 5′ flanking region and intronic sequences are shown in lowercase letters. Nucleotides are numbered relative to the presumed transcription start site identified by the primer extension [14,15,19,20,25]. The transcription start site is designated by an asterisk and the corresponding TATA-box identified by the Intelligenetics PC Gene program is double underlined. The cathelin domain is underlined with a dotted line, with residues identical to those described by Ritonja, et al. [2] shown in bold type. The NG and RV residues shown at the 3′ ends of Exon II and the 5′ ends of Exon III were previously shown to exist in cDNA for protegrins [16,22] and for PR-39 [22]. The first nucleotide of the start codon (ATG) is numbered as base 1. Predicted mature peptide sequences are shown in boldface and are doubly underlined. The stop codon is indicated by the asterisks. The start codon is followed by a putative signal peptide which is single underlined, as is the consensus polyadenylation signal.

three exons in myeloid defensin genes, and two exons in enteric defensin genes [1]. As expected, the search yielded the large family of cDNAs corresponding to cathelin-associated bovine, porcine and rabbit leukocyte peptides mentioned in section 1.

To date, three protegrin mature peptides and four protegrin cDNAs have been identified [15–16,22]. To assess protegrin-related genes further, we screened a porcine genomic library of approximately 2.3×10^5 clones in EMBL-3 SP6/T7 with the ^{32}P-labeled protegrin cDNA, and identified 45 hybridizing clones. Southern blot analysis was carried out with purified DNA from positive clones by hybridization with protegrin cDNA and a protegrin specific oligonucleotide complementary to nt 403–429 of protegrin cDNA sequences. Although all of the clones hybridized with the complete cDNA probe, only about a half of them hybridized with the protegrin-specific probe (Fig. 4). A specific oligonucleotide probe for porcine prophenin, another cathelin-associated porcine leukocyte-derived antimicrobial peptide, hybridized to several of the non-protegrin clones (data not shown). These results confirm (i) that the conserved proregion homologous to cathelin is present within the same gene as the mature antimicrobial peptides and is not added on by posttranscriptional events, and (ii) that the protegrins account for about half of the cathelin-related genes in the pig.

To determine the clustering and approximate number of protegrin genes or pseudogenes, four protegrin clones and seven other related clones were digested with different restriction enzymes and hybridized with protegrin-specific (complementary to nt 403–429 of PG1–3 cDNA) and more general cathelin-detecting oligonucleotide probes. Several bands hybridized with the protegrin-specific probes in each of the protegrin clones (Fig. 5). The observed patterns indicate that at least four protegrin genes (or pseudogenes) exist in an individual pig, and further demonstrate that several protegrin-related genes (or

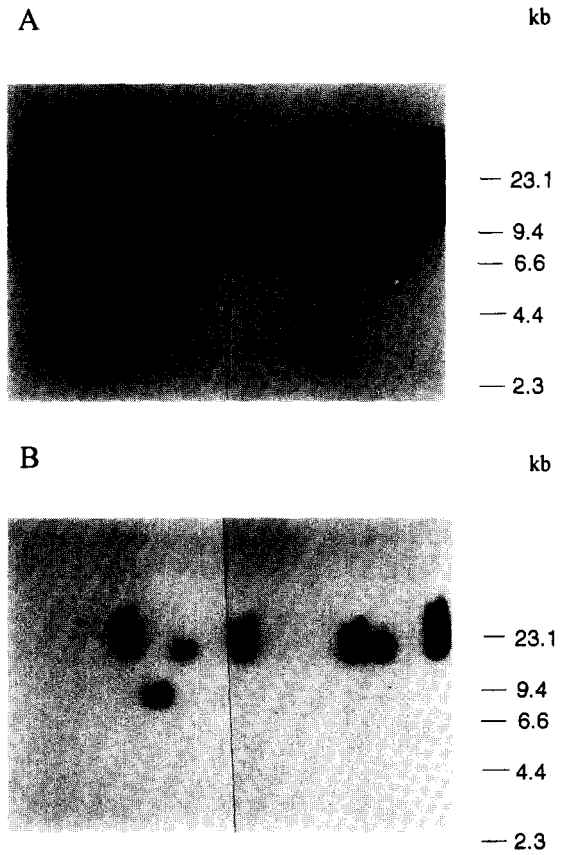

Fig. 4. Southern blots of purified genomic clones for protegrin and other cathelin-associated genes. After screening a genomic library with a 691 bp PG-3 cDNA probe, DNA from 15 hybridizing clones was digested by *Xho*I, size-fractionated on an agarose gel, transferred to a nylon membrane, and hybridized with the following radiolabeled probes: (A) a ^{32}P-labeled 691 bp PG-3 cDNA; and (B) a ^{32}P-labeled 27 base oligonucleotide complementary to nt 403–429 of the PG1-3 cDNA sequences. The membrane used with probe (A) (top) was washed (in 0.4 N NaOH) and reprobed with (B) (bottom).

pseudogenes) are densely clustered within the 15–21 kb of porcine DNA carried within each phage insert.

Acknowledgements: We thank John Chang for help in doing some DNA sequence. This work was supported, in part, by a grant from the NIH, AI 22839.

References

[1] Ganz, T. and Lehrer, R.I. (1994) Current Opinion in Immunol. 6, 584–589.
[2] Ritonja, A., Kopitar, M., Jerala, R. and Turk, V. (1989) FEBS Lett. 255, 211–214.
[3] Sal, G.D., Storici, P., Schneider, C., Romeo, D. and Zanetti, M. (1992) Biochem. Biophys. Res. Commun. 187, 467–472.
[4] Storici, P., Sal, G.D., Schneider, C. and Zanetti, M. (1992) FEBS Lett. 314, 187–190.
[5] Gennaro, R., Skerlavaj, B. and Romeo, D. (1989) Infect. Immunol. 57, 3142–3146.
[6] Zanetti, M., Sal, G.D., Storici, P., Schneider, C. and Romeo, D. (1993) J. Biol. Chem. 268, 522–526.
[7] Frank, R.W., Gennaro, R., Schneider, K., Przybylski, M. and Remeo, D. (1990) J. Biol. Chem. 265, 18871–18874.
[8] Scocchi, M., Romeo, D. and Zanetti, M. (1994) FEBS Lett. 352, 197–200.
[9] Ooi, C.E., Weiss, J., Levy, O. and Elsbach, P. (1990) J. Biol. Chem. 265, 15956–15962.
[10] Levy, O., Weiss, J., Zarember, K., Ooi, C.E. and Elsbach, P. (1993) J. Biol. Chem. 268, 6058–6063.
[11] Tossi, A., Scocchi, M., Skerlavaj, B. and Gennaro, R. (1994) FEBS Lett. 339, 108–112.
[12] Larrick, J.W., Morgan, J.G., Palings, I., Hirata, M. and Yen, M.H. (1991) Biochem. Biophys. Res. Commun. 179, 170–175.
[13] Agerberth, B., Lee, J.-Y., Bergman, T., Carlquist, M., Boman, H.G., Mutt, V. and Jornvall, H. (1991) Eur. J. Biochem. 202, 849–854.
[14] Storici, P. and Zanetti, M. (1993) Biochem. Biophys. Res. Commun. 196, 1058–1065
[15] Kokryakov, V.N., Harwig, S.S.L., Panyutich, E.A., Shevchenko, A.A., Aleshina, G.M., Shamova, O.V., Korneva, H.A. and Lehrer, R.I. (1993) FEBS Lett. 327, 231–236.
[16] Zhao, C., Liu, L. and Lehrer, R.I. (1994) FEBS Lett. 346, 285–288.
[17] Harwig, S.S., Kokryakov, V.N., Swiderek, K.M., Aleshina, G.M., Zhao, C. and Lehrer, R.I. (1995) FEBS Lett. 362, 65–69.
[18] Pungercar, J., Strukelj, B., Kopitar, G., Renko, M., Lenarcic, B., Gubensek, F. and Turk, V. (1993) FEBS Lett. 336, 284–288.
[19] Zanetti, M., Storici, P., Tossi, A., Scocchi, M. and Gennaro, R. (1994) J. Biol. Chem. 269, 1–4.
[20] Paola, S., Scocchi, M., Tossi, A., Gennaro, R. and Zanetti, M. (1994) FEBS Lett. 337, 303–307.
[21] Mount, S.M. (1982) Nucleic Acid Res. 10, 459–472.
[22] Storici, P. and Zanetti, M. (1993) Biochem. Biophys. Res. Commun. 196, 1363–1368.
[23] Mirgorodskaya, O.A., Schevchenko, A.A., Adballa, K.O., Chernushevich, I.V., Egorov, T.A., Musoliamov, A.X., Kokryakov, V.N. and Shamova, O. (1993) FEBS Lett. 330, 339–342.
[24] Harwig S.S.L., Swiderek, K., Lee, T.D. and Lehrer, R.I. (1995) J. Peptide Sci. in press.
[25] Tossi, A., Scocchi, M., Zanetti., Storici, P. and Gennaro, R. (1995) Eur. J. Biochem. 228, 941–946.
[26] Faisst, S. and Meyer, S. (1992) Nucleic Acids Res. 20, 3–26.
[27] Lenardo, M.J., Kuang, A., Gifford, A. and Baltimore, D. (1988) Hamatol. Bluttransfus. 32, 411–415.
[28] Yang, Z., Sugawara, M., Ponath, P.D., Wessendorf, L., Banerji, J., Li, Y. and Strominger, L. (1990) Proc. Natl. Acad. Sci. USA 87, 9226–9230.

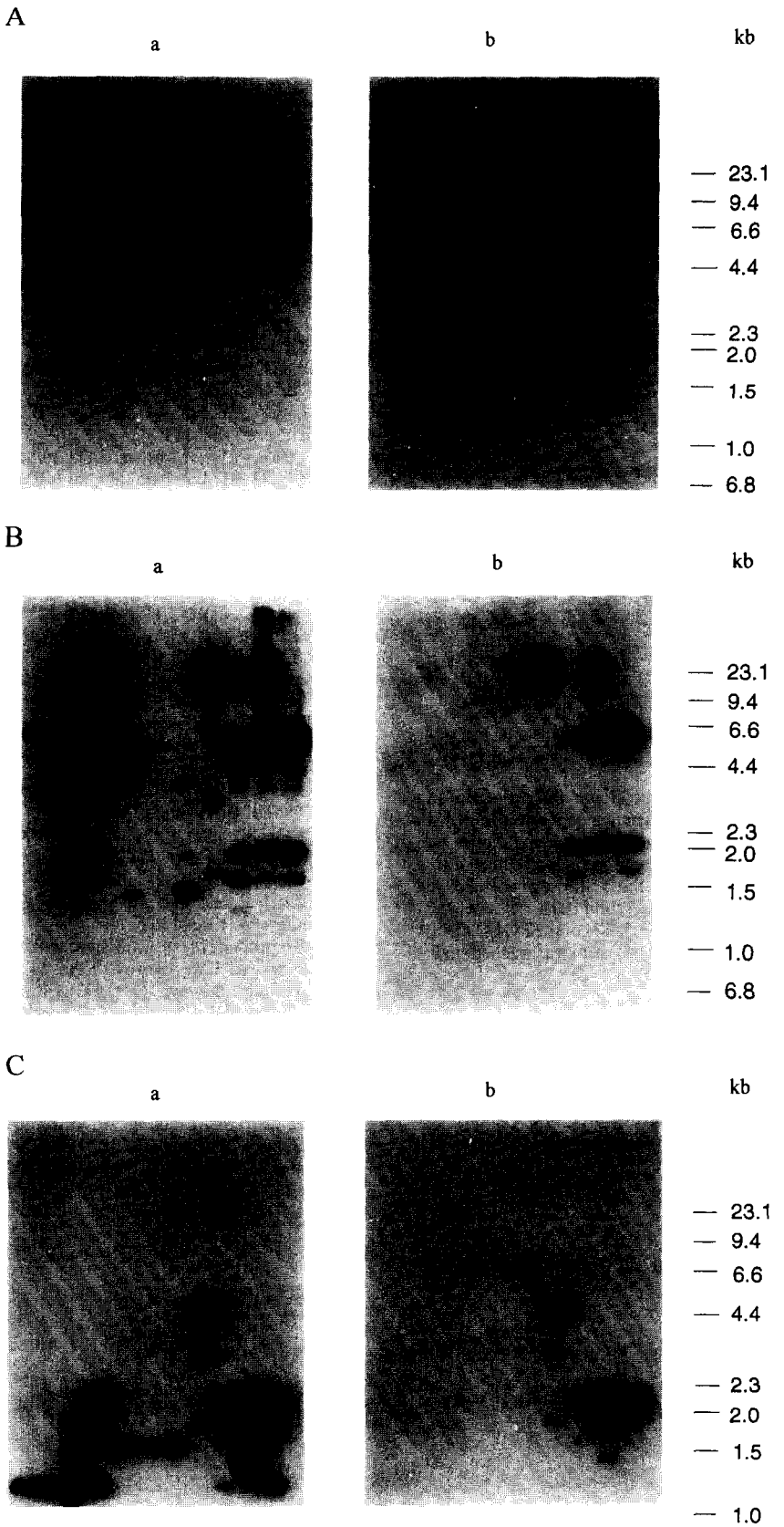

Fig. 5. Southern blot analysis of protegrin genes. Four genomic protegrin clones (the four right lanes of each blot) and seven cathelin-associated non-protegrin clones were digested, fractionated on agarose and transferred to nylon membranes. The DNA digests were first hybridized with a protegrin PG-3 cDNA probe (left panel, (a), of each), then washed and re-probed with the PG-specific probe used in Fig. 4 (right panel, (b), of each). The following restriction enzymes were used for the digestions: A (top pair) *Bam*HI; B (middle pair) *Bam*HI+*Eco*RI; C (bottom pair) *Bam*HI+*Pst*I.