FEBS Letters 376 (1995) 130–134 FEBS 16308

Structures of genes for two cathelin-associated antimicrobial peptides: prophenin-2 and PR-39

Chengquan Zhao, Tomas Ganz, Robert I. Lehrer*

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

Received 16 October 1995

Abstract We characterized genes for prophenin (PF)-2 and PR-39, two cathelin-associated antimicrobial peptides found in porcine leukocytes. Both contained 4 exons and 3 introns and were compact, contiguous and highly homologous. Exons I–III encoded most of their cathelin domains. Exon IV specified the final few cathelin residues, including its conserved C-terminal valine, followed by the mature PR-39 peptide or a PF-2 precursor. The highly conserved 5′ flanking sequences of this gene family contained NF-κB, IL-6, GM-CSF and NF-1 binding motifs and the introns were unusually conserved. These data suggest that the panoply of porcine cathelin-associated antimicrobial peptides arose relatively recently via gene reduplications and exon shuffling, and that in vivo expression of cathelin-associated antimicrobial peptides may respond to mediators generated early during infection.

Key words: Protegrin; Antimicrobial peptide; Cathelin; Prophenin; PR-39

1. Introduction

Endogenous antimicrobial peptides produced by leukocytes are fundamental contributors to innate immunity against microbial infections [1]. Recent cDNA cloning studies have revealed that many structurally and functionally diverse leukocyte antimicrobial peptides share a highly conserved pro-region, known as the 'cathelin' domain. Among these are bovine indolicidin, cyclic dodecapeptide, Bac 5 and Bac 7 [2–7]; rabbit 'P15' and CAP 18 [8–11]; porcine PR-39 [12,13], protegrins [14–16], prophenins [17–19], PMAP-23, PMAP-36 and PMAP-37 [20–22]; and human FALL-39/CAP-18 [23,24].

Several cathelin-associated antimicrobial peptides, including porcine PR-39 [12], protegrins PG1-3 [14], and prophenin [17], have been isolated from leukocytes. Others, including porcine PMAP-23, PMAP-36, PMAP-37 and protegrins PG-4 and 5 have been synthesized based on their cDNA or gene sequences. In some cases, both peptide isolation and cDNA cloning have been accomplished. For example, Pungercar et al. obtained two cDNA clones from porcine bone marrow, 'clone-12' and 'clone-6', that encoded nearly identical proline-rich antimicrobial peptides [18,19] and we independently purified the corresponding mature peptides, Prophenin (PF)-1 and 2, from porcine leukocytes [17].

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

Sequence data have been deposited with EMBL/GenBank Libraries with the following accession numbers X89201 *S. scrofa* βPR-39 gene, X89202 *S. scrofa* PF-2 gene

To date, only genes for porcine protegrins and PR-39 have been described [25,26]. We now report the structures of genes for prophenin-2 (PF-2) and a variant form of PR-39.

2. Materials and methods

2.1. Genomic screening and Southern blot analysis

A porcine liver genomic library in EMBL3 SP6/T7 (Clontech, Palo Alto, CA) was screened as previously described [25]. Genomic DNA was purified from hybridizing clones, digested with various restriction endonucleases (New England Biolabs, Beverly, MA) and resolved by electrophoresis in 0.8% agarose gels. Overnight capillary transfer to Gene Screen Plus membranes (Dupont, Boston, MA) was followed by prehybridization in a Rap-hyb buffer with 100 μg/ml of denatured salmon sperm DNA. Subsequently, the filters were hybridized for 2 h at 60°C with radio-labeled cDNA or at 42°C with radiolabelled synthetic oligonucleotides. The final wash was performed in 0.1 × SSC with 0.1% SDS, at either 60°C or 42°C, respectively.

2.2. Polymerase chain reaction (PCR)

Oligonucleotide primers derived from PF-2 (clone-12), PR-39, and protegrin cDNA sequences [13,15,16,18] were used (Table 1). Reactions were carried out in a total volume of 50 μl containing 50 ng of selected genomic library DNA (templates), 25 μl 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) overlaid with 20 μl 'Chill-out' liquid wax. Thirty cycles were performed, as follows: 1 min denaturation at 94°C, 1 min primer annealing at 55°C, 2 min primer extension at 72°C and 10 min final extension at 72°C.

2.3. Gene cloning and sequencing

PCR DNA products of interest and restriction fragments of genomic library clones were fractionated by agarose gel electrophoresis, excised, purified with a QIAEX Gel Extraction kit (QIAGEN Inc, Chatsworth, CA), and ligated into pBluescript KS+ vector or pCR-Script SK+ cloning vector. DNA for sequencing was prepared with a QIAGEN plasmid kit, and sequenced by the dideoxynucleotide chain termination method using a Sequenase version 2.0 kit (U.S. Biochemical, Cleveland, OH), and/or a Bst DNA sequencing kit (Bio-Rad Laboratories, Melville, NY), and pBluescript universal primers, or specific primers based on PF, and PR-39 cDNA sequences. DNA sequences were analyzed using PC-Gene (Intelligenetics, Palo Alto, CA).

3. Results and discussion

By screening a pig genomic library of $\approx 2.3 \times 10^5$ clones with protegrin (PG) cDNA that contained the conserved cathelin region, we identified 45 hybridizing clones [25]. About half of these did not hybridize with our protegrin-specific oligonucleotide probe, P4 (Table 1), indicating that they contained non-protegrin, cathelin-associated genes. Ten P4-negative clones were digested with restriction enzymes. Three of these (Fig. 1) hybridized with oligonucleotide probe-P5, which was specific for prophenin, and one of these (Clone-13) also hybridized with probes-P7 (Fig. 1) and P6 (data not shown), which were specific for PR-39.

SSDI 0014-5793(95)01237-0

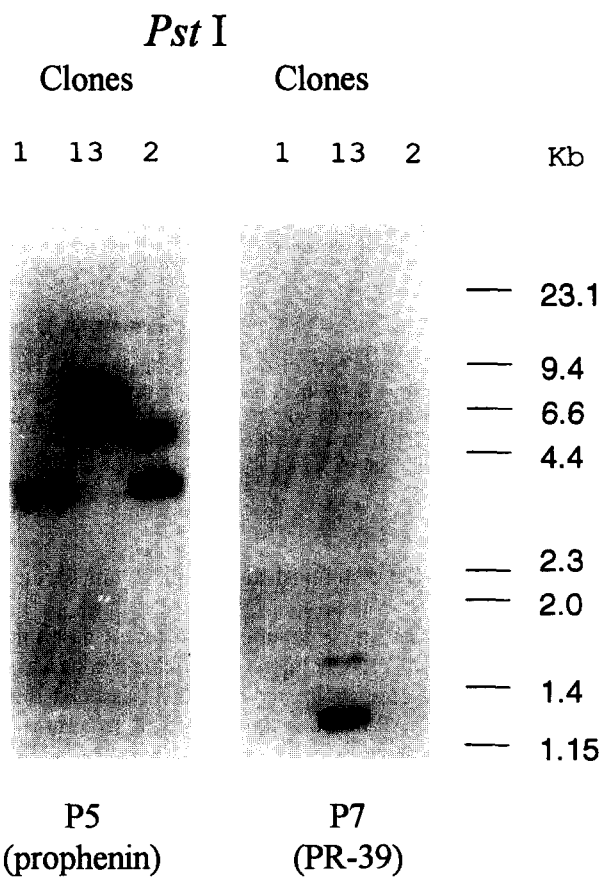

Fig. 1. Southern blot analysis of prophenin (PF) and PR-39 genes. Three PF-positive genomic clones were digested with *Pst*I, fractionated on agarose gels and transferred to nylon membranes. The membranes were first hybridized with the prophenin (PF)-specific probe-P5, then washed and re-probed with the PR-39 specific probe-P7.

Fig. 1 also indicates that clone 2 contained at least two PF genes, and that clone 13 contained at least one PF gene and one PR-39 gene. We also hybridized 15 PG containing clones with probe-P5 (PF specific) and probe-P7 (PR-39 specific). Although no PF-hybridizing clone was found, 1 of these clones reacted with the PR-39 probe, indicating that PR-39 and protegrin genes are contiguous (data not shown).

PCR amplification of clones 1, 2 and 13 was done with different primer pairs (Fig. 2) and the 7 PCR products were subcloned into pBluescript vector and sequenced. Products 1 and 2 contained the PF-2 coding region; product 3 contained the coding region of a novel PR-39 gene which we called βPR-39. Products 4–6 encoded the mature PF-2 peptide and its 3′-untranslated region; product 9 coded for the mature PR-39 peptide and its 3′-untranslated region (confirming that clone 13 contained both PF-2 and βPR-39 genes). To obtain the 5′-flanking region of the PF-2 gene, clone 1 DNA was cut with *Pst*I, and probed with the oligonucleotide P8 (Table 1). A hybridizing 1.2 kb fragment was identified, inserted into pBluescript KS vector and sequenced. It comprised 691 bp of the 5′-flanking region, exon I and most of intron I of the PF-2 gene. To obtain the 5′-flanking region of βPR-39 gene, PCR amplification of our clone 13 DNA was done with primers P9 and P10 (Table 1). The 1.1 kb PCR product was inserted into pBluescript vector and sequenced, revealing that it contained the 5′-flanking region (142 bp) exon I, intron I, exon II, intron II and exon III of the βPR-39 gene.

The genes for PF-2 and βPR-39 both consisted of four exons and three introns (Figs. 3 and 4, Table 2). A very short 5′-untranslated segment was contained in Exon I, along with the codons of the first 66 amino acids, including a 29 residue signal peptide and the first 37 cathelin residues. Exons II and III contained the next 60 cathelin residues. Exon IV contained the final 2 or 4 cathelin residues and the complete sequences of mature PF-2 or βPR-39, followed by their 3′-untranslated regions and polyadenylation sites, respectively. Both the cDNA (18 and unpublished data) and gene structures (Fig. 4) of PF-2 indicate that its post-cathelin portion of PF-2 should begin from the seventh bp in exon IV. However, when mature PF-1 purified from pig leukocytes, it lacked the predicted 18 residue N-terminal segment and the C-terminal Gly-Arg-Arg sequence, which had evidently been removed during post-translational processing [17].

The amino acid sequence corresponding to our PF-2 gene was identical to that deduced previously from the cDNA of Pungercar's clone-12 [18]. The βPR39 genomic clone was almost identical to the cDNA sequence of PR-39 reported by Storici and Zanetti [13], differing principally by a base pair and codon change (GCA to GGA) in Exon 1. This would cause a Gly^{21} for Ala^{21} substitution in the signal sequence, without altering the mature PR-39 peptide. The cathelin domain encoded by our βPR-39 gene corresponded precisely to the cathelin sequence reported by Storici and Zanetti [13]. In contrast, the PR-39 gene recently reported by Gudmundson et al. [26]

Table 1
Oligonucleotides for PCR and southern blot studies

Primers	Sequences (5′-3′)	cDNA	Location	references
P1	GTCGGAATTCATGGAGACCCAGAGG(or A)GCCAG	S**	1-20(+)*	[15]
P2	GTCGTCTAGAC(or G)GTTTCACAAGAATTTATTT	S	672-691(-)*	[15]
P3	ATGGGCTCACCTGGGCACC	PR-39	1-17(+)	[13]
P4	GCAGAACCTACGCCTACAATAGCACAG	PG	403-429(-)	[15]
P5	CTCAAGCTTGTCAGGAGATTTCCCTGGTGG	PF	384-404(+)	[18]
P6	GTCGGAATTCAGGAGACGTCCCCGAC	PR-39	408-423(+)	[13]
P7	CCAAGGCTCCCACCAAGGATCAT	PR-39	465-485(+)	[13]
P8	CACCACGAGTC(orG)CCAGCAGCAGAAG	S	49-72(-)	[15]
P9	CTCATTACAGGAGATGTCCAG	PR-39	375-395(-)	[13]
P10	CCCACTGCGCAGCAATCCCCTGA	PR-39 gene	686-708(+)	[26]

*(+) and (−) indicate positive and negative strands, respectively. **S indicates that the sequences of PF, PR-39 and PG are identical.

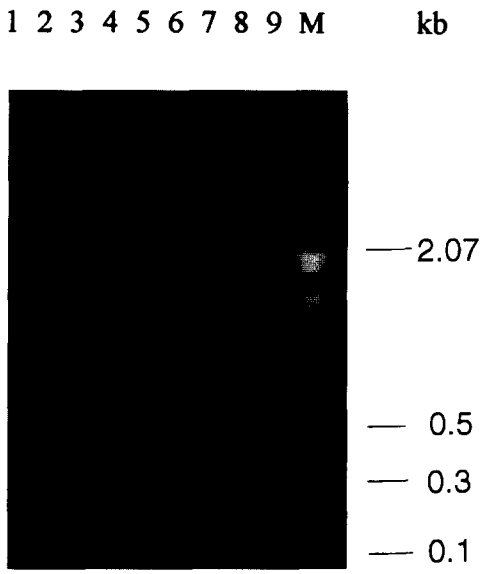

Fig. 2. PCR amplification of prophenin and PR-39 genes. PCR products obtained using a clone 1 template are shown in lanes 1, 4 and 7. PCR products from a clone 2 template are shown in lanes 2, 5 and 8. PCR products from a clone 13 template are in lanes 3, 6 and 9. Primer (P) pairs, as described in Table 2, were used as follows: lanes 1, 2: P1 & P2; lane 3: P3 & P2; lanes 4, 5, 6: P5 & /P2; lanes 7, 8, 9: P6 & P2.

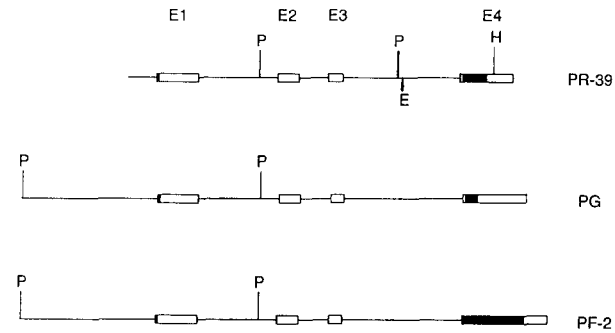

Fig. 3. Structures of prophenin PF-2, βPR-39, and protegrin (PG) genes. Open boxes denote the four exons (E). The extreme left portion (shaded) of exon I (E1) denotes the 5′-untranslated region, and the black box in exon 4 (E4) represents the sequences encoding the mature peptides. The restriction sites are *Pst*I (P), *Eco*RI (E), *Hin*dIII (H). The map is drawn to scale.

differed from Storici and Zanetti's cDNA by 7 base pair changes (and 5 amino acid substitutions) in the cathelin domain segments encoded within Exons II and III.

We scanned the 5′-flanking region of PF-2 for eukaryotic promoter elements. Fig. 4 shows the consensus TATA-like box (ATAAA) identified in the region of −32 to −28. The sequence differs from the consensus TATA(TA)A by first base deletion, but it meets the −30 location requirement. Other potential regulatory motifs shown in Fig. 4 include sites for NF-κb, NF-IL6, IL-6RE, GMCSF and NFI. Corresponding sites were present in the previously reported protegrin and PR-39 genes [25,26]. The 142 bps sequenced in the 5′-region of the βPR39 gene were identical to their counterparts in PF-2 and PG genes.

Genes encoding PF-2 and βPR-39 had identical layouts, with four exons and three introns (Fig. 3). The sequences encoding the mature peptides were located in exon IV, which began with the final 2 or 4 residues fo the cathelin domain. Table 2 show the resemblance of PF-2 and βPR-39 genes to each other and to protegrin genes. From the *Pst*I cut site in its 5′-flanking region to the end of intron III, only 17 nucleotide substitutions, 1 nucleotide deletion and a 6 bp difference in the length of intron III distinguished the PF-2 and protegrin genes.

The present studies, in conjunction with recent descriptions of protegrin [25] and PR-39 [26] genes provide a coherent picture of the porcine genes that encode cathelin-associated antimicrobial peptides. All of them have identical structures that differ substantially from those of defensin genes [27,28]. Their homology and nearby chromosomal locations suggests that the remarkably large family of cathelin-containing antimicrobial peptides expanded via gene reduplication and insertion or substitution of novel sequences into exon IV of previous family members. The conserved intron sequences of the genes could reflect: (a) the recency of this expansion; (b) the occurrence of specific gene conversion events that assured overall sequence preservation; (c) evolutionary selection occasioned by the functional significance of the intron sequences; or (d) combinations of the aforementioned possibilities. Intron I of the Ha-*ras* genes of humans and rodents is highly conserved and contains at least 3 *cis*-acting regulatory elements that modulate gene expression [29]. Regulatory elements also exist in intron II of the KI-*ras* gene [30].

Since one of our cloned 15–21 kb inserts contained both PF and PR-39 genes and another contained both PG and PR-39 genes, it is apparent that the multiple PG, PR-39, and PF are densely clustered, presumably on porcine chromosome 13 [26]. Our βPR-39 gene and the PR-39 gene recently described by Gudmundsson et al. [26] differed on average by 3.6%, reflecting occasional base pair substitutions scattered throughout the gene.

Table 2
Homology of PF-2, βPR-39 and PG-1 genes

Gene base pairs	5′ side	ExonI	IntronI	ExonII	IntronII	ExonIII	IntronIII	ExonIV
PF-2	691	211	404	108	152	72	601	424
PG-1	690	211	404	108	152	72	595	313
βPR-39	142	215**	404	108	146	72	591	263
Comparisons	Homology of the gene sequences (% identity)							
PG-1/PF-2	99.0	100	99.8	99.1	98.0	100	99.0	85.3*
βPR39/PG-1	99.3	99.5	99.5	99.1	86.8	76.4	69.4*	58.1*

*Gaps (≥ 10 base pair deletions) were omitted from the calculation. **This difference between βPR-39 and PF-2 or PG-1 reflects an alternative transcriptional start site.

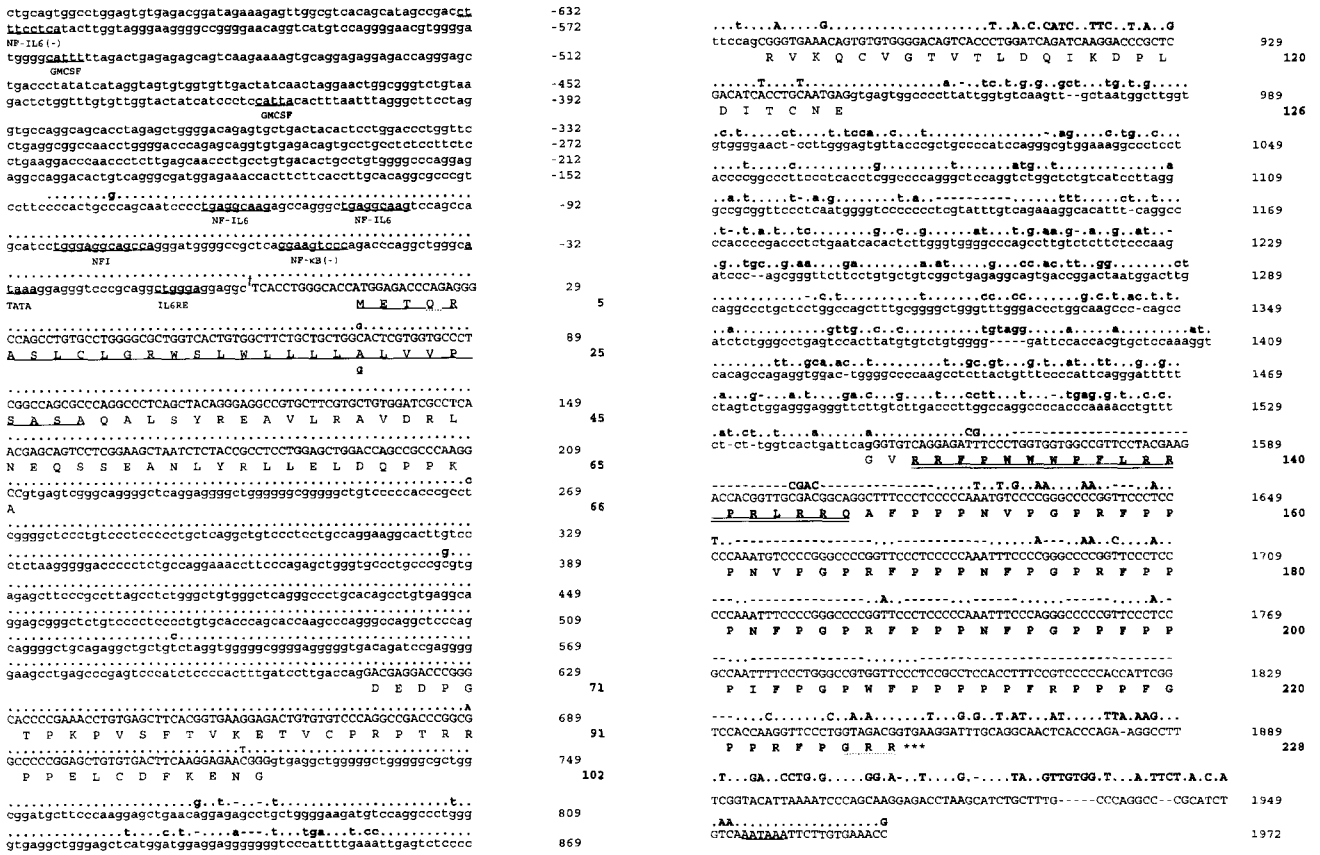

Fig. 4. Nucleotide sequences of prophenin-2 and βPR-39 genes. The sequence of the PF-2 gene is shown in full, whereas nucleotides of the βPR-39 gene are shown in bold lower or upper case letters above the PF-2 sequence only when they differ from PF-2. Identical residues are indicated by dots. Some gaps have been introduced for alignment, and are shown by dashes. Exon nucleotides are capitalized, and 5′-flanking region and intronic sequences are shown in lower case letters. Nucleotides are numbered according to the presumed transcription start site in PF-2, which is designated by the dagger symbol. Amino acid numbering is according to the pig PF-2 (clone 12) cDNA [18]. The signal peptide is indicated by underlined capital letters, and the polyadenylation signal (AATAAA) has also been underlined. The putative PF-2 precursor, prior to its final processing by proteolytic removal of the doubly underlined 18 N-terminal residues and conversion of its C-terminal amidation sequence (dotted underlining) is shown in bold capital letters. A TATA-like box and several potential binding sites (− means negative strand) for regulatory factors are underlined and named. The bold numbers refer to the amino acid transcript of PF-2. Gaps (−) introduced into the PF-2 sequence for alignment purposes were not counted when numbering the bases.

The 5′ upstream regions of the genes contained clusters of potential transcriptional regulatory elements. Within 130 bp upstream of the transcriptional start sites of all of these genes were one NF-κb recognition site, one NFI motif, two NF-IL6 motifs and one IL-6 response element (IL6-RE) (Fig. 4). Similar binding sites are present in the promoter regions of other genes that participate in a variety of immune responses [e.g. 31,32]. The presence of multiple upstream regulatory elements, including NF-κB and NF-IL-6, has also been found in genes that encode inducible antimicrobial peptides in invertebrates [33–35] and in the gene of a bovine β-defensin produced by respiratory tract epithelial cells [36]. The presence of similar upstream regulatory elements in the genes for cathelin-associated peptides suggests that their in vivo production is poised to respond to signals, such as cytokines, that are generated early during infectious challenges.

Acknowledgements: We thank John Chang for help in DNA sequencing. This work was supported, in part, by NIH grants AI22839 and AI 37945.

References

[1] Boman, H.G. (1991) Cell. 65, 205–207.
[2] Sal, G.D., Storici, P., Schneider, C., Romeo, D. and Zanetti, M. (1992) Biochem. Biophys. Res. Commun. 187, 467–472.
[3] Storici, P., Sal, G.D., Schneider, C. and Zanetti, M. (1992) FEBS Lett. 314, 187–190.
[4] Gennaro, R. Skerlavaj, B. and Romeo, D. (1989) Infect. Immun. 57, 3142–3146.
[5] Zanetti, M., Sal, G.D., Storici, P., Schneider, C. and Romeo, D. (1993) J. Biol. Chem. 268, 522–526.
[6] Frank, R.W., Gennaro, R., Schneider, K. Przybylski, M. and Remeo, D. (1990) J. Biol. Chem. 265, 18871–18874.
[7] Scocchi, M., Romeo, D. and Zanetti, M. (1994) FEBS Lett. 352, 197–200.
[8] Ooi, C.E., Weiss, J., Levy, O. and Elsbach, P. (1990) J. Biol. Chem. 265, 15956–15962.
[9] Levy, O., Weiss, J., Zarember, K., Ooi, C.E. and Elsbach, P. (1993) J. Biol. Chem. 268, 6058–6063.
[10] Tossi, A., Scocchi, M., Skerlavaj, B. and Gennaro, R. (1994) FEBS Lett. 339, 108–112.
[11] Larrick, J.W., Morgan, J.G., Palings, I., Hirata, M. and Yen, M.H. (1991) Biochem. Biophys. Res. Commun. 179, 170–175.
[12] Agerberth, B., Lee, J-Y., Bergman, T., Carlquist, M., Boman,

H.G., Mutt, V. and Jörnvall, H. (1991) Eur. J. Biochem. 202, 849–854.
[13] Storici, P. and Zanetti, M. (1993) Biochem. Biophys. Res. Commun. 196, 1058–1065.
[14] 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.
[15] Zhao, C., Liu, L. and Lehrer, R.I. (1994) FEBS Lett. 346, 285–288.
[16] Storici, P. and Zanetti, M. (1993) Biochem. Biophys. Res. Commun. 196, 1363–1368.
[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] Strukelj, B., Pungercar, J., Kopitar, G., Renko, M., Lenarcic, B., Berbic, S. and Turk, V. (1995) Biol. Chem. Hoppe-Seyler 376, 507–510.
[20] Zanetti, M. Storici, P., Tossi, A., Scocchi, M. and Gennaro, R. (1994) J. Biol. Chem. 269, 1–4.
[21] Paola, S., Scocchi, M., Tossi, A., Gennaro, R. and Zanetti, M. (1994) FEBS Lett. 337, 303–307.
[22] Tossi, A., Scocchi, M., Zanetti, M., Storici, P. and Gennaro, R. (1995) Eur. J. Biochem. 228, 941–946.
[23] Agerberth, B., Gunne, H., Odeberg, J., Kogner, P. and Boman, H.G. (1995) Proc. Natl. Acad. Sci. USA 92, 195–199.
[24] Larrick, J.W., Hirata, M., Balint, R.F., Lee, J., Zhong, J. and Wright, S.C. (1995) Infect. Immun. 63, 1291–1297.
[25] Zhao, C., Ganz, T. and Lehrer, R.I. (1995) FEBS Lett. 368, 197–202.
[26] Gudmundsson, G.H., Magnusson, K.P., Chowdhary, B.P., Johansson, M., Andersson, L. and Boman, H.G. (1995) Proc. Natl. Acad. Sci. USA 92, 7085–7089.
[27] Ganz, T. and Lehrer, R.I. (1994) Curr. Opin. Immunol. 6, 584–589.
[28] Linzmeier, R., Michaelson, D., Liu, L. and Ganz, T. (1993) FEBS Lett. 321, 267–273.
[29] Telliez, J-B., Plumb, M., Balmain, A. and Bailleul, B. (1995) Mol. Carcinogen. 12, 137–145.
[30] Chen, B., Johanson, L., Weist, J.S., Anderson, M.W. and Tou, M. (1994) Proc. Natl. Acad. Sci. USA 91, 1589–1593.
[31] Baeuerle, P.A. (1991) Biochim. Biophys. Acta 1072, 63–80.
[32] Ohmori, Y. and Hamilton, T.A. (1994) Pharmacol. Ther. 63, 235–264.
[33] Reichhart, J-M., Merster, M., Dimarcq, J-L., Zachary, D., Hoffmann, D., Ruiz, C., Richards, G. and Hoffmann, J.A. (1992) EMBO J. 11, 1469–1477.
[34] Sun, S.-C. and Faye, I. (1992) Eur. J. Biochem. 204, 885–892.
[35] Sun, S.-C., Lindstrom, I., Lee, J-Y. and Faye, I. (1991) Eur. J. Biochem. 196, 247–254.
[36] Diamond, G., Jones, D.E. and Bevins, C.L. (1993) Proc. Natl. Acad. Sci. USA 90, 4596–4600.