FEBS Letters 346 (1994) 285–288

FEBS 14107

Identification of a new member of the protegrin family by cDNA cloning

Chengquan Zhao, Lide Liu, Robert I. Lehrer*

Department of Medicine, UCLA Center for the Health Sciences, Los Angeles, CA 90024, USA

Received 20 April 1994

Abstract

The porcine leukocyte protegrins are a family of cysteine-rich antimicrobial peptides the primary structures of which combine features of defensins and tachyplesins. We cloned three protegrins from porcine bone marrow mRNA by PCR, including one (PG-4) that was previously unknown. The 691 bp protegrin cDNAs were > 98.8% identical, and each was surrounded by highly conserved 5′ and (in some instances) 3′ sequences present in structurally dissimilar antimicrobial and LPS-binding peptides of animal leukocytes.

Key words: Protegrin; cDNA cloning; PG-4; Cathelin; Antimicrobial peptide

1. Introduction

Antimicrobial polypeptides play an important role in the defense mechanisms of phagocytes. Unlike the leukocyte-generated oxidants which also participate in phagocyte-mediated host defense, the antimicrobial peptides and proteins of leukocytes show marked inter-species differences in structure, spectrum and potency. Recent reports have shown that porcine, bovine and rabbit leukocytes synthesize several structurally diverse antimicrobial peptides on precursors that contain a conserved 'cathelin' domain [1,2]. Included among these cathelin-associated peptides are: porcine PR-39 [3,4], porcine protegrin PG-2, [5,6], porcine PMAP-36 [7], rabbit CAP18 and p15 [8,9], bovine indolicidin [10,11], bovine cyclic dodecapeptide [12,13], and bovine Bac5 [14,15].

We recently reported the primary sequences of three cysteine-rich antimicrobial peptides (protegrins) that were purified from porcine leukocytes [5]. Soon thereafter, Storici and Zanetti cloned protegrin PG-2, and reported that its precursor contained a cathelin domain [4]. We now describe the cDNA cloning of three other protegrins from porcine bone marrow cells: PG-1, PG-3 and the newly discovered PG-4.

2. Materials and methods

2.1. cDNA generation and PCR amplification

Total RNA was extracted from the bone marrow cells of a young red Duroc pig with guanidinium thiocyanate [16]. 1 μg of total RNA was used to synthesize the first strand cDNA, with 20 pmol oligo(dT) primer and 200 U Moloney-murine leukemia virus (M-MLV) reverse transcriptase (Clontech Laboratory, Palo Alto, CA) in a total reaction volume of 20 μl. Two PCR primers were prepared. The sense primer (5′-GTCGGAATTCATGGAGACCCAGAG (A or G) GCCAG-3′) corresponded to the 5′ regions of PG-2 and PR-39 cDNA and contained an *Eco*RI restriction site. The antisense primer (5′-GTCGTCTAGA (C or G) GTTTCACAAGAATTTATTT-3′) was complementary to 3′ ends of PG-2 and PR-39 cDNA immediately preceding their poly(A) tails and contained an *Xba*I restriction site. PCR was carried out in a 50 μl volume using a 1/10 vol. of the above pig cDNA as template, 25 pmol primers and 2.5 U AmpliTaq DNA polymerase (Perkin Elmer-Cetus). The reaction was run for 30 cycles, with 1 min denaturation (94°C) and annealing (60°C) steps and a 2 min extension step (72°C) per cycle.

2.2. cDNA cloning and sequencing

The amplified cDNA was fractionated by preparative agarose electrophoresis and stained with ethidium bromide. The main fragment was cut out, digested with *Eco*RI and *Xba*I endonucleases (New England Biolabs, Beverly, MA), subcloned into a M13mp18 bacteriophage vector, and transformed into *E. coli* XL1-Blue MRF′ competent cells (Stratagene, La Jolla, CA). DNA sequencing was performed with a kit (US Biochemical Corp., Cleveland, OH). Nucleotide and protein sequences were analyzed with PC-GENE (Intelligenetics, Palo Alto, CA).

2.3. Northern blots

10 μg of total RNA was denatured in 50% formamide, separated by electrophoresis through 1% agarose gels in 0.62 M formaldehyde, and blotted onto GeneScreen Plus membranes (DuPont, Boston, MA) by capillary transfer. The membrane was baked at 80°C for 2 h, and hybridized with ^{32}P-labeled probe in rapid hybridization buffer (Amersham, Arlington Height, IL).

3. Results and discussion

Storici and Zanetti recently cloned protegrin PG-2 and reported that its 5′ end contained a cathelin domain that was nearly identical to the corresponding region in PR-39 and several other antimicrobial peptides [6]. Since PG-1 and PG-3 closely resemble PG-2 in primary structure, we expected that their mRNAs would have similar sequences at their 5′ and 3′ ends. We amplified reverse transcribed porcine bone marrow cell RNA, using an upstream primer that corresponded to the 5′ ends of PG-2 and PR-39 cDNA and a downstream primer matching the region immediately preceding the poly(A) tail. The resulting ≈0.7 kb PCR product was subcloned

*Corresponding author. Fax: (1) (310) 206-8766.

SSDI 0014-5793(94)00493-F

	2	4	6 8	10 12		14	16 18	
PG-1	RGG	R	LCYCR	RRF	C	V	CVGR	(NH_2)
PG-2	RGG	R	LCYCR	RRF	C	I	CV	(NH_2)
PG-3	RGG	G	LCYCR	RRF	C	V	CVGR	(NH_2)
PG-4	RGG	R	LCYCR	GWI	C	F	CVGR	(G)

Fig. 1. Primary sequences of protegrins. The amino acid sequences of protegrins PG-1, -2 and -3 were previously described [5]. The sequence of PG-4 is translated from the cDNA sequences described in this report. The C-terminal glycine will probably be represented by an amide in the mature peptide, as reported for PG-1, -2 and -3 [17].

into M13mp18, and recombinant plaques were chosen for single-strand DNA purification and sequencing. These revealed cDNA sequences for three different protegrins, two of which corresponded to the previously described PG-1 and PG-3 [5], and the other (represented by 4 clones) to a novel protegrin, PG-4. Curiously, none of the 20 sequenced clones corresponded to PG-2. The amino acid sequences of all four protegrins are shown in Fig. 1.

The cDNA sequences of protegrins (PG-1, -3 and -4) contained 691 bases, as had previously been shown for PG-2 [6]. The nucleotide and deduced amino acid sequences of all 4 protegrins are shown in Fig. 2. PG-1, PG-3 and PG-4 each contained complete 447 nucleotide open reading frames, with a stop codon (TGA) at position 448, corresponding to a protein with a mass of approximately 16.6 kDa, that contained 149 amino acid residues. Although PG-1 and PG-3 differed by only one nucleotide substitution, this occurred in the mature peptide region and led to the arginine vs. glycine seen at position 4 of the mature peptide (position 134 of the precursor). There were 5 nucleotide substitutions in the cDNA sequences of PG-1 and PG-4. Four of these were located in a region that encoded the mature peptide. All were in the first nucleotide of the codon and led to amino acid changes in the mature peptide. The fifth difference was in 3′ untranslated region.

The cDNA sequences of PG-1 and PG-2 differed only by four nucleotide substitutions. One was situated at codon 144 in the mature peptide coding region, and was responsible for the single amino acid difference (Val→Ile) at position 15 in the mature peptides. Two nucleotide substitutions were present in the 3′ end untranslated region. The final difference replaced the CGA found at position 442 in PG-1, -3 and -4 cDNA with a stop codon (TGA) in PG-2. This change explains the premature truncation of PG-2 (which has only 16 residues), relative to other protegrins, which have 18.

The deduced carboxy-terminal sequence of each protegrin included a carboxy-terminal glycine that was not found in mature PG-1, -2 and -3 (PG-4 has not yet been isolated). However, since protegrins PG 1–3 are known to have amidated carboxy-terminal arginine residues [17] the 'missing' glycine residue is the probable donor of the C-terminal amide nitrogen atom [18,19]. Northern blot analysis of porcine bone marrow RNA by with a

```
        10         20         30         40         50
ATGGAGACCCAGAGAGCCAGCCTGTGCCTGGGGCGCTGGTCACTGTGGCTTCTGCTGCTG   60
METGluThrGlnArgAlaSerLeuCysLeuGlyArgTrpSerLeuTrpLeuLeuLeuLeu   20

GCACTCGTGGTGCCCTCGGCCAGCGCCCAGGCCCTCAGCTACAGGGAGGCCGTGCTTCGT  120
AlaLeuValValProSerAlaSerAlaGlnAlaLeuSerTyrArgGluAlaValLeuArg   40

GCTGTGGATCGCCTCAACGAGCAGTCCTCGGAAGCTAATCTCTACCGCCTCCTGGAGCTG  180
AlaValAspArgLeuAsnGluGlnSerSerGluAlaAsnLeuTyrArgLeuLeuGluLeu   60

GACCAGCCGCCCAAGGCCGACGAGGACCCGGGCACCCCGAAACCTGTGAGCTTCACGGTG  240
AspGlnProProLysAlaAspGluAspProGlyThrProLysProValSerPheThrVal   80

AAGGAGACTGTGTGTCCCAGGCCGACCCGGCAGCCCCCGGAGCTGTGTGACTTCAAGGAG  300
LysGluThrValCysProArgProThrArgGlnProProGluLeuCysAspPheLysGlu  100

AACGGGCGGGTGAAACAGTGTGTGGGGACAGTCACCCTGGATCAGATCAAGGACCCGCTC  360
AsnGlyArgValLysGlnCysValGlyThrValThrLeuAspGlnIleLysAspProLeu  120
                                                 G3         G4
GACATCACCTGCAATGAGGTTCAAGGTGTCAGGGGAGGTCGCCTGTGCTATTGTAGGCGT  420
AspIleThrCysAsnGluValGlnGlyValArgGlyGlyArgLeuCysTyrCysArgArg  140
                                                 Gly3       Gly4
      A2
T4 A4 T4          T2
AGGTTCTGCGTCTGTGTCGGACGAGGATGACGGTTGCGACGGCAGGCTTTCCCTCCCCCA  480
ArgPheCysValCysValGlyArgGly---                                 149
Trp4Ile4 Phe4       ---2
         Ile2

ATTTTCCCGGGGCCAGGTTTCCGTCCCCCAATTTTTCCGCCTCCACCTTTCCGGCCCGCA  540
                                        A2 G2
CCATTCGGTCCACCAAGGTTCCCTGGTAGACGGTGAAGGATTTGCAGGCAACTCACCCAG  600
          C4
AAGGCCTTTCGGTACATTAAAATCCCAGCAAGGAGACCTAAGCATCTGCTTTGCCCAGGC  660

CCGCATCTGTCAAATAAATTCTTGTGAAACC                               691
```

Fig. 2. cDNA sequences of protegrins. The PG-1 sequence is shown on the main rows. The other protegrin sequences are shown only where they differ from PG-1, and are marked by an identifying superscript ([2] PG-2, [3] PG-3, [4] PG-4). Predicted mature peptide sequences are shown in boldface. The stop codon is indicated by dashes.

```
PG-1     METQRASLCL GRWSLWLLLL ALVVPSASAQ ALSYREAVLR AVDRLNEQSS EANLYRLLEL
PG-2     .......... .......... .......... .......... .......... ..........
PG-3     .......... .......... .......... .......... .......... ..........
PG-4     .......... .......... .......... .......... .......... ..........
PR-39    .......... .......... .......... .......... .......... ..........
Cathelin ---------- ---------- ---------- E.R....... .......... ..........
C-12     .......... .......... .......... .......... .......... ..........

PG-1     DQPPKADEDP GTPKPVSFTV KETVCPRPTR QPPELCDFKE NGRVKQCVGT VTLDQIKDPL
PG-2     .......... .......... .......... .......... .......... ..........
PG-3     .......... .......... .......... .......... .......... ..........
PG-4     .......... .......... .......... .......... .......... ..........
PR-39    .......... .......... .......... .......... .......... ...NPSIHS.
Cathelin .......... .......... .......... .......... ----...... ...NPSIHS.
C-12     .......... .......... .......... R......... .......... ..........

PG-1     DITCNEVQGV RGGRLCYCRR RFCVCVGRG
PG-2     .......... RGGRLCYCRR RFCICVG
PG-3     .......... RGGGLCYCRR RFCVCVGRG
PG-4     .......... RGGRLCYCRG WICFCVGRG
PR-39    ..S...I.S. RRRPRPPYLP RPRPPPFFPP RLPPRIPPGF PPRFPPRFPG KR
Cathelin ..S...I.S.
C-12     ......GVRR FPWWWPFLRR PRLRRQAFPP PNVPGPRFPP PNVPGPRFPP PNFPGPRFPP

C-12     PNFPGPRFPP PNFPGPPFPP PIFPGPWFPP PPPFRPPPFG PPRFPGRR
```

Fig. 3. Alignment of porcine leukocyte peptides. The sequences of protegrin PG-2 [6], protegrins PG-1, -3, -4 (this paper), PR-39 [3,4], cathelin [1,2] and clone C-12 [20] are shown. The sequences are compared to PG-1, with identities indicated by dots, differences by the substituted residue, and gaps by dashes. Mature peptides are shown in boldface.

^{32}P-labeled synthetic probe complementary to a sequence unique to protegrins revealed a ≈0.75 kb mRNA that was absent when pig alveolar macrophage RNA was probed (data not shown), suggesting the differential expression of protegrins by granulocytes and mononuclear phagocytes. As indicated in Fig. 3, the N-terminal prepro-regions of protegrins are 100% identical to each other (disregarding deletions) and show 93% and 88% identity to the corresponding regions in PR-39 and cathelin, respectively .

Overall, 89% (618/691) of the nucleotides in PG-1 cDNA are identical to the corresponding residues in the recently described clone C12 [20]. Fig. 4 aligns the cDNA 3′ to the respective stop codons of these porcine bone marrow peptides. Note that the bases occupying positions 517–691 in protegrin cDNA are 98.9% identical to those found between positions 630–804 in cDNA from clone C12. Protegrin cDNA also shows definite, but more limited 3′ homology with cDNA for bovine Bac5 (Table 1). In contrast to the considerable 5′ cDNA ho-

```
PG1      - ----------CGAGGATGACGGTTGCGACGGCAGGCTTTC -471
                     ||| ||  ||||||||||||||||||||||
C-12     - CCTGGTGGTGGCCGTTCCTACGAAGACCACGGTTGCGACGGCAGGCTTTC -446

PG1      - CCTCCCCCAATTTTCCCGGGGCCAGGTTTCCGTCCCCCAATTTTT----- -516
           |||||||||| | |||| |||||  | |||| ||||||||| | |
C-12     - CCTCCCCCAAATGTCCCCGGGCCCCGGTTCCCTCCCCCAAATGTCCCCGG -496

PG1      - ---------------------------------------------------- -516

C-12     - GCCCCGGTTCCCTCCCCCAAATTTCCCCGGGCCCCGGTTCCCTCCCCCAA -546

PG1      - ---------------------------------------------------- -516

C-12     - ATTTCCCCGGGCCCCGGTTCCCTCCCCCAAATTTCCCAGGGCCCCCGTTC -596

PG1      - ----------------------------------CCGCCTCCACCTTTCCG -533
                                              |||||||||||||||||
C-12     - CCTCCGCCAATTTTCCCTGGGCCGTGGTTCCCTCCGCCTCCACCTTTCCG -646

PG1      - GCCCGCACCATTCGGTCCACCAAGGTTCCCTGGTAGACGGTGAAGGATTT -583
             ||| ||||||||||||||||||||||||||||||||||||||||||||
C-12     - TCCCCCACCATTCGGTCCACCAAGGTTCCCTGGTAGACGGTGAAGGATTT -696

PG1      - GCAGGCAACTCACCCAGAAGGCCTTTCGGTACATTAAAATCCCAGCAAGG -633
           ||||||||||||||||||||||||||||||||||||||||||||||||||
C-12     - GCAGGCAACTCACCCAGAAGGCCTTTCGGTACATTAAAATCCCAGCAAGG -746

PG1      - AGACCTAAGCATCTGCTTTGCCCAGGCCCGCATCTGTCAAATAAATTCTT -683
           ||||||||||||||||||||||||||||||||||||||||||||||||||
C-12     - AGACCTAAGCATCTGCTTTGCCCAGGCCCGCATCTGTCAAATAAATTCTT -796

PG1      - GTGAAACC -691
           ||||||||
C-12     - GTGAAACC -804
```

Fig. 4. A comparison of the sequences 3′ to the stop codons of porcine PG-1 and clone C-12 [20]. The **TGA** signifies the stop codons.

Table 1
Homology of protegrins to other cathelin-associated antimicrobial peptides

Region	Peptide (species)	PG-1 (pig)	Clone C12 (pig)	PR-39 (pig)	Indolicidin (cattle)	Bac 5 (cattle)	Cyclic dodecapeptide (cattle)	CAP18 (rabbit)
5′	Base no.	1–336	3–338	18–353	13–348	15–350	13–348	1–336
	Identical	336/336	334/336	332/336	298/336	291/336	285/336	253/336
	%	100.0	99.4	98.8	88.7	86.7	84.8	75.3
3′	Base no.	517–691	630–804	638–658	529–548	528–661	553–579	No homology
	Identical	175/175	173/175	19/20	19/20	92/134	22/27	
	%	100.0	98.9	95.0	95.0	68.7	81.5	

The nucleotide sequences 5′ and 3′ to the mature protegrin PG-1 were matched with the comparable mRNA regions of the following antimicrobial peptides or their precursors: clone C12 [20], PR-39 [3,4], indolicidin [10,11], Bac5 [14,15], cyclic dodecapeptide [12,13] and CAP 18 [8]. Although all precursors showed considerable homology to PG-1 in their 5′ cathelin-containing regions, only clone C12 (prophenin) and, to a lesser extent, Bac 5, also showed extensive homology in the 3′ untranslated region.

mology between protegrins and these peptides, its 3′ homology to porcine PR-39, bovine indolicidin and bovine cyclic dodecapeptide is limited to the respective polyadenylation signal regions (Table 1).

We recently purified several peptides that we call 'prophenins', which proved to be mature products of the 'clone C-12' peptide family (Harwig et al., unpublished). Although we will describe their structure and properties in another communication, the striking 5′ and 3′ cDNA homology of protegrins and prophenins (e.g. clone C-12) cDNA, suggests that one of these peptide families probably originated by insertion of its exon into a gene originally specifying the other. Because clone C-12 is homologous to the protegrin cDNA, both 3′ *and* immediately 5′ to the bases encoding the protegrin sequence, and because protegrin cDNA retains with almost perfect fidelity the bases that encode the last 19 amino acids of the unprocessed prophenin peptide plus the prophenin stop codon (TGA), it is virtually certain that prophenin gene(s) existed first and the protegrin exon was the interloper. The almost complete conservation of the noncoding 3′ prophenin nucleotides in protegrin cDNA suggests that this insertional event occurred very recently, on an evolutionary time scale. Studies on the genomic level should further clarify the evolutionary origins and relationships of these intriguing peptide families.

Acknowledgements: We thank Tomas Ganz for valuable suggestions and assistance, and Marcelo Couto for helping to prepare the bone marrow cDNA library. This work was supported, in part, by a grant (AI 22839) from the National Institutes of Health.

References

[1] Ritonja, A., Kopitar, M., Jerala, R. and Turk, V. (1989) FEBS Lett. 255, 211–214.
[2] Lenarcic, B., Ritonja, A., Dolenc, I., Stoka, V., Berbic, S., Pungercar, J., Strukelj, B. and Turk, V. (1993) FEBS Lett. 336, 289–292.
[3] Agerberth, B., Lee, J.Y., Bergman, T., Carlquist, M., Boman, H.G., Mutt, V. and Jornvall, H. (1991) Eur. J. Biochem. 202, 849–854.
[4] Storici, P. and Zanetti, M. (1993) Biochem. Biophys. Res. Commun. 196, 1058–1065.
[5] 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.
[6] Storici P. and Zanetti, M. (1993) Biochem. Biophys. Res. Commun. 196, 1363–1368.
[7] Zanetti, M., Del Sal, P., Storici, P., Schneider, C. and Romeo, D. (1992) Biochem. Biophys. Res. Commun. 187: 467–472.
[8] Larrick, J.W., Morgan, J.G., Palings, I., Hirata, M. and Yen, M.H. (1991) Biochem. Biophys. Res. Commun. 179, 170–175.
[9] Levy, O., Weiss, J., Zarember, K., Ooi, C.E. and Elsbach, P. (1993) J. Biol. Chem. 268, 6058–6063.
[10] Selsted, M.E., Novotny, M.J., Morris, W.L., Tang, Y.Q., Smith, W. and Cullor, J.S. (1992) J. Biol. Chem. 267, 4292–4295.
[11] Zanetti, M., Del Sal, G., Storici, P., Schneider, C. and Romeo, D. (1993). J. Biol. Chem. 268, 522–526.
[12] Romeo, D., Skerlavaj, B., Bolognesi, M. and Gennaro, R. (1988) J. Biol. Chem. 263, 9573–9575.
[13] Storici, P., Del Sal, G., Schneider C. and Zanetti, M. (1992) FEBS Lett. 314, 187–190.
[14] Gennaro, R., Skerlavaj, B. and Romeo, D. (1989) Infect. Immun. 57, 3142–3146.
[15] Zanetti, M., Del Sal, G., Storici, P., Schneider, C. and Romeo, D. (1993) J. Biol. Chem. 268, 522–526.
[16] Chomczinski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[17] Mirgorodskaya, O.A., Schevchenko, A.A., Adballa, K.O., Chernushevich, I.V., Egorov, T.A., Musoliamov, A.X., Kokryakov, V.N. and Shamova, O.V. (1993) FEBS Lett. 330, 339–342.
[18] Bradbury, A.F., Finnie, M.D.A. and Smyth, D.G. (1982) Nature 298, 686–688.
[19] Merkler, D.J., Kulathila R., Consalvo, A.P., Yound, S.D. and Ash, D.E. (1992) Biochemistry 31, 7282–7288.
[20] Pungercar, J., Strukelj, B., Kopitar, G., Renko, M., Lenarcic, B., Gubensek, F. and Turk, V. (1993) FEBS Lett. 336, 284–288.