Volume 336, number 2, 284–288 FEBS 13415 December 1993

Molecular cloning of a putative homolog of proline/arginine-rich antibacterial peptides from porcine bone marrow

Jože Pungerčar*, Borut Štrukelj, Gregor Kopitar, Metka Renko, Brigita Lenarčič, Franc Gubenšek, Vito Turk

Department of Biochemistry and Molecular Biology, Jožef Stefan Institute, Jamova 39, PO Box 100, 61111 Ljubljana, Slovenia

Received 23 October 1993; revised version received 8 November 1993

Screening of a porcine bone marrow cDNA library with a PCR-derived probe from rabbit LPS-binding protein CAP18 led to the discovery of two closely related clones. The longer, full-length cDNA clone encodes a 228 amino acid residue protein similar to the family of antibacterial/LPS-binding cationic peptides. In contrast to other hitherto discovered precursors of Pro/Arg-rich peptides from this family, they have a novel, unique structure of the C-terminal region of 100 amino acid residues with a repeating sequence of ten residues (FPPPNXPGPR, where X = V or F). These precursors could represent a part of the antibacterial peptide repertoire of porcine bone marrow.

Antibacterial peptide; Bone marrow; Leukocyte; cDNA

1. INTRODUCTION

In the last decade, it has become obvious that antibacterial peptides are important constituents of a host-defense system widely distributed among living organisms [1]. Early reports on insect cecropins and mammalian defensins (for reviews see [2] and [3]) at the beginning of the 80s were followed by the discovery of a number of new types of antibacterial peptides. Recently, it has been demonstrated that in mammals, some different types of antibacterial peptides, such as LPS-binding protein CAP18 from rabbit granulocytes [4], indolicidin [5], bac5 [6], cyclic dodecapeptide [7], all from bovine neutrophils, and LPS-binding isoproteins p15 from rabbit polymorphonuclear leukocytes [8], are synthesized in bone marrow as larger precursor molecules which show a high level of amino acid identity in their N-terminal region. It has been suggested that all these proteins are members of a novel family of leukocyte proteins with antibacterial, LPS-binding and proteinase-inhibitory activities [8]. The N-terminal region is similar to the protein sequence of pig cathelin which had previously been isolated from pig blood leukocytes [9]. Therefore, it has been speculated that cathelin is the N-terminal fragment of a pig homolog of antibacterial peptides [8].

*Corresponding author. Fax: (386) (61) 273 594.

Abbreviations: cDNA, complementary DNA; kDa, kilodalton; LPS-binding protein, lipopolysaccharide-binding protein; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate-buffered saline; RT-PCR, reverse transcription-polymerase chain reaction; SDS, sodium dodecyl sulfate.

In order to find such a precursor in pig, a porcine bone marrow cDNA library was prepared and screened with a PCR-derived probe obtained by amplification of the propeptide region of a rabbit CAP18 cDNA. During cloning, we discovered that two of the positive clones encode a precursor, similar to cathelin, which has in its C-terminal part a novel 100 amino acid residue long polypeptide similar to proline/arginine-rich antibacterial peptides. In comparison with other known antibacterial peptides from this group, the putative homolog shows relatively little amino acid identity.

2. MATERIALS AND METHODS

2.1. *Materials*

Restriction endonucleases and DNA-modifying enzymes were purchased from Boehringer-Mannheim (Germany) or Pharmacia (Sweden). All other chemicals of analytical grade or higher were from Sigma (USA) and Serva (Germany). Oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesizer (USA), precipitated by sodium acetate/ethanol and purified by PAGE. [α-^{35}S]dCTP used for hybridization and nucleotide sequencing was obtained from Amersham (UK). The bacterial strain used as a host for bacteriophage λgt11 *E. coli* Y1090 *hsd*R was from Amersham, while the one used for cloning the plasmid pUC19 (Pharmacia) was *E. coli* DH5α from Gibco BRL (USA).

2.2. *Preparation of cDNA library*

Total RNA was isolated from the bone marrow of a 6-week old pig by the guanidinium thiocyanate/cesium trifluoroacetate method [10,11] and poly(A)$^+$ RNA purified by affinity chromatography on oligo(dT)-cellulose [12]. Complementary DNA was synthesized using a cDNA synthesis system of Amersham, fractionated by gel chromatography and size-enriched cDNA of more than 500 bp used for the construction of a cDNA library in λgt11 using a cloning kit from the same manufacturer.

Published by Elsevier Science Publishers B.V.

2.3. *Screening of cDNA library*

The cDNA library was screened with a ^{35}S-random primer-labeled [13] rabbit PCR probe of about 300 bp. In order to obtain the probe, total RNA was isolated from rabbit bone marrow and the first strand of cDNA synthesized with reverse transcriptase using an oligo(dT)$_{16}$ primer. The RT-PCR was carried out using a commercial kit and the instructions of Perkin-Elmer Cetus (USA). The single-stranded cDNA was amplified using the oligonucleotides 5′-GAC GAATTC TAC CGG GAG GCT GTG C-3′ and 5′-GAG AAGCTT CTC TTG GGC CCT GTT GC-3′ corresponding to the propeptide region of rabbit CAP18 [4]. After plating the cDNA library, plaques were transferred to a nylon membrane (Hybond-N, Amersham), DNA fixed by UV irradiation and hybridized in 6 × SSC, pH 7.0 (1 × SSC = 150 mM NaCl, 15 mM sodium citrate), 5 × Denhardt's solution [14], 50% deionized formamide, 0.5% SDS, 100 μg/ml denatured salmon sperm DNA and ^{35}S-labeled probe (8×10^8 cpm/μg) overnight at 42°C. The filters were subsequently washed twice, for 1 min each, in 2 × SSC/0.1% SDS and 1 × SSC/0.1% SDS at room temperature, thoroughly dried and exposed to Kodak X-Omat S film. Positive plaques were picked, replated and re-hybridized under the same conditions.

2.4. *Nucleotide sequencing and sequence analysis*

Bacteriophage DNA was isolated from positive clones using a standard procedure [14], cDNA excised with *Bam*HI and ligated into pUC19. Plasmid cDNA clones were sequenced by the dideoxy chain-termination method [15] using a T7-sequencing kit of Pharmacia with standard dNTP mixes and deaza G/A sequencing mixes. The final nucleotide sequence of both strands was obtained after subcloning and sequencing of the *Ava*I restriction fragments, and by using internal sequencing primers deduced from the previously determined sequence. Nucleotide and protein sequences were analyzed on a computer by DNASIS (Pharmacia) and PC/GENE (IntelliGenetics, USA).

2.5. *SDS-PAGE and immunoblotting*

SDS-PAGE was conducted on a 15% gel. Immunoblotting was performed on nitrocellulose membranes NC2 (Serva), non-specific binding was blocked with 5% skimmed milk in PBS. As primary antibodies, we used rabbit antiserum raised against pig cathelin and purified on protein A-Sepharose (Pharmacia). The bound antibodies were detected by alkaline phosphatase conjugated goat anti-rabbit IgG (Jackson Immunoresearch, USA). Controls with non-immune serum and competition assay with cathelin were also performed.

3. RESULTS

First screening of about 10^5 recombinant plaques with a rabbit probe yielded more than 100 positive clones. Ten of them were chosen for plaque purification. After isolation of recombinant λDNA, their cDNA inserts were subcloned into pUC19. Nucleotide sequencing revealed that in most of these clones, part of the 5′ end was missing. The longest cDNA clone C12 of 820 bp was shown to be complete. The 5′ untranslated end precedes the initiation codon by just two nucleotides. It contains an open reading frame of 684 bp coding for a 228 amino acid residue protein with a stop codon at position 687. In the 3′ untranslated end of 115 nucleotides, a putative signal for polyadenylation was found at position 786 and a polyadenylation site 18 nucleotides further downstream (Fig. 1). Although its overall G + C content is only about 63%, the cDNA sequence

```
CCATGGAGACCCAGAGGGCCAGCCTGTGCCTGGGGCGCTGGTCACTGTGGCTTCTGCTGCTGGCACTCGTGGTGCCCTCGGCCAGCGCC    89
  M  E  T  Q  R  A  S  L  C  L  G  R  W  S  L  W  L  L  L  L  A  L  V  V  P  S  A  S  A      29

CAGGCCCTCAGCTACAGGGAGGCCGTGCTTCGTGCTGTGGATCGCCTCAACGAGCAGTCCTCGGAAGCTAATCTCTACCGCCTCCTGGAG   179
 Q  A  L  S  Y  R  E  A  V  L  R  A  V  D  R  L  N  E  Q  S  S  E  A  N  L  Y  R  L  L  E      59

CTGGACCAGCCGCCCAAGGCCGACGAGGACCCGGGCACCCCGAAACCTGTGAGCTTCACGGTGAAGGAGACTGTGTGTCCCAGGCCGACC   269
 L  D  Q  P  P  K  A  D  E  D  P  G  T  P  K  P  V  S  F  T  V  K  E  T  V  C  P  R  P  T      89

CGGCGGCCCCCGGAGCTGTGTGACTTCAAGGAGAACGGGCGGGTGAAACAGTGTGTGGGGACAGTCACCCTGGATCAGATCAAGGACCCG   359
 R  R  P  P  E  L  C  D  F  K  E  N  G  R  V  K  Q  C  V  G  T  V  T  L  D  Q  I  K  D  P      119

CTCGACATCACCTGCAATGAGGGTGTCAGGAGATTTCCCTGGTGGTGGCCGTTCCTACGAAGACCACGGTTGCGACGGCAGGCTTTCCCT   449
 L  D  I  T  C  N  E  G  V  R  R  F  P  W  W  W  P  F  L  R  R  P  R  L  R  R  Q  A  F  P      149

CCCCCAAATGTCCCCGGGCCCCGGTTCCCTCCCCCAAATGTCCCCGGGCCCCGGTTCCCTCCCCCAAATTTCCCCGGGCCCCGGTTCCCT   539
 P  P  N  V  P  G  P  R  F  P  P  P  N  V  P  G  P  R  F  P  P  P  N  F  P  G  P  R  F  P      179

CCCCCAAATTTCCCCGGGCCCCGGTTCCCTCCCCCAAATTTCCCAGGGCCCCCGTTCCCTCCGCCAATTTTCCCTGGGCCGTGGTTCCCT   629
 P  P  N  F  P  G  P  R  F  P  P  P  N  F  P  G  P  P  F  P  P  P  I  F  P  G  P  W  F  P      209

CCGCCTCCACCTTTCCGTCCCCCACCATTCGGTCCACCAAGGTTCCCTGGTAGACGGTGAAGGATTTGCAGGCAACTCACCCAGAAGGCC   719
 P  P  P  P  F  R  P  P  P  F  G  P  P  R  F  P  G  R  R  *                                     228

TTTCGGTACATTAAAATCCCAGCAAGGAGACCTAAGCATCTGCTTTGCCCAGGCCCGCATCTGTCAAATAAATTCTTGTGAAACCAAAAA   809

AAAAAAAAAAA                                                                                    820
```

Fig. 1. The cDNA sequence of porcine C12 protein. A stop codon is represented by an asterisk. The presumed signal and Pro/Arg-rich peptides are underlined. These data have been submitted to the GenBank and have been assigned accession number X75438.

contains a few regions with a high content of G + C nucleotides which had to be resolved by using 7-deaza-dATP and 7-deaza-dGTP in sequencing mixes. One of the sequenced incomplete cDNAs of approximately 770 bp, clone C6, showed about 98% nucleotide identity with the C12 clone. Nucleotide exchanges responsible for the difference between the two clones are spread throughout the entire sequence. Most do not change the amino acid sequence. However, one of the rare non-silent mutations which led to an exchange of Val, present in C12, for Phe at position 163 is located in one of the interesting repeats in the C-terminal region, which are discussed below.

In order to inspect further possible differences between primary clones, we re-screened them using the C12 cDNA as a probe. The observed signals can be roughly divided into two groups on the basis of their intensity. The first group is closely related to the C12 clone (as for example clone C6); the second one, obviously more different in nucleotide sequence, has not been further studied so far.

The deduced protein sequence of the C12 clone shows a typical signal peptide of 30 residues at the N-terminus followed by 198 residues of a putative precursor protein. The precursor has a calculated molecular mass of about 22.7 kDa. No putative N-glycosylation sites could be found in the sequence, whereas an amidation site was predicted at Pro^{225}. Based on a hydropathy plot (data not shown) and comparison with other similar proteins (Fig. 2), its protein sequence can be divided into two halves.

The N-terminal 98 residue 'pro-region' is similar to cathelin and other known precursors of antibacterial peptides from leukocytes (Fig. 2). It contains all four conserved cysteines and has a net charge of −3. This region shows the highest level of amino acid identity to pig cathelin (84%) [9] followed by those of the corresponding regions of precursors of bovine indolicidin (78%) [5], bac5 (75%) [6], cyclic dodecapeptide (74%) [7], rabbit CAP18 (65%) [4] and of the sequence of rabbit p15 isoproteins (31%) [8]. Comparison also revealed that, as with other similar proteins, there is no gap of four amino acid residues in the region between the second and third Cys residue, such as that seen in pig cathelin [9]. Additionally, a gap of two residues was introduced just at the end of the 'pro-region' of C12 to obtain a better alignment.

The C-terminal half of 100 residues is very basic, is rich in Pro, Arg and Phe residues, and is unique in its structure. Searching was performed throughout several

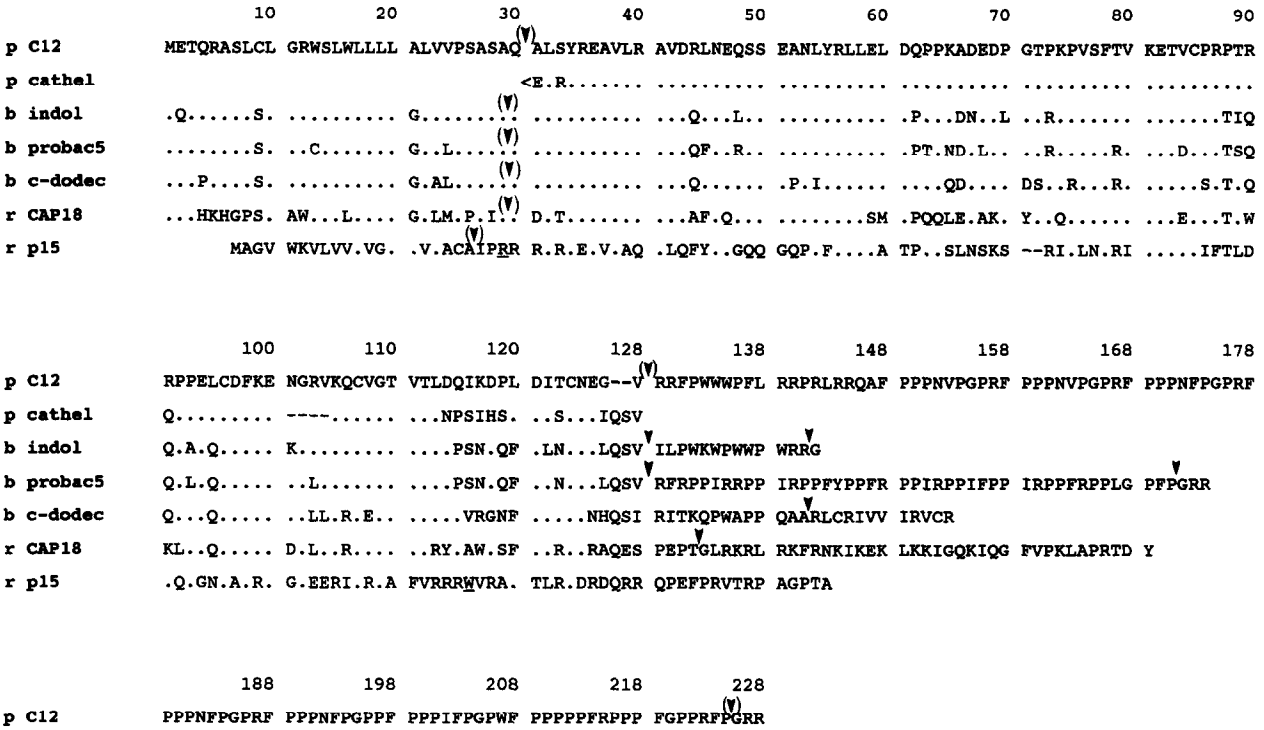

Fig. 2. Alignment of porcine C12 protein (p C12) with pig cathelin (p cathel), precursors of bovine indolicidin (b indol), bovine bac5 (b probac5), bovine cyclic dodecapeptide (b c-dodec), rabbit CAP18 (r CAP18) and rabbit p15 isoproteins (r p15). See the main text for references. Identical amino acid residues are indicated by a dot. The underlined residues represent variable amino acids in rabbit p15 isoforms. Gaps introduced to optimize the alignment are denoted by a dash. Arrowheads denote the known positions of a subsequent processing of precursor proteins. Arrowheads in parenthesis indicate the putative cleavage site for signal peptidase and putative processing sites in C12 protein. The blocked N-terminal residue of pig cathelin, pyroglutamic acid, is shown by <E. No attempts were made to align different sequences of the C-terminal region of these proteins. Numbering is according to porcine C12 protein.

nucleotide and protein databases. All attempts to find a reasonable alignment of this region with the sequences of other similar antibacterial peptides were unsuccessful. However, it looks most similar to the proline/arginine-rich group of cationic antibacterial peptides, such as bovine bactenectins bac5 [19,6] and bac7 [19], and porcine PR-39 [21]. The Pro/Arg region of C12 and C6 shows some similarities to other members of the group, but also several differences. These similarities are a high percentage of non-polar residues (71%) including Pro (45%) and Phe (16%) and a highly basic net charge (+15). On the other hand, it is much longer than all known Pro/Arg peptides, it includes three consecutive Trp residues at the beginning and it contains a lower percentage of Arg residues (15%) and a completely different repeating motif. The repeating sequence in C12 and C6 proteins, especially obvious in the middle part of the sequence, is a stretch of ten residues FPPPNXPGPR (X = V or F) tandemly repeating four times (Fig. 3).

Western blot analysis of SDS gel electrophoretic fractions using polyclonal antibodies against cathelin corresponding to the N-terminal half of the putative precursor protein revealed the presence of several immunochemically related proteins with molecular masses of approximately 14 kDa, which corresponds to the position of native cathelin under the same conditions, 16 kDa and a few higher molecular mass proteins with 23 kDa, 30 kDa and 31 kDa (Fig. 4). These higher molecular mass proteins are, however, not present in leukocytes isolated from peripheral blood.

4. DISCUSSION

Screening of our porcine bone marrow cDNA library with a rabbit homolog belonging to the novel family of antibacterial/LPS-binding peptides [4] indicate that

129	R	R	F	P	W	W	W	P											136
137			F	L	R	R	P	R	L	R	R	Q	A						147
148			F	P	P	P	N		V	P	G	P	R						157
158			F	P	P	P	N		V*	P	G	P	R						167
168			F	P	P	P	N		F	P	G	P	R						177
178			F	P	P	P	N		F	P	G	P	R						187
188			F	P	P	P	N		F	P	G	P	P						197
198			F	P	P	P	I		F	P	G	P	W						207
208			F	P	P	P	P	P	F	R	P	P	P						218
219			F	G	P	P	R		F	P	G	R	R						228

Fig. 3. Alignment of internal repeats in the Pro/Arg-rich C-terminal region of porcine C12 protein. Identical amino acid residues are boxed. Numbers indicate amino acid positions. In C6, Phe instead of Val residue is present at position 163 (shown by an asterisk).

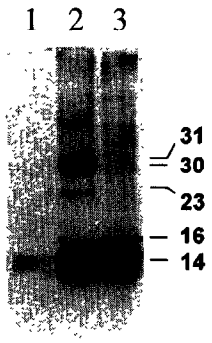

Fig. 4. Western blot analysis of proteins reacting with antibodies against cathelin from porcine bone marrow (lane 2) and from peripheral blood (lane 3). Numbers on the right indicate the approximate molecular mass of the proteins in kDa. For comparison, native cathelin is shown in lane 1.

these proteins are encoded by about 0.1% of the total population of mRNA molecules. Additionally, as shown by Western blot analysis, several similar proteins of different molecular masses are apparently also present in bone marrow cells lysate. One could speculate that at least some of the higher molecular mass proteins could represent precursor forms which are further processed in bone marrow probably during maturation of leukocytes. Another possibility would be that there is a group of proteins sharing similar epitopes with cathelin. It has been demonstrated that mature leukocytes isolated from peripheral blood no longer contain these higher molecular mass proteins. As was expected, the deduced protein sequences of two porcine cDNA clones (C12 and C6) share the highest degree of identity with pig cathelin followed by precursors of bovine indolicidin, bac5, cyclic dodecapeptide, rabbit CAP18 and p15 isoproteins [9,5–7,4,8].

The differences between C12, C6 and cathelin clearly indicate that they are the products of different genes. The protein sequence of pig cathelin [9] begins with a pyroglutamic acid. Additionally, Ala was determined as the first amino acid residue by protein sequencing of one of the cathelin isoforms (A. Ritonja, personal communication). Therefore, we suspect that the cleavage site for signal peptidase in the clones C12 and C6 is likely to be between Gln^{30}-Ala^{31}, although this position was predicted from the generally accepted rule [16,17] only as a putative minor cleavage site. The putative major cleavage site was recognized between Ala^{29}-Gln^{30}. However, the putative length of the precursor of the two proteins, beginning with Ala residue, is 198 residues.

The 98 amino acid residue long 'pro-region' of the C12 precursor ends with a valine, as in the case of cathelin [9] and precursors of bovine indolicidin [5] and bac5 [6]. It has been reported that these bovine inactive precursors are activated by cleavage with neutrophil elastase just after this valine [18]. If this is also the case for porcine C12 and C6 precursor proteins of 198 amino acids, their activation would result in a 98 residue

propeptide, which is similar to cathelin, and a 100 residue Pro/Arg-rich peptide. Similarly to bac5 [19,6], the C-terminus of this cationic peptide could be further processed by cleavage of the C-terminal tripeptide, Gly^{226}-Arg^{227}-Arg^{228}. The final mature Pro/Arg-rich peptide would thus end with a proline which may have an α-amide group. In other words, the C12 and probably also C6 protein are likely to be synthesized as preproproteins of 228 amino acids which could be processed to result finally in a 97 residue mature Pro/Arg peptide (see Fig. 2).

The N-terminal region of the C12 and C6 Pro/Arg-rich peptides contains three consecutive Trp residues. Multiple tryptophanes have also been reported for bovine tridecapeptide amide indolicidin and suspected to be important for its antibacterial activity [20]. One could also speculate that the C12 and C6 peptides are further processed into smaller antibacterial peptides, similar in size to indolicidin. So far, there is no evidence for this. The repeating sequence of ten amino acids FPPPNXPGPR, where X can be Val or Phe, tandemly repeated four times in the middle region of the C12 and C6 peptides, can also be found in the preceding and subsequent parts of the peptide, although no longer being perfect. Furthermore, almost every tenth or fifth residue is Phe, indicating that the original repeating unit might be a pentapeptide. It would be interesting to know how these repeats are encoded in the porcine genome. To our knowledge, such a structure of the peptides appears to be unique. Obviously, there are similarities with other proline/arginine-rich proteins from the family of antibacterial/LPS-binding peptides, such as bovine bac5 and bac7 [19,6], and an antibacterial peptide of 39 residues (PR-39) recently isolated from porcine intestine [21]. All these peptides are very basic due to the presence of Arg residues, with a high percentage of non-polar residues including Pro and Phe. In bac7, for example, a tridecamer sequence is tandemly repeated three times. In contrast, the short repeating motifs that usually include Pro, Arg and Phe (or some other hydrophobic residue) seem to be different in these cationic peptides. At the moment, it is difficult to predict whether their secondary and tertiary structures are similar. It has been reported that the mechanism of action of PR-39 [22] seems to be different from that of bactenectins [23], although the final result is the same, killing bacteria. In conclusion, as was expected, the porcine C12 and C6 proteins belong to the novel family of leukocyte proteins with antibacterial and LPS-binding activities sharing a high degree of amino acid identity in the N-terminal region. However, it remains to be experimentally verified if the two proteins, which are the longest of all known from the group, also possess such activities against bacteria as was established for other similar Pro/Arg-rich antibacterial peptides, which is very likely due to the unique structural motifs in the C-terminal region.

Acknowledgements: This work was supported by a grant from the Ministry for Science and Technology of Slovenia.

REFERENCES

[1] Boman, H. (1991) Cell 65, 205–207.
[2] Boman, H.G., Faye, I., Gudmundsson, G.H., Lee, J.-Y. and Lidholm, D.-A. (1991) Eur. J. Biochem. 201, 23–31.
[3] Lehrer, R.I. and Ganz, T. (1990) Blood 76, 2169–2181.
[4] Larrick, J.W., Morgan, J.G., Palings, I., Hirata, M. and Yen, M.H. (1991) Biochem. Biophys. Res. Commun. 179, 170–175.
[5] Del Sal., G., Storici, P., Schneider, C., Romeo, D. and Zanetti, M. (1992) Biochem. Biophys. Res. Commun. 187, 467–472.
[6] Zanetti, M., Del Sal, G., Storici, P., Schneider, C. and Romeo, D. (1993) J. Biol. Chem. 268, 522–526.
[7] Storici, P., Del Sal., G., Schneider, C. and Zanetti, M. (1992) FEBS Lett. 314, 187–190.
[8] Levy, O., Weiss, J., Zaremberger, K., Ooi, C.E. and Elsbach (1993) J. Biol. Chem. 268, 6058–6063.
[9] Ritonja, A., Kopitar, M., Jerala, R. and Turk, V. (1989) FEBS Lett. 255, 211–214.
[10] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[11] Mirkes, P.E. (1985) Anal. Biochem. 148, 376–383.
[12] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1409–1412.
[13] Feinberg, A.P. and Vogelstein, B. (1984) Anal. Biochem. 137, 266–267.
[14] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[15] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[16] Perlman, D. and Halvorson, H.O. (1983) J. Mol. Biol. 167, 391–409.
[17] Von Heine, G. (1983) Eur. J. Biochem. 133, 17–21.
[18] Scocchi, M., Skerlavaj, B., Romeo, D. and Gennaro, R. (1992) Eur. J. Biochem. 209, 589–595.
[19] Frank, R.W., Gennaro, R., Schneider, K., Przybylski, M. and Romeo, D. (1990) J. Biol. Chem. 265, 18871–18874.
[20] 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.
[21] Agerberth, B., Lee, J.-Y., Bergman, T., Carlquist, M., Boman, H.G., Mutt., V. and Joernvall, H. (1991) Eur. J. Biochem. 202, 849–854.
[22] Boman, H.G., Agerberth, B. and Boman, A. (1993) Infect. Immunol. 61, 2978–2984.
[23] Skerlavaj, B., Romeo, D. and Gennaro, R. (1990) Infect. Immunol. 58, 3724–3730.