FEBS Letters 412 (1997) 144–148 FEBS 18917

cDNA cloning of three cecropin-like antimicrobial peptides (Styelins) from the tunicate, *Styela clava*

Chengquan Zhao, Lilian Liaw, In Hee Lee, Robert I. Lehrer*

Department of Medicine, UCLA School of Medicine, Los Angeles, CA 90095-1690, USA

Received 2 June 1997

Abstract We cloned precursors of three new antimicrobial peptides, Styelins C, D and E, from a pharyngeal cDNA library of a tunicate, *Styela clava*. Preprostyelins resembled dipteran preprocecropins, while the mature domain of Styelin C resembled Cecropin P1, an antimicrobial peptide purified from the porcine intestine. Beginning with the last 6 residues of their signal sequences, Styelin C and Cecropin 1 from *Drosophila virilis* had 8/11 identical amino acids (72.7%). Moreover, 4 of the last 6 residues of their mature peptide domains were also identical. Styelins were shorter, by 8 residues, than dipteran cecropins and preprostyelins contained a conserved, polyanionic C-terminal extension that was absent in preprocecropins. Delineation of cecropin-like antimicrobial peptides in a protochordate supports the antiquity of this family as effectors of innate immunity in animals and it increases the likelihood that additional cecropin-like peptides will be found among other evolutionary descendants of protochordates — vertebrates.

Key words: Antimicrobial peptide; cDNA; Cecropin; Styelin; *Styela clava*; Tunicate

1. Introduction

Increasing evidence suggests that endogenous antimicrobial peptides play important roles in the host defense mechanisms of phagocytes [1] and epithelial cells [2] of mammals and are key effector molecules of innate immunity in animals [3]. Tunicates (sea squirts) are simple marine invertebrates that belong to the Phylum Chordata. As a bridge between invertebrates and vertebrates, tunicates afford opportunities to examine systems ancestral to those found in mammals. To date, we have identified two families of antimicrobial peptides in the hemocytes of *Styela clava*, Clavanins [4,5] and Styelins [6]. Clavanins are histidine-rich, C-terminally amidated, α-helical peptides that contain 23 amino acids, including (in some) a methylated tyrosine residue. Their structures and antimicrobial properties were recently described [4,5,7].

Styelins A and B are larger than Clavanins and also contain unusual amino acids, including hydroxylysine residues [6]. Because we had originally succeeded only in identifying 18 of the first 20 N-terminal amino acid residues of Styelins A and B [6], we tried to clone these peptides in order to obtain more complete structural information. Instead, we cloned precursors of three new members of the Styelin family, one of which (Styelin C) was highly homologous to Styelins A and B. Analysis of these Styelin precursors indicated that Styelins were related to insect Cecropins, well characterized antimicrobial peptides found in diptera (flies) and lepidoptera (moths and butterflies). As such, Styelins provide a previously missing link between the cecropins of invertebrates and Cecropin P1, a cecropin-like antimicrobial peptide previously isolated from the small intestine of the pig.

2. Materials and methods

2.1. RNA isolation and 3′-RACE analysis

Tunicate pharyngeal and hemocyte total RNAs were purified from freshly harvested, live *Styela clava* using a RNA separator kit (Clontech, Palo Alto, CA). First-strand cDNA synthesis and Styelin 3′-side cDNA amplifications were carried out with a 3′-RACE kit (Gibco BRL, Gaithersburg, MD). Total pharyngeal RNA (1 μg) and 1 μl of adapter primer (10 μg) were used to obtain the first-strand cDNA. A 'guess primer', that corresponded to amino acids 3–7 of Styelin A (GTCGGAATTCTTTGGAAAAGCTTTT, Phe–Gly–Lys–Ala–Phe) was synthesized. Its *Eco*RI restriction site is underlined. PCR was performed in a total volume of 50 μl with 10% of the cDNA synthesized above, 10 pmol of each AUAP and guess primers, and 5 U of pfu DNA polymerase. The reaction was run for 35 cycles in a GeneAmp PCR System 2400 (Perkin, Elmer), with the following cycle temperatures and times: 94°C, 20 s; 43°C, 20 s; and 68°C, 2 min. After electrophoresis on a 1.2% agarose gel, a 320 bp PCR product was obtained and purified.

2.2. Styela clava cDNA library screening

The 320 bp PCR product was used to screen a *Styela clava* pharyngeal cDNA library constructed in λ TripIEx™ by Clontech Laboratories, using *E. coli* strain XL1-Blue as a host. Phage DNA was transferred to nylon membrane filters (Dupont, Boston, MA), which was hybridized overnight at 48°C with the ^{32}P-labeled 320 bp PCR product in Rapid-hyb buffer (Amersham, Arlington Heights, IL). After several washes, finally at 60°C in 0.1×SSC and 0.1% SDS, the filters were exposed to X-ray films at −70°C with an intensifying screen. The positive clones were subjected to additional rounds of plaque screening at low density.

2.3. DNA amplification and sequencing

Positive phage plaques were picked to undergo PCR amplification, which was performed with LD-Insert Screening Amplimers (Clontech Lab, Palo Alto, CA) and Pfu DNA polymerase in a GeneAmp PCR system 2400. Amplified PCR products were sequenced directly by the fluorescein-labeled dideoxynucleotide terminator method, and analyzed on an Applied Biosystem 373A DNA Sequencer (Perkin-Elmer, Palo Alto, CA) at the UCLA DNA Sequencing Facility.

2.4. Northern blot analysis

Total RNA (10 μl) from tunicate pharyngeal tissues and hemocytes was isolated in a 1.2% formaldehyde agarose gel using RNA size marker I (Boehringer Mannheim, Germany). After the gel was washed, the RNA was transferred to a Gene Screen Plus nylon membrane (DuPont, Boston, MA) and the membrane was hybridized with ^{32}P-labeled synthesized oligonucleotide. The filter was washed at 60°C in 0.1×SSC, 0.1% SDS and then used for autoradiography.

3. Results and discussion

We recently purified two similar antimicrobial peptides,

*Corresponding author. Fax: +1 (310) 206-8766.
E-mail: rlehrer@med1.medsch.ucla.edu

PII S0014-5793(97)00769-2

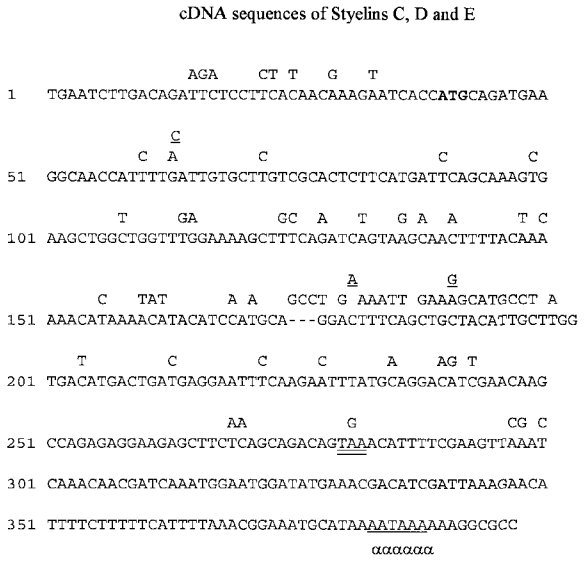

Fig. 1. cDNA sequences of Styelins C, D and E. The main sequence is that of Styelin C. Styelin D is shown above it, only where its sequence differs from that of Styelin C. Styelin E differed from Styelin D at only three bases, which are shown above Styelin D and underlined. The start codon (ATG) is bolded, and the stop codon (TAA) is doubly underlined. The polyadenylation signal is indicated by αααααα.

Styelins A and B, from hemocytes of the solitary tunicate, *S. clava* and identified 18 of their 20 N-terminal amino acids. We used this information to design a short 'guess primer', and performed 3′-RACE analyses at low annealing temperatures on *S. clava* pharyngeal tissue RNA. The resulting 320 bp PCR product was used to screen a *S. clava* pharyngeal cDNA library of approximately 1.0×10^5 clones. Thirteen apparently positive clones were identified and sequenced. Eight of these were identical and encoded the peptide designated Styelin C in Fig. 1. Another two clones encoded Styelins D and E, respectively. The remaining three clones contained unrelated sequences.

The cDNA sequences of Styelins C and D contained 397 and 385 bp, respectively, and had the same structure (Fig. 1). The Styelin E clone was shorter (292 bp), because (relative to Styelin D) it lacked nucleotides 1, 2 and 293–385. In Styelin C, the initiation codon was preceded by 39 5′-untranslated nucleotides, and it was followed by a single 240 bp open reading frame which encoded 80 amino acid residues. The open reading frame contained a putative signal sequence (residues 1–22) that was followed by the mature peptide, which ended with a glycine (residues 23–54) and was followed by a polyanionic extension (residues 55–80). The amino acid sequences of preproStyelins C, D and E are shown in Fig. 2. We assume that Gly^{54} participates in forming a C-terminal amide, and that the polyanionic extension is removed during processing.

Before proteolytic processing or other post-translational modifications, the 80 residue-Styelin C precursor would have a mass of 9248 Da. Styelins D and E have open reading frames of 243 bp that encode 81 amino acids, one more than Styelin C. Their respective masses are 9560 Da (Styelin D) and 9588 Da (Styelin E). The calculated isoelectric points of Styelin D (p*I* 9.12) and Styelin E (p*I* 9.54). are higher than that of Styelin C (p*I* 5.05) because their mature peptide domains contain 2–3 more cationic (lysine or arginine) residues.

The previously reported partial sequences of Styelins A and B [6] are also shown in Fig. 2. Styelins A and B are identical in at least 17 of their first 20 N-terminal amino acids (2 residues were unidentified in each). Sixteen of the 18 identified residues (88.9%) of Styelin A were identical in Styelin C, as were 15/18 (83.3%) of the identified residues in Styelin B. These data establish that Styelins A, B and C are homologues.

It can be seen in Fig. 1 that the cDNA sequences of Styelins C, D and E are highly conserved. Overall, Styelins C and D are 82% identical, and Styelins D and E are 99% identical. Even though the signal peptides of Styelins C and D have identical amino acid sequences, at the nucleotide level, their 5′-side untranslated part and signal sequences differ by several base pairs. The 41% difference between the sequences encoding the mature Styelin C and D domain accounts for their many amino acid differences, which are especially prominent in the C-terminal portions of the mature peptides. The polyanionic C-terminal extensions were more highly conserved, showing only 10/78 (12.8%) base substitutions. The stop codons of Styelin C (TAA) and Styelins D and E (TGA) were different, but their 3′-side untranslated sequences were highly conserved, differing in only 3 bp. Although our Styelin E clone lacked 91 bp in its 3′-side untranslated part, the remaining sequences of Styelins D and E contained only 3 bp differences, one in the signal sequence, and two in the mature peptide domain. This suggests that the five presently identified Styelins have diverged into two branches. One of these contains Styelins A, B and C and the other contains Styelins D and E.

Fig. 3 shows the intriguing similarities between the mature domains of Styelins A, B and C and Cecropin P1, which was isolated from pig intestine [8]. If Styelin C is amidated, then both peptides would contain 31 amino acid residues and 9 (52.9%) of their first 17 N-terminal residues would be identical or similar. Unfortunately, Cecropin P1 has not yet been

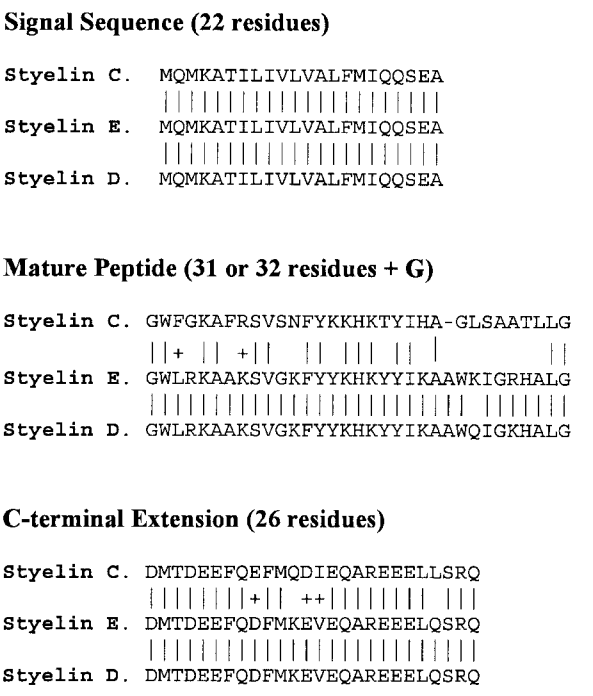

Fig. 2. Primary structures of Styelin precursors. The domains are shown separately.

Stylins A-C and Cecropin P1 (*Sus scrofa domesticus*)

```
Stylin A      GXFGλ AFXSV SNFAλ λHλTA
Stylin B      GXFGP AFHSV SNFAλ λHλTA
Stylin C      GWFGK AFRSV SNFYK KHKTY IHAGL SAATL L
Cecropin P1   SWLSK TAKKL ENSAK KRISE GIAIA IQGGP R
```

Stylin C and Cecropin Dv1 (*Drosophila virilis*)

```
Stylin C      IQQSEAGWFGKAFRSVSNFYKKHKTYIH--------AGLSAATLLG
              + ||||||+ |  + +     + ++           |   |||  |
Cecropin Dv1  LGQSEAGWLKKIGKKIERIGQHTRDATIQGLGIAQQAANVAATARG
```

Fig. 3. Comparison of Stylins and Cecropins. The top series compares the sequences of Stylins A and B [6] and Stylin C to Cecropin P1 [8], which was isolated from porcine intestines. Identical residues are bolded and similar residues are underlined. λ signifies hydroxylysine, a post-translational modification of lysine. Below it, Stylin C has been aligned with Cecropin 1 from *D. virilis* [10]. A single 8 residue gap has been introduced for alignment purposes. Identical residues are bolded and similarities are indicated by plus signs.
←

cloned, so we could not compare its precursor to those of Stylins C–E.

However, when we compared the precursors of Stylins and insect Cecropins, many additional similarities were noted (Fig. 4). Insect cecropins vary in size from 35 to 39 residues and are typically amidated. The mature (i.e. processed) peptides have a strongly basic NH_2-terminal region and a more hydrophobic C-terminal region. Insect preprocecropins are of two types: dipteran and lepidoteran. Dipteran precursors, exemplified by *Drosophila melanogaster* and *D. virilis*, have a 23

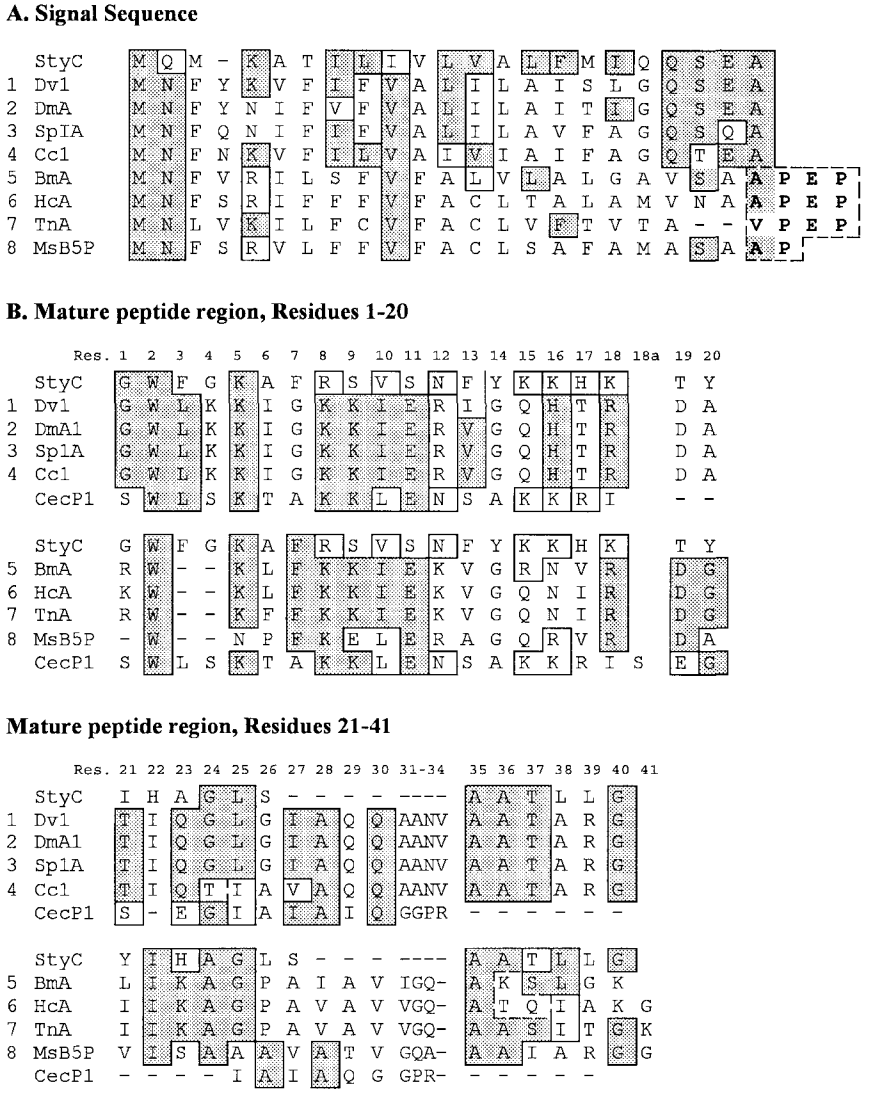

Fig. 4. Comparison of Styelin C, Cecropin P1 and insect cecropins. Sequences 1–4 are of representative dipteran cecropins. Sequences 5–8 are of representative lepidopteran cecropins. References for the cecropin sequences are shown in brackets. StyC, Styelin C from *S. clava*; CecP1[8], Cecropin P1 from the domestic pig, *Sus scrofa domestica*; Dv1, Cecropin I from *D. virilis* [10]; DmA1, Cecropin A1 from the fruitfly, *D. melanogaster* [13]; SpIA, Cecropin (Sarcotoxin) 1A from the fleshfly, *Sarcophaga peregrina* [14]; Cc1, Cecropin 1from the 'medfly', *Ceratitis capitata* [15]; BmA, Cecropin A from the silkworm, *Bombyx mori* [16]; HcA, Cecropin A from the cecropia moth, *Hyalophora cecropia* [17]; TnA, Cecropin A from the moth, *Trichoplusia ni* [18]; MsB5, Cecropin B5 from the tobacco hornworm, *Manduca sexta* [19]. Residues of Styelin C or Cecropin P1 and the identical residue in an insect cecropin are boxed and shaded. Residues of Styelin C or Cecropin P1 that are either a conservative substitutions for the typical insect cecropin residue, or are identical to each other are boxed but unshaded. The short propiece residues of lepidopteran cecropins are bolded (all but MsB5 with experimental support).

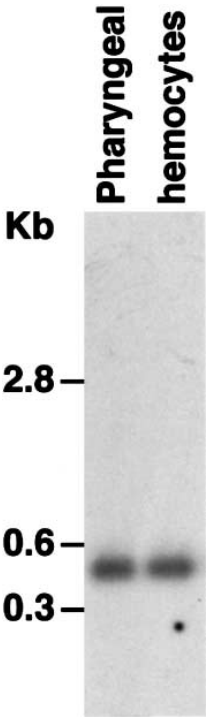

Fig. 5. Northern analysis. Total RNA was prepared from pharyngeal tissue and hemocytes of *S. clava* and probed with ^{32}P-labeled antisense nucleotides complimentary to nt 161–191 of Styelin C cDNA. Both tissues contained hybridizing mRNA species of ≈0.5 kb. Size markers are shown on the left.

residue signal sequence that ends with a conserved tetrapeptide motif (Gln–Ser/Thr–Glu/Gln–Ala) that is followed immediately by a ≈40 residue sequence which encodes the mature peptide. The mature *Drosophila* cecropin domains typically begin with Gly–Trp–Leu–Lys–Lys and end with a conserved 13 residue motif, Ala–Gln–Gln–Ala–Ala–Asn–Val–Ala–Ala–Thr–Ala–Arg–Gly. The mature domains of procecropins of the fleshfly, *Sarcophaga peregrina*, and the Mediterranean fruitfly (or 'medfly'), *Ceratitis capitata*, also conform to this pattern.

When the precursors of *D. virilis* Cecropin 1 and Styelin C were compared (Fig. 3), 12/38 (31.6%) of the residues were identical and 6/38 (15.8%) were similar. Because the mature *D. virilis* Cecropin 1 peptide contained 40 residues (including the C-terminal glycine), a single 8 residue gap was introduced for alignment purposes into the Styelin C sequence shown in Figs. 3 and 4.

Lepidopteran cecropins have a shortened signal sequence that does not end with the aforementioned tetrapeptide motif and is followd by a short $(X\text{-pro})_{1-2}$ propeptide sequence that is absent in prepro-dipteran cecropins and preproStyelins. In *H. cecropia*, the 4 residue prosequence is removed stepwise by a dipeptidyl peptidase [9]. Lepidopteran preprocecropins have quite variable C-termini, which do not necessarily end with a glycine. Evolutionary aspects of lepidopteran and dipteran cecropins were comprehensively reviewed from an evolutionary standpoint by Zhou et al. [10].

From all of these considerations, we conclude that Styelin precursors conform more closely to dipteran cecropins than to those of lepidoptera. The similarities include the conserved terminal signal sequence tetrapeptide, identity of 3 or 4 of the first 5 residues of the mature cecropin peptide, including the invariant tryptophan at position 2. Although residue 3 of Styelin C is phenylalanine, and not leucine as typically found in dipteran cecropins, Styelins D and E contain leucine in this position. Like dipteran cecropins, and unlike those of lepidoptera, the mature Styelin sequences directly follow the signal peptide without interposition of a 2 or 4 residue propiece. PreproStyelins also contain echoes of the conserved C-terminal motif of dipteran cecropins in that 5/11 (45.5%) of the C-terminal residues of Styelin C are identical to those found in dipteran propeptides.

While there are remarkable similarities between tunicate Styelins and dipteran cecropins, some significant differences exist. Styelins have only 31 or 32 residues, whereas cecropins have 35–39. Styelins, but not cecropins, contain a central cationic tetrapeptide domain (KKHK) and they lack the 2 acidic residues (Asp and Glu) present in almost all cecropins. Finally, their precursors end with a remarkable, 26 residue C-terminal extension that contains 10 acidic residues (7–8 glutamics and 2–3 aspartics), which would provide charge balance for the highly cationic styelin domains, which contain 7 (Styelin C), 10 (Styelin D) and 11 (Styelin E) positively-charged residues (Lys+Arg+His) without acidic residues.

Overall, perhaps, the major differences between dipteran cecropins and tunicate styelins are the loss (or gain) of 8 sequential amino acid residues and the loss (or gain) of a polyanionic C-terminal extension that follows the domain containing the mature peptide. Deciding whether tunicates or insects have been the losers or gainers in these respects must await the discovery of cecropin-like peptides in a species, perhaps a protist, that can claim common ancestry to both tunicates and insects.

Although we identified precursors of three new Styelins in these studies, the ones we originally sought — precursors of Styelins A and B — were not found. Perhaps this resulted from our having prepared the cDNA library from pharyngeal tissue rather than from hemocytes, even though hemocyte Clavanins were successfully cloned from the same library [5]. A 31 nucleotide antisense probe, complementary to nucleotides 161–191 of the Styelin C cDNA sequence recognized a message of about 0.5 kb in both pharyngeal tissue and hemocytes found by Northern analysis (Fig. 5), so that in the future, when these studies are performed with a hemocyte cDNA library, the 'missing' Styelin clones may be found. Tissue specific expression of different members of an antimicrobial peptide families is well recognized for defensins and β-defensins [11] and for cecropins [10,12].

Acknowledgements: This research was supported, in part, by a grant (AI 22839) from the National Institutes of Health.

References

[1] Martin, E., Ganz, T. and Lehrer, R.I. (1995) J Leuk Biol 58, 128–136.
[2] Bevins, C.L. (1994) Ciba Foundation Symp 186, 250–269.
[3] Boman, H.G. (1995) Ann Rev Immunol 13, 61–92.
[4] Lee, I.H., Zhao, C., Cho, Y., Harwig, S.S.L., Cooper, E.L. and Lehrer, R.I. (1997) FEBS Lett 400, 158–162.
[5] C. Zhao, L. Liaw, I.H. Lee, and R.I. Lehrer, 1997 submitted.
[6] I.H. Lee, Y. Cho, and R.I. Lehrer, Comp. Biochem. Physiol B, 1997 in press.
[7] I.H. Lee, Y. Cho, and R.I. Lehrer, Infect Immun, 1997 in press.
[8] Lee, J.Y., Boman, A., Sun, C.X., Andersson, M., Jornvall, H.,

Mutt, V. and Boman, H.G. (1989) Proc Natl Acad Sci USA 86, 9159–9162.
[9] Boman, H.G., Boman, I.A., Andreu, D., Li, Z.Q., Merrifield, R.B., Schlenstedt, G. and Zimmermann, R. (1989) J Biol Chem 264, 5852–5860.
[10] Zhou, X., Nguyen, T. and Kimbrell, D.A. (1997) J Mol Evol 44, 272–281.
[11] Zhao, C., Wang, I. and Lehrer, R.I. (1996) FEBS Lett 396, 319–322.
[12] Samakovlis, C., Kylsten, P., Kimbrell, D.A., Engstrom, A. and Hultmark, D. (1991) EMBO J 10, 163–169.
[13] Kylsten, P., Samakovlis, C. and Hultmark, D. (1990) EMBO J 9, 217–224.
[14] Iwai, H., Nakajima, Y., Natori, S., Arata, Y. and Shimada, I. (1993) Eur J Biochem 217, 639–644.
[15] Rosetto, M., Manetti, A.G., Marchini, D., Dallai, R., Telford, J.L. and Baldari, C.T. (1993) Gene 134, 241–243.
[16] Morishima, I., Suginaka, S., Ueno, T. and Hirano, H. (1990) Comp Biochem Physiol 95B, 551–554.
[17] Gudmundsson, G.H., Lidholm, D.A., Asling, B., Gan, R. and Boman, H.G. (1991) J Biol Chem 266, 11510–11517.
[18] Kang, D., Liu, G., Gunne, H. and Steiner, H. (1996) Insect Biochem Mol Biol 26, 177–184.
[19] Dickinson, L., Russell, V. and Dunn, P.E. (1988) J Biol Chem 263, 19424–19429.