Structure of genes for dermaseptins B, antimicrobial peptides from frog skin

Exon 1-encoded prepropeptide is conserved in genes for peptides of highly different structures and activities

Véronique Vouille, Mohamed Amiche*, Pierre Nicolas

Laboratoire de Bioactivation des Peptides, Institut Jacques Monod, Université Paris 7, 2 Place Jussieu, 75251 Paris Cedex 01, France

Received 2 June 1997; revised version received 17 July 1997

Abstract **We cloned the genes of two members of the dermaseptin family, broad-spectrum antimicrobial peptides isolated from the skin of the arboreal frog *Phyllomedusa bicolor*. The dermaseptin gene *Drg2* has a 2-exon coding structure interrupted by a small 137-bp intron, wherein exon 1 encoded a 22-residue hydrophobic signal peptide and the first three amino acids of the acidic propiece; exon 2 contained the 18 additional acidic residues of the propiece plus a typical prohormone processing signal Lys–Arg and a 32-residue dermaseptin progenitor sequence. The dermaseptin genes *Drg2* and *Drg1g2* have conserved sequences at both untranslated ends and in the first and second coding exons. In contrast, *Drg1g2* comprises a third coding exon for a short version of the acidic propiece and a second dermaseptin progenitor sequence. Structural conservation between the two genes suggests that *Drg1g2* arose recently from an ancestral *Drg2*-like gene through amplification of part of the second coding exon and 3′-untranslated region. Analysis of the cDNAs coding precursors for several frog skin peptides of highly different structures and activities demonstrates that the signal peptides and part of the acidic propieces are encoded by conserved nucleotides encompassed by the first coding exon of the dermaseptin genes. The organization of the genes that belong to this family, with the signal peptide and the progenitor sequence on separate exons, permits strikingly different peptides to be directed into the secretory pathway. The recruitment of such a homologous 'secretory' exon by otherwise non-homologous genes may have been an early event in the evolution of amphibian.**

Key words: Antimicrobial peptide; Frog skin; Gene structure; Dermaseptin; Opioid peptide

1. Introduction

In recent years, the skin secretions of the arboreal frogs *Phyllomedusa sauvagei* and *P. bicolor* have become a pivotal model for the discovery of new mammalian-like hormones and neuropeptides [1], including dermorphin, dermenkephalin and deltorphins, D-amino acid-containing heptapeptides which exhibit exceptional affinity and selectivity for the μ- and δ-opioid receptors, respectively [2]. The granular glands of the skin also contain several α-helical cationic (lysine-rich) peptides, 24–34 residues long, termed dermaseptins (Fig. 1). These peptides are broadly microbicidal, killing Gram-positive and Gram-negative bacteria as well as yeast, protozoan and fungi at the micromolar level [3,4]. Other frog species produce distinct repertoires of antimicrobial peptides whose members differ from dermaseptins with respect to their chain length, hydrophobicity, charge distribution and spectrum of activity (reviewed in [5]). All these peptides act completely unlike traditional antibiotic agents by permeating the microbial membrane and impairing ability to carry out anabolic processes. Therefore, antimicrobial peptides from frog skin are of potential clinical interest since they are unlikely to induce the acquisition of antibiotic resistance genes by bacterial pathogens.

Although much is known about the function and structure of dermaseptins and related peptides [5], little is known about them at the gene level. We recently reported on the isolation and characterization of cDNAs encoding precursors of dermaseptin B1 and dermaseptin B2 [6,7] and showed that preprodermaseptins, preprodermorphin and preprodeltorphins all contain an almost identical amino-terminal domain including a 22-residue signal peptide and a 20–23-residue acidic propiece. The extensive similarities between the prepro regions of precursors encoding end products with very different structures and biological activities led to the suggestion that the genes encoding these peptides are all members of a same family. We report here the cloning and sequencing of two genes for dermaseptins which represent the first sequence information from *P. bicolor* genome and a comparison of the dermaseptin gene sequences with cDNAs coding precursors for several frog skin peptides originating from various amphibian species.

2. Materials and methods

2.1. Genomic library construction

High molecular weight genomic DNA was isolated from *P. bicolor* blood as described [8] and partially digested with Sau3A (Bethesda Research Laboratories). The fragments produced by this partial digestion were separated on a 10–40% sucrose velocity gradient. Fragments containing the 10–20-kb fragments were pooled and dialyzed overnight against 4 l of TE buffer to remove sucrose. After ethanol precipitation, the sized DNAs were partially filled in using dATP and dGTP to create ends that are compatible with those of 1 Fix II supplied in the Lambda Fix II/*Xho*I partial fill-in vector kit (Stratagene). The ligated DNAs were packaged in vitro using the protocol and reagents supplied with Gigapack II Gold packaging system (Stratagene). At this step, two different experiments were carried out: first, 132 000 clones arising from a single encapsidation reaction were

*Corresponding author. Fax: (44) 27-59-94.
E-mail: amiche@infobiogen.fr

Abbreviations: bp, base pairs; nt, nucleotide; DRS, dermaseptin; cDNA, complementary DNA; kb, kilobases; TATA, thymine–adenine–thymine–adenine; UTR, untranslated region; SSC, standard sodium citrate; SDS, sodium dodecyl sulfate

PII S0014-5793(97)00972-1

screened (see below); second, 1.4×10^6 clones from five encapsidation reactions achieved with the same ligation reaction were plated on NZYCM plates to perform amplification of the library. The plates were overlaid with 10 ml of SM buffer and incubated overnight under gentle rocking at 4°C. After incubation, the bacteriophage suspension was recovered from each plate and pooled in a sterile container. After adding chloroform (5% final concentration), and incubating the mixture for 15 min at room temperature, the cell debris were removed by centrifugation at $2000\times g$ for 10 min. The recovered supernatant was transferred into a sterile glass bottle and 2 ml aliquots in 7% DMSO were stored at −80°C.

2.2. Genomic library screening

A total of 2×10^6 clones from the amplified library were screened on duplicate Nytran-Plus filters (Schleicher and Schuell) with ^{32}P-labeled cDNA coding for preprodermaseptin B2 [6]. Prehybridization was carried out for 3 h at 65°C in 0.5 M sodium phosphate buffer, pH 7.2, containing 1 mM EDTA, 1% bovine serum albumin and 7% SDS. Hybridization was achieved overnight at 65°C with gentle shaking in 20 ml of the same buffer, added with 2.5×10^6 cpm/ml of ^{32}P-labeled cDNA. Filters were then washed at 65°C in 2×SSC containing 0.5% SDS for 10 min, then in 2×SSC with 0.1% SDS for 2×20 min, then in 0.2×SSC plus 0.1% SDS for 2×20 min and exposed to Kodak Biomax AR film overnight with an intensifying screen at −80°C. Positive clones were subjected to two additional rounds of plaque purification at low density.

2.3. Subcloning and sequencing

Phage DNAs, prepared from positive plaques according to Sambrook et al. [8] were digested with various restriction endonucleases (Bethesda Research Laboratories). The resulting fragments were characterized by Southern blotting [9] using the same protocol as in the screening for prehybridization and hybridization. Based on this analysis, a 2.9-kb *Sal*I–*Eco*RI fragment, a 2.3-kb *Not*I fragment and a 4.5-kb *Eco*RI–*Bam*HI fragment were subcloned into BlueScript pSK-vector, leading respectively to pSK3, pSK9 and pSKJ plasmids. Both strands of pSK3, pSK9 and pSKJ DNAs were sequenced by the dideoxy method [10] using the Thermo Sequenase radiolabeled terminator sequencing kit (Amersham Life Science), pBlueScript universal primers and various gene-specific oligomer primers based on dermaseptin B1 and B2 cDNAs. The cycling termination reactions were carried out by cycling 15 times, each with 30 s of denaturation at 95°C, 30 s of primer annealing at primer's T_m minus 2°C and 1 min of primer extension at 72°C. Computer analysis and alignments of the DNA sequences were performed using Clustal W [11].

3. Results and discussion

Twelve DNA genomic clones in the amplified genomic library were found that strongly hybridized to the complete sequence of dermaseptin B2 cDNA. To determine whether these clones represented different dermaseptin genes, or alleles, they were digested with *Bam*H1, *Not*I or *Eco*R1 and subjected to Southern blot analysis. All of the positive clones gave a single and distinct band, ranging in size from 1.9 to 12 kb, with the three enzymes, alone or in combination (not shown). Since there are no *Bam*H1, *Not*I or *Eco*R1 site in the cDNA and known intronic gene sequences (see below), the observed patterns indicate, according to sequencing of PCR phage DNA (unpublished results) that at least seven related dermaseptin genes, or pseudogenes, exist in an individual frog. They further demonstrate that these genes are not densely clustered since a unique hybridizing band was ever observed in each 14–20-kb phage insert. This multigenes organization provides the frog with a large gene dosage that contributes to a high abundance of the dermaseptins in the skin. Two of these clones which exhibited very strong hybridization signals were subcloned into pBlueScript SK-vector to yield recombinant plasmids pSK9 and pSKJ. The latter exhibited 95% nucleotide identity with pSK3, a genomic clone originating from an unamplified DNA library made from another specimen of *P. bicolor*.

A

```
                    .          10         20         30
Dermaseptin B2      GLWSKIKEVGKEAAKAAAKAAGKAALGAVSEAV(NH2)
Dermaseptin Drg1    GLWSNIKTAGKEAAKAALKAAGKAALGAVTDAV(NH2)
Dermaseptin Drg2    GLWSKIKEAG----KAVLTAAGKAALGAVSDAV(NH2)
                    **** **  *    **   **********  **
```

B

```
Temporin-H          LSPNLLKSLL(NH2)
Gaegurin-5          FLGALFKVASKVLPSVFCAITKKC
Brevinin-2E         GIMDTLKNLAKTAGKGALQSLVKMASCKLSGQC
Esculentin-1        GIFSKLAGKKLKNLLISGLKNVGKEVGMDVVRTGIDIAGCKIKGEC
```

Fig. 1. A: Amino acid sequences of dermaseptins B. The sequence of dermaseptin B2 (also termed adenoregulin [26]) was determined from the mature peptide [4,26] and cDNA cloning [6]. Those of DRg1 and DRg2 were inferred from the gene sequences reported here. The amino acids that are identical in the sequences are indicated by an asterisk. Gaps (−) have been introduced to maximize sequence similarities. B: Amino acid sequences of representative members of the brevinin (brevinin-2E), esculentin (esculentin-1), temporin (temporin H) and gaegurin (gaegurin-5) families of antimicrobial peptides [15–20]. Cys residues that form a disulfide bridge are underlined.

3.1. Structure of dermaseptin gene Drg2

The nucleotide and deduced amino acid sequences of dermaseptin gene *Drg2* present in the insert of the recombinant plasmid pSK9 are shown in Fig. 2. The gene is small (984 bp[1]), containing a 297-bp 5′- and 317-bp 3′-untranslated regions and a 368-bp protein-coding region. A comparison of the preprodermaseptin B2 cDNA [7] with the corresponding genomic sequence revealed that the coding region of the gene consisted of two exons, ranging from 74 (exon 1) to 157 bp (exon 2), interrupted by a small A+T-rich (66%) intron of 137 bp. The intron–exon splice sites exhibited excellent adherence to 5′- and 3′-splice site consensus rule [12]. The first exon also contained the short 30-bp 5′-untranslated region and codons for the first 25 amino acids of the preprodermaseptin, comprising a 22-residue signal peptide and the first three glutamic residues of the acidic spacer peptide. Exon 2 contained the sequence coding the remaining of the acidic propiece plus a typical prohormone processing signal Lys–Arg and a 29-residue dermaseptin sequence which is immediately followed by an amidation consensus sequence Gly–Glu–Gln, a stop codon and a 3′-untranslated region.

The mRNA sequence as deduced from the genomic DNA is slightly different (90% identity) from that obtained from previously characterized dermaseptin B2 cDNA [6]. They differ in only 14 punctual nucleotide substitutions within their coding sequences, eight of which occurring at third base positions of codons and preserve the amino acid sequence, and in a short stretch of 12 nt that were absent from the gene sequence. Accordingly, the predicted amino acid sequence corresponding to the signal peptide, the acidic propiece and the di-basic endoproteolytic cleavage site encoded in the gene were identical to that deduced from dermaseptin B2 mRNA. The dermaseptin progenitor sequence encoded in the genomic structure is almost identical to that of the cDNA differing by nine nucleotides which cause a Ala^9 for Val, Leu^{18} for Ala, Thr^{19} for Lys and Asp^{31} for Glu substitutions in the dermaseptin B2 sequence (Fig. 1). They also differ by a deletion of 12 nucleotides in the gene sequence corresponding to Lys^{11}–Glu–Ala–Ala^{14} of the dermaseptin B2 sequence. Whether these differ-

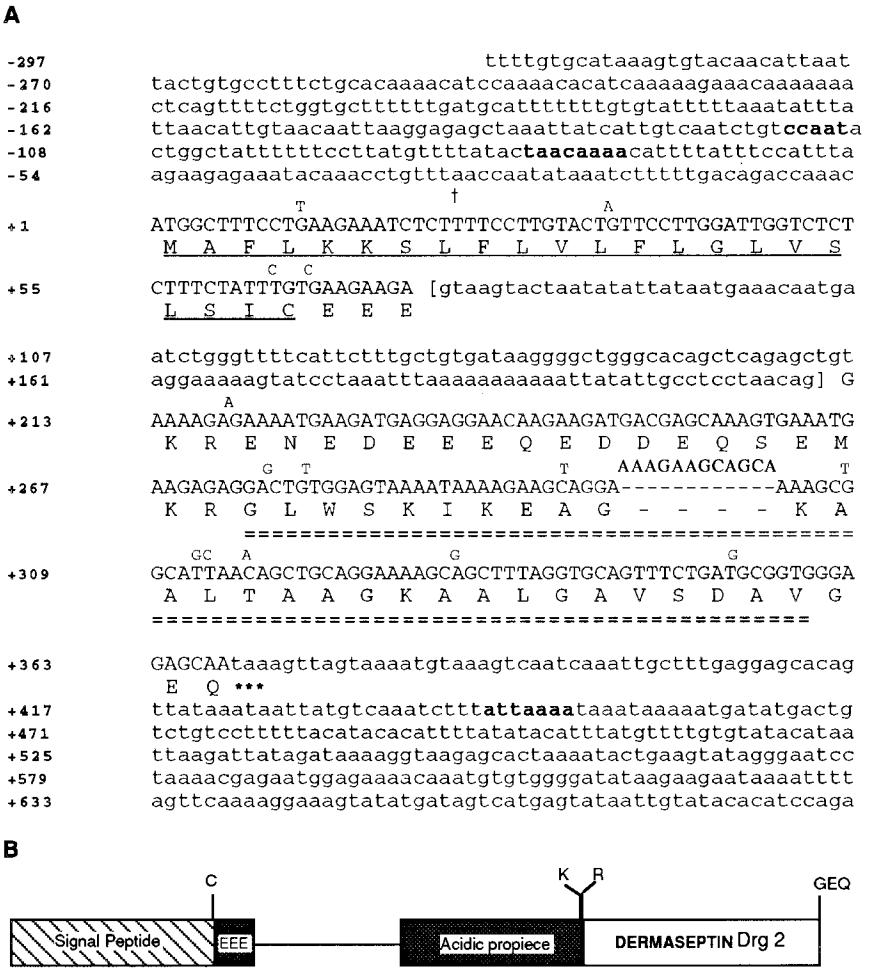


Fig. 2. The dermaseptin *Drg2* gene. A: Nucleotide sequences of coding exons are indicated by capital letters. The intronic sequence (in parentheses), 5′- and 3′-untranslated sequences are in lowercase letters. The transcription start site is tentatively indicated by a dagger. Nucleotides are numbered relative to the first nucleotide of the ATG start codon. The putative TATA box, CAAT box and polyadenylation signal are in bold letters. Predicted mature preprodermaseptin sequence is shown under the coding DNA. The predicted signal peptide is underlined and mature Drg2 dermaseptin sequence is double underlined. The stop codon is indicated by asterisks. Nucleotide coding sequence of dermaseptin B2 cDNA is indicated on the top row only when it differs from that of the gene. Deletions are shown by dashes (1/nucleotide). B: Schematic representation of the *Drg2* gene. Coding exons are drawn as rectangle, the intron as a thick line.

ences represented allelic or polymorphic variants remain to be elucidated.

Two canonical polyadenylation sites were identified 71 (ATTAAA) and 184 (AATAAA) bp downstream of the TAA stop codon in the 3′-flanking region of the gene (Fig. 2). Comparison with the 86-nt 3′-sequence of dermaseptin B2 cDNA containing a poly(A) tail demonstrated that the first site was used for poly(A) tailing. It is followed after ≈20 bp by GT- and TT-rich sequences present in most genes 3′ of the polyadenylation site [13]. Even though further experiments will be requested to confirm whether the transcription initiation occurs at, or near, the site indicated by the 5′-terminus of the dermaseptin B2 cDNA (indicated by a dagger in Fig. 2), analysis of the nucleotide sequence of the gene immediately upstream of the 5′-end of the cDNA start revealed a variant TATA-like element located at −42 from the putative transcription start site, where the TATA box is expected to be. One CCAAT box was found farther upstream at −79. Interestingly, the ATG translation site is within a sequence (AACATGGC) which fulfil Kozak's criteria for strong initiation site except for the nucleotide in position 2 [14].

3.2. *Exon 1 is shared by genes encoding peptides with sequences and activities*

Homology search of dermaseptin gene against the EMBL/GenBank identified a large family of cDNAs corresponding to precursors of opioid and antimicrobial peptides originating from the skin secretions of various amphibian species. Brevinins [15,16], esculentins [16,17], gaegurins [18,19] and temporins [20] are 10–46-residue-long antimicrobial peptides (Fig. 1) isolated from the skin secretions of *Rana brevipoda, R. esculenta, R. rugosa* and *R. temporaria,* respectively. Dermorphin, Tyr–D-Ala–Phe–Gly–Tyr–Pro–Ser–NH_2 skin [21], deltorphin A, Tyr–D-Met–Phe–His–Leu–Met–Asp–NH_2 [22], also named dermorphin gene-associated peptide [23] or dermenkephalin [24], and the deltorphins B and C [25], Tyr–D-Ala–Phe–Xaa–Val–Val–Gly–NH_2 (where Xaa is either Asp or Glu) are opioid heptapeptide amides isolated from *Phyllomedusa* ssp. skin. All these peptides are synthesized as part of precursor proteins whose amino terminal domain includes a signal peptide of 22 residues, followed by an acidic leader peptide domain containing 16–25 residues and a progenitor sequence of variable length, coding either for an antimicrobial or an

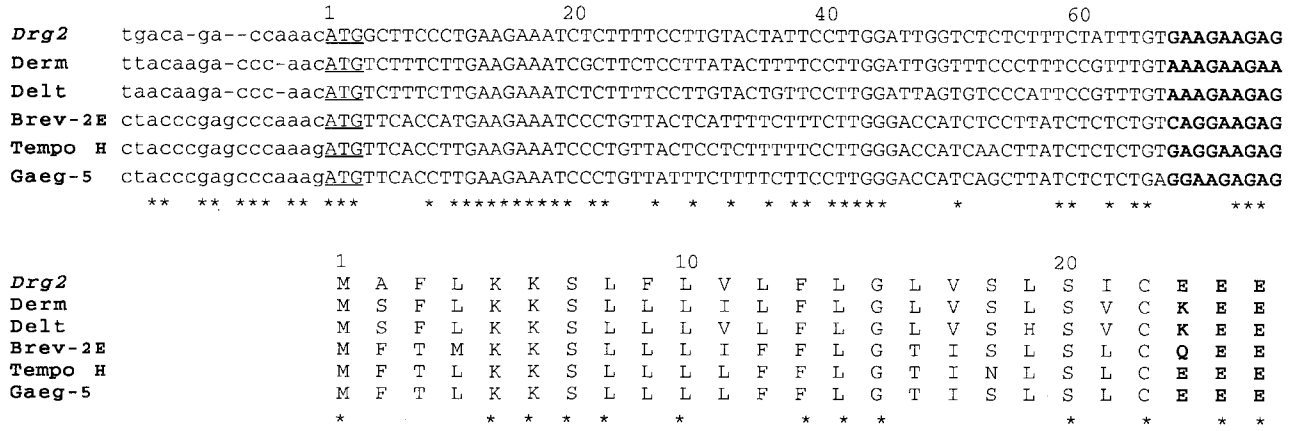


Fig. 3. Nucleotide (top) and amino acid (bottom) sequence alignments of predicted signal peptides (underlined) and amino end of the acidic pro pieces encoded by exon 1 of gene *Drg2* and by preprodermorphin (DRM) [27], preprodeltorphins (DELT) [28], preprobrevinin-2E (Brev-2E) [17], preprotemporin H (Tempo H) [20] and preprogaegurin-5 (Gaeg-5) [19] mRNAs. Asterisks denote identical nucleotides or amino acids between sequences. Part of the 5′-untranslated regions of the respective gene or cDNA sequences are also shown in lower case letters. Gaps (−) have been introduced to maximize similarity. The ATG start codon is underlined. Nucleotides and amino acids corresponding to the beginning of the acidic propiece are in bold letters.

opioid peptide. As depicted in Fig. 3, the coding nucleotide sequence in dermaseptin genes exon 1 comprising the 22-residue signal peptide and the first three residues of the acidic propiece shares 48% identity (50% at the amino acid level) with the corresponding regions of the esculentins, brevinins, temporins, gaegurins, dermorphin and deltorphins cDNAs. Also, 5′-region in the gene is ≈60% identical to 5′-untranslated regions of cDNAs at least over the first 13 bp immediately upstream of the ATG initiation codon, but sequences diverge beyond that point. This suggests that the dermaseptin, gaegurin, brevinin, esculentin, temporin, dermorphin and deltorphin gene subfamilies have evolved with a conserved 'secretory cassette' exon followed by exon(s) encoding end products with stickingly different sequences and biological activities. The duplication and recombitional events that promoted the association of such a homologous secretory exon with nucleotide sequences coding for a variety of end products in various amphibian species remain to be elucidated but are likely to have occurred at the very early stages of evolution.

3.3. Structure of dermaseptin gene Drg1g2

Dermaseptin gene *Drg1g2* presents in the insert of the recombinant plasmid pSK3 was sequenced, including over 1830 and 487 bp of upstream and downstream untranslated sequences, respectively. As shown in Fig. 4, *Drg1g2* and *Drg2* are nearly identical at both ends. Within the first 300 nt of their 5′-flanking regions there were 46-bp differences, 33 of these are located more than 100 nt upstream of the presumed mRNA start. The TATA-like elements and CCAAT boxes occurred in identical positions in the two genes. The overall similarity between the genes is also present for the 317 nt forming the 3′-untranslated part of *Drg2*, (75% identity) including a poly(A) signal at identical positions (Fig. 4). The coding structure of the *Drg1g2* gene comprised two exons homologous to those of *Drg2* (98% and 90% nucleotide identity, respectively), interrupted by a short intron, then a putative second intron of 65 nt and finally a third potential exon encoding part of an acidic propiece followed by a second dermaseptin progenitor sequence. The first two exons encoded identical signal peptides and acidic propieces in both genes, and predicted dermaseptin progenitor sequences differ in four amino acids and insertion of four residues in the central core of the mature peptide issued from *Drg1g2* (Fig. 1). Surprisingly, the intron that interrupts the first two exons exhibits a strong sequence conservation along its entire length in both genes (83% nucleotide sequence identities). This suggests that the two genes expanded recently without sufficient time to allow the intron sequence to drift.

The nucleotide sequence of the putative intron downstream from exon 2 in *Drg1g2* was found to be largely similar (≈66% identity) to the first 66 nt of the 3′-untranslated regions of both genes. Whereas this intronic sequence also closely corresponds to the 3′-untranslated part of mature preprodermaseptin B2 mRNA (70% identity), no equivalent of the poly(A) sites located 71 (*Drg2*) or 55 bp (*Drg1g2*) downstream of the TAA stop codon in the 3′-flanking regions of the genes can be found at the end of the putative intronic sequence. The third potential coding exon in gene *Drg1g2* comprises 120 bp and contains the genetic information for a small part of the acidic propiece plus one copy of a dermaseptin progenitor sequence which is identical to that encoded in gene *Drg2*. Altogether, these findings suggest that the structure of *Drg1g2* has arisen from internal duplication and fusion of an individual segment of a *Drg2*-like ancestral gene, encompassing coding nucleotides of its second exon and ≈60 bp of the 3′-flanking region. The differences between the sequences

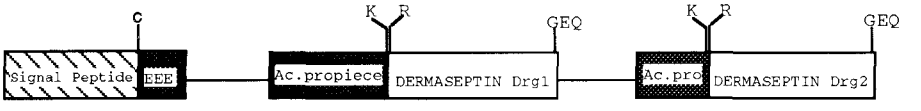
→

Fig. 4. The dermaseptin *Drg1g2* gene. A: Nucleotide sequences of coding exons are indicated by capital letters. The intronic sequence, 5′- and 3′-untranslated sequences are shown in lowercase letters. The putative transcription start site is indicated by a dagger. Nucleotides are numbered relative to the first nucleotide of the ATG start codon. The putative TATA box, CAAT box and polyadenylation signal are in bold letters. Predicted mature preprodermaseptin sequence is shown under the coding DNA. The predicted signal peptide is underlined and mature Drg1 and Drg2 dermaseptin sequences are double underlined. The stop codons are indicated by asterisks. B: Schematic representation of the *Drg1g2* gene. Coding exons are drawn as rectangle, introns as thick lines.

A

```
-1820                cgatcctcagatgacactatttaatcccttcctgacagcc
-1780          agttataataatatggctcttcatttttttattcccctccttccaagagcaat
-1726          agcttaatatttatccatctacataaacctttagggcttttacttcacggtac
-1672          aagatgtactttctattggtgtcattgactgtaacatagaatgaactgaaaaat
-1618          gcagaaaaaaattatgggcaacatgtatgtattttggcgctcagacaccatttt
-1564          tttgttttggattagtagcctagaatgtgcctgaattctataaacctgttctag
-1510          tgcgggtcttagatttagcgctttctaactacaaagcgaaaattctagctcagg
-1456          taaacccctttaacgtgctcttcctgctccagcagtaccgcgcctggttttgc
-1402          cattcaaaagtcattgatgtgtagaatgatgcaatttagtgtgacactcctca
-1348          agcatgccccattctcagcattacaatgagcctaatattggagtgctcctgat
-1294          taaaaataaagcgaacagggctgcgcagttttcgggctgaaaagactttagaag
-1240          ccagccccatacataaatgttgccctatgtgttttatgtattctgtaggtagat
-1186          acaatttaggcaatagcaaatatgtagtgttttctttagtaccatgaaaaaatc
-1132          caaatatcctctatttcttaaatctgcactgcagcgccatttttctgcaaccaat
-1078          aacttttttgtaccttcacctctgagtgaggatgcattttctcaggctgatgtt
-1024          aatttgtttggtaccattttgtgatatatatgacatttggatcacttttttattc
-970           caatttttagaggacttgaaacgaccaaaaaaaataaaatttttgcgtttgtaa
-916           ttttttttcttctacagtgttcagtctacaggataaatcaggttacattgtaata
-862           gatcagacttctacagacttggtaatacaaattatgtgtaacctttttatctttg
-808           tgtatgtaaaatggcaaaaattggttgattagcttatttcaaagttccctcatc
-754           tcattttcagatatagtaatttcttgtgtgtaaagtatggtataattgatcatt
-700           ctttgcttagagaagtggttctcagcctctaatgcaattatgccactggcctgt
-646           ctaaattaattactctaaagcagctgctacaatactacttcaacctctaggcta
-592           gaacatagatctatgtcactattatcattttataaagagggcaacttttcttca
-538           ttatattacttactcctagacaggtagagggattggagggagtggagtcaaaga
-484           tacctgaagcatgtctctctcttagagcctcgtccaaaatttacattacagata
-430           tgtttaaatgtcttctgcctgtctggcagcaagtaataggctgatatagatatg
-376           tgttagattccacaaaactattaaaaaaagaagcaaatgtctgttactctgctt
-322           tcacacgtaaatttctgacaaaaaagtaccaaaaatgctccaacattaattact
-268           gtgcctttctgcacaaaaacatccaaagacacataaaaaagagccaaaaaagct
-214           cagttttattgctttttggatgcatttttgtgtatttttaaatatttattaac
-160           attgcaacaattaagtagagctaaattgttatcattgtcaacctat ccaat ct
-107           tagctttttttttcctttatattttatac taataaaa catttttatttttcaatta
-54            agaagagaaatccaaaacctgtttaactaatataaatctttttgacagaccaaaa
                                          †
+1            ATGGCTTCCCTGAAGAAATCTCTTTTCCTTGTACTATTCCTTGGATTGGTCTCT
               M  A  S  L  K  K  S  L  F  L  V  L  F  L  G  L  V  S
+55           CTTTCTATATGTGAAGAAGA [gtaagtactaatatattataatgaaacaatga
               L  S  I  C  E  E  E
+107          atctgggttttcattctttgctgtgataagggctgggcccagctcagagctgta
+161          ggaaagtatcataagttttaaaaaaaattatattgcctcctaacacag]   G
+211          AAAAGAGAAAATGAAGATGAGGAGGAACAAGAAGATGACGAGCAAAGTGAAATG
               K  R  E  N  E  D  E  E  E  Q  E  D  D  E  Q  S  E  M
+265          AAGAGAGGGCTGTGGAGTAACATAAAAACAGCAGGAAAAGAAGCAGCAAAAGCT
               K  R  G  L  W  S  N  I  K  T  A  G  K  E  A  A  K  A
                  ==============================================
+319          GCATTAAAAGCTGCAGGAAAAGCAGCTTTAGGTGCAGTTACTGATGCGGTGGGA
               A  L  K  A  A  G  K  A  A  L  G  A  V  T  D  A  V  G
               ================================================
+373          GAGCAAtaagttagtaaaatgtaaaatcaatcaaattgctctgaggaacacagt
               E  Q  ***
+427          tatacataattatgccaaatgacgagcaaag TGAAATGAAGAGAGGGCTGTGG
                                               E  M  K  R  G  L  W
                                                  ========
+480          AGTAAAATTAAAGAAGCAGGAAAAGCTGTATTAACAGCTGCAGGAAAAGCAGCT
               S  K  I  K  E  A  G  K  A  V  L  T  A  A  G  K  A  A
               ================================================
+534          TTAGGTGCAGTTTCTGATGCGGTGGGAGAGCAAtaaagttagtaaaattgccct
               L  G  A  V  S  D  A  V  G  E  Q  ***
               ======================
+588          gagaaacacaattatcaataattatgtcaaatctat attaaaa taaattaaaa
+641          tgatatgactgtctgtccttttacatatacattttatattgtatacataatta
+695          agattatagataaaaggtaagagcactaaaatgctgaagtatagggaatcataa
+749          gacaagagtggagaaaacaaatgtgtggggacataagaagaataaaactttagt
+803          ttaaaaaggaaaattatatgatagtcatgagtttaaattgtatataacatccag
+857          aaagtaggtttctaccattggaggagggagttttttttttttttagatgataat
+911          aacttgggggttcatgggttcacacataccatagtctaaagaaagactaggctt
+965          atatataaagttacaaaagtcattgtgtgtggcttctagtccttttaaaattat
+1019         aatatggtaactaaatcattgcaagattcaacacttaaca
```

B



of the first intron in both genes consist of only a few nucleotides. Also, strong conservation of the sequences exist between the second putative intron in *Drg1g2* and 3′-flanking regions of both genes, as well as with the 3′-untranslated part of dermaseptin B2 mRNA. It thus appears likely that expansion of *Drg2* and *Drg1g2* dermaseptin genes through duplication as well as internal amplification were very recent events in the evolutionary process.

Although we have not confirmed the transcription and subsequent translation of *Drg1g2* gene, regions encompassing coding exons 1 and 2 may well be expressed through incomplete or alternate splicing of the transcribed pre-mRNA (Fig. 4). For instance, skipping exon 3 by using donor site at position 393 in intron 2 and 3′-acceptor site in position 590 within the 3′-flanking region would yield a prepro form containing a signal peptide, a complete acidic propiece and a single copy of a dermaseptin Drg1 progenitor sequence. The sequence encoding the second dermaseptin copy is located in exon 3, which begins with the final two residues for the acidic propiece plus the Lys–Arg doublet. Although numerous possible scenarios could offer a possibility to express this dermaseptin copy (for instance, skipping of exon 2 by using 5′-donor site at position 75 in intron 1 and 3′-acceptor site in position 457 in intron 2), the resulting prepro form should only contain five among 21 residues of the acidic pro segment, i.e. EEDEM. Such a shorter version of a dermaseptin precursor has not been yet isolated. Meanwhile, even though these models are plausible and the sequence of gene *Drg1g2* has been obtained from two genomic banks both in agreement, the possibility that this gene is a pseudogene, or that exon 3 represents a dermaseptin exon which has been silenced during evolution cannot be ruled out.

Acknowledgements: This work was supported by funds from the CNRS and the Fondation pour la Recherche Médicale. This paper is dedicated to Prof. Vittorio Erspamer.

References

[1] Lazarus, L.H. and Attila, M. (1993) Prog. Neurobiol. 41, 473–507.
[2] Erspamer, V. (1992) Int. J. Dev. Neurosci. 10, 3–30.
[3] Mor, A., Hani, K. and Nicolas, P. (1995) J. Biol. Chem. 269, 31635–31640.
[4] Mor, A., Amiche, M. and Nicolas, P. (1994) Biochemistry 33, 6642–6650.
[5] Nicolas, P. and Mor, A. (1995) Ann. Rev. Microbiol. 49, 277–304.
[6] Amiche, M., Ducancel, F., Boulain, J.C., Lajeunesse, C., Menez, A. and Nicolas, P. (1993) Biochem. Biophys. Res. Commun. 191, 983–990.
[7] Amiche, M., Ducancel, F., Mor, A., Boulain, J.C., Menez, A. and Nicolas, P. (1994) J. Biol. Chem. 269, 17847–17852.
[8] J. Sambrook, E.F. Fritch, and T. Maniats, Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
[9] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[10] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[11] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucl. Acid. Res. 22, 4673–4680.
[12] Prestridge, D.S. (1991) CABIOS 7, 203–206.
[13] Birnstiel, M.L., Busslinger, M. and Sturb, K. (1985) Cell 41, 349–359.
[14] Kozak, M. (1986) Cell 44, 283–292.
[15] Morikawa, N., Hagiwara, K. and Nakajima, T. (1992) Biochem. Biophys. Res. Commun. 189, 184–190.
[16] Simmaco, M., Mignogna, G., Barra, D. and Bossa, F. (1993) FEBS Lett. 324, 159–161.
[17] Simmaco, M., Mignogna, G., Barra, D. and Bossa, F. (1994) J. Biol. Chem. 269, 11956–11961.
[18] Park, J.M., Jung, J.E. and Lee, B.J. (1994) Biochem. Biophys. Res. Commun. 215, 948–954.
[19] Park, J.M., Lee, J.Y., Moon, H.M. and Lee, B.J. (1995) Biochim. Biophys. Acta 1264, 23–25.
[20] Simmaco, M., Mignogna, G., Canofeni, S., Miele, R., Mangone, M.L. and Barra, D. (1996) Eur. J. Biochem. 242, 788–792.
[21] Montecucchi, P.C., De Castiglione, R., Piani, S., Gozzini, L. and Erspamer, V. (1981) Int. J. Pept. Protein Res. 17, 275–283.
[22] G. Kreil, D. Barra, M. Simmaco, V. Erspamer, G. Falconieri-Erspamer, L. Negri, C. Severini, R. Corsi, and P. Melchiorri, Eur. J. Pharmacol., 1989.
[23] Lazarus, L.H., Wilson, W.E., De Castiglione, R. and Guglietta, A. (1989) J. Biol. Chem. 264, 3047–3050.
[24] Mor, A., Delfour, A., Sagan, S., Amiche, M., Pradelles, P., Rossier, J. and Nicolas, P. (1989) FEBS Lett. 255, 269–274.
[25] Erspamer, V., Melchiorri, P., Falconieri-Erspamer, G., Negri, L., Corsi, R., Severini, C., Barra, D., Simmaco, M. and Kreil, G. (1989) Proc. Natl. Acad. Sci. USA 86, 5188–5192.
[26] Daly, W.J., Caceres, J., Moni, W.R., Gusovski, F., Moos, M., Seamon, B.K., Milton, K. and Myers, W.C. (1992) Proc. Natl. Acad. Sci. 89, 10960–10963.
[27] Richter, K., Egger, R. and Kreil, G. (1987) Science 238, 200–202.
[28] Richter, K., Egger, R., Negri, L., Corsi, R., Severini, C. and Kreil, G. (1990) Proc. Natl. Acad. Sci. USA 87, 4836–4839.