Volume 233, number 2, 282–288 FEB 05937 June 1988

Relationship of promagainin to three other prohormones from the skin of *Xenopus laevis*: a different perspective

Lois T. Hunt and Winona C. Barker

Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC 20007, USA

Received 28 March 1988

We observed a striking sequence similarity between precursors for promagainin and procaerulein type I (excluding the caerulein peptide region). Additional comparisons of the promagainin precursor with those of other procaeruleins, proxenopsin, and peptide-Gly-Leu-amide revealed that all possess one or more copies of a structurally similar spacer module, from which an amphiphilic spacer peptide is cleaved. Promagainin yields the magainins, spacer peptides with antimicrobial activity; we suggest other spacer peptides may have similar activity. We propose that the genes for the four kinds of hormones were derived from a common ancestral gene through gene and exon duplications and that the procaerulein and proxenopsin genes are mosaic genes in which the original 3'-ends were replaced by exon shuffling.

Promagainin; Procaerulein; Domain homology; Exon shuffling; Exon duplication; (*Xenopus laevis*)

1. INTRODUCTION

Many small peptides have been found in the secretions from amphibian cutaneous glands; some have been characterized as to pharmacological activities, but for most, biological roles are only suggested [1,2]. The determination of nucleotide sequences for several prohormones from *Xenopus laevis*, the African clawed frog, has made possible the correlation of a number of the small peptides with their precursors [1–3]. These include complete and partial gene sequences for precursors of procaeruleins [4] and cDNA sequences for the precursors of procaeruleins [3,5,6], proxenopsin [7] and the peptide-Gly-Leu-amide (PGLa) [8,9]. There is also evidence for a second proxenopsin gene [2,4]. Most recently the partial cDNA sequence of the promagainin precursor was determined [10].

We find that the promagainin sequence possesses an overall close similarity to the procaerulein type I precursor [6], excluding the carboxyl-terminal caerulein peptide. Furthermore, the precursors for proxenopsin (excluding the xenopsin peptide) and PGLa exhibit similarity to the amino-terminal regions of promagainin and procaerulein precursors.

Heretofore, active peptides from amphibian skin secretions have been classified by their similarities to active peptides from mammals, other vertebrates, and some invertebrates. As evidence from recent studies [1,2,6] suggests and as our analyses show, this approach is inadequate to describe the relationships among the genes for these four kinds of prohormones. Comparison of the information from the protein and nucleic acid sequences of these prohormones creates a different picture; each possesses one or more copies of a structurally similar small domain. We propose that these prohormones should be grouped together on the basis of the type of domain that they all share.

2. COMPUTER ANALYSES AND RESULTS

We searched for sequences similar to the promagainin precur-

Correspondence address: L.T. Hunt, Protein Identification Resource, National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington, DC 20007, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)

sor, using the FASTP program of Lipman and Pearson [11] and the Protein Sequence Database of the Protein Identification Resource (National Biomedical Research Foundation, Washington, DC). We found that three other kinds of frog prohormone precursors, those for caerulein, xenopsin and PGLa, obtained initial scores above 50 and higher optimized scores with promagainin. No other sequences in the database gave scores in this range. The highest initial and optimized scores were 123 and 209 for the match with the procaerulein type I precursor, and 93 and 124 for the match with the proxenopsin precursor; such scores are high enough to indicate a probable relationship [11]. The match of the promagainin with the PGLa precursor got lower scores of 66 and 75. However, in a FASTP search of the PGLa precursor, it got initial and optimized scores of 177 and 177 (no gaps inserted) with the proxenopsin precursor.

To visualize the extent of the regions of similarity among these prohormones, we generated dotmatrix graphic plots, using the Mutation Data Matrix [12] with a minimum score of 20 and a window size (segment length) of 20. This type of comparison is especially useful for displaying duplicated regions in sequences, such as those in promagainin [10] and procaerulein [3–6]. Fig.1 shows the plots of the promagainin precursor with the precursors of procaerulein type I, proxenopsin and PGLa; duplicated regions form the repeated short diagonals. In contrast, the plot of the proxenopsin and PGLa precursors is a single complete diagonal line (not shown). In the plot of promagainin and procaerulein type I, the region representing the carboxyl-terminal 20 residues of the latter sequence (where the single caerulein peptide is found) shows no similarity to the promagainin sequence.

We also compared the sequences of the precursors for promagainin and procaerulein type I with our program ALIGN [12], using a bias of 6, a gap penalty of 6, and 100 randomizations. The alignment (see fig.2) was essentially the same as that from the FASTP search; however, we omitted from procaerulein the first five residues (of the signal peptide) and the carboxyl-terminal 18 residues (containing the caerulein peptide), which were unmatched in the search, in order to compare sequences of approximately the same length. The ALIGN score was 13.9 standard deviation units (SD). The two sequences have 38% identity and nearly 74% similarity. Most of the identities, aside from those in the signal peptide, are somewhat clustered in three areas, as are many of the conservative substitutions. The identities and similarities fall into three repeating regions of approx. 50 residues in both proteins (see also fig.1). We have constructed alignments showing the sequence similarities in the signal and propeptides (fig.3) and in the repeating regions (fig.4). Because of the very close resemblance of the first 21 residues of the promagainin precursor fragment to the signal peptide and propeptide of the other prohormones, we assume that they correspond, that only the first five residues of the signal peptide are missing and that the cleavage site is after the first proline. As shown in fig.4, the precursors of proxenopsin and PGLa contain only one copy each of the repeating region [2,7,8]. The small regions consisting of the xenopsin and caerulein active peptides (with their processing sequences) are not, of course, included in the alignment (fig.4); the position of one of these regions in a sequence is indicated by a slash before or after a repeating region.

3. DISCUSSION

We have shown that extensive sequence similarities are present between the precursors for promagainin and three other kinds of prohormones from the African clawed frog. The results of our computer-assisted comparisons strongly support our proposal that these sequences had a common origin. The resemblance of the promagainin and procaerulein type I precursors is especially striking, as the protein sequences are also almost the same length, excluding the carboxyl-terminal caerulein peptide module (see fig.2).

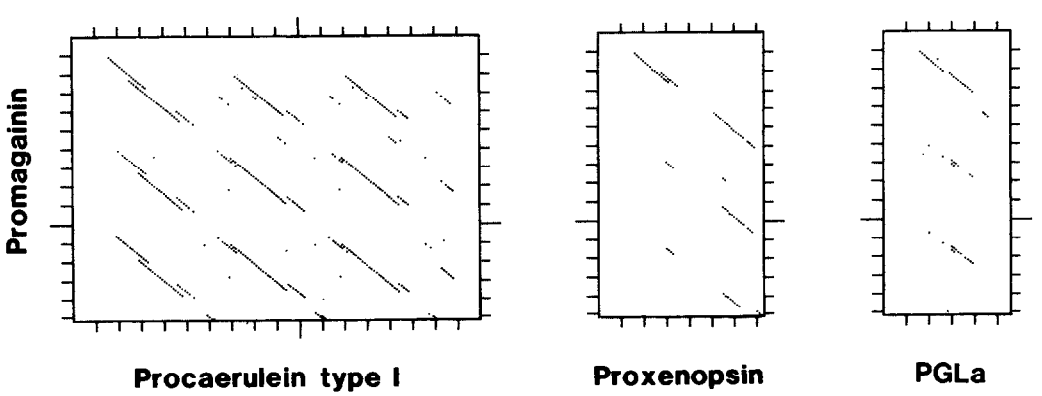

Fig.1. Comparison of the promagainin precursor sequence with those of the precursors for procaerulein type I, proxenopsin and PGLa. The dotmatrix plots were generated using the Mutation Data Matrix with a minimum score of 20 and a window size of 20. Grid marks represent 10-residue lengths. The short diagonal lines represent the three repeats in the promagainin and procaerulein sequences. The extension of the top lefthand diagonal line in each plot indicates the similar signal peptides.

```
       1        10            20            30            40            50            60            70

M1     /FICSLIAVICANALPQPEASADEDMDEREVRGIGKFLHSAGKFGKAFVGEIMKSKRDA-EAVGPEAFADQ

C1  ...LLCVLFAVLSANPLSQPEGFADE---ERDVRGLASFLGKALKAGLK GAHLLGGAPQQREANDERRFADD

       71        80            90           100           110           120           130          140

M1     D--LDEREVRGIGKFLHSAKKFGKAFVGEIMNSKRDA-EAVGPEAFADE--DLDEREVRGIGKFLHSAKK

C1     DDDVNERDVRGFASFLGKALKAALK GANMLGGTPQQREANDERRFADDEDDVNERDVRGFGSFLGKALK

      141       150           160

M1     FGKAFVGEIMNSKRDA-EAVDDRRWVE

C1     AALK GANALGGSPQQREANDERRFADG...

Total score = 1204, 6 breaks
59 identities out of 157 possible matches between residues
100 random runs
Alignment score =   13.92 SD   Standard deviation =   16.83   Mean =  969.78
```

Fig.2. Alignments of corresponding regions of the promagainin (M1) and procaerulein type I (C1) precursors produced by the ALIGN program, using the Mutation Data Matrix with a bias of 6 and a gap penalty of 6. Identities between the sequences are indicated by paired dots. For the procaerulein sequence, the signal peptide cleavage site and intron locations are indicated by an arrow and triangles, respectively.

These precursors can be thought of as composed of two or more kinds of modules, or domains, arranged in a basically similar manner. Repeats of these modules, especially in procaeruleins, are arranged to produce different combinations. This concept was proposed [4,6] to explain the variety of structures found in the family of procaerulein hormones, and perhaps also the structures of proxenopsin and PGLa precursors [4]; we are extending it to include the promagainin precursor.

In all of these precursors the amino-terminal module includes a signal peptide of 20 residues and a 6-residue propeptide (see fig.3). All also have at least one copy of a module in which a segment of 22–27 predominantly hydrophobic residues, but containing several lysines, is bounded on each side by a short segment of mostly hydrophilic residues (see fig.4). This module, or domain, was termed a spacer region [1,2] or a spacer segment [4] and, in the case of procaeruleins, an intercaerulein segment [5]. The central hydrophobic segment was called a spacer peptide [2]. The precursors of proxenopsin and PGLa have one copy of the spacer module [7,8], procaeruleins have two or more very similar copies [4–6], and promagainin has three nearly identical copies [10]. Similarities among

```
                                          1                   2↓
                     1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6

1 PROMAGAININ                /F I C S L I A V I C A N A L P Q P E A S A...
2 PROXENOPSIN        M Y K G I F L C V L L A V I C A N S L A T P S S D A...
3 PGLa               M Y K Q I F L C L I I A A L C A T I M A E A S A F A...
4 PROCAERULEIN 1     M F K G I L L C V L F A V L S A N P L S Q P E G F A...

Conserved            M Y K G I F L C   L   A V   C A N   L     P       A
```

Fig.3. Alignment of the amino ends of the four precursor sequences. Conserved residues are shown below the alignment; residues identical in all four sequences are marked. An arrow indicates the proposed cleavage site for the signal peptide; residues 21–26 form the proposed propeptide.

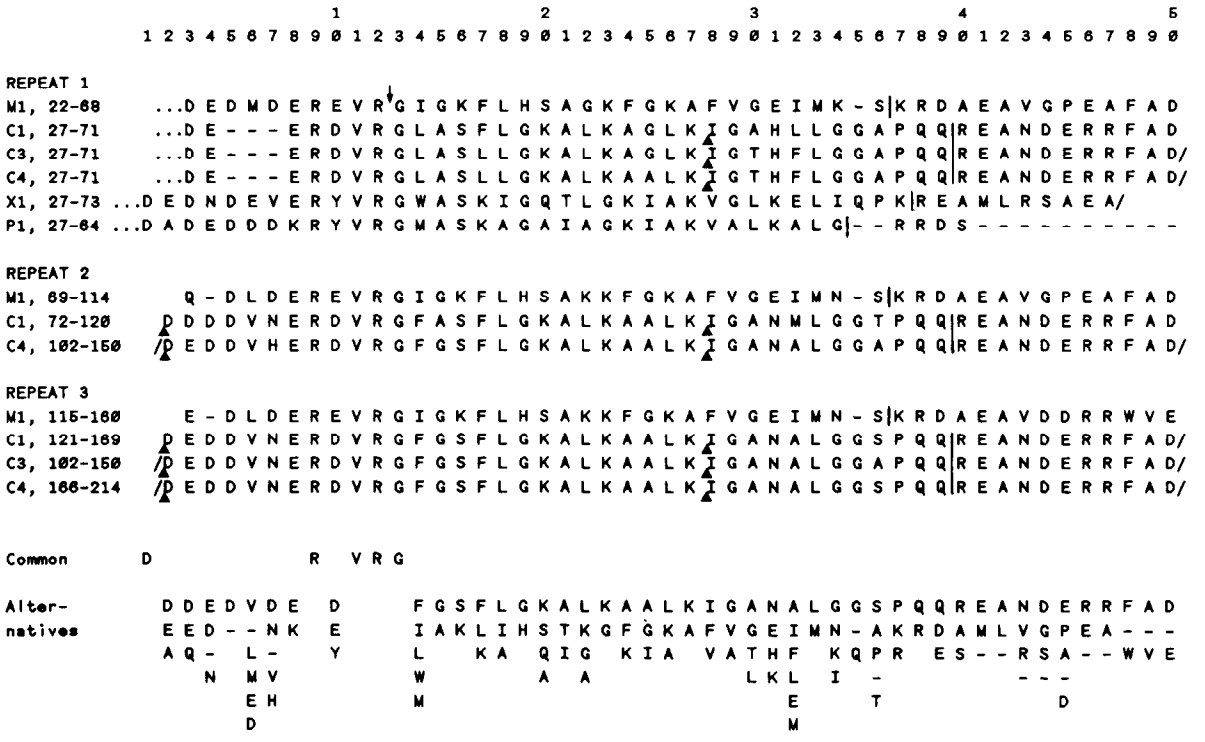

Fig.4. Alignment of spacer modules from promagainin (M1), procaerulein type I (C1), procaerulein type III (C3), procaerulein type IV (C4), proxenopsin (X1) and PGLa (P1). Residue numbers for the sequence segments are shown also. Common and alternative residues are shown below the alignment. The arrow between positions 12 and 13 indicates a common amino-terminal cleavage site for the hydrophobic segments (i.e., magainins and equivalents); carboxyl-ends are marked by vertical bars (see text). Intron positions in procaerulein sequences are represented by triangles; all are between the first and second nucleotide in the codon (see text).

most of these spacer peptides have also been pointed out by others [1,2,7,13,14], although the overall high degree of similarity between the promagainin and procaerulein precursors was not recognized. In addition, the sequence around the Arg-Gly cleavage site (positions 12–13 in the fig.4 alignment) is highly conserved in all modules, not only in the procaerulein modules [6], and may form the recognition site for a specific proteinase [6,9]. In the partially sequenced genes for procaerulein precursors [4], each spacer module is encoded by two exons, except that the signal and propeptide are also part of exon 2 (the first protein-coding exon). Positions of the introns, all of which fall between the first and second nucleotide in a codon and so are phase 1 introns in Patthy's [15] classification, are indicated in figs 2 and 4. We predict that phase 1 introns will be found in corresponding positions in the genes of the three other precursors.

The central hydrophobic segments of the spacer modules from these prohormones have been studied by Williams and co-workers [1,2] and by Kreil and co-workers [6,9]. These segments have been found as free peptides in the skin secretions of the frog. It was proposed that the segments could form amphipathic helices and have potential membrane-crossing capability and that secondary cleavages before lysine residues could affect activity (e.g., deactivating a previously active peptide). Although no function was established for any of these peptides, the degree of sequence similarity suggested that the peptides would also show functional similarity. The antimicrobial activity reported for natural and synthetic magainin peptides [10,16], which form the central hydrophobic segment of the spacer modules in promagainin, supports the possibility of similar activity for the other peptides. Although the magainins do not have a direct lytic effect, they disturb normal mem-

brane function(s) by an as yet undefined mechanism that appears to be dependent on membrane-crossing capability; removal of four to six amino-terminal residues resulted in loss of activity, presumably because the peptide was too short to form a membrane-spanning helix [16].

A third type of module, or domain, consists of a neuroactive peptide sequence framed by short processing sequences; this type of module is seen only in the proxenopsin [7,17] and procaerulein precursors [4–6,18]. One to many copies of this type of module may be present. If one copy (as in proxenopsin), it is located at the carboxyl-end of the precursor, and if more than one (as in most procaeruleins), they are interspersed with the spacer modules and may be doubled [4,5]. The gene for the type I procaerulein precursor was found to contain a potential caerulein exon, lying between the exons encoding spacer modules 1 and 2, as well as a partial caerulein exon following the exons for spacer module 2 [4]. It was proposed that the arrangements of the different types of modules observed in the various procaerulein genes is the result of a combination of genetic events [4]. Originally there was a simple gene composed of a 5′-noncoding exon, an exon coding for a signal peptide, propeptide and part of a spacer region, an exon for the rest of the spacer region, an exon for the caerulein peptide (with processing sequences) and the 3′-noncoding region, and three introns. The first event was a duplication of most of this simple gene including introns starting midway through the second exon. Subsequently, other mechanisms, such as deletion or nonexpression of exons, duplication of single exons, duplication of entire genes, and possibly gene conversion events, produced the family of procaerulein genes now observed [4].

Although the active peptides contained in the third type of module may be related to various mammalian active peptides (e.g., caerulein to gastrin and cholecystokinin, and xenopsin perhaps to neurotensin), as well as to cholecystokinin-like or neurotensin-like peptides in other tissues of the frog [2,19], there is no similarity between the sequences of the spacer modules and the remainder (the amino-end) of the sequences of the mammalian hormone precursors [1,3,6]. However, there is similarity of gene structure in that the genes for the precursors of procaeruleins and proxenopsin [4], for human progastrin [20–22] and for rat procholecystokinin [23,24] all appear to have only phase 1 introns [15]. We assume that the amphibian genes for cholecystokinin-like and

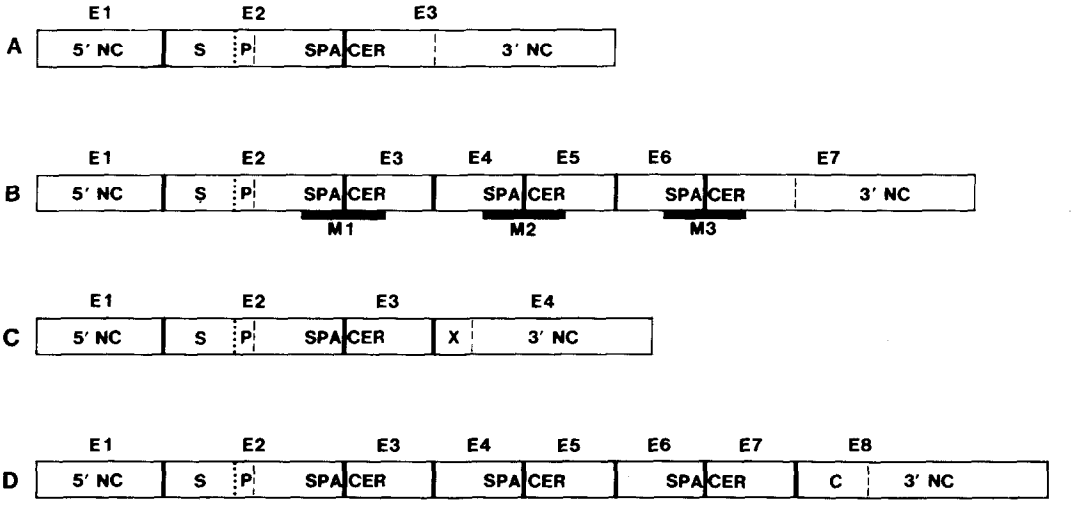

Fig.5. Diagram of proposed exon-intron arrangement in an ancestral gene compared with suggested arrangements in the prohormone genes. (A) Ancestral gene and possibly PGLa precursor gene. (B) Promagainin precursor gene. (C) Proxenopsin precursor gene. (D) Procaerulein (type I) precursor gene. Intron positions (exon boundaries) are marked by wide solid vertical lines; the boundary between signal peptide and propeptide is indicated by a dotted line, and other domain boundaries are indicated by dashed lines. Symbols for features are as follows: exons, E1, E2, etc.; 5′-noncoding region, 5′ NC; 3′-noncoding region, 3′ NC; signal peptide, S; propeptide, P; magainin peptides, M1, M2, M3; xenopsin peptide module, X; caerulein peptide module, C.

neurotensin-like proteins also have phase 1 introns. This type of gene structure, in which all introns have the same phase class, is correlated with exon shuffling (including exchange between originally nonhomologous genes) and exon duplication in gene assembly, according to Patthy [15].

As striking as the protein sequence similarity is, the relationship between the promagainin and procaerulein type I precursors is most likely indirect rather than direct. The near identity of the repeated spacer modules within each sequence, as opposed to over 60% difference between the two, argues for an independent development of the repetitive structure from a smaller ancestral gene. We envision this ancestral gene, from which the four kinds of prohormones were derived, as having several exons corresponding to the 5′-noncoding region, the signal peptide and propeptide plus part of the spacer module, the remainder of the spacer module and the 3′-noncoding region (see fig.5). Duplications produced several copies of the entire gene before the different prohormones evolved. The gene for the PGLa precursor may retain this basic form. Subsequent duplications of the exons encoding the spacer module could produce a gene encoding a protein like the promagainin precursor (see fig.5). On the other hand, the evidence indicates that the proxenopsin and procaerulein precursors are encoded by mosaic genes produced by exon shuffling with two other kinds of genes. The 3′-end of the original gene was replaced in each by an exon, for a neuroactive peptide module and a 3′-noncoding region, derived from the 3′-end of another hormone gene in the frog's genome (see fig.5). For example, by this process the simple procaerulein gene described by Vlasak et al. [4] would have derived its 3′-exon from a cholecystokinin-like or gastrin-like gene; duplications and other genetic mechanisms would produce the various procaerulein precursors [4]. In our view the exon shuffling events are superimposed (in the genes for proxenopsins and procaeruleins) on a gene structure common to all four kinds of prohormones; thus the four are basically homologous. Furthermore, the homologous spacer domains in the four kinds of prohormones are processed to release similar peptides; the magainin peptides were shown to have antimicrobial activity [10,16], so it is probable that all these spacer peptides will have functional as well as structural similarity.

ADDENDUM

After submission of this paper we learned of the contribution of Soravia et al. [25], which demonstrates that the amphiphilic spacers, identified here as analogs of magainin, have antimicrobial activity.

Acknowledgements: We thank Deanna Davalos for expert technical support and James K. Bair for invaluable editorial and technical assistance in the preparation of the manuscript. This research was supported by NIH grants CA-40474 from the National Cancer Institute and RR-01821 from the Division of Research Resources.

REFERENCES

[1] Gibson, B.W., Poulter, L., Williams, D.H. and Maggio, J.E. (1986) J. Biol. Chem. 261, 5341–5349.
[2] Giovannini, M.G., Poulter, L., Gibson, B.W. and Williams, D.H. (1987) Biochem. J. 243, 113–120.
[3] Hoffmann, W., Bach, T.C., Seliger, H. and Kreil, G. (1983) EMBO J. 2, 111–114.
[4] Vlasak, R., Wiborg, O., Richter, K., Burgschwaiger, S., Vuust, J. and Kreil, G. (1987) Eur. J. Biochem. 169, 53–58.
[5] Wakabayashi, T., Kato, H. and Tachibana, S. (1984) Gene 31, 295–299.
[6] Richter, K., Egger, R. and Kreil, G. (1986) J. Biol. Chem. 261, 3676–3680.
[7] Sures, I. and Crippa, M. (1984) Proc. Natl. Acad. Sci. USA 81, 380–384.
[8] Hoffmann, W., Richter, K. and Kreil, G. (1983) EMBO J. 2, 711–714.
[9] Andreu, D., Aschauer, H., Kreil, G. and Merrifield, R.B. (1985) Eur. J. Biochem. 149, 531–535.
[10] Zasloff, M. (1987) Proc. Natl. Acad. Sci. USA 84, 5449–5453.
[11] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[12] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
[13] Cannon, M. (1987) Nature 328, 478.
[14] Cannon, M. (1987) Nature 329, 494.
[15] Patthy, L. (1987) FEBS Lett. 214, 1–7.
[16] Zasloff, M., Martin, B. and Chen, H.-C. (1988) Proc. Natl. Acad. Sci. USA 85, 910–913.
[17] Araki, K., Tachibana, S., Uchiyama, M., Nakajima, T. and Yasuhara, T. (1973) Chem. Pharm. Bull. 21, 2801–2804.
[18] Anastasi, M., Bertaccini, G., Cei, J.M., De Daro, G., Erspamer, V., Impicciatore, M. and Roseghini, M. (1970) Br. J. Pharmacol. 38, 221–228.
[19] Dimaline, R. (1983) Peptides 4, 457–462.

[20] Ito, R., Sato, K., Helmer, T., Jay, G. and Agarwal, K. (1984) Proc. Natl. Acad. Sci. USA 81, 4662–4666.
[21] Wiborg, O., Berglund, L., Boel, E., Norris, F., Norris, K., Rehfeld, J.F., Marcker, K.A. and Vuust, J. (1984) Proc. Natl. Acad. Sci. USA 81, 1067–1069.
[22] Kato, K., Hayashizaki, Y., Takahashi, Y., Himeno, S. and Matsubara, K. (1983) Nucleic Acids Res. 11, 8197–8203.
[23] Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. and Matsubara, K. (1985) Proc. Natl. Acad. Sci. USA 82, 1931–1935.
[24] Deschenes, R.J., Haun, R.S., Funckes, C.L. and Dixon, J.E. (1985) J. Biol. Chem. 260, 1280–1286.
[25] Soravia, E., Martini, G. and Zasloff, M. (1988), 337–340.