Biochemical and Biophysical Research Communications **277,** 7–12 (2000)
doi:10.1006/bbrc.2000.3629, available online at http://www.idealibrary.com on IDEAL®

Genomic Organization and Splicing Variants of a Peptidylglycine α-Hydroxylating Monooxygenase from Sea Anemones[1]

Michael Williamson, Frank Hauser,[2] and Cornelis J. P. Grimmelikhuijzen[3]
Department of Cell Biology, Zoological Institute, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

Received August 21, 2000

Cnidarians are primitive animals that use neuropeptides as their transmitters. All the numerous cnidarian neuropeptides isolated, so far, have a carboxy-terminal amide group that is essential for their actions. This strongly suggests that α-amidating enzymes are essential for the functioning of primitive nervous systems. In mammals, peptide amidation is catalyzed by two enzymes, peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidyl-α-hydroxyglycine α-amidating lyase (PAL) that act sequentially. These two activities are contained within one bifunctional enzyme, peptidylglycine α-amidating monooxygenase (PAM), which is coded for by a single gene. In a previous paper (F. Hauser *et al., Biochem. Biophys. Res. Commun.* 241, 509–512, 1997) we have cloned the first known cnidarian PHM from the sea anemone *Calliactis parasitica*. In the present paper we have determined the structure of its gene (CP1). CP1 is >12 kb in size and contains 15 exons and 14 introns. The last coding exon (exon 15) contains a stop codon, leaving no room for PAL and, thereby, for a bifunctional PAM enzyme as in mammals. Furthermore, we found a CP1 splicing variant (CP1-B) that contains exon-9 instead of exon-8, which was present in the previously characterized PHM cDNA (CP1-A). CP1-A and -B have 97% amino acid sequence identity, whereas both splicing variants have around 42% sequence identity with the PHM part of rat PAM. Essential amino acid residues for the catalytic activity and the 3D structure of PHM are conserved between CP1-A, -B and the PHM part of rat PAM. Furthermore, eight introns in CP1 occur in the same positions and have the same intron phasing as eight introns in the rat PAM gene, showing that the sea anemone PHM is not only structurally, but also evolutionarily related to the PHM part of rat PAM.

Cnidarians are primitive animals such as sea anemones, corals, jellyfishes, and *Hydra*. Cnidarians are interesting for several reasons. First, they are useful model organisms for developmental biologists, because they are simple, easy to manipulate, and have a strong regeneration capacity (1–3). Second, cnidarians are interesting from an evolutionary point of view, because they are the lowest animal group having a nervous system and it was probably within cnidarians or a closely related ancestor phylum that nervous systems first evolved (4).

The primitive nervous systems of cnidarians use neuropeptides as their transmitters (5). All the numerous (~30) cnidarian neuropeptides that we and other groups have isolated, so far, have an amide group at their carboxy-termini (5–9). This amide group prevents degradation by unspecific carboxypeptidases, but is also essential for the biological actions of the peptides (5, 8).

In mammals, peptide amidation is known to occur in two steps. In the first step, a peptidylglycine propeptide is hydroxylated at the α-C atom of glycine. This reaction is catalyzed by peptidylglycine α-hydroxylating monooxygenase (PHM), which requires O_2, Cu ions, and ascorbate. The next step is a cleavage reaction, where the enzyme peptidyl α-hydroxyglycine α-amidating lyase (PAL) performs an N-C cleavage, thereby releasing peptidylamide and glyoxylic acid. In mammals, both PHM and PAL activities are contained within one bifunctional enzyme, peptidylglycine α-amidating monooxygenase (PAM) that is coded for by a single gene (10–13). From X-ray studies it is known that the PHM part of rat PAM

[1] The nucleotide sequences reported in this paper have been submitted to the GenBank Data Bank with Accession Nos. AY006478, AY006479, and AY007173-AY007186.

[2] Present address: Lundbeck A/S, Ottiliavej 9, DK-2500 Valby, Denmark.

[3] To whom correspondence should be addressed. E-mail: cgrimmelikhuijzen@zi.ku.dk. Homepage: http://www.zi.ku.dk/cellbiology/.

0006-291X/00 $35.00

```
A CCT GAT CAC TTG GCT GCA ATT TTG TTG CTA GCC ACT GGT GGT GAA ATT CCT GCA CAA AAA AAA G   774
  Pro Asp His Leu Ala Ala Ile Leu Leu Leu Ala Thr Gly Gly Glu Ile Pro Ala Gln Lys Lys     211

A CCT CAT CAC TTG GCT GCA ATT TTA GTT CTA ACA AAC AGA GAG CGG GCT GCA ATT CCA CCA CAG AAA GAA G 780
  Pro His His Leu Ala Ala Ile Leu Val Leu Thr Asn Arg Glu Arg Ala Ala Ile Pro Pro Gln Lys Glu   213
```

FIG. 1. Upper panel: cDNA and encoded amino acid sequence of exon-9. Lower panel: cDNA and encoded amino acid sequence of exon-8. The numbering refers to Fig. 3 of this paper and Fig. 2 of Ref. (16).

contains two opposing domains (A and B, or I and II) that each contains a Cu ion (Cu^{+} or Cu^{2+}) necessary for catalytic activity bound to three essential His residues (in domain A) and two His and one Met residue (in domain B) (14, 15). The two opposing domains are each held together by three Cys–Cys bridges (in the A domain) and two Cys–Cys bridges (in the B domain) and are connected by a hinge region (14, 15).

In a previous paper we described the cloning of the first cnidarian PHM (CP1-A) from the sea anemone *Calliactis parasitica* (16). This cnidarian PHM cDNA contained a stop codon at its 3′ end, suggesting that the cnidarian PAL enzyme must be coded for by a different gene. In the present paper we have determined the genomic organization of the cnidarian PHM gene (CP1) and established the presence of splicing variants.

MATERIALS AND METHODS

Preparation of poly(A^{+}) RNA and cDNA from *C. parasitica*, the generation of the initial PCR products (Fig. 2A), 5′/3′-RACE, as well as DNA sequencing and sequence analyses were described earlier (16).

For cloning of the gene, a genomic library from *C. parasitica* was constructed in λ FIX II by Stratagene. This library was screened with the initial PCR products described above. From 750,000 plaques screened, 51 were positive. Two of them were used to compose the CP1 gene (Fig. 2E).

Northern blots were performed as descibed earlier (17), using a radioactive probe (2×10^6 cpm/ml), corresponding to nucleotide positions 592–1216 of Fig. 2 from (16), and 5 μg of mRNA. For Southern blots, a BioRad ZetaProbe membrane was used, following the protocol of the manufacturer, and 12 μg of genomic DNA was digested with either one of the restriction enzymes (Fig. 5). A radioactive probe corresponding to nucleotide positions 717–1156 of Fig. 2 from (16) was used at 2×10^6 cpm/ml.

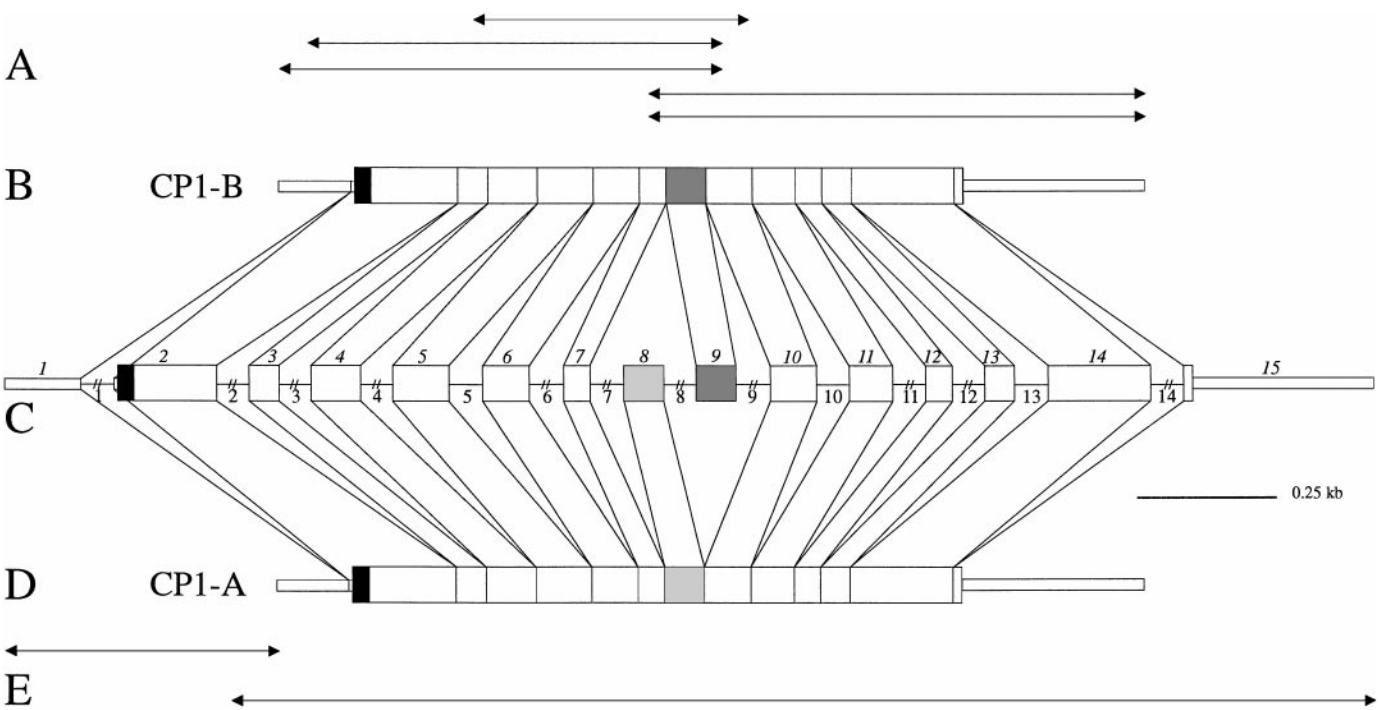

FIG. 2. Schematic representation of the CP1 gene, its two splicing variants, and the positions of the genomic and cDNA clones. (A) Positions of the cDNA clones used to clone CP1-B. The upper clone is the initial PCR product, the two lower clones are the products of 3′-RACE, and the two middle clones are the products of 5′-RACE PCR. (B) Organization of the splicing variant CP1-B that contains exon-9, but not exon-8 of the gene. (C) Organization of the CP1 gene. Introns are represented by lines (not drawn to scale) and marked 1–14 (in roman), exons are represented by boxes, being broad for the coding and small for the noncoding regions and marked 1–15 (in italic). The region coding for the signal sequence is highlighted in black. Exon-8 is highlighted in light grey, and exon-9 in dark grey. (D) Organization of the splicing variant CP1-A that contains exon-8, but not exon-9 of the gene. (E) Positions of the genomic clones used to clone CP1.

TABLE 1

Intron/Exon Boundaries of CP1

Intron	5′ donor	Intron size (bp)	3′ acceptor	Intron phase
1	G gtaagtt . . .	~3600	. . . accacag AA	—
2	AAG gtaagtt . . .	~1800	. . . cttgcag CAT	3
	Lys		His	
3	G gtatgtt . . .	~460	. . . cattcag TT	1
	Val		Val	
4	TG gtgagta . . .	~530	. . . atttag G	2
	Trp		Trp	
5	G gtaacaa . . .	121	. . . tctttag AT	1
	Asp		Asp	
6	A gtaagtt . . .	~240	. . . attgcag AT	1
	Asn		Asn	
7	AG gttgcca . . .	~1280	. . . aatccag A	2
	Arg		Arg	
8	AG gtaggat . . .	~550	. . . actacag A	2
	Arg		Arg	
9	G gttaaac . . .	~610	. . . tgtacag CT	1
	Ala		Ala	
10	G gtaagtt . . .	85	. . . attgcag GC	1
	Gly		Gly	
11	CAG gtattca . . .	~460	. . . gttgcag GCT	3
	Gln		Ala	
12	TTG gtaagta . . .	~240	. . . attttag GCT	3
	Leu		Ala	
13	GG gtaagat . . .	139	. . . cattcag A	2
	Gly		Gly	
14	G gttagtt . . .	~470	. . . tccgtag GC	1
	Gly		Gly	

Note. The sequence of each of the intron–exon boundaries is shown, as well as the codons for the amino acid residues. Uppercase and lowercase letters represent nucleotides in the exons and introns, respectively. The sequences of some of the introns can be retrieved from our GenBank submission, Accession Nos. AY006479, and AY007173-AY007186. Other introns have not been fully sequenced. The overall positions of the introns are shown in Figs. 2 and 3. Introns 2–5, 7, and 9–11 occur in the same positions and have the same intron phasings on eight introns in the rat gene (20).

RESULTS

PHM cDNA's. To find out whether other *Calliactis* PHM cDNA's existed in addition to the one published previously (CP1-A; 16), we sequenced additional PCR clones from our initial PCR reaction (16). This resulted in the identification of a new cDNA (CP1-B) that was identical to CP1-A with the exception of a continuous sequence of 65 nucleotides, coding for 21 amino acid residues located in the middle of the PHM (Figs. 1 and 2).

Alternative splicing. To establish whether CP1-A and -B originated by alternative splicing from the same gene, we screened a genomic library from *Calliactis* with the PCR products from our initial PCR reaction (16). This resulted in several positive genomic clones that were sequenced and turned out to be the gene (CP1) coding for both CP1-A and -B. CP1 is >12 kb in size and has 14 introns and 15 exons (Fig. 2). The introns varied in size from 85 bp to ~3600 bp (Table 1). Exon-8 is present in CP1-A but not in -B, whereas exon-9 is present in CP1-B, but absent in -A, showing the existence of alternative splicing of the CP1 gene transcripts (Fig. 2).

Comparison of CP1-B with other PHMs. CP1-B has 97% amino acid sequence identity with CP1-A (Fig. 3). Both PHMs have about 40% sequence identity with rat, *Xenopus, Drosophila,* and *Lymnaea* PHMs (Fig. 3, see also Discussion). Eight introns in the CP1 gene occur at the same positions and have the same intron phasing as eight introns in the rat gene, showing that the sea anemone PHMs are not only structurally, but also evolutionarily related to the PHM part of rat PAM (Fig. 3, Table 1).

Northern blot analysis. Northern blot analysis of *Calliactis* mRNA, using a hybridization probe covering most of the coding region of CP1-A, showed a prominent band at 1.6 kb (Fig. 4), fitting very well with the size of CP1-A mRNA, which is 1562 bp (16). CP1-B mRNA will hybridize within the same band, because it has a large nucleotide sequence identity with CP1-A (98%) and the size of exon 9 is comparable to that of exon 8 (Fig. 3). In addition to the strong band at 1.6 kb, there is a weak band at 5.1 kb that we can not interpret at present.

Southern blot analysis. We carried out a Southern blot analysis, using *Calliactis* genomic DNA digested

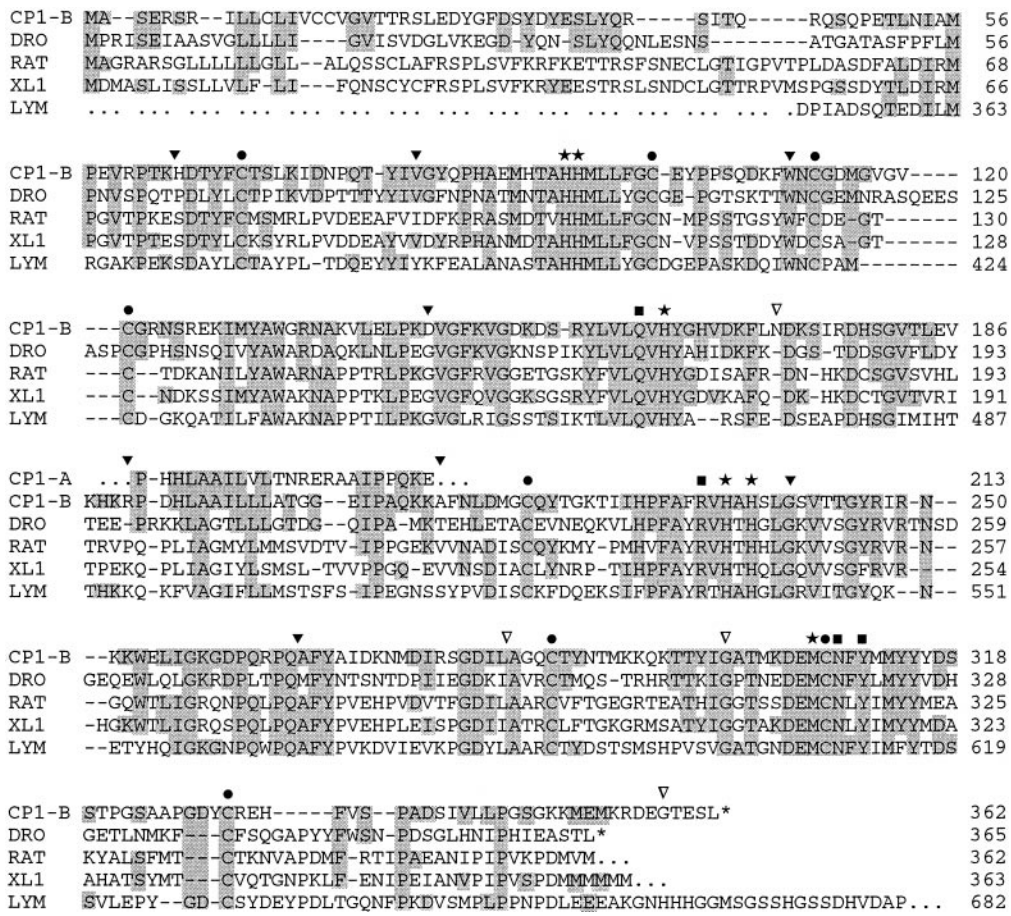

FIG. 3. Amino acid residues comparison of the PHMs from *C. parasitica* (CP1-A and CP1-B; from CP1-A only the amino acid residues coded by exon-8 are given), *Drosophila* (DRO), *Lymnaea* (LYM; here only PHM variant 2 (21) is shown), *Xenopus laevis* (XL1), and rat. Amino acid residue positions are given at the right. Residues that are identical between CP1-B and at least one of the other PHMs are highlighted in grey. The five His and one Met residue that bind two Cu ions are indicated by asterisks. The eight conserved Cys residues that form four Cys–Cys bridges are indicated by filled circles. The four conserved substrate binding sites (see text) are indicated by filled squares. The positions of introns in CP1 are marked by open or filled triangles. Eight introns in the CP1 gene occur at the same positions and have the same intron phasing as eight introns in the rat PHM (or PAM) gene. These introns are marked by filled triangles. Spaces are introduced to optimize alignments. The A-domain of rat PHM stops at Cys-186 and the B-domain starts at Cys-227 (14). The data from the CP1-A variant and from the rat, *Drosophila*, *Xenopus*, and *Lymnaea* PHMs are from (14, 16, 20–24).

with five different restriction enzymes, and a hybridization probe covering a large portion of the coding region of CP1. This yielded single, strongly hybridizing bands, accompanied, however, by weakly hybridizing bands (Fig. 5). These findings suggest the presence of a single PHM gene, although the presence of another CP1-related gene in *Calliactis* can not be excluded.

DISCUSSION

As mentioned in the Introduction, several amino acid residues in the PHM part of rat PAM have been recognized as being important for its 3-D structure and catalytic activity. These are a Cu ion binding site in the A-domain (consisting of three His residues that occupy three of the four coordination positions in a tetrahedron) and a Cu ion binding site in the B domain (consisting of two His and one Met residue, again in a tetrahedral configuration) that each bind a Cu ion (CuA and CuB), which are necessary for catalytic activity of PHM (14, 15, 18). These six amino acid residues are conserved between the rat and sea anemone PHMs (marked with asterisks in Fig. 3). Furthermore, the C-terminal carboxyl group of the Gly-extended propeptide makes hydrogen bridges with Arg-240 and Tyr-318, and the C-terminal peptidyl amide bond makes a hydrogen bridge with Asn-316 of the rat PHM (15, 18). Also these three direct C-terminal substrate binding sites are conserved between the rat and sea anemone PHMs (marked with filled squares in Fig. 3). In addition, a water molecule bridges the C-terminal carboxyl group of the propeptide substrate and Gln-170 (18). Again, this residue (marked with a filled square in Fig. 3) is conserved between the rat and the sea anem-

one PHMs. Finally, eight out of ten Cys residues present in the rat PHM (making three Cys–Cys bridges in the A- and two Cys–Cys bridges in the B-domain (14)), are conserved between the rat and the sea anemone proteins (marked with filled circles in Fig. 3). All the eighteen above-mentioned amino acid residues are also conserved between the sea anemone and *Xenopus, Drosophila,* and *Lymnaea* PHMs (Fig. 3), showing that these residues are essential for the enzyme and that the sea anemone protein is likely to have PHM activity.

In addition to the above-mentioned residues, large blocks of amino acid residues are conserved between the five species. These blocks are often flanking the Cu ion binding sites (marked with asterisks in Fig. 3). One important step in the PHM catalytic reaction is the transfer of an electron from CuA^{+} (in the A domain) to CuB^{2+} (in the opposing B domain) (15, 18). The two Cu ions are held 11 Å apart on either side of the interdomain cleft and it was recently proposed that the electron is transferred from CuA^{+} *via* His-108 (referring to the rat PHM, see Fig. 3), Gln-170, the C-terminus of the propeptide substrate, and His-242 to CuB^{2+} (18) (these residues are marked with either an asterisk or a filled square in Fig. 3). The striking conservation of several other amino acid residues near the Cu ion binding sites, however, suggests that also these additional residues must play a crucial role, either in the electron transfer reaction, substrate binding, or in the precise conservation of the 3D structure of the catalytic pocket of PHM. Other blocks of conserved amino acid residues further away from the catalytic core might be involved in the overall 3D structure of the enzyme.

Eight introns in the sea anemone gene occur in the same positions and have the same intron phasing as eight introns in the PHM part of the rat PAM gene (Table 1; filled triangles in Fig. 3). Thus, independently from the striking structural similarities between the rat and sea anemone PHMs, their genomic organizations further confirm that the two proteins are clearly evolutionarily related.

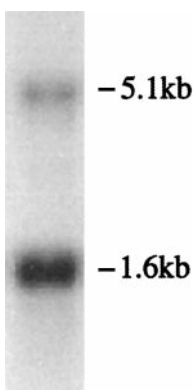

FIG. 4. Northern blot analysis of *Calliactis* mRNA (5 μg/lane) with a cDNA fragment coding for nucleotide positions 592–1216 of Fig. 2 from Ref. (16). A prominent band at 1.6 kb, corresponding very well to the cloned sea anemone PHM (CP1-A/CP1-B) cDNA can be seen, as well as a weak band at 5.1 kb that presently cannot be interpreted.

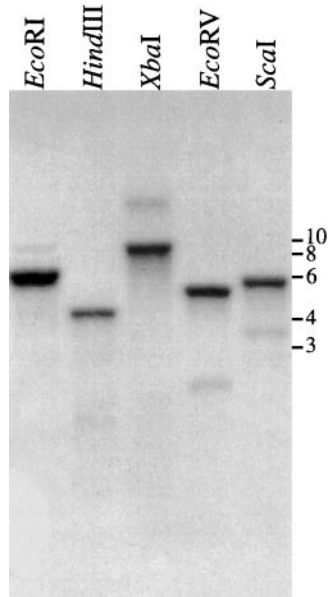

FIG. 5. Southern blot analysis. Genomic DNA (12 μg) from *C. parasitica* was digested with *Eco*RI, *Hin*dIII, *Xba*I, *Eco*RV, or *Sca*I. The hybridization probe corresponds to nucleotide sequences 717–1156 of Fig. 2 from Ref. (16). Single strongly hybridizing and weakly hybridizing bands appeared in each lane. The size of the hybridizing bands is given in kb.

The CP1 gene yields two splicing variants CP1-A and -B, which differ by the alternative use of exons-8 and -9 (Figs. 2 and 3). Alternative splicing in the rat PHM has only been found for the gene regions coding for the C-terminal parts of PHM (or PAM) (13, 19, 20) and not for the middle regions of PHM, as is the case with sea anemones. Exons-8 and -9 code for regions located in the hinge, connecting the two domains of PHM (this region corresponds to the hinge region between residues Ser-187 and Ser-226 of rat PHM, see Fig. 3 and Ref. 14). The alternative use of the two exons in CP1-A and -B, therefore, does not directly effect the substrate binding sites mentioned above. The slightly different size and the different number of Pro residues in the two resulting hinge parts, however, could indirectly affect the 3D structure in parts of the sea anemone PHMs and, thereby, lead to two enzymes with different properties, e.g., different substrate specificities.

ACKNOWLEDGMENTS

We want to thank Drs. Mario Amzel, Betty Eipper, and Richard Mains for discussions, Lotte Steffensen for typing the manuscript, and the Danish Natural Science Research Council and Carlsberg Foundation for financial support.

REFERENCES

1. Gierer, A. (1977) *Curr. Top. Dev. Biol.* **11,** 17–59.
2. Gierer, A., Berking, S., Bode, H., David, C. N., Flick, K., Hansmann, G., Schaller, H., and Trenckner, E. (1972) *Nature New Biol.* **239,** 98–101.
3. Minobe, S., Koizumi, O., and Sugiyama, T. (1995) *Dev. Biol.* **172,** 170–181.
4. Mackie, G. O. (1990) *Am. Zool.* **30,** 907–920.
5. Grimmelikhuijzen, C. J. P., Leviev, I., and Carstensen, K. (1996) *Int. Rev. Cytol.* **167,** 37–89.
6. Leitz, T., Morand, K., and Mann, M. (1994) *Dev. Biol.* **163,** 440–446.
7. Takahashi, T., Muneoka, Y., Lohmann, J., Lopez de Haro, M. S., Solleder, G., Bosch, T. C. G., David, C. N., Bode, H. R., Koizumi, O., Shimizu, H., Hatta, M., Fujisawa, T., and Sugiyama, T. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1241–1246.
8. Yum, S., Takahashi, T., Koizumi, O., Ariura, Y., Kobayakawa, Y., Mohri, S., and Fujisawa, T. (1998a) *Biochem. Biophys. Res. Commun.* **248,** 584–590.
9. Takahashi, T., Koizumi, O., Ariura, Y., Romanovitch, A., Bosch, T. C. G., Kobayakawa, Y., Mohri, S., Bode, H. R., Yum, S., Hatta, M., and Fujisawa, T. (2000) *Development* **127,** 997–1005.
10. Bradbury, A. F., Finnie, M. D. A., and Smyth, D. G. (1982) *Nature* **298,** 686–688.
11. Perkins, S. N., Husten, E. J., and Eipper, B. A. (1990) *Biochem. Biophys. Res. Commun.* **171,** 926–932.
12. Bradbury, A. F., and Smyth, D. G. (1991) *Trends Biochem. Sci.* **16,** 112–115.
13. Eipper, B. A., Stoffers, D. A., and Mains, R. E. (1992) *Annu. Rev. Neurosci.* **15,** 57–85.
14. Kolhekar, A. S., Keutmann, H. T., Mains, R. E., Quon, A. S. W., and Eipper, B. A. (1997) *Biochemistry* **36,** 10901–10909.
15. Prigge, S. T., Kolhekar, A. S., Eipper, B. A., Mains, R. E., and Amzel, L. M. (1997) *Science* **278,** 1300–1305.
16. Hauser, F., Williamson, M., and Grimmelikhuijzen, C. J. P. (1997) *Biochem. Biophys. Res. Commun.* **241,** 509–512.
17. Eriksen, K. K., Hauser, F., Schiøtt, M., Pedersen, K.-M., Søndergaard, L., and Grimmelikhuijzen, C. J. P. (2000) *Genome Res.* **10,** 924–938.
18. Prigge, S. T., Kolhekar, A. S., Eipper, B. A., Mains, R. E., and Amzel, L. M. (1999) *Nature Struct. Biol.* **6,** 976–983.
19. Eipper, B. A., Green, C. B.-R., Campbell, T. A., Stoffers, D. A., Keutmann, H. T., Mains, R. E., and Quafik, L. (1992) *J. Biol. Chem.* **267,** 4008–4015.
20. Quafik, L., Stoffers, D. A., Campbell, T. A., Johnson, R. C., Bloomquist, B. T., Mains, R. E., and Eipper, B. A. (1992) *Mol. Endocrinol.* **6,** 1571–1584.
21. Spijker, S., Smit, A. B., Eipper, B. A., Malik, A., Mains, R. E., and Geraerts, W. P. M. (1999) *FASEB J.* **13,** 735–748.
22. Kolhekar, A. S., Roberts, M. S., Jiang, N., Johnson, R. C., Mains, R. E., Eipper, B. A., and Taghert, P. H. (1997) *J. Neurosci.* **17,** 1363–1376.
23. Mizuno, K., Ohsuye, K., Wada, Y., Fuchimura, K., Tanaka, S., and Matsuo, H. (1987) *Biochem. Biophys. Res. Commun.* **148,** 546–552.
24. Ohsuye, K., Kitano, K., Wada, Y., Fuchimura, K., Tanaka, S., Mizuno, K., and Matsuo, H. (1988) *Biochem. Biophys. Res. Commun.* **150,** 1275–1281.