Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 11647–11651, December 1995
Neurobiology

Molecular cloning of a preprohormone from sea anemones containing numerous copies of a metamorphosis-inducing neuropeptide: A likely role for dipeptidyl aminopeptidase in neuropeptide precursor processing

(development/neurotransmitter/peptide hormone/posttranslational modification/cnidaria)

ILIA LEVIEV AND CORNELIS J. P. GRIMMELIKHUIJZEN

Department of Cell Biology and Anatomy, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

Communicated by John Kendrew, Cambridge, United Kingdom, August 15, 1995

ABSTRACT **Neuropeptides are an important group of hormones mediating or modulating neuronal communication. Neuropeptides are especially abundant in evolutionarily "old" nervous systems, such as those of cnidarians, the lowest animal group having a nervous system. Cnidarians often have a life cycle including a polyp, a medusa, and a planula larva stage. Recently, a neuropeptide, <Glu-Gln-Pro-Gly-Leu-Trp-NH_2, has been isolated from sea anemones that induces metamorphosis in a hydroid planula larva to become a hydropolyp [Leitz, T., Morand, K. & Mann, M. (1994) *Dev. Biol.* 163, 440–446]. Here, we have cloned the precursor protein for this metamorphosis-inducing neuropeptide from sea anemones. The precursor protein is 514-amino acid residues long and contains 10 copies of the immature, authentic neuropeptide (Gln-Gln-Pro-Gly-Leu-Trp-Gly). All neuropeptide copies are preceded by Xaa-Pro or Xaa-Ala sequences, suggesting a role for dipeptidyl aminopeptidase in neuropeptide precursor processing. In addition to these neuropeptide copies, there are 14 copies of another, closely related neuropeptide sequence (Gln-Asn-Pro-Gly-Leu-Trp-Gly). These copies are flanked by basic cleavage sites and, therefore, are likely to be released from the precursor protein. Furthermore, there are 13 other, related neuropeptide sequences having only small sequence variations (the most frequent sequence: Gln-Pro-Gly-Leu-Trp-Gly, eight copies). These variants are preceded by Lys-Arg, Xaa-Ala, or Xaa-Pro sequences, and are followed by basic cleavage sites, and, therefore, are also likely to be produced from the precursor. Thus, there are at least 37 closely related neuropeptides localized on the precursor protein, making this precursor one of the most productive preprohormones known so far. This report also shows that unusual processing sites are common in cnidarian preprohormones.**

Cnidarians have the simplest nervous system in the animal kingdom, and it was probably within this group of animals, or in a closely related ancestor phylum, that nervous systems first evolved (1). The primitive nervous systems of cnidarians are strongly peptidergic. From a single sea anemone species, *Anthopleura elegantissima*, we have recently isolated 16 different neuropeptides (for review, see ref. 2). These peptides are all structurally related and have the C-terminal sequence Arg-Xaa-NH_2 or Lys-Xaa-NH_2 in common, where Xaa is Ala, Asn, Ile, Phe, Pro, and Trp. All peptides are located in dense-core secretory vesicles of neurons and have excitatory or inhibitory actions on muscle preparations or isolated muscle cells, suggesting that they are neurotransmitters or neuromodulators (2–4).

Cnidarians often have a life cycle including a polyp, a medusa, and a planula larva stage. Recently, Leitz and coworkers (5, 6) have isolated a neuropeptide from *A. elegantissima* that induces metamorphosis in planula larvae of the marine hydroid *Hydractinia echinata*. This peptide, <Glu-Gln-Pro-Gly-Leu-Trp-NH_2 [metamorphosin A (MMA)], has an interesting structure because it does not belong to the large family of Arg-Xaa-NH_2 or Lys-Xaa-NH_2 neuropeptides present in sea anemones. The work of Leitz *et al.* (5, 6) is also interesting because it shows that neuropeptides in cnidarians, in addition to being transmitters, also can be hormones that control developmental processes such as metamorphosis. In the present study, we have cloned the preprohormone for this metamorphosis-inducing neuropeptide.*

METHODS

cDNA Library of Sea Anemones. *A. elegantissima* was obtained from Biomarine Laboratories (Venice, CA). We used an amplified cDNA library that was derived from a nonamplified cDNA library described earlier (7).

Radioactive Labeling of DNA Probes. DNA was labeled with [γ-^{32}P]dATP (Amersham; specific activity, 110 TBq/mmol) using T4 polynucleotide kinase from Amersham (8). The specific activity was usually $> 1 \times 10^9$ cpm/μg of DNA.

Screening of cDNA Library. The cDNA library of *A. elegantissima* was screened with a mixed pool of ^{32}P-labeled oligonucleotides CA(A/G)CA(A/G)CCIGGI(C/T)TITGGGG, which is able to hybridize with all possible DNA sequences corresponding to the amino acid sequence Gln-Gln-Pro-Gly-Leu-Trp-Gly. Plaque lifting and processing of nitrocellulose filters were done as described in ref. 8. Dry filters were baked for 2 hr at 80°C and prehybridized for 4 hr at 37°C in a solution containing 6× SSC (1× SSC is 150 mM NaCl/15 mM sodium citrate, pH 7.0)/5× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin/0.02% polyvinylpyrrolidone/0.02% Ficoll)/0.1% SDS/0.01% herring sperm DNA. Filters were hybridized for 18 hr at 37°C in the same buffer containing the radioactive probe at 5×10^5 cpm/ml. Washing was three times for 45 min in 2× SSC/0.1% SDS at room temperature. The cDNA library was also screened with a labeled cDNA insertion of clone P41 (Fig. 1). When this probe was used, filters were hybridized overnight at 42°C in 50% (vol/vol) formamide/5× SSPE (1× SSPE is 150 mM NaCl/10 mM sodium diphosphate/1 mM EDTA, pH 7.4)/5× Denhardt's solution/0.1% SDS and washed twice for 1 hr at 65°C with 1× SSC/0.1% SDS.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: DPAP, dipeptidyl aminopeptidase; LWamide, Leu-Trp-NH_2; MMA, metamorphosin A.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. U34781).

```
CAATGAACTGAGTGGAACACAAGTAATACATATTCTTCACTTCGGTTGATA ATG GCC CTC AAG TGT CAT CTA GTT CTA CTG      81
                                                     Met Ala Leu Lys Cys His Leu Val Leu Leu      10

GCC ATT ACT TTA CTA TTA GCA CAG TGT TCA GGG TCA GTA GAC AAG AAG GAT AGT ACG ACG AAT CAC TTA     150
Ala Ile Thr Leu Leu Leu Ala Gln Cys Ser Gly Ser Val Asp Lys Lys Asp Ser Thr Thr Asn His Leu      33

GAT GAG AAG AAA ACA GAT TCC ACA GAA GCA CAT ATT GTA CAA GAA ACA GAC GCG TTA AAA GAA AAT TCT     219
Asp Glu Lys Lys Thr Asp Ser Thr Glu Ala His Ile Val Gln Glu Thr Asp Ala Leu Lys Glu Asn Ser      56

TAT CTT GGC GCC GAG GAG GAA TCT AAA GAA GAA GAC AAG AAG AGA TCC GCC GCT CCT CAG CAG CCT GGC     288
Tyr Leu Gly Ala Glu Glu Glu Ser Lys Glu Glu Asp Lys Lys Arg Ser Ala Ala Pro Gln Gln Pro Gly      79

CTC TGG GGG AAA CGC CAG AAA ATA GGA CTA TGG GGA AGA TCC GCT GAC GCA GGA CAG CCA GGC CTC TGG     357
Leu Trp Gly Lys Arg Gln Lys Ile Gly Leu Trp Gly Arg Ser Ala Asp Ala Gly Gln Pro Gly Leu Trp     102

GGC AAA CGA CAA AGT CCC GGA TTA TGG GGA AGA TCC GCT GAC GCA GGA CAG CCA GGC CTC TGG GGC AAA     426
Gly Lys Arg Gln Ser Pro Gly Leu Trp Gly Arg Ser Ala Asp Ala Gly Gln Pro Gly Leu Trp Gly Lys     125

CGT CAA AAT CCC GGA TTA TGG GGA AGA TCC GCT GAC GCA GGA CAG CCA GGC CTC TGG GGC AAA CGT CAA     495
Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Ala Gly Gln Pro Gly Leu Trp Gly Lys Arg Gln     148

AAT CCC GGA TTA TGG GGA AGA TCG GCT GAC GCA GGA CAG CCA GGC CTC TGG GGC AAA CGT CAA AAT CCC     564
Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Ala Gly Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro     171

GGA TTA TGG GGA AGG TCC GCT GAC GCA AGA CAA CCC GGA CTC TGG GGC AAA CGT GAA ATC TAC GCA TTA     633
Gly Leu Trp Gly Arg Ser Ala Asp Ala Arg Gln Pro Gly Leu Trp Gly Lys Arg Glu Ile Tyr Ala Leu     194

TGG GGA GGA AAA CGT CAA AAT CCC GGA CTT TGG GGA AGA TCC GCT GAT CCA GGA CAG CCC GGC CTC TGG     702
Trp Gly Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Pro Gly Gln Pro Gly Leu Trp     217

GGC AAA CGT GAA CTC GTC GGA TTA TGG GGG GGA AAA CGT CAA AAC CCC GGA TTG TGG GGA AGA TCG GCT     771
Gly Lys Arg Glu Leu Val Gly Leu Trp Gly Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala     240

GAA GCA GGA CAG CCA GGA CTT TGG GGA AAA CGC CAA AAA ATA GGA TTG TGG GGA CGT TCG GCT GAC CCA     840
Glu Ala Gly Gln Pro Gly Leu Trp Gly Lys Arg Gln Lys Ile Gly Leu Trp Gly Arg Ser Ala Asp Pro     263

CTT CAG CCT GGC CTC TGG GGC AAA CGT CAA AAT CCC GGA TTA TGG GGA AGA TCT GCT GAC CCG CAG CAG     909
Leu Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Pro Gln Gln     286

CCT GGC CTC TGG GGC AAA CGT CAA AAT CCC GGA TTA TGG GGA AGA TCT GCT GAC CCG CAG CAG CCT GGC     978
Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Pro Gln Gln Pro Gly     309

CTC TGG GGC AAA CGT CAA AAT CCC GGA TTA TGG GGA AGA TCT GCT GAC CCG CAG CAG CCT GGC CTC TGG    1047
Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Pro Gln Gln Pro Gly Leu Trp     332

GGC AAA CGT CAA AAT CCC GGA TTA TGG GGA AGA TCT GCT GAC CCG CAG CAA CCT GGC CTC TGG GGC AAA    1116
Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Asp Pro Gln Gln Pro Gly Leu Trp Gly Lys     355

AGC CCC GGT TTA TGG GGA CGA TCC GCT GAC CCA CAA CAG CCT GGA CTT TGG GGG AAA CGC CAA AAT CCC    1185
Ser Pro Gly Leu Trp Gly Arg Ser Ala Asp Pro Gln Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro     378

GGA TTT TGG GGA AGA TCT GCT GAC CCG CAG CAG CCT GGC CTC TGG GGC AAA CGT CAA AAT CCC GGA TTA    1254
Gly Phe Trp Gly Arg Ser Ala Asp Pro Gln Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu     401

TGG GGA AGA TCT GCT GAC CCG CAG CAA CCT GGC CTC TGG GGC AAA CGT CAA AAT CCC GGA TTA TGG GGA    1323
Trp Gly Arg Ser Ala Asp Pro Gln Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly     424

AGA TCT GCT GAC CCG CAG CAA CCT GGC CTC TGG GGC AAA CGT CAA AAC CCC GGT TTA TGG GGA CGA TCC    1392
Arg Ser Ala Asp Pro Gln Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser     447

GCT GAC CCA CAA CAG CCT GGA CTT TGG GGG AAA CGC CAA AAT CCA GGA CTA TGG GGA AGA AGT GCT GGC    1461
Ala Asp Pro Gln Gln Pro Gly Leu Trp Gly Lys Arg Gln Asn Pro Gly Leu Trp Gly Arg Ser Ala Gly     470

TCC GGT CAA CTC GGA CTT TGG GGT AAA AGG CAA TCA CGC ATT GGA TTA TGG GGA AGA TCT GCC GAG CCT    1530
Ser Gly Gln Leu Gly Leu Trp Gly Lys Arg Gln Ser Arg Ile Gly Leu Trp Gly Arg Ser Ala Glu Pro     493

CCA CAA TTT GAA GAT TTA GAA GAT TTA AAG AAA AAA TCA GCA ATT CCC CAA CCA AAA GGA CAA TGA TAA    1599
Pro Gln Phe Glu Asp Leu Glu Asp Leu Lys Lys Lys Ser Ala Ile Pro Gln Pro Lys Gly Gln Stop Stop   514

TATCCTAGGATCTTCAAAAGTTATCCCGATCATCAATCCCCGGACAAGAGATATTTTAATTTCTGCCGCACGATTGACAGTTCCATTCCAT    1690
TACGAAGAACAAAAAGCTACGTTTCTTTAAGATAATAAATCAAATTCAATATTGTTTGAAGCAATGCACTTCAGGTTTTCACACAAAACTA    1781
ATACAAAAGTTATAAACATAAATAATAAACAAAAAAGGGGTAAGAAACCTGGTTTTTCGTTTTAGACTTTTCAAATTGGTCCTGCATGTCA    1872
TTGTAAGAGGTCGGCAGTAGAAGCTCTTTGAAACACTGTAAGTAGTCATTGAATGCACATTTATGGCAATAAAGATCACGCAGTGGTCTGC    1963
GGCCGCGAA                                                                                       1972
```

FIG. 1. cDNA and deduced amino acid sequence of the MMA precursor from *A. elegantissima*. The cDNA is from clone P41. Nucleotides are numbered from the 5′ to 3′ end, and the amino acid residues are numbered starting with the first ATG of the open reading frame. All authentic, immature MMA sequences are underlined and printed in boldface type. Highly likely, but putative neuropeptide sequences are underlined only. Amino acid sequences that might not give rise to mature neuropeptides or residues that might not be contained in the mature peptides are underlined by a dashed line. Possible polyadenylylation signals are underlined twice.

DNA Sequencing and Sequence Analyses. cDNA insertions from positive λgt11 clones were excised from the λgt11 arms with *Eco*RI and subcloned into pBluescript (Stratagene). Their sequences were determined by the dideoxynucleotide chain-termination method (9) using the Sequenase sequencing kit from United States Biochemical and [α-^{35}S]αATP from Amersham. All subclones were sequenced in both directions. Central regions of the long insertion of clone P41 were sequenced after construction of deletion clones in pBluescript using the *Exo* III/S1 kit from Pharmacia. DNA sequence compilation, nucleotide and amino acid sequence comparisons, and data base searches were done by using the DNASTAR program (DNAstar, Madison, WI).

Northern Blots. For Northern blots, we isolated mRNA from *A. elegantissima* using the Oligotex Direct mRNA kit from Qiagen. Poly(A)$^+$ RNA was denatured with formaldehyde and electrophoresed on 1.2% agarose gels as described in ref. 8. Further steps were done as in ref. 7.

RESULTS

Isolation and Characterization of the cDNA Encoding the Precursor. We screened 5×10^5 phages of an amplified λgt11 cDNA library of *A. elegantissima* using a mixed pool of oligonucleotides corresponding to Gln-Gln-Pro-Gly-Leu-Trp-Gly, which is the presumed amino acid sequence of unprocessed MMA. We obtained three positive clones, all having an identical insertion of 1972-bp coding for the neuropeptide precursor, suggesting that they originated from a single, primary clone in the nonamplified cDNA library. Rescreening of our amplified cDNA library (5×10^5 phages) with the cDNA insertion of one of these clones (clone P41) yielded 12 positive clones that, again, had insertions fully identical with that of clone P41. The nucleotide sequence of clone P41 is shown in Fig. 1. The cDNA region coding for the neuropeptide precursor protein extends from nucleotide 52 to nucleotide 1593. The coding region is preceded by a 51-bp leader sequence containing several stop codons and is followed by a trailer sequence of 379 bp. At nucleotide positions 1905–1910 and 1940–1945 are two consensus sequences for mRNA polyadenylylation (10).

The cDNA insertion of clone P41 was used as a hybridization probe for a Northern blot analysis using poly(A)$^+$ mRNA isolated from *A. elegantissima* (whole animal). This analysis gave a single band corresponding to mRNA of about 2100 bases (data not shown because of space limitations). This result indicates that the cDNA of clone P41 represents most of the mRNA coding for the neuropeptide precursor.

Primary Structure of Precursor Protein. The precursor protein deduced from the cDNA of Fig. 1 is 514-amino acid residues long. The N terminus of the protein has a hydrophobic signal sequence needed for translocation across the membrane of the endoplasmic reticulum. The signal sequence is probably cleaved off at Ala-17 (11).

The precursor protein contains 10 copies of unprocessed MMA (Gln-Gln-Pro-Gly-Leu-Trp-Gly) at amino acid positions 76–82, 285–291, 306–312, 327–333, 348–354, 367–373, 388–394, 409–415, 430–436, and 451–457 (underlined and in boldface type in Fig. 1). The last 9 copies are regularly distributed over the C-terminal half of the protein and are separated by spacers of about 20 amino acid residues long. All 10 immature neuropeptide sequences have basic residues (Lys-Arg or Lys) at their C termini (Table 1), which are established sites for endoproteolytic precursor cleavage (12). At their N termini, however, 9 neuropeptide copies are preceded by the sequence Ser-Ala-Asp-Pro, and one copy is preceded by the sequence Ser-Ala-Ala-Pro (Fig. 1; Table 1). As authentic <Glu-Gln-Pro-Gly-Leu-Trp-NH_2 has been isolated from sea anemone extracts, this clearly proves that there must be processing at the C-terminal sides of Pro residues. The most likely enzyme catalyzing this cleavage is dipeptidyl aminopeptidase (DPAP), which cleaves at the C-terminal side of N-terminal Xaa-Pro and Xaa-Ala sequences (13). DPAP would remove the N-terminal elongations of MMA in two steps, cleaving first after Ser-Ala and subsequently after the remaining Asp-Pro or Ala-Pro sequences (Table 1).

In addition to the 10 copies of immature MMA, there is a large number of other, putative neuropeptide sequences that are closely related to authentic MMA (Table 1). For reasons of simplicity, we have named the most frequent, putative peptide Antho-LWamide I (14 copies; LWamide is Leu-Trp-NH_2), the authentic metamorphosis-inducing peptide (MMA) isolated by Leitz *et al.* (5), Antho-LWamide II (10 copies), a third peptide occurring in high frequency Antho-LWamide III (6 copies), and other, closely related peptides, Antho-LWamides IV–IX (Table 1). The 14 immature Antho-LWamide I sequences are followed by single basic amino acid residues (Arg) and preceded by the processing sequence Lys-Arg (Table 1). Thus, although Antho-LWamide I is still a putative peptide that has not been isolated yet, it probably exists and is released in a very high copy number from its precursor. The Antho-LWamides IV, VII, and VIII have exactly the same processing sites as Antho-LWamide I, and it is likely that also these peptides will be released from the Antho-LWamide precursor. All mature Antho-LWamides I, II, IV, VII, and VIII will have an N-terminal < Glu group and the C-terminal structure Gly-Leu-Trp-NH_2 (Table 1).

There are other, putative peptides (the Antho-LWamides III, V, VI, and IX) that are very similar, or nearly identical, to the authentic peptide MMA (Antho-LWamide II). Their immature sequences are followed by Lys-Arg residues, and they are preceded by basic residues and Xaa-Ala or Xaa-Pro sequences (Table 1), so they will probably be released from the precursor and processed. Their final structures, however, are

Table 1. N- and C-terminal extensions of MMA and related, putative neuropeptide sequences

N- and C-terminal extensions and neuropeptide sequence	Copy no.	Name
Arg↓Ser-Ala-Asp-Pro-**Gln-Gln-Pro-Gly-Leu-Trp-Gly**-Lys-Arg↓[a]	8	MMA (Antho-LWamide II)
Arg↓Ser-Ala-Asp-Pro-**Gln-Gln-Pro-Gly-Leu-Trp-Gly**-Lys↓[b]	1	MMA (Antho-LWamide II)
Arg↓Ser-Ala-Ala-Pro-**Gln-Gln-Pro-Gly-Leu-Trp-Gly**-Lys-Arg↓[c]	1	MMA (Antho-LWamide II)
Lys-Arg↓Gln-Asn-Pro-Gly-Leu-Trp-Gly-Arg↓[d]	14	Antho-LWamide I
Lys-Arg↓Gln-Ser-Pro-Gly-Leu-Trp-Gly-Arg↓[e]	1	Antho-LWamide VII
Lys-Arg↓Gln-Lys-Ile-Gly-Leu-Trp-Gly-Arg↓[f]	2	Antho-LWamide IV
Lys-Arg↓Gln-Ser-Arg-Ile-Gly-Leu-Trp-Gly-Arg↓[g]	1	Antho-LWamide VIII
Arg↓Ser-Ala-Gly-Ser-Gly-Gln-Leu-Gly-Leu-Trp-Gly-Lys-Arg↓[h]	1	Antho-LWamide IX
Arg↓Ser-Ala-Asp-Ala-Gly-Gln-Pro-Gly-Leu-Trp-Gly-Lys-Arg↓[i]	4	Antho-LWamide III
Arg↓Ser-Ala-Glu-Ala-Gly-Gln-Pro-Gly-Leu-Trp-Gly-Lys-Arg↓[j]	1	Antho-LWamide III
Arg↓Ser-Ala-Asp-Pro-Gly-Gln-Pro-Gly-Leu-Trp-Gly-Lys-Arg↓[k]	1	Antho-LWamide III
Arg↓Ser-Ala-Asp-Pro-Leu-Gln-Pro-Gly-Leu-Trp-Gly-Lys-Arg↓[l]	1	Antho-LWamide V
Arg↓Ser-Ala-Asp-Ala-Arg-Gln-Pro-Gly-Leu-Trp-Gly-Lys-Arg↓[m]	1	Antho-LWamide VI
Lys↓Ser-Pro-Gly-Leu-Trp-Gly-Arg↓[n]	1	
Lys-Arg↓Glu-Leu-Val-Gly-Leu-Trp-Gly-Gly-Lys-Arg↓[o]	1	
Lys-Arg↓Glu-Ile-Tyr-Ala-Leu-Trp-Gly-Gly-Lys-Arg↓[p]	1	
Arg↓Ser-Ala-Glu-Pro-Pro-Gln-Phe-Glu-Asp-Leu-Glu-Asp-Leu-Lys-Lys-Lys↓[q]	1	

Sites of initial cleavage at tribasic, dibasic, or monobasic residues are indicated by arrows. MMA copies are underlined and printed in boldface type. Highly likely, but putative, peptide sequences are underlined only. Uncertain mature sequences or residues are underlined by a dashed line. The underlined neuropeptide sequences can be found in Fig. 1 at the following amino acid positions: a, 285, 306, 327, 367, 388, 409, 430, 451; b, 348; c, 76; d, 127, 148, 169, 200, 231, 273, 294, 315, 336, 376, 397, 418, 439, 460; e, 106; f, 85, 252; g, 481; h, 470; i, 97, 118, 139, 160; j, 243; k, 212; l, 264; m, 181; n, 358; o, 221; p, 190; q, 492.

still unclear (see *Discussion*). Finally, there are other, putative peptide sequences flanked by basic residues, but it is not certain whether these will yield intact peptides (underlined by a dashed line in Table 1) (see *Discussion*).

The amino acid sequence of the Antho-LWamide precursor and its corresponding cDNA show no significant sequence similarly to other amino acid or nucleotide sequences contained in the GenBank or EMBL data bases.

DISCUSSION

We have cloned the precursor protein for the metamorphosis-inducing neuropeptide MMA that has recently been isolated from *A. elegantissima* by Leitz and coworkers (5, 6). The precursor protein contains 10 copies of authentic, immature MMA (Table 1). Nine of these neuropeptide copies are followed by pairs of basic residues, which are established cleavage sites for endoproteolytic precursor processing (12). One authentic, immature MMA copy (at positions 348–354 of Fig. 1; see also Table 1) is followed by a single basic residue. Processing at single basic residues is also known, and it is especially frequent in certain invertebrate prohormones (7, 12, 14–18). After comparing various prohormones that are cleaved at single basic residues, Devi (14) has proposed several rules that have to be fulfilled in order to obtain cleavage. Many of these requirements are satisfied for the immature MMA copy that is followed by the single basic residue. One important rule, however, says that at position −3, −5, or −7 of the basic residue where cleavage occurs, a second basic residue should be present. This rule is not followed. Nevertheless, we think that the MMA copy at positions 348–354 will be released from its precursor by cleavage at the C-terminal side of its C-terminally-located Lys residue. This is for the following reasons: (*i*) the surroundings of this MMA copy is nearly identical to that of the other nine copies of authentic MMA, and it lies in the middle of a regular row of MMA copies (Fig. 1; Table 1), (*ii*) the 14 copies of Antho-LWamide I that are nearly identical to MMA (one Gln has been exchanged for the conserved amino acid residue Asn) are followed by a single basic amino acid residue (Arg), and also here Devi's rule (14) does not hold. Nevertheless, these copies have to be released from the precursor, in order to release the peptides containing the copies of authentic MMA (Fig. 1; also see below). The −3, −5, −7 rule of Devi (14), therefore, might not be generally applicable.

All 10 copies of authentic MMA are preceded by Xaa-Ala and Xaa-Pro sequences (Table 1), suggesting that DPAP might be responsible for the final processing of the immature neuropeptides. So far, DPAP has not been recognized as a neuronal processing enzyme for neuropeptide precursors. DPAP, however, is known to be involved in the final processing of immature, N-terminal extended yeast α-mating factor, honey bee melittin, cecropin from moth, and caerulein and xenopsin from frog skin (13). It is interesting that the Xaa residue in the Xaa-Ala and Xaa-Pro sequences of the N-terminally extended yeast, insect, and frog skin peptides very often is Asp or Glu (13). In pro-α-mating factor of *Saccharomyces cerevisiae*, for example, all three Xaa residues in each N-terminal Xaa-Ala-Xaa-Ala-Xaa-Ala extension of the four copies of α-mating factor are acidic (19), and in pro-melittin 9 out of 11 Xaa residues are either Asp or Glu (20). This fact fits very well with the presence of Asp-Pro sequences in the C-terminal halves of 9 out of 10 N-terminal extensions of the immature MMA sequences (Table 1). Furthermore, the N-terminal extension of immature xenopsin is Ser-Ala-Glu-Ala, and N-terminally elongated xenopsin is known to be processed by DPAP (13, 21). The Ser-Ala dipeptide sequence can be found back in the N-terminal extensions of all 10 immature MMA copies (Table 1). All this, then, clearly points to a role of DPAP in the final trimming of immature MMA.

In addition to the 10 copies of immature MMA, 14 copies of immature Antho-LWamide I, two copies of immature Antho-LWamide IV, and one copy of each Antho-LWamide VII and VIII are present on the precursor protein (Table 1). These 18 immature neuropeptide copies are preceded by the dibasic sequence Lys-Arg and followed by a single Arg residue. It is fortunate that many of the Antho-LWamide I sequences are sandwiched between two MMA sequences (Fig. 1). Because authentic MMA sequences have to be released from the precursor, this implies that the Antho-LWamide I sequences also should be released. This release of Antho-LWamide I sequences probably occurs by endoproteolytic cleavage at the C-terminal sides of the Lys-Arg sequences (to liberate the C termini of immature MMA) and at the C-terminal sides of the single Arg residues (to liberate the N termini of Ser-Ala-Xaa-Pro-MMA) (Table 1, Fig. 1). Therefore, it is very likely that processing will occur at single Arg residues, despite the fact that here also these residues do not obey the −3, −5, −7 rule of Devi (14). As most of the Antho-LWamide I sequences have to be excised from the precursor protein, the other Antho-LWamide I sequences and the Antho-LWamide IV, VII, and VIII sequences will probably also be liberated because they have the same N- and C-terminal processing sites (Table 1) and they lie in regular rows, with similar spacer sequences, in the precursor protein (Fig. 1).

The precursor protein also contains six copies of Antho-LWamide III and one copy each of Antho-LWamide V, VI, and IX (Table 1). These immature peptide copies are followed by Lys-Arg sequences and preceded by the sequences Ser-Ala, Ser-Ala-Xaa-Pro or Ser-Ala-Xaa-Ala, where Xaa is an acidic residue (Table 1). These N-terminal extensions are, again, very similar or identical (Ser-Ala-Glu-Ala) to the N termini of other peptide sequences where DPAP is known to be responsible for the final peptide trimming (13). Therefore, the Antho-LWamides III, V, VI, and IX are probably released from the precursor and subsequently processed into their final structures. It is unclear, however, what the final structures of Antho-LWamide III, V, VI, and IX are. They might be N-terminally elongated by a single Gly, Leu, or Arg residue, by the sequence Gly-Ser-Gly (Table 1), or start like the other Antho-LWamides, with an N-terminal <Glu group. If the last possibility is correct, there should be processing at N-terminally located Gly, Leu, and Arg residues (Table 1). We favor this latter possibility, but we have, in fact, no good arguments for this choice, other than that this would yield N-terminal <Glu groups of mature neuropeptides (e.g., <Glu-Pro-Gly-Leu-Trp-NH_2, eight copies) that are nearly identical to authentic MMA. Peptide isolation and sequence analysis would clarify this point.

Finally, there are four peptide sequences with a somewhat uncertain status (Table 1). The first peptide sequence Ser-Pro-Gly-Leu-Trp-Gly (at positions 356–361 of Fig. 1) is nearly identical with Antho-LWamide VII, except that it does not contain an N-terminal <Glu group. Therefore, its N terminus will be degraded by DPAP (Ser-Pro) and, if N-terminal processing at Gly and Leu residues exists (see above), nearly the whole peptide sequence will be digested. Two other peptide sequences (at positions 190–196 and 221–227 of Fig. 1; see also Table 1) have the sequence Gly-Leu-Trp-Gly or Leu-Trp-Gly in common with the Antho-LWamides. These two peptide sequences are flanked by dibasic (Lys-Arg) residues. However, from many other cnidarian prohormones it is known that processing can occur at acidic residues (7, 16–18). Therefore, the N-terminal Glu residues could possibly be removed, and if processing at N-terminal Leu (or Ile) exists, much of their N termini would be degraded (Table 1). The last uncertain peptide structure is Ser-Ala-Glu-Pro-Pro-Gln-Phe-Glu-Asp-Leu-Glu-Asp-Leu (at positions 490–502 of Fig. 1). This sequence is flanked by basic amino acid residues (Table 1). If the cnidarian processing enzymes specific for acidic

Table 2. Established and putative, mature neuropeptides that could be released from the MMA precursor protein

Copy no.	Structure	Name
14	<Glu-Asn-Pro-Gly-Leu-Trp-NH_2	Antho-LWamide I
10	<Glu-Gln-Pro-Gly-Leu-Trp-NH_2	MMA (Antho-LWamide II)
6	Gly-Gln-Pro-Gly-Leu-Trp-NH_2	Antho-LWamide III
1	Leu-Gln-Pro-Gly-Leu-Trp-NH_2	Antho-LWamide V
1	Arg-Gln-Pro-Gly-Leu-Trp-NH_2	Antho-LWamide VI
1	<Glu-Ser-Pro-Gly-Leu-Trp-NH_2	Antho-LWamide VII
2	<Glu-Lys-Ile-Gly-Leu-Trp-NH_2	Antho-LWamide IV
1	<Glu-Ser-Arg-Ile-Gly-Leu-Trp-NH_2	Antho-LWamide VIII
1	Gly-Ser-Gly-Gln-Leu-Gly-Leu-Trp-NH_2	Antho-LWamide IX
37		

Amino acid residues identical with those of MMA are boxed. Residues that possibly are not contained in the mature peptides are underlined with a dashed line.

amino acid residues (7, 16–18) are present in the neurosecretory vesicles containing the Antho-LWamide precursor, the peptide sequence would be Ser-Ala-Glu-Pro-Pro-Gln-Phe-Glu. DPAP would remove the N-terminal Ser-Ala sequence of both possible peptide structures, but this enzyme could not digest the remaining peptides any further, because the two consecutive Pro residues at the peptide amino acid positions 2 and 3 form a block that protects the peptide against both DPAP and many other nonspecific aminopeptidases (22).

In summary, the Antho-LWamide precursor contains 37 immature neuropeptide copies that are likely to be released from the precursor and, in addition, 4 neuropeptide sequences that have a more uncertain status. The 37 immature neuropeptide copies probably yield nine different, mature neuropeptides that are closely related: the Antho-LWamides I–IX. Table 2 shows the putative structures of these mature neuropeptides and the extent to which these peptides are structurally related. All neuropeptides have the C-terminal sequence Gly-Leu-Trp-NH_2 in common, and also in the N-terminal parts there are many sequence similarities. Leitz and coworkers (5) have found that the metamorphosis-inducing potency of MMA (Antho-LWamide II) resides in the C-terminal part of the molecule, as N-terminal variants can still induce metamorphosis in *Hydractinia* planula larvae. Thus, many, or perhaps all, of the peptides of Table 2 may have a metamorphosis-inducing capacity.

We thank Astrid Juel Jensen for typing the manuscript and the Danish Natural Science Research Council for financial support.

1. Mackie, G. O. (1990) *Am. Zool.* **30,** 907–920.
2. Grimmelikhuijzen, C. J. P., Carstensen, K., Darmer, D., Moosler, A., Nothacker, H.-P., Reinscheid, R. K., Schmutzler, C., Vollert, H., McFarlane, I. D. & Rinehart, K. L. (1992) *Am. Zool.* **32,** 1–12.
3. Westfall, J. A. & Grimmelikhuijzen, C. J. P. (1993) *Biol. Bull.* **185,** 109–114.
4. McFarlane, I. D., Anderson, P. A. V. & Grimmelikhuijzen, C. J. P. (1991) *J. Exp. Biol.* **156,** 419–431.
5. Leitz, T., Morand, K. & Mann, M. (1994) *Dev. Biol.* **163,** 440–446.
6. Leitz, T. & Lay, M. (1995) *Roux's Arch. Dev. Biol.* **204,** 276–279.
7. Schmutzler, C., Darmer, D., Diekhoff, D. & Grimmelikhuijzen, C. J. P. (1992) *J. Biol. Chem.* **267,** 22534–22541.
8. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
9. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
10. Wickens, M. (1990) *Trends Biochem. Sci.* **15,** 277–281.
11. von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
12. Sossin, W. S., Fisher, J. M. & Scheller, R. H. (1989) *Neuron* **2,** 1407–1417.
13. Kreil, G. (1990) *Trends Biochem. Sci.* **15,** 23–26.
14. Devi, L. (1991) *FEBS Lett.* **280,** 189–194.
15. Taussig, R. & Scheller, R. H. (1986) *DNA* **5,** 453–461.
16. Darmer, D., Schmutzler, C., Diekhoff, D. & Grimmelikhuijzen, C. J. P. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 2555–2559.
17. Schmutzler, C., Diekhoff, D. & Grimmelikhuijzen, C. J. P. (1994) *Biochem. J.* **299,** 431–436.
18. Reinscheid, R. K. & Grimmelikhuijzen, C. J. P. (1994) *J. Neurochem.* **62,** 1214–1222.
19. Fuller, R. S., Sterne, R. E. & Thorner, J. (1988) *Annu. Rev. Physiol.* **50,** 345–362.
20. Vlasak, R., Unger-Ullman, C., Kreil, G. & Frischauf, A.-M. (1983) *Eur. J. Biochem.* **135,** 123–126.
21. Sures, I. & Crippa, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 380–384.
22. Carstensen, K., Rinehart, K. L., McFarlane, I. D. & Grimmelikhuijzen, C. J. P. (1992) *Peptides* **13,** 851–857.