Biochem. J. (1994) 299, 431–436 (Printed in Great Britain)

The primary structure of the Pol-RFamide neuropeptide precursor protein from the hydromedusa *Polyorchis penicillatus* indicates a novel processing proteinase activity

Cornelia SCHMUTZLER,* Dagmar DIEKHOFF* and Cornelis J. P. GRIMMELIKHUIJZEN*†‡

*Centre for Molecular Neurobiology (ZMNH), University of Hamburg, Martinistrasse 52, D-20246 Hamburg, Federal Republic of Germany, and †Department of Cell Biology and Anatomy, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

Neuropeptides containing the C-terminal sequence Arg-Phe-NH_2 (RFamide) occur throughout the Animal Kingdom and are abundant in evolutionarily 'old' nervous systems such as those of cnidarians. From the hydromedusa *Polyorchis penicillatus* we have previously isolated two neuropeptides, Pol-RFamide I (<Glu-Leu-Leu-Gly-Gly-Arg-Phe-NH_2) and Pol-RFamide II (<Glu-Trp-Leu-Lys-Gly-Arg-Phe-NH_2). Here we describe the cloning of a common precursor protein for these peptides from *P. penicillatus*. The precursor protein contains one copy of Pol-RFamide I, 11 copies of Pol-RFamide II and one putative neuropeptide sequence. The Pol-RFamide I sequence is flanked by pairs of basic residues (Arg-Lys). At the C-termini of all Pol-RFamide II sequences, single basic residues (Arg) occur. Paired and single basic residues are established sites for post-translational precursor cleavage. At the N-termini of the Pol-RFamide II sequences, however, basic residues are lacking and, instead, either single Asp (in eight cases) or single Asn residues (in three cases) occur. This means that processing must take place at Asp and/or Asn residues. This is firm evidence for the presence of one or more unconventional processing enzymes. The first type of processing enzyme could be an endoproteinase or aminopeptidase hydrolysing at the C-terminal side of Asp residues. Proteolytic cleavage at acidic amino acid residues has previously been inferred from other cnidarian neuropeptide precursors. The second type of processing enzyme could be an endoproteinase or aminopeptidase hydrolysing at the C-terminal side of Asn residues.

INTRODUCTION

Cnidarians are interesting animals from an evolutionary point of view, because it was probably within this group, or a closely related ancestor, that nervous systems first evolved (Mackie, 1990). In the primitive nervous systems of cnidarians, peptides play an important role [see Grimmelikhuijzen et al. (1992a) for a review]. By using a radioimmunoassay for the C-terminal sequence Arg-Phe-NH_2 (RFamide), we have isolated various RFamide neuropeptides from cnidarians, among them <Glu-Gly-Arg-Phe-NH_2 (Antho-RFamide; where <Glu is pyroglutamic acid) from sea anemones and sea pansies, which belong to the class of Anthozoa, and <Glu-Leu-Leu-Gly-Gly-Arg-Phe-NH_2 (Pol-RFamide I) and <Glu-Trp-Leu-Lys-Gly-Arg-Phe-NH_2 (Pol-RFamide II) from the jellyfish *Polyorchis penicillatus*, which is a member of the Hydrozoa and which is often used as a model animal for electrophysiological investigations (Grimmelikhuijzen and Graff, 1986; Grimmelikhuijzen and Groeger, 1987; Grimmelikhuijzen et al., 1988, 1992b; Anderson and Spencer, 1989). These and several other neuropeptides from other cnidarian species have the C-terminal sequence Gly-Arg-Phe-NH_2 in common and also additional sequence similarities. Antho-RFamide is located in sensory motor neurons of sea anemones and sea pansies and has an excitatory action on various muscle groups of these animals, suggesting that it is a transmitter or modulator at neuromuscular junctions (McFarlane et al., 1987; Anctil and Grimmelikhuijzen, 1989; McFarlane and Grimmelikhuijzen, 1991; Grimmelikhuijzen et al., 1992a). Neurons producing the Pol-RFamides have a widespread distribution in *Polyorchis* and are located both in the periphery, where they are associated with smooth-muscle fibres, and in the two nerve rings of the bell margin, where they appear to be presynaptic to the 'swimming motor neurons' (Grimmelikhuijzen and Spencer, 1984; Grimmelikhuijzen et al., 1988). Recent experiments using voltage clamp of isolated, cultured 'swimming motor neurons' of *Polyorchis* have shown that Pol-RFamide II blocks a voltage-dependent K^+ current (Grigoriev and Spencer, 1992). All these data indicate that cnidarian RFamide peptides are involved in neurotransmission.

We have recently cloned the Antho-RFamide precursor proteins from several anthozoan species. In the sea anemone *Calliactis parasitica*, the precursor contains 19 copies of immature Antho-RFamide and seven other, putative neuropeptide sequences (Darmer et al., 1991). The Antho-RFamide precursor from the sea anemone *Anthopleura elegantissima* contains 13 copies of immature Antho-RFamide and 20 other neuropeptide sequences (Schmutzler et al., 1992). Established cleavage sites (basic residues) follow each Antho-RFamide sequence in both precursor proteins, whereas, at the N-terminus of each Antho-RFamide sequence, acidic residues occur. These acidic residues must be the cleavage sites for a new type of processing enzyme occurring in neurons (Darmer et al., 1991; Schmutzler et al., 1992).

In the present day study we have investigated the biosynthesis

Abbreviations used: RFamide, Arg-Phe-NH_2; Pol-RFamide I, <Glu-Leu-Leu-Gly-Gly-Arg-Phe-NH_2; Pol-RFamide II, <Glu-Trp-Leu-Lys-Gly-Arg-Phe-NH_2; <Glu, pyroglutamic acid; poly(A)$^+$, polyadenylated; DMSO, dimethyl sulphoxide; 1 × SSC, 150 mM NaCl/15 mM sodium citrate, pH 7.0; 1 × Denhardt's solution, 0.02% BSA/0.02% polyvinylpyrrolidone/0.02% Ficoll.

‡ To whom correspondence should be addressed, at the Copenhagen address.

The nucleotide sequence reported in this paper has been submitted to the GenBank and EMBL Nucleotide Sequence Databases under the accession number L14777.

of the Pol-RFamides in the hydrozoan jellyfish *Polyorchis*. Our results show that the acidic processing sites that we previously discovered in neuropeptide precursor proteins of anthozoans do also occur in neuropeptide precursors of hydrozoans and, therefore, in two different and remote classes of Cnidaria. Furthermore, we found evidence for a second, novel type of processing site, suggesting that the cnidarian nervous systems use a variety of 'non-classical' processing signals for the maturation of their neuropeptide precursor proteins.

MATERIALS AND METHODS

cDNA library

Polyorchis penicillatus were kindly collected by Dr. A. N. Spencer (University of Alberta, Edmonton, Alberta, Canada) in Bamfield Inlet, Vancouver Island, Canada, and starved for 24 h. Bell margins were subsequently removed and immediately homogenized in a solution of 4 M guanidine isothiocyanate, 0.1 M mercaptoethanol, 0.01 M EDTA, 2% *N*-lauroylsarcosine sodium salt and 0.05 M Tris/HCl, pH 5. The mixture was then sent to Hamburg. Total RNA was isolated as described by Chirgwin et al. (1979). Polyadenylated [poly(A)$^+$] RNA was prepared by oligo(dT)–cellulose chromatography (Aviv and Leder, 1972). Double-stranded cDNA was synthesized with an oligo(dT) primer using a cDNA synthesis kit from Pharmacia. The cDNA was ligated to *Eco*RI-digested, dephosphorylated λgt11 arms (Promega) and recombinant λ DNA was packaged with Gigapack II Gold extracts from Stratagene.

Screening of the cDNA library

The cDNA library was screened with an 8-fold degenerate pool of 23-mer oligonucleotides. They were synthesized on the basis of the sequence Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly which corresponds to the amino acid sequence of unprocessed Pol-RFamide II. The nucleotide sequence was 5′-CC-A/GAA-ICG-ICC-C/TTT-IAG-CCA-C/TTG-3′. Only four out of six possible codons for both Arg and Leu (CGN and CTN, respectively; N = A,C,T,G) were represented in this oligonucleotide pool. The oligonucleotides were labelled to a specific radioactivity of 2.5×10^8 c.p.m./μg using T4 polynucleotide kinase (GIBCO/BRL) and [γ-^{32}P]ATP (Amersham; specific radioactivity 185 TBq/mmol). Labelled oligonucleotides were applied at a concentration of 5×10^5 c.p.m./ml of hybridization solution. Plaque lifting and hybridization procedures were carried out as described by Maniatis et al. (1982) and Sambrook et al. (1989). The final washing steps were at 45 °C.

DNA sequencing and sequence analysis

DNA sequences were determined by the dideoxynucleotide chain-termination method of Sanger et al. (1977). DNA sequence compilation, nucleotide and amino acid sequence alignment, and database searches were performed using the DNASTAR programme (DNASTAR Inc.).

Northern-blot analysis

A cDNA insertion representing nucleotides 53-983 of Figure 2 (below) was used as a probe for Northern-blot analysis. It was labelled with [α-^{32}P]dCTP using the oligolabelling kit from Pharmacia (Feinberg and Vogelstein, 1983). Poly(A)$^+$ RNA was denatured with glyoxal and dimethyl sulphoxide (DMSO) and electrophoresed on 1.2%-agarose gels as described by Sambrook et al. (1989). The RNA was then blotted on to nylon membranes (Hybond N, Amersham International) by capillary transfer and fixed to the membrane by u.v. irradiation as recommended by the manufacturer. Blots were prehybridized at 42 °C in a solution containing 6 × SSC (1 × SSC is 150 mM NaCl/15 mM sodium citrate, pH 7.0), 1 × Denhardt's solution (0.02% BSA/0.02% polyvinylpyrrolidone/0.02% Ficoll), 0.1% SDS, 0.01% herring sperm DNA and 50% (v/v) formamide. Hybridization was in the same solution with the radioactive probe added at a concentration of 5×10^5 c.p.m./ml. Final washing was 10 min at 42 °C in 2 × SSC/0.1% SDS.

RESULTS

Isolation of cDNAs coding for the Pol-RFamide precursor protein

A λgt11 cDNA library of bell margins of *P. penicillatus* was constructed, and about 3×10^5 recombinant phages were screened with an 8-fold degenerate pool of 23-mer oligonucleotides corresponding to Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly, which is the presumed amino acid sequence of unprocessed Pol-RFamide II (see the Materials and methods section). This yielded 38 hybridizing clones, which, during further purification, showed a surprising sensitivity to stringent washing conditions. Although a 'melting' temperature of about 60–66 °C was expected according to the sequence of the oligonucleotide probe, washing at 50 °C instead of 45 °C significantly reduced the intensity of the hybridization signals. This could be explained later by a codon usage in *Polyorchis*, which was different from the one that we assumed (see the Discussion section).

The cDNA insertions of 26 purified λ clones were transferred to pUC19 and partially sequenced. All cDNA insertions coded for both Pol-RFamide I and II. Two recombinants, clones 12 and 17, were selected for a detailed sequence analysis. The cDNA insertion of clone 12 provided most sequence information for the 3′ end (Figure 1b), whereas that of clone 17 provided most information for the 5′ end (Figure 1c). The overlapping parts of the cDNAs contained in the two clones were completely identical. The cDNA insertions of clones 12 and 17 were cleaved with *Rsa*I and *Hae*III (Figure 1a), their fragments ligated into pUC19 and

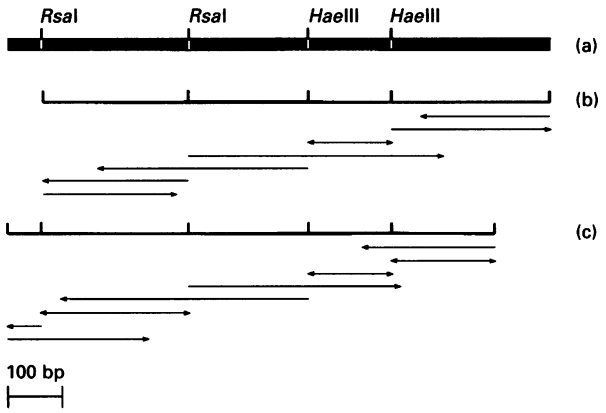

Figure 1 Restriction map and sequencing strategy of cDNA clones coding for the Pol-RFamide precursor

(**a**) Partial restriction map of the composite cDNA assembled from clones 12 (nucleotides 66–1020) and 17 (nucleotides 1–915); (**b**) sequence information contributed by clone 12 (thick line) and the sequencing strategy using subclones generated after digestion with *Rsa*I and *Hae*III (arrows); (**c**) sequence information contributed by clone 17 (thick line) and the sequencing strategy using subclones generated after digestion with the same restriction enzymes (arrows).

```
                                                                CAAAGAGAGACACAGTGGAAGCAGTCGAGAATATTAAA   38

ATG AAT CTA ATA ACA CTA CTT GTA CTT GGC GTG TCT ACT TGT TTA ATA TAT GGT ATT GAA   98
Met Asn Leu Ile Thr Leu Leu Val Leu Gly Val Ser Thr Cys Leu Ile Tyr Gly Ile Glu   20

GCT GAC GAA AAA ACA TCA AGT GCT CTT GAA AAT GAA ATT GTG GAA ATC TTA AAT GGA AAT  158
Ala Asp Glu Lys Thr Ser Ser Ala Leu Glu Asn Glu Ile Val Glu Ile Leu Asn Gly Asn   40

TTT AAA AAT GAA AAA AAA TCT ATA GAA ACA TCA GAT CAG TGG TTA AAA GGG CGA TTT GGT  218
Phe Lys Asn Glu Lys Lys Ser Ile Glu Thr Ser Asp Gln Trp Leu Lys Gly Arg Phe Gly   60

CGT GAA GTG AAC CAA TGG TTA AAA GGG AGA TTT GGT CGC GAG TTG TCA GAT CAG TGG TTA  278
Arg Glu Val Asn Gln Trp Leu Lys Gly Arg Phe Gly Arg Glu Leu Ser Asp Gln Trp Leu   80

AAA GGG AGG TTT GGT CGC GAG TTG TCA GAT CAG TGG TTA AAA GGA AGA TTT GGA CGA GAG  338
Lys Gly Arg Phe Gly Arg Glu Leu Ser Asp Gln Trp Leu Lys Gly Arg Phe Gly Arg Glu  100

GTA CTC GAT CAG TGG TTG AAA GGA AGA TTT GGT CGT GAT GCA TCA AAC CAA TGG TTG AAA  398
Val Leu Asp Gln Trp Leu Lys Gly Arg Phe Gly Arg Asp Ala Ser Asn Gln Trp Leu Lys  120

GGG AGA TTT GGT CGC GAG TTG TCA GAT CAA TGG TTA AAA GGA AGA TTT GGT CGT GAA GGA  458
Gly Arg Phe Gly Arg Glu Leu Ser Asp Gln Trp Leu Lys Gly Arg Phe Gly Arg Glu Gly  140

TCA AAC CAA TGG TTA AAA GGA AGA TTT GGT CGT GAA GCA TCA AAG AAC GAT CTG GAA AAA  518
Ser Asn Gln Trp Leu Lys Gly Arg Phe Gly Arg Glu Ala Ser Lys Asn Asp Leu Glu Lys  160

CAA AAT GGT AGG GGT GAT TCG GAC CAG TGG CTT AAA GGA AGA TTT GGC CGT GAA GCA AGG  578
Gln Asn Gly Arg Gly Asp Ser Asp Gln Trp Leu Lys Gly Arg Phe Gly Arg Glu Ala Arg  180

AAG CAG TTA TTG GGA GGA AGA TTT GGG CGT AAA GAT ATG AAT CAG TTA TTA GCA GAA CGA  638
Lys Gln Leu Leu Gly Gly Arg Phe Gly Arg Lys Asp Met Asn Gln Leu Leu Ala Glu Arg  200

CAT GGG CGT GAA ACA TCC GAT CAA TGG TTA AAG GGG AGG TTT GGA CGA CAA CTT TCA GAT  698
His Gly Arg Glu Thr Ser Asp Gln Trp Leu Lys Gly Arg Phe Gly Arg Gln Leu Ser Asp  220

CAA TGG TTG AAA GGA AGA TTC GGC CGT GAA GTA AAA AAT GAT AAA AAT AAT CCA TTT CGC  758
Gln Trp Leu Lys Gly Arg Phe Gly Arg Glu Val Lys Asn Asp Lys Asn Asn Pro Phe Arg  240

AGT CGT TAC ACA GGA GAT TCA ACA CAA CTG CAG CGA GAG AAC AAC CAA CCT ATT GAA GAA  818
Ser Arg Tyr Thr Gly Asp Ser Thr Gln Leu Gln Arg Glu Asn Asn Gln Pro Ile Glu Glu  260

TTA AGA GAC AAT ACA GAA AAA GTG TCG ATA GAA AAT AAA CCA ATC ATG AAA AAA ACT AGC  878
Leu Arg Asp Asn Thr Glu Lys Val Ser Ile Glu Asn Lys Pro Ile Met Lys Lys Thr Ser  280

GTC AAA ATT AGC AAA ACT GTT TAA GTAGAAAATAAAGTTTTATAAAGCTTGTTCTTTACTATGAAACACAG  949
Val Lys Ile Ser Lys Thr Val  *                                                    287

TGTTGAAAAGATTTTAGATCGAGTAGTAATTAATACTTAATACTATGTTTAACAATTTACCCAAATAATTT        1020
```

Figure 2 Composite cDNA and deduced amino acid sequence of the Pol-RFamide precursor from *Polyorchis penicillatus*

Nucleotide residues are numbered from the 5′ to 3′ end and the amino acid residues are numbered starting with position 1 in the open reading frame. Pol-RFamide I and II sequences are underlined and printed **bold-face** type, whereas a putative neuropeptide sequence is underlined only. The nucleotide sequence in this Figure is a compilation of data obtained from clones 12 (nucleotides 66–1020) and 17 (nucleotides 1–915).

the resulting subclones were sequenced in both directions to avoid sequencing artefacts (Figure 1). The composite cDNA obtained in this way is presented in Figure 2.

The cDNA sequence of Figure 2 is 1020 bp long. It contains a 5′ untranslated leader of 38 bp and a 3′ untranslated region of 121 bp which starts with nucleotide 900. A potential polyadenylation signal (AATAAA) is located at positions 909–914. However, it is too far away from the 3′ end to be functional in the corresponding mRNA (Wickens, 1990). We also did not find a poly(A) tail in any of our clones (Figure 2).

Each of the 26 clones that we sequenced contained a cDNA insertion that was completely identical with a region of the cDNA given in Figure 2.

Primary structure of the Pol-RFamide precursor protein

The cDNA of Figure 2 has an open reading frame of 861 bp extending over nucleotides 39–899. The Pol-RFamide precursor protein deduced from this sequence consists of 287 amino acids. The N-terminus has a hydrophobic signal sequence for endoplasmic-reticulum membrane translocation, which is probably cleaved off at Ala-21 (von Heijne, 1986, 1990).

The precursor protein contains 11 copies of unprocessed Pol-RFamide II (Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly) at amino acid positions 53–176 and 208–228 and one copy of Pol-RFamide I (Gln-Leu-Leu-Gly-Gly-Arg-Phe-Gly) at positions 182–189 (Figures 2 and 3). These sequences contain an N-terminal Gln and a C-terminal Gly residue, which are known to be converted into an N-terminal < Glu and a C-terminal amide group (Fischer and Spiess, 1987; Bradbury and Smyth, 1991; Pohl et al., 1991).

The single copy of immature Pol-RFamide I is flanked by pairs of basic amino acid residues (Arg-Lys). These residues are established sites for post-translational precursor cleavage (Sossin et al., 1989). Furthermore, each of the Pol-RFamide II sequences is followed by a single basic residue (Arg). Also, processing at single basic residues is well-known both in vertebrates and invertebrates (Devi, 1991). At the N-terminus of each Pol-RFamide II sequence, however, basic residues are lacking and are replaced by single Asp (in eight cases) or single Asn (in three cases) residues. Acidic amino acid residues (Asp or Glu) have recently been recognized as sites where a new type of processing enzyme must cleave the Antho-RFamide precursors of sea anemones (Darmer et al., 1991; Schmutzler et al., 1992). The present finding of Asp residues preceding eight out of 11 Pol-RFamide II sequences in the Pol-RFamide precursor of a hydrozoan jellyfish suggests the general occurrence of such an acidic-residue-specific processing enzyme in cnidarians. Three

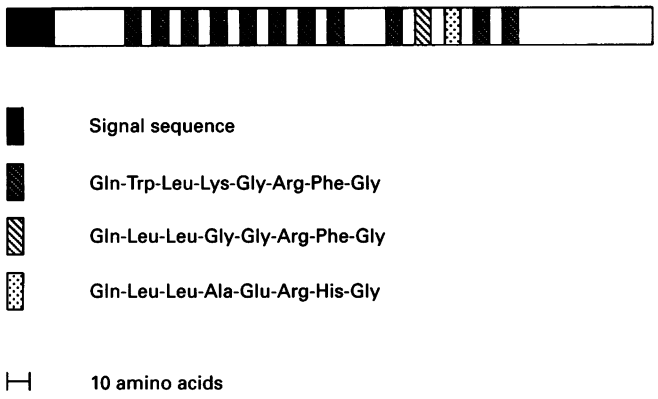

Figure 3 Schematic representation of the Pol-RFamide precursor protein from *Polyorchis penicillatus*

The black box at the N-terminus indicates the hydrophobic signal sequence. Hatched boxes represent the established neuropeptide sequences and the stippled box a putative neuropeptide sequence.

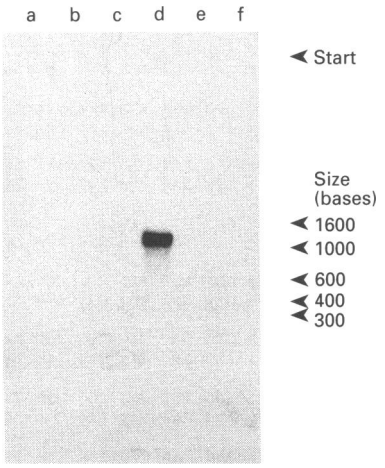

Figure 4 Northern-blot analysis

Poly(A)$^+$ RNA isolated from several animal species (5 μg/lane) was hybridized with a cDNA insertion coding for nearly the whole Pol-RFamide precursor from *Polyorchis* (nucleotides 53–983). a, *Cyanea lamarckii* (subumbrella); b, *Hydra magnipapillata* (whole animal); c, rat (brain); d, *Polyorchis penicillatus* (bell margin); e, *Calliactis parasitica* (whole animal); f, *Anthopleura elegantissima* (whole animal).

copies of Pol-RFamide II are preceded by an Asn residue. Provided that these copies do also give rise to mature Pol-RFamide II, another 'non-classical' processing enzyme is likely to be used in *Polyorchis*.

In addition to the two established peptides Pol-RFamides I and II, a third, hitherto unknown, neuropeptide sequence might possibly be present on the Pol-RFamide precursor. The peptide is located at amino acid positions 195–202 (Figure 2) and has the immature sequence Gln-Leu-Leu-Ala-Glu-Arg-His-Gly. This sequence is surrounded by processing signals which are also found for Pol-RFamide II (Figure 2; Table 1), which suggests that it might be processed to a mature neuropeptide, < Glu-Leu-Leu-Ala-Glu-Arg-His-NH_2. This peptide would be closely related to Pol-RFamide I (63% sequence identity). However, as the sequence has an internal Glu residue (at amino acid position 199 of Figure 2), it might be cleaved if the acidic-residue-specific processing enzyme mentioned above is an endoproteinase. This makes the status of the neuropeptide uncertain.

The amino acid sequence of the Pol-RFamide precursor and its corresponding cDNA show no significant sequence similarity to any other amino acid or nucleotide sequence contained in the GenBank or EMBL databases.

Northern-blot analysis

Figure 4 shows a Northern blot of poly(A)$^+$ RNA isolated from the jellyfish *Cyanea lamarckii* (Scyphozoa), the freshwater polyp

Table 1 N- and C-terminal extensions of established and putative neuropeptide sequences found within the Pol-RFamide precursor protein of *Polyorchis penicillatus*

Initial cleavage sites are indicated by arrows. Unusual processing sites are printed **bold-face** type. Neuropeptide sequences are underlined.

N- and C-terminal extensions and neuropeptide sequence	Name	Copy no
Arg-Lys↓Gln-Leu-Leu-Gly-Gly-Arg-Phe-Gly-Arg-Lys↓	Pol-RFamide I	1
Arg↓Glu-Leu-Ser-**Asp**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	3
Arg↓Gln-Leu-Ser-**Asp**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Arg↓Glu-Thr-Ser-**Asp**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Lys-Lys↓Ser-Ile-Glu-Thr-Ser-**Asp**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Lys↓Gln-Asn-Gly-Arg-Gly-Asp-Ser-**Asp**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Arg↓Glu-Val-Leu-**Asp**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Arg↓Glu-Gly-Ser-**Asn**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Arg↓Asp-Ala-Ser-**Asn**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Arg↓Glu-Val-**Asn**-Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly-Arg↓	Pol-RFamide II	1
Arg-Lys↓Asp-Met-Asn-Gln-Leu-Leu-Ala-Glu-Arg-His-Gly-Arg↓		1

Hydra magnipapillata (Hydrozoa), the jellyfish *Polyorchis penicillatus* (Hydrozoa), the sea anemones *Anthopleura elegantissima* and *Calliactis parasitica* (Anthozoa) and the rat as an example of a mammalian species. The hybridization probe was a cDNA insertion that represented nucleotides 53–983 of Figure 2. The lane containing the *Polyorchis* RNA (Figure 4d) shows a distinct band indicating the presence of mRNA with a length of about 1200 bases. Assuming that the mRNA coding for the Pol-RFamide precursor contains a poly(A) tail of about 200 bases, as is usually the case in eukaryotes, the length of the mRNA (Figure 4d) corresponds well with the length of the compiled cDNA given in Figure 2. This means that we have cloned most of the mRNA coding for the Pol-RFamide precursor.

We also found a faint signal in the lane containing the *Hydra* mRNA, which, however, was only visible after strong over-exposure (results not shown). No hybridization signals were found for the other species, although we used hybridization conditions of relatively low stringency.

DISCUSSION

The Pol-RFamide precursor protein from *Polyorchis penicillatus* contains 11 copies of Pol-RFamide II and one copy of Pol-RFamide I (Figure 2; Table 1). This ratio corresponds well with our data from peptide purification, where we have always found much more Pol-RFamide II than Pol-RFamide I in tissue extracts (this ratio varied between 5:1 and 12:1 in different preparations and purification steps, but it is difficult to give an accurate number for the initial acetic acid extract as we do not have radioimmunoassays specific for either Pol-RFamide I or II; Grimmelikhuijzen et al., 1988, 1992b).

We started to screen our cDNA library of *Polyorchis* with an oligonucleotide probe coding for Pol-RFamide II instead of Pol-RFamide I because we found more Pol-RFamide II than I in extracts of *Polyorchis*, suggesting that more hybridizing targets for a Pol-RFamide II probe would be present. In addition, an oligonucleotide pool designed on the basis of immature Pol-RFamide II (Gln-Trp-Leu-Lys-Gly-Arg-Phe-Gly) is less degenerate than a probe for Pol-RFamide I (Gln-Leu-Leu-Gly-Gly-Arg-Phe-Gly), owing to the more favourable amino acid composition. In order to reduce further the degeneracy of the oligonucleotide pool, we did not include all possible codons for the Pol-RFamide II sequence. On the basis of codon usage for the Antho-RFamide precursors in two sea-anemone species (Darmer et al., 1991; Schmutzler et al., 1992) we assumed that most codons for Leu and Arg would be of the CTN or CGN (N = A, C, G, T) type respectively, and, therefore, we omitted the TTA/G and AGA/G variants. We found, however, that ten out of 11 Leu codons and ten out of 11 Arg codons in the cDNA coding for the Pol-RFamide II sequences were of the TTA/G and AGA/G type respectively (Figure 2). Codons of the CTN or CGN type never occur within the same peptide coding region (Figure 2). Thus the oligonucleotides gave at least one mismatch with each of their target sequences in the cDNA, which explains the unexpectedly low melting temperature of the duplex.

In addition to the A and T preference in the first position of the Arg and Leu codons in the peptide coding regions, there is a general over-representation of A and T nucleotides in the whole coding sequence of the Pol-RFamide precursor cDNA (Figure 2): in 73% of the codons where there is a choice between A/T and G/C at the third position, A/T is preferred. A similar bias in codon usage also has been described for genes of *Hydra*, another hydrozoan [82% (Bosch et al., 1989); 78% (Fisher and Bode, 1989)]. The high A/T content, therefore, might be a general property of genes in the Hydrozoa.

The maturation of the Pol-RFamide precursor involves both classical and non-classical mechanisms. The Pol-RFamide I sequence, for example, is flanked by two Arg-Lys sequences (amino acid positions 179–211 in Figure 2; Table 1). Paired basic residues are established sites for peptide precursor cleavage. Several processing enzymes that cleave at such sites have been extensively characterized such as the Kex2 gene product from the yeast *Saccharomyces cerevisiae*, which is involved in the generation of α-mating factor, and PC1/PC3 and PC2 from mammals which are involved in the generation of peptide hormones (Julius et al., 1984; Smeekens and Steiner, 1990; Seidah et al., 1990, 1991; Smeekens et al., 1991). It is interesting that an enzyme showing similarity to PC3 has recently been cloned from *Hydra* (Chan et al., 1992). This suggests that a PC3-related proteinase might be involved in the processing of the paired basic residues near Pol-RFamide I.

In addition to the Arg-Lys sequences at both sides of Pol-RFamide I, single basic residues (Arg) occur at the C-terminal side of each Pol-RFamide II sequence. Also these residues are established cleavage sites which are frequently used for the generation of neuropeptides, especially in invertebrates (Schaefer et al., 1985; Nambu et al., 1988; Schneider and Taghert, 1988; Linacre et al., 1990; Devi, 1991; Darmer et al., 1991; Rosoff et al., 1992; Schmutzler et al., 1992). Devi (1991) has proposed several rules (consensus sequences) for processing at monobasic sites, the most important being the requirement of a second basic residue at position −3, −5 or −7 of the basic residue where cleavage occurs. This and the other requirements of Devi (1991) are perfectly fulfilled for all Pol-RFamide II sequences (Figure 2; Table 1).

The processing sites at the N terminus of each Pol-RFamide II sequence, however, are very different from what is known so far from prohormones of higher invertebrates or vertebrates (Figure 2; Table 1): instead of basic residues, single Asp (in eight cases) or single Asn (in three cases) residues occur. Acidic residues (Asp or Glu) have also been found at the N-termini of most of the Antho-RFamide sequences in the Antho-RFamide precursors of sea anemones and must be the cleavage sites for a new type of processing enzyme occurring in neurons (Darmer et al., 1991; Schmutzler et al., 1992). From our present data (Table 1), this novel type of enzyme is most likely to be an endoproteinase, because the simplest way to generate Pol-RFamide II is by endoproteolytic cleavage. The enzyme could possibly be an aminopeptidase, but in this case other proteinases must be involved to remove the additional residues that are located N-terminally of the Asp residues (Table 1).

Three Pol-RFamide II sequences are preceded by an Asn residue (Table 1). Because we have never been able to isolate N-terminally extended Pol-RFamide II, we assume that also these sequences are converted into authentic (non-extended) Pol-RFamide II. If this is true, then an additional processing enzyme hydrolysing at the C-terminal side of Asn residues has to be postulated. This processing enzyme could either be an amino-peptidase or an endoproteinase. Again, in the case of an Asn-residue-specific aminopeptidase, other proteinases would be needed.

Processing at Asn or Asp residues has not been demonstrated so far for neuropeptide precursor proteins of higher organisms. However, cleavage at the C-terminal side of Asn residues takes place during the maturation of the precursor for a-factor, which is a mating pheromone from the yeast *Saccharomyces cerevisiae*, and for M-factor, which is a mating pheromone from *Schizosaccharomyces pombe* (Brake et al., 1985; Davey, 1992). Similarly, the substrate specificity of the acidic-residue-specific proteinase from *Polyorchis* might be the same as that of the V8

proteinase from *Staphylococcus aureus* (Drapeau et al., 1972; Rydén et al., 1974). Thus the possibility exists that the neuronal processing enzymes that we presently postulate in cnidarians may have their counterparts in yeast and bacteria. This would be a situation similar to that of subtilisin (from *Bacillus subtilis*), the Kex-2 gene product (from *S. cerevisiae*), PC1/PC3, PC2 and furin (from higher eukaryotes) which are all members of the same family of serine proteinases (Barr, 1991).

In conclusion, we have found that the acidic processing sites that we have previously observed in the Antho-RFamide precursor proteins of sea anemones do also occur in the Pol-RFamide precursor of *Polyorchis*. The jellyfish *Polyorchis* (Hydrozoa) and sea anemones (Anthozoa) belong to different classes of Cnidaria and are only distantly related. This suggests that acidic residues are generally used for neuropeptide precursor processing in cnidarians. We have also found evidence for cleavage of the Pol-RFamide precursor at Asn residues, suggesting that cnidarian nervous systems may use a variety of unconventional processing signals to generate neuropeptides from their neuropeptide precursor proteins.

We thank Dr. Andrew N. Spencer for providing *Polyorchis*, Birgit Hepke for typing the manuscript, and the Bundesministerium für Forschung und Technologie, Deutsche Forschungsgemeinschaft (Gr762/10-1) and Fonds der Chemischen Industrie for financial support.

REFERENCES

Anctil, M. and Grimmelikhuijzen, C. J. P. (1989) Gen. Pharmacol. **20**, 381–384
Anderson, P. A. V. and Spencer, A. N. (1989) J. Neurobiol **20**, 435–457
Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 1408–1412
Barr, P. J. (1991) Cell **66**, 1–3
Bosch, T. C. G., Unger, T. F., Fisher, D. A. and Steele, R. E. (1989) Mol. Cell. Biol. **9**, 4141–4151
Bradbury, A. F. and Smyth, D. G. (1991) Trends Biochem. Sci. **16**, 112–115
Brake, A. J., Brenner, C., Najarian, R., Laybourn, P. and Merryweather, J. (1985) in Protein Transport and Secretion (Gething, M. J., ed.), pp. 103–108, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Chan, S. J., Oliva, A. A., Lamendola, J., Grenns, A., Bode, H. and Steiner, D. F. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 6678–6682
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry **18**, 5294–5299
Darmer, D., Schmutzler, C., Diekhoff, D. and Grimmelikhuijzen, C. J. P. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 2555–2559
Davey, J. (1992) EMBO J. **11**, 951–960
Devi, L. (1991) FEBS Lett. **280**, 189–194
Drapeau, G. R., Boily, Y. and Houmard, J. (1972) J. Biol. Chem. **247**, 6720–6726
Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. **132**, 6–13
Fischer, W. H. and Spiess, J. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 3628–3632
Fisher, D. A. and Bode, H. R. (1989) Gene **84**, 55–64
Grigoriev, N. and Spencer, A. N. (1992) Soc. Neurosci. Abstr. **18**, 471
Grimmelikhuijzen, C. J. P. and Graff, D. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 9817–9821
Grimmelikhuijzen, C. J. P. and Groeger, A. (1987) FEBS Lett. **211**, 105–108
Grimmelikhuijzen, C. J. P. and Spencer, A. N. (1984) J. Comp. Neurol. **230**, 361–371
Grimmelikhuijzen, C. J. P., Hahn, M., Rinehart, K. L. and Spencer, A. N. (1988) Brain Res. **475**, 198–203
Grimmelikhuijzen, C. J. P., Carstensen, K., Darmer, D., Moosler, A., Nothacker, H.-P., Reinscheid, R. K., Schmutzler, C., Vollert, H., McFarlane, I. D. and Rinehart, K. L. (1992a) Am. Zool. **32**, 1–12
Grimmelikhuijzen, C. J. P., Rinehart, K. L. and Spencer, A. N. (1992b) Biochem. Biophys. Res. Commun. **183**, 375–382
Julius, D., Brake, A., Blair, L., Kunisawa, R. and Thorner, J. (1984) Cell **37**, 1075–1089
Linacre, A., Kellet, E., Saunders, S., Bright, K., Benjamin, P. R. and Burke, J. F. (1990) J. Neurosci. **10**, 412–419
Mackie, G. O. (1990) Am. Zool. **30**, 907–920
Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
McFarlane, I. D. and Grimmelikhuijzen, C. J. P. (1991) J. Exp. Biol. **155**, 669–673
McFarlane, I. D., Graff, D. and Grimmelikhuijzen, C. J. P. (1987) J. Exp. Biol. **133**, 157–168
Nambu, J. R., Murphy-Erdosh, C., Andrews, P. C., Feistner, G. J. and Scheller, R. H. (1988) Neuron **1**, 55–61
Pohl, T., Zimmer, M., Mugele, K. and Spiess, J. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 10059–10063
Rosoff, M. L., Bürglin, T. R. and Li, C. (1992) J. Neurosci. **12**, 2356–2361
Rydén, A. C., Rydén, L. and Phillipson, L. (1974) Eur. J. Biochem. **44**, 105–114
Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Schaefer, M., Picciotto, M. R., Kreiner, T., Kaldany, R.-R., Taussig, R. and Scheller, R. H. (1985) Cell **41**, 457–467
Schmutzler, C., Darmer, D., Diekhoff, D. and Grimmelikhuijzen, C. J. P. (1992) J. Biol. Chem. **267**, 22534–22541
Schneider, L. E. and Taghert, P. H. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 1993–1997
Seidah, N. G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M. and Chrétien, M. (1990) DNA Cell Biol. **9**, 415–424
Seidah, N. G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M. G., Lazure, C., Mbikay, M. and Chrétien, M. (1991) Mol. Endocrinol. **5**, 111–122
Smeekens, S. P. and Steiner, D. F. (1990) J. Biol. Chem. **265**, 2997–3000
Smeekens, S. P., Avruch, A. S., LaMendola, J., Chan, S. J. and Steiner, D. F. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 340–344
Sossin, W. S., Fisher, J. M. and Scheller, R. H. (1989) Neuron **2**, 1407–1417
Von Heijne, G. (1986) Nucleic Acids Res. **14**, 4683–4690
Von Heijne, G. (1990) J. Membr. Biol. **115**, 195–201
Wickens, M. (1990) Trends. Biochem. Sci. **15**, 227–281

Received 15 October 1993; accepted 3 November 1993