Volume 257, number 1, 31-34 FEB 07713 October 1989

Cloning and sequence analysis of cDNA for precursor of a crustacean hyperglycemic hormone

Wolfgang Weidemann, Jörg Gromoll and Rainer Keller

Institut für Zoophysiologie der Rheinischen Friedrich-Wilhelms-Universität, D-5300 Bonn, FRG

Received 28 August 1989

Crustacean hyperglycemic hormone (CHH) from *Carcinus maenas*, a 72 amino acid neuropeptide, originates in neurosecretory perikarya in the eyestalk ganglia. Poly (A)RNA was isolated from these perikarya and a cDNA library was prepared. Screening of 20000 clones with a 26-mer oligonucleotide, corresponding to a partial sequence of CHH, yielded one positive clone with an insert of approximately 2000 bp, which contained the complete coding sequence for a pre-pro CHH. This precursor consists of a putative 26 amino acid signal sequence, a 38 amino acid peptide of unknown function (Peptide C), and the CHH sequence at the carboxyl end. The CHH-sequence is flanked N-terminally by a Lys-Arg cleavage site and C-terminally by the tetrapeptide Gly-Arg-Lys-Lys which is followed by the stop codon.

Hyperglycemic hormone; Prohormone; DNA, complementary; Neuropeptide; (*Carcinus maenas*, Crustacean)

1. INTRODUCTION

Neurosecretory structures in the optic ganglia of the crustacean eyestalk are the source of a number of neuropeptides which are of great importance for the regulation of a variety of physiological functions (see [1,2] for reviews). Among them is a neuropeptide with potent hyperglycemic action and, possibly, other metabolic effects as well. This 'crustacean hyperglycemic hormone' (CHH) has been isolated from the shore crab, *Carcinus maenas*, and its amino acid sequence has been determined both by Edman degradation [3] and cDNA cloning (this paper). With 72 amino acids, it is the second-biggest invertebrate neuropeptide elucidated thus far. Its sequence does not resemble that of any other known neuropeptide. Preliminary structural data from studies on eyestalk peptides indicate that CHH is a member of a novel family of neuropeptides [3]. Although the number of invertebrate neuropeptides with known amino acid sequences is growing rapidly [4], our knowledge concerning prohormone structure is still very limited, and there has not been any such information for a crustacean neuropeptide. The cDNA cloning approach offered the possibility to elucidate the precursor of CHH. We took advantage of the fact that the hormone is synthesized in a subgroup of large perikarya which form a cluster, the medulla terminalis X-organ (XO) [5]. It is connected via an axon tract to a neurohemal organ, the sinus gland (SG). Due to its visibility under the dissecting microscope, the XO can be collected and used for mRNA isolation. In this paper, we describe the preparation of a cDNA library and the complete sequence of a CHH precursor, deduced from one cDNA clone.

Correspondence address: J. Gromoll, Institut für Zoophysiologie der Universität Bonn, Endenicher Allee 11-13, D-5300 Bonn 1, FRG

Abbreviations: cDNA, complementary DNA; mRNA, messenger RNA; HPLC, high performance liquid chromatography; Tris, tris (hydroxymethyl) aminomethane; SDS, sodium dodecyl sulfate

2. MATERIALS AND METHODS

Clusters of neurosecretory X-organ perikarya were dissected from eyestalk ganglia of 500 *Carcinus maenas* and stored at −70°C. Total RNA was prepared by the guanidinium thiocyanate/CsCl procedure [6] and poly (A) RNA was isolated by chromatography on oligo d (T) cellulose [7]. A cDNA library was constructed according to the method of Okayama and Berg [8] from 1 μg of poly (A) RNA. The cDNA clones were screened by colony hybridization [9] with a mixture of ^{32}P-end-labelled oligonucleotide probes that were constructed as shown in fig.1. Oligonucleotides were synthesized on a Cyclone TM DNA synthesizer (Biosearch) and purified by reversed phase HPLC [10]. Hybridizations were carried out at 60°C and filters were washed 4 times with 900 mM NaCl, 6 mM EDTA and 0.1% SDS in 180 mM Tris, pH 8, at 65°C. Plasmid DNA of the positive clone was analyzed by restriction mapping with the enzymes *Bgl*II and *Pst*I. Fragments were subcloned in the pUC 18 vector and nucleotide sequences were determined by the dideoxy sequencing method [11].

3. RESULTS

The cDNA library, which we constructed from 1 μg of poly (A) RNA, consisted of approximately 100 000 clones. Screening of the first 20 000 clones with the probe mixture shown in fig.1 yielded one positive clone with an insert length of approximately 2000 bp. Diges-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50 © 1989 Federation of European Biochemical Societies

```
Protein    Met Met Asp Glu Phe Asp Gln Tyr Ala
m-RNA  5' AUG AUG GAU GAA UUU GAU CAA UAU GC  3'
                  C   G   C   C   G   C
Probe  3' TAC TAC CTA CTT AAA CTA GTT ATA CG  5'
                  G   C   G   G   C   G
```

Fig.1. Partial sequence of CHH (see fig.2), the corresponding segment of the mRNA and the synthetic 26-mer oligonucleotides (64-fold degenerated) used in this study.

tion of the insert with *Bgl*II restriction enzyme gave two fragments of 500 and 1500 bp, respectively. By Southern blotting, the hybridization signal was localized to the 500 bp fragment. Both fragments were subcloned in the pUC 18 vector, and the 500 bp fragment was sequenced. A *Pst*I restriction cleavage site at position 238 (fig.2) was utilized for further subcloning in pUC 18 and confirmation of the sequence. The cDNA fragment shown in fig.2 contains the complete coding sequence for a CHH precursor, preceded by a 5'-flanking sequence and followed by a 3'-untranslated part, of which only the beginning is shown. The remaining 3'-untranslated sequence is represented by the 1500 bp fragment which we have not yet fully sequenced. Preliminary results indicate a poly (A) tail.

The open reading frame includes a putative 26 amino acid signal peptide, the 38 amino acid peptide C and the CHH sequence of 72 residues which occupies the carboxyl end. It is separated from the C-peptide by Lys-Arg, and flanked C-terminally by the tetrapeptide Gly-Arg-Lys-Lys which is followed by the stop codon.

```
                                    -153
                                -5' CAGTTCGACCAGGAATTCGCAGAAGGAAGACGT

-120  ACACCTCCTCCTATAGTGAACCTTCTAGAATTGTACCCTACCATGTATAGCAAAACTATT
                                                  -26  MetTyrSerLysThrIle

 -60  CCCGCCATGCTAGCAATCATCACCGTAGCCTACCTATGCGCACTCCCGCACGCACACGCA  -1
 -20  ProAlaMetLeuAlaIleIleThrValAlaTyrLeuCysAlaLeuProHisAlaHisAla  -1

   1  CGCTCCACGCAAGGCTACGGACGCATGGATAGGATTCTGGCGGCCTTGAAAACCTCGCCA
   1  ArgSerThrGlnGlyTyrGlyArgMetAspArgIleLeuAlaAlaLeuLysThrSerPro
      PEPTIDE C ──►

  61  ATGGAGCCCAGCGCAGCCCTAGCGGTGGAGAATGGAACTACACACCCGTTGGAAAAGAGG
  21  MetGluProSerAlaAlaLeuAlaValGluAsnGlyThrThrHisProLeuGluLysArg
                                                  38

 121  CAAATTTACGACACGTCCTGCAAGGGTGTTTACGACCGTGCTCTGTTCAATGACTTGGAG
  41  GlnIleTyrAspThrSerCysLysGlyValTyrAspArgAlaLeuPheAsnAspLeuGlu
      CHH ──►

 181  CACGTGTGTGACGATTGTTACAACCTCTACAGAACCTCCTATGTTGCCTCGGCCTGCAGA
  61  HisValCysAspAspCysTyrAsnLeuTyrArgThrSerTyrValAlaSerAlaCysArg

 241  TCAAACTGCTATAGCAACTTGGTGTTCCGGCAATGCATGGATGACCTTTTAATGATGGAC
  81  SerAsnCysTyrSerAsnLeuValPheArgGlnCysMetAspAspLeuLeuMetMetAsp

 301  GAGTTTGACCAATATGCCAGAAAGGTACAGATGGTTGGCAGGAAGAAGTAAACAACAGAT  -3'
 101  GluPheAspGlnTyrAlaArgLysValGlnMetValGlyArgLysLysEND
                                          72
```

Fig.2. Part of the coding strand of the cDNA clone that contains the complete coding sequence for the CHH precursor, and the deduced amino acid sequence. The sequences of the mature CHH (plus the putative amide-donating Gly) and peptide C are underlined. The numbers of nucleotides and amino acid residues appear at the beginning of each line, beginning with 1 at the start of the coding sequence for peptide C. Proteolytic processing sites are boxed. The partial CHH-sequence used for oligonucleotide probe construction is market by a dotted line. The first and last residues of peptide C and CHH are numbered for easy reference.

4. DISCUSSION

A cDNA library, prepared from poly (A) RNA which was isolated from neurosecretory XO-cell somata, contained one clone from which the complete primary structure of the precursor of the crustacean hyperglycemic hormone (CHH) of *Carcinus maenas* could be deduced. The open reading frame, consisting of 426 nucleotides, encodes a putative 26 amino acid signal peptide, a 38 amino acid peptide (peptide C), and the CHH sequence of 72 amino acids. The CHH sequence is flanked by Lys-Arg N-terminally, and C-terminally by the tetrapeptide Gly-Arg-Lys-Lys.

The putative signal peptide displays well-known features of such sequences, in particular the central hydrophobic region. The signal sequence cleavage site is after Ala in position 26 (numbered −1 in fig.2). This is consistent with the structure of peptide C. The same site would have been predicted by use of the rules formulated by von Heijne [12].

The isolation of peptide C from the sinus gland (SG), which consists of the terminals of the XO cells, proves not only the cleavage after Ala^{26}, but also the expected processing at Lys-Arg betwen peptide C and CHH. The sequence of peptide C, as determined in our laboratory (unpublished results), is in agreement with that deduced here from the cDNA. A peptide showing 50% homology, peptide H, had previously been isolated from SG of the crab, *Cardisoma carnifex* [13]. This homology was also revealed by a search at GENBANK (The Los Alamos National Laboratory), whereas no significant homology between the CHH-coding region and other sequences was found. Neither for peptide C nor H has a biological function been demonstrated as yet. It was hypothesized that peptide H is a biosynthetic intermediate, from which smaller, potentially bioactive peptides arise by proteolysis at two internal sites [13]. At these cleavage sites, a Ser^{11} (an unconventional one) and an Arg^{17} have been proposed (13). In peptide C, there is an Arg^{11} and a Lys^{17}, the former being an even more likely site of cleavage than Ser. However, such cleavages are very slow (13), and it remains to be shown whether fragments are produced in vivo to any significant extent. The occurrence of intact C-peptide in the SG of *Carcinus maenas* in relatively large amounts and in proportion to CHH (unpublished) appears to argue against a significant role as a biosynthetic intermediate.

As to the sequence of the mature CHH, the results of Edman degradation (3) and cDNA analysis are in complete agreement. The residue Gln^{1} in the precursor is consistent with $pyroGlu^{1}$ in the mature neuropeptide, and Gly^{73} is the expected amide group donor for the terminal Val^{72}. The processing site Lys-Arg between C-peptide and CHH is cleaved and removed as expected, and the C-terminal 3 basic residues Arg-Lys-Lys of the precursor are also removed.

It is noteworthy that 3 basic residues are inserted between the amide-donating Gly and the translational stop signal. A similar situation is found in the precursor of the mammalian corticotropin releasing factor (CRF), although in this case there is only one Lys between the Gly and the stop codon. The removal of one or more basic residues may in some way be linked to the amidation reaction that involves the preceding Gly.

The dibasic site Arg-Lys (positions 67 and 68) close to the C-terminus is another potential cleavage site which could, if processed incompletely, generate two forms of CHH. However, there is as yet no evidence of a truncated form in the SG of *Carcinus* [3]. It is interesting to discuss the present results with reference to former studies on the synthesis and processing of neuropeptide precursors in the XO-SG system. In one earlier study [14], using 2D-Page, polypeptides of 19–21 kDa were predominantly labelled in XO cells after in vitro incubation of whole eyestalk ganglia of crayfish (*Orconectes virilis*) and blue crab (*Callinectes sapidus*) with [^{3}H]-amino acids. In the SG, the label was mainly associated with several 7 kDa peptides. In more recent, detailed pulse-chase experiments on the XO-SG-system of *Cardisoma carnifex* by Stuenkel [15], a 14 kDa polypeptide was identified as the primary biosynthetic product. Three more 14 kDa polypeptides appeared to arise from the first, presumably by posttranslational transformation. Apparently, the primary 14 kDa precursor gave rise to peptide H (M_r 4.5 kDa) and three HPLC-separable 6 kDa peptides. In view of the discrepancy in the molecular mass relationship, it was suggested that the precursor gives rise to peptide H and only one 6 kDa peptide, and that the two other 6 kDa forms were due to posttranslational modification. In another study [13], it was suggested that peptide H was synthesized as part of a common precursor together with two forms of CHH and the related moult inhibiting hormone (MIH).

The result reported here appears to resolve this problem. From our cDNA data, the calculated M_r of the precursor is 13.3 kDa. This is close to Stuenkel's results [15] and confirms his conclusion that peptide H (peptide C in our case) and one 6 kDa-peptide (CHH) are on one precursor. How can the appearance of several 14 kDa precursors and three different 6 kDa peptides be explained? We should like to suggest that the different 14 kDa-polypeptides are not due to posttranslational modification, but are translation products from several genes, and that the different 6 kDa peptides come, accordingly, from different precursors. The existence of several SG-peptides of between 6 and 8 kDa with different biological activities is now well documented [3,16]. We assume that the CHH precursor described here is the product of one gene of a family of related genes.

Acknowledgments: This study was supported by a grant from the Deutsche Forschungsgemeinschaft (Ke 206/8-1). We thank Dr. K. Olek, Bonn, for the Okayama-Berg vector and for use of his

facilities for oligonucleotide synthesis. Thanks are also due to Dr. J.W. Slootstra, Free University Amsterdam, for performing the GENBANK search.

REFERENCES

[1] Beltz, B.A. (1988) in: Endocrinology of Selected Invertebrate Types (Laufer, H. and Downer, R.G.H. eds.) pp. 235–258, Liss, New York.
[2] Webster, S.G. and Keller, R. (1988) in: Neural Hormones in Invertebrates, Cambridge Univ. Press, Cambridge, NY, pp. 173–196.
[3] Kegel, G., Reichwein, B., Weese, S., Gaus, G., Peter-Katalinic, J. and Keller, R. (1989) FEBS Lett. 255, 10–14.
[4] Thorndyke, M.C. and Goldsworthy, G.J. (eds.) (1988), Neurohormones in Invertebrates, Cambridge University Press, Cambridge.
[5] Dircksen, H., Webster, S.G. and Keller, R. (1988) Cell Tissue , Res. 251, 3–12.
[6] Chirgwin, J.M. Przybyla, A.E., Mac Donald, R.J. and Rutter, , W.J. (1979) Biochemistry 18, 5294–5299.
[7] Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
[8] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.
[9] Hanahan, D. and Meselson, M. (1983) Methods Enzymol. 100, 333–342.
[10] Pingoud, A. and Fliess, A. (1988) GIT Fachz. Lab. 5, 475–478.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Von Heijne, G. (1986) Nucl. Acids Res. 14, 4683–4691.
[13] Newcomb, R.W. (1987) J. Neurochem. 49, 574–583.
[14] Andrew, R.D. and Saleuddin, A.S.M. (1979) J. Comp. Physiol. B. 134, 303–313.
[15] Stuenkel, E.L. (1986) Peptides 7, 397–406.
[16] Huberman, G. and Aguilar, M.B. (1989) Comp. Biochem. Physiol. 93B, 299–305.