Vol. 204, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1994 Pages 292-296

ISOLATION AND STRUCTURE OF THE *DROSOPHILA* CORAZONIN GENE

Jan A. Veenstra*

Department of Entomology and Center for Insect Science, The University of Arizona, Tucson, AZ 85721

Received August 25, 1994

Summary: A recombinant DNA clone containing the corazonin gene was isolated from a genomic library of *Drosophila melanogaster*. Its nucleotide sequence predicts a preprocorazonin consisting of an 19 amino acid putative signal peptide, the 11 amino acid corazonin sequence, a Gly used for amidation, a Lys-Arg proteolytic processing site, and a 39 amino acid corazonin-precursor-related peptide (CPRP). CPRP has an internal Arg-Arg sequence and thus could possibly be further processed into a tripeptide and a 34-mer. Neither CPRP or its possible products are structurally related to any known neuropeptide, and their physiological function is unknown. The structure of the predicted preprocorazonin is remarkably similar to the preprohormone of adipokinetic hormone, which suggests that corazonin and adipokinetic hormone have a common evolutionary origin.

Neuropeptides are structurally the most diverse group of neurochemical mediators and play important and diverse roles in the regulatory physiology of higher organisms [1]. Since they are almost the directly translated products of genes, and not, like classical neurotransmitters, the products of enzymes, their structures have more potential for rapid evolutionary change than those of other neurotransmitters or hormones [2]. Evolution of some peptide families appears to have been fast indeed. The arthropod peptide family characterized by insect AKH and crustacean RPCH is an intriguing example of neuropeptide evolution. While from crustaceans so far only a single molecular form has been characterized [3], at least 18 different molecular species have already been found in insects [4]. It has been suggested, that the evolution of this peptide family within the class of insects is related to a functional evolution, i.e., that changes in the energy substrates that are mobilized by these peptides are accompanied by changes in the structures of these peptides [5].

Corazonin (pGlu-Thr-Phe-Gln-Tyr-Ser-Arg-Gly-Trp-Thr-Asn-NH_2) is an insect neuropeptide which appears to be generally present in insects and shows both structural and

*Correspondence should be addressed to: J.A. Veenstra, Department of Entomology, 410 Forbes, bldg. # 36, The University of Arizona, Tucson, AZ 85721.

Abbreviations: AKH, adipokinetic hormone; RPCH, red pigment concentrating hormone.

0006-291X/94 $5.00

functional similarities to members of the AKH/RPCH family [6,7]. Its chemical characterization from four different insect species suggests that its sequence has been well preserved during evolution [5,6]. However, since the insect members of the AKH/RPCH peptide family are so structurally diverse, it is not clear whether these similarities reflect an evolutionary relationship of corazonin with the AKH/RPCH family. The structures of several cDNA's as well as that of a single gene coding for members of the AKH/RPCH family are already known [8-11]. Therefore, the structure of a corazonin gene might provide evidence in favor or against such a relationship. I report here that the structure of the corazonin gene from *Drosophila melanogaster* supports an evolutionary relationship between corazonin and the AKH/RPCH peptide family.

MATERIALS AND METHODS

Screening of a Drosophila genomic library: A 35-mer oligonucleotide 5'-CA(A/G)ACITT-(T/C)CA(A/G)TA(T/C)TCI(C/A)GIGGITGGACIAA(T/C)GG-'3, corresponding to the amino acids and the C-terminal amide of corazonin, was synthetized by the Division of Biotechnology of the University of Arizona. The oligonucleotide was end-labelled with T4 polynucleotide kinase using [γ-^{32}P]ATP to a specific activity of at least 1500 Ci/mmol. Fifty thousand plaques of a genomic EMBL-3 library of *Drosophila melanogaster* (a gift from Dr. J. Tamkun, University of California, Santa Cruz) were lifted on nitrocellulose filters and screened using moderately stringent conditions (5 X Denhardt, 8% formamide, 0.1% yeast RNA, 5 X SSC, 0.1% SDS, at 40 °C). Preparation of λ-DNA, restriction analysis, agarose gel electrophoresis, southern blotting, subcloning, and sequence analysis were all performed using standard techniques [12].

RESULTS AND DISCUSSION

A total of ten potentially positive plaques were picked and of these four independent positive clones were plaque purified. Restriction analysis suggested that they belonged to two different genes. One gene was represented by three independent clones, but sequence analysis showed its hybridization with the probe to be due to a perfect match of the 13 3'-bases of the oligonucleotide with a sequence present in this gene, which did not code for corazonin or a related peptide. However, the other clone contained a perfect match with the oligonucleotide used for screening the library (Fig. 1).

The nucleotide sequence of this clone contains a TATA-box an arthropod initiator sequence [13]. The first possible start codon after the arthropod initiator sequence is the start of an open reading frame which predicts a 72-amino acid precursor. Its first 19 amino acids are those of a typical signal peptide, having a single basic amino acid near the N-terminal, a hydrophobic stretch, and ending in a possible cleavage site for a signal peptide [14]. Immediately after the putative signal peptide follows Gln-Thr-Phe-Tyr-Ser-Arg-Gly-Trp-Thr-Asn-Gly and a pair of dibasic amino acids. Since N-terminal pyroglutamate is a secondary modification of Gln [15], and C-terminal amides found in neuropeptides are the result of a secondary modification of a C-terminal Gly [16], this sequence is exactly as expected for a corazonin precursor. The last part of the precursor sequence predicts a 39 amino acid residue peptide with an internal Arg-Arg sequence. This peptide may therefore be further processed into a tripeptide and a 34 amino acid residue peptide. The structure of the 39 amino acid residue peptide is unrelated to any known peptide, and its function is unknown. The structures of such precursor-related peptides from the

```
taatctgggtaagtgtgtgttctggagggcattgaagcaaggattctggaggaaaacgcagacacatcacgcatc     75

ctttggcatcaatccttaatggctatctaacaacgatataaaagcccgtcgaggagcatcaatagttcagacgca    150

gttgtgattcgaacgcgaaaggaaccaaccgagattaccgagtgtcctgcaaacaggactaacttctgccgaaac    225

ATG TTG CGC CTC CTG CTG CTG CCC CTC TTC CTC TTC ACG CTC TCC ATG TGC ATG GGC    282
Met Leu Arg Leu Leu Leu Leu Pro Leu Phe Leu Phe Thr Leu Ser Met Cys Met Gly

<---------- Signal peptide ---------->

CAG ACC TTC CAG TAC TCC CGC GGA TGG ACC AAC GGC AAG AGG TCC TTT AAC GCC GCA    339
Gln Thr Phe Gln Tyr Ser Arg Gly Trp Thr Asn Gly Lys Arg Ser Phe Asn Ala Ala
                                              *** ***
<---------- Corazonin ----------> <----------

TCT CCC CTC CTG GCC AAC GGC CAT CTC CAT CGG GCC AGC GAG CTG GGA CTC ACG GAT    396
Ser Pro Leu Leu Ala Asn Gly His Leu His Arg Ala Ser Glu Leu Gly Leu Thr Asp

---------- Corazonin-precursor related peptide ----------

CTC TAC GAT TTG CAG GAT TGG AGC AGC GAT CGT AGG CTC GAG CGG TAA gtttcatcgaa    455
Leu Tyr Asp Leu Gln Asp Trp Ser Ser Asp Arg Arg Leu Glu Arg ---
                                        *** ***
---------- ? ? ---------->

agaataacagaaatatacaactttatttttaaaatacataaagaatatgataacccaaaaacattccattagata    530

cctattaaaatcatataactaaggaattatatattataattgccctacttatacacttcattcctacttattttc    605

ttgtttttatgtattaaagtaatattttaaaattttcctccagctgtctatcgcagctccaacgttcacagatt    680
```

Fig. 1. Nucleotide and deduced amino acid sequence of *Drosophila* preprocorazonin. The sequences of the signal peptide, corazonin and the corazonin-precursor-related peptide are underlined. The TATA-box, arthorpod initiator sequence and polyadenylation homologies are in capitals and putative dibasic proteolytic processing sites are indicated by asteriks. Bases 283 through 317 match perfectly with the oligonucleotide used for screening.

AKH/RPCH family have not been conserved during evolution, except between very closely related species [8-11]. It has been suggested elsewhere, that their sole function may be to give sufficient length to the preprohormone for proper processing [11]. A variant polyadenylation site (ATTAAA) is present 84 base pairs downstream of the stop codon.

When corazonin was first isolated, a structural similarity was noted with peptides of the AKH/RPCH-family, most strongly with the hypertrehalosemic hormone from *Heliothis zea* [5]. Except for this structural similarity between corazonin and the adipokinetic hormone family there is also a functional similarity, as both corazonin and at least some of the insect members of the AKH/RPCH family stimulate the rate of heart beat [6,17]. In insects this arthropod peptide family has gone through rapid evolution, resulting in what is perhaps the most diverse neuropeptide family. In crustaceans on the other hand, only a single peptide has been found so far [3]. Sequences of several AKH/RPCH-related peptide-cDNAs and genes have now been determined [8-11]. The predicted structures of all the peptide precursors is very similar, a short signal peptide followed immediately by the sequence of the AKH/RPCH-related peptide, a glycine, a pair of dibasic amino acids and remaining amino acids of the precursor. This structure is remarkably similar to that of the *Drosophila* corazonin gene. Two facts in particular point this

out: (1) there is only a single copy of the peptide encoded by the gene, and not a larger number of related peptides, as in other other insect genes encoding small neuropeptides [18-21], and (2) the N-terminal of the peptide arises from the precursor by cleavage of the signal peptide, and not by cleavage at another pair of dibasic amino acids. The lengths of the signal peptides and the remaining parts of the precursor peptides are also similar, although the RPCH-precursor is significantly longer [11]. This reinforces the notion that corazonin is evolutionarily related to the AKH/RPCH peptide family. Indirect evidence in support of this hypothesis also comes from the localization of these peptides within the neuroendocrine system. Homologous peptides like crustacean RPCH and the insect AKHs can be expected to be expressed in homologous cells [22]. However the expression of AKH/RPCH-related peptides in insects and crustaceans is markedly different. In insects the expresssion of AKH-related peptides is strictly limited to the glandular cells of the corpus cardiacum, as demonstrated both by immunocytochemistry in several insect species [23], and by analysis of mRNA in a more limited number of species [8-10]. RPCH, on the other hand, is expressed in pairs of neurons in the segmental ganglia, as well as in neurosecretory cells with neurohemal release sites in the sinus gland [24], a pattern which is similar to that recently described for corazonin in the American cockroach, *Periplaneta americana* [25]. This suggests that during evolution, the AKH/RPCH family may have split in two branches, one, the insect AKH-related peptides being expressed exclusively in the the corpus cardiacum, the other, corazonin, being expressed within the nervous system.

ACKNOWLEDGMENTS

I thank John Tamkun for an aliquot of his excellent genomic *Drosophila melanogaster* libray, and Henry H. Hagedorn and René Feyereisen for their interest and critical reading of the manuscript. Supported by a grant from the Agricultural Experiment Station at the University of Arizona (ARZT-136327-H-31-021) to Henry H. Hagedorn and in part by a grant from the National Institutes of Health (AI 33429) to the author.

REFERENCES

1. Krieger, D.T. (1983) Science 222, 975-985.
2. Joosse, J. (1986) in *Comparative Endocrinology: Developments and Directions* (C.L. Ralph, Ed) pp 13 - 32, Alan R. Liss, Inc.
3. Gaus, G., Kleinholz, L.H., Kegel, G., and Keller, R. (1990) J. Comp. Physiol. B 160, 373-379.
4. Gäde, G. (1991) Biol. Chem. Hoppe-Seyler 372, 193-201.
5. Veenstra, J.A., and Camps, F. (1990) Neuropeptides 15, 107-109.
6. Veenstra, J.A. (1989) FEBS Lett. 250, 231-234.
7. Veenstra, J.A. (1991) Peptides 12, 1285-1289.
8. Schulz-Aellen, M.F., Roulet, E., Fisher-Lougheed, J., and O'Shea, M. (1989) Neuron 2, 1369-1373.
9. Bradfield, J.Y., and Keeley, L.L. (1989) J. Biol. Chem. 264, 12791-12794.
10. Noyes, B.E., and Schaffer, M.H. (1990) J. Biol. Chem. 265, 483-489.
11. Linck, B., Klein, J.M., Mangerich, S., Keller, R., and Weidemann, W.M. (1993) Biochem. Biophys. Res. Comm. 195, 807-813.
12. Sambrook, J. Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
13. Cherbas, L., and Cherbas, P. (1993) Insect Biochem. Molec. Biol. 23, 81-90.
14. Von Heijne, G. A (1986) Nucleic Acids Res. 14, 4683-4690.

15. Smyth, D.G., Stein, W.H., and Moore, S. (1963) J. Biol. Chem. 238, 227-234.
16. Bradbury, A.F., Finnie, M.D.A., and Smyth, D.G. (1982) Nature 298, 686-688.
17. Scarborough, R.M., Jamieson, G.C., Kalish, F., Kramer, S.J., McEnroe, G.A., Miller, C.A., and Schooley, D.A. (1984) Proc. Natl. Acad. Sci. USA 81, 5575-5579.
18. Schneider, L.E., and Taghert, P.H. (1988) Proc. Natl. Acad. Sci. USA 85, 1993-1997.
19. Nichols, R., Schneeuwly, S.A., and Dixon, J.E. (1988) J. Biol. Chem. 263, 12167-12170.
20. Kawano, T., Kataoka, H., Nagasawa, H., Isogai, A., and Suzuki, A. (1992) Biochem. Biophys. Res. Comm. 189, 221-226.
21. Donly, B.C., Ding, Q., Tobe, S.S., and Bendena, W.G. (1993) Proc. Natl. Acad. Sci. USA 90, 8807-8811.
22. Veenstra, J.A. (1988) Neuropeptides 12, 49-54.
23. Schooneveld, H., Romberg-Privee, H.M., and Veenstra, J.A. (1987) J. Insect Physiol. 33, 167-176.
24. Mangerich, S., Keller, R., and Dircksen, H. (1986) Cell Tissue Res. 245, 377-386.
25. Veenstra, J.A., and Davis, N.T. (1993) Cell Tissue Res 274, 57-64.