Volume 226, number 1, 8–12 FEB 05400 December 1987

Insect immunity: cDNA clones coding for the precursor forms of cecropins A and D, antibacterial proteins from *Hyalophora cecropia*

Dan-Anders Lidholm, Gudmundur H. Gudmundsson, Kleanthis G. Xanthopoulos and Hans G. Boman

Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden

Received 2 November 1987

Using synthetic probes cDNA clones were isolated correponding to the precursor forms of cecropins A and D. The sequences obtained were compared to earlier data for preprocecropin B. A processing scheme in three or four steps is discussed.

cDNA clone; Cecropin A; Cecropin D; Nucleotide sequence; Amino acid sequence; (*Hyalophora cecropia*)

1. INTRODUCTION

The cecropins are potent antibacterial proteins that constitute a main part of the cell-free immunity in higher insects (for a recent review see [1]). Cecropins A and B are 37 and 35 amino acid residues long, respectively, and they both have broad spectrum antibacterial properties. The D-form is 36 amino acid residues long and it has a narrow spectrum with activity only against *E. coli* and two other bacterial species. Cecropins and other immune proteins can be selectively induced in diapausing pupae by an injection of live non-pathogenic bacteria [1]. Thus, the immune genes in Cecropia are a useful model system for selective gene activation.

Sequences of cDNA clones indicate that cecropin B from the Cecropia moth [2] and cecropin IA from the meat fly *Sarcophaga* [3]are made as larger precursor proteins. However, the proforms differed in these two cases. Here, we now report the precursor sequences for cecropins A and D from the Cecropia moth.

Correspondence address: H.G. Boman, Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden

2. MATERIALS AND METHODS

The over-all procedure for the isolation of cDNA clones was the one described by Maniatis et al. [4]. Two different cDNA libraries both based on poly(A) containing immune RNA were used, one was constructed in pBR322 [5], and the other in pUC9 according to the method of Heidecker and Messing [6]. Two probes were used, one based on the amino acids 28–38 of cecropin A (5′-CCITTIGCIATITGIGTIGCITGICCIACIAC-3′; where I is deoxyinosine, forming hydrogen bonds to A, T or C) [7] and the second based on the first 26 nucleotides encoding the precursors of cecropin A and B (5′-ATGAAYTTYTCIAGXATITTIT-TYTT-3′; where Y is either thymidine or cytidine; X is either adenosine or guanosine; and I is deoxyinosine). Both probes were kindly prepared by KabiGen, Stockholm. They were labeled with ^{32}P either 5′ [4] or 3′ [8] and separately used to screen the two libraries. A total of 8×10^3 colonies were

Published by Elsevier Science Publishers B.V. (Biomedical Division)

screened utilizing nylon filters (Amersham, Hybond-N). Hybridization was done in 6 × SSC, 5 × Denhardts solution, 0.5% $NaDodSO_4$ and 100 μg/ml of Herring sperm DNA at 42°C for 18 h, with the probe at 5×10^5–1×10^6 cpm/ml. (1 × SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0; Denhardt's solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin.) Filters were washed several times in 6 × SSC at room temperature, then washed at 50°C in 6 × SSC for 3 min and autoradiographed against Fuji RX-film at −70°C using intensifying screens. Colonies that hybridized to the probes were rescreened twice and selected for sequencing. Clones pCP9A1 and pCP9A2 were isolated with the probe corresponding to amino acids 28–38 while clone pCP9D1 was obtained with the N-terminal probe. DNA sequences were determined by the method of Maxam and Gilbert [8]. Computer alignment of DNA sequences (fig.1) was done by 'Bestfit' in the Wisconsin UWGCG program.

Fig.1. Nucleotide sequences of cDNA clones corresponding to the precursors of cecropin A, B and D (pCP901–pCP902 are overlapping cecropin B clones, pCP9A1–pCP9A2 are overlapping cecropin A clones and pCP9D1 is a cecropin D clone). Nucleotide homology (A/B, A/D) is indicated by dots (:) and homology (B/D) where A differs is underlined. Start codons, stop codons and poly(A) signals are written in bold letters. Gaps are indicated by dotted lines. The sequences were aligned by computer using the Wisconsin UWGCG 'Bestfit' program.

3. RESULTS

3.1. *Nucleotide sequences of cDNA clones for cecropins A and D*

Three of the isolated clones corresponding to cecropins A and D (pCP9A1, pCP9A2 and pCP9D1) contain the complete protein encoding regions. Fig.1 gives the DNA sequences of these clones together with the overlapping sequences of the two cecropin B clones characterized earlier [2]. The beginning and the end of the open reading frames are marked by bold letters. The polypeptides coded for are 64 (the A form) and 62 (the D form) amino acid residues long, respectively. Thus, like the B form [2], cecropins A and D are translated as precursors about twice the size of the mature cecropins. The three precursors are quite similar (59–71% homology within the coding regions), while for A and D beyond the stop signal there was only 31% homology. In this area the AT content was 82%. Clone pCP9D1 contains a poly(A) signal, AATAAA [9] but it is not polyadenylated. Clone pCP9A2 has two closely positioned poly(A) signals, 14 nucleotides upstream of the poly(A) site while pCP9A1 is truncated and lacks these signals. For the two cecropin A clones, the overlapping regions sequenced are identical.

3.2. *Translated protein sequences for preprocecropins A and D*

Fig.2 shows the translation of the open reading frames of the clones for cecropins A and D. Included are also earlier data from two clones for cecropin B [2]. The structure of the precursor proteins was divided into four regions of which the mature protein with a C-terminal extension corresponds to region 4. The sequences in this part of fig.2 confirm the protein sequences previously published for cecropins A and D [10,11]. The leader peptides are divided into three parts which were compared for homologies (table 1). The first 12 amino acid residues in the N-terminus (region 1) are identical in the A and B forms while 7 of these residues are identical in the D form. The 5 residues altered in region 1 of the precursor of cecropin D are conservative replacements. Region 1 is conserved also in cecropin IA from the meat fly *S. peregrina* (6 of 12 residues identical, 5 of the 6 altered residues are conservative replacements) [3].

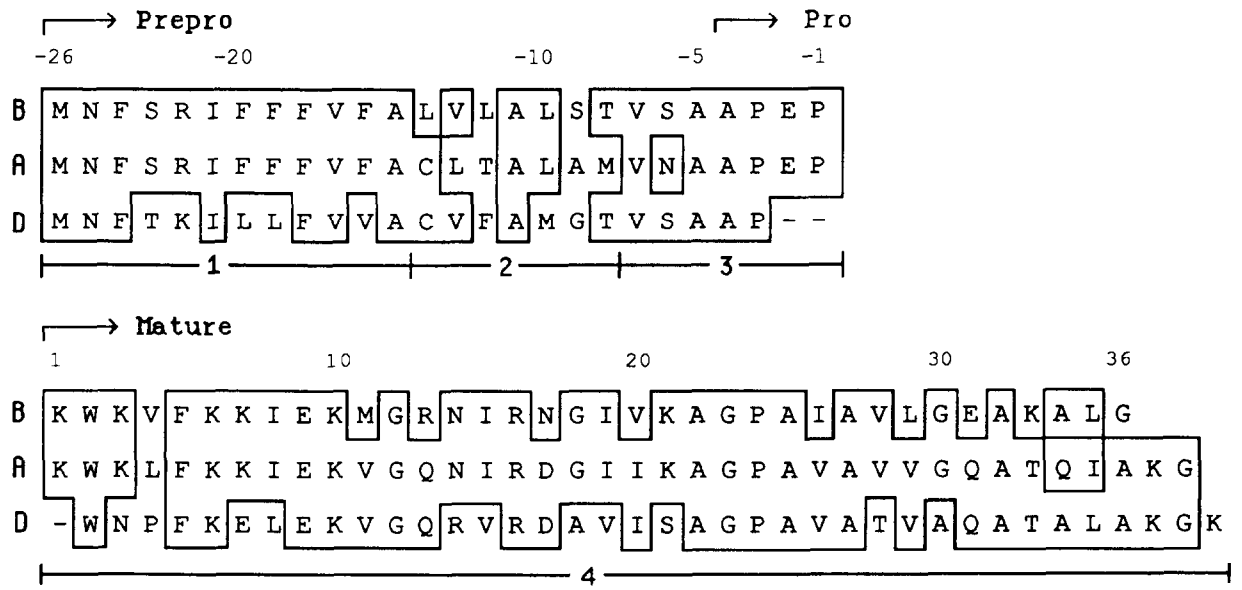

Fig.2. Amino acid sequences of the prepro-forms of cecropins A, B and D. Identical amino acid residues in one letter code are boxed. Prepro-, pro- and mature forms are indicated by arrows. The mature cecropins do not include the C-terminal Gly in cecropins A and B and C-terminal Gly-Lys in cecropin D. These glycine residues are believed to be the donors of an amide group to the preceding residue. The processing of the precursors and the regions indicated are discussed in the text.

Table 1

Comparison of amino acid sequences of preprocecropin A, B and D (regions as presented in fig.2)

Region no.	% homology		
	A/B	A/D	B/D
1	100	58	58
2	29	29	43
3	86	80	100
4	66	63	40
1–4	71	60	48

Thus, the N-terminus is highly conserved in all known prepro-sequences. Region 2 is clearly less conserved than any other part of the molecule and can serve as a control for the rate of evolution. Region 3 again is strongly conserved, presumably because it contains the recognition structures needed by the processing enzyme(s).

4. DISCUSSION

We have previously suggested [2] that the precursor of cecropin B is processed in four different steps by three different enzymes: cleavage by a signal peptidase that removes residues −26 to −5, stepwise action by a dipeptidylpeptidase that removes two Pro containing dipeptides (residues −4 to −1) and finally cleavage of the C-terminal Gly residue to give an amidation of the next last residue. The close similarity between the precursors of the A and B forms strongly suggests that they are both processed in the same manner. The argument for involving a dipeptidylpeptidase is firstly that no signal peptidase has been found to cleave a Pro-Lys bond (between residues −1 and +1) [12]. Secondly, the sequence Ala-Pro-Glu-Pro is identical to the first four residues of the pro-part of the melittin precursor which is known to be processed stepwise by a dipeptidylpeptidase IV [13]. Assuming these processing steps lead to the prediction that the signal peptidase should cleave the Ala-Ala bond between residues −5 and −4, a suggestion in agreement with the known specificity of signal peptidases [12]. In consonance is also the fact that this bond is conserved in all three preprocecropins from the Cecropia moth. The cecropin precursor in *Sarcophaga* lacks Pro containing dipeptides [3]. However, in this case the border between the leader peptide and the mature protein is an Ala-Gly bond which is known to be cleaved by signal peptidases [12].

The prepro-form of cecropin D differs somewhat from the corresponding molecules for the A and B form. Firstly, there is only one Pro containing dipeptide. However, since also in this case no signal peptidase has been found to cleave a Pro-Trp bond [12] we believe that a dipeptidylpeptidase is involved. Secondly, in the C-terminus there is beyond the Gly residue an additional Lys. The present in vitro data suggest that the amidating enzyme works only on a terminal Gly [14]. It is unlikely that the enzyme would accept a terminal Gly-Lys and for the processing of cecropin D we must therefore involve the preceding action of a carboxypeptidase H as found to be the case for some neurohormones [15].

Some small gene coded peptides are made as polyproteins and this applies to the magainins, the antibacterial peptides made in frog skin [16]. For other small exported proteins there appears to be a lower size limit imposed by the biosynthetic machinery. The cecropin precursors are only 62 and 64 amino acid residues long and they are among the smallest molecules found to be primary translation products that are secreted. In the case of other small precursors like M13 procoat protein [17], prepromelittin [18] and frog skin prepropeptide GLa [19] available data indicate a post-translational export mechanism that is ATP-dependent but independent of a signal recognition particle and a docking protein. The key feature suggested for this mechanism is that the signal sequence and the mature protein form a special conformation that is recognized by the translocation machinery. Since the size of the cecropin precursors are equal to the smallest of the above preproproteins we suggest that the same export mechanism could also be used for the cecropins. An alternative would be a mechanism of receptor-mediated transport in which the first 12 residues of the N-terminus are used for recognition.

ACKNOWLEDGEMENTS

We like to thank R. Gan and D. Hultmark for advice on different experimental methods. The

work was supported by The Swedish Natural Science Research Council by grant BU 2453.

REFERENCES

[1] Boman, H.G. and Hultmark, D. (1987) Annu. Rev. Microbiol. 41, 103–126.
[2] Von Hofsten, P., Faye, I., Kockum, K., Lee, J.-Y., Xanthopoulos, K.G., Boman, I.A., Boman, H.G., Engström, Å., Andreu, D. and Merrifield, R.B. (1985) Proc. Natl. Acad. Sci. USA 82, 2240–2243.
[3] Matsumoto, N., Okada, M., Takahasi, H., Ming, Q.X., Nakajima, Y., Nakanishi, Y., Komano, H. and Natori, S. (1986) Biochem. J. 239, 717–722.
[4] Maniatis, T., Fritsch, I.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[5] Lee, J.-Y., Edlund, T., Ny, T., Faye, I. and Boman, H.G. (1983) EMBO J. 2, 577–581.
[6] Heidecker, G. and Messing, J. (1983) Nucleotide Acid Res. 11, 4891–4906.
[7] Ohtsuka, I., Matsuki, S., Ikehara, M., Takahasi, Y. and Matsubara, K. (1985) J. Biol. Chem. 260, 2605–2608.
[8] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[9] Proudfoot, N.J. and Brownlee, J.J. (1976) Nature 263, 211–214.
[10] Steiner, H., Hultmark, D., Engström, Å., Bennich, H. and Boman, H.G. (1981) Nature 292, 246–248.
[11] Hultmark, D., Engström, Å., Bennich, H., Kapur, R. and Boman, H.G. (1982) Eur. J. Biochem. 127, 207–217.
[12] Von Heijne, G. (1985) J. Mol. Biol. 184, 99–105.
[13] Kreil, G., Haiml, L. and Suchanek, G. (1980) Eur. J. Biochem. 111, 49–58.
[14] Eipper, B.A., Mains, R.E. and Glembotski, C.C. (1983) Proc. Natl. Acad. Sci. USA 80, 5144–5148.
[15] Lynch, D.R. and Snyder, S.H. (1986) Annu. Rev. Biochem. 55, 773–799.
[16] Zasloff, M. (1987) Proc. Natl. Acad. Sci. USA 84, 5449–5453.
[17] Watts, C., Wickner, W. and Zimmermann, R. (1983) Proc. Natl. Acad. Sci. USA 80, 2803–2813.
[18] Müller, G. and Zimmermann, R. (1987) EMBO J. 6, 2099–2107.
[19] Schlenstedt, G. and Zimmermann, R. (1987) EMBO J. 6, 699–703.