Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2240–2243, April 1985
Biochemistry

Molecular cloning, cDNA sequencing, and chemical synthesis of cecropin B from *Hyalophora cecropia*

(insect immunity/solid-phase peptide synthesis/COOH-terminal amide/cecropin B precursor/antibacterial peptides)

P. VAN HOFSTEN*, I. FAYE*, K. KOCKUM*, J.-Y. LEE*, K. G. XANTHOPOULOS*, I. A. BOMAN*, H. G. BOMAN*, Å. ENGSTRÖM†, D. ANDREU‡, AND R. B. MERRIFIELD‡

*Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden; †Department of Immunology, Biomedical Center, University of Uppsala, S-751 23 Uppsala, Sweden; and ‡The Rockefeller University, New York, NY 10021

Contributed by R. B. Merrifield, November 26, 1984

ABSTRACT **Two cDNA clones containing coding information for cecropin B from the Cecropia moth (*Hyalophora cecropia*) were identified by means of a synthetic probe. Sequencing of the two inserts showed that cecropin B is processed from a 62-amino acid residue precursor molecule including a 26-residue leader peptide and a COOH-terminal glycine residue. The latter presumably donates the nitrogen of the amide group present on the COOH-terminal leucine residue of the mature cecropin B. The sequence deduced for the mature cecropin B differed in the COOH-terminal region from the tentative structure previously determined by carboxypeptidase digestion. To settle the discrepancy, cecropin B was synthesized according to the cDNA sequence with an amidated COOH-terminal leucine. Natural and synthetic cecropin B were found to be indistinguishable with respect to electrophoretic mobility and antibacterial activity against seven different bacteria. The COOH-terminal tetrapeptides were isolated from both natural and synthetic cecropin B and found to be indistinguishable. The correct sequence for cecropin B is Lys-Trp-Lys-Val-Phe-Lys-Lys-Ile-Glu-Lys-Met-Gly-Arg-Asn-Ile-Arg-Asn-Gly-Ile-Val-Lys-Ala-Gly-Pro-Ala-Ile-Ala-Val-Leu-Gly-Glu-Ala-Lys-Ala-Leu-NH_2.**

Humoral immunity can be induced in many insects by an injection of live nonpathogenic bacteria. This phenomenon has been analyzed at the molecular level by using diapausing pupae of the Cecropia moth (*Hyalophora cecropia*) as a model system (1, 2). After a short period of RNA synthesis the insects respond to the bacteria by the production of a potent antibacterial activity which is due to the composite action of lysozyme and two new classes of bactericidal proteins, the cecropins (3, 4) and the attacins (5–7). The cecropins are small basic proteins with a comparatively long hydrophobic region. The three principal cecropins, A, B, and D, were isolated from immune Cecropia hemolymph. Sequence work revealed a high degree of homology among these three forms, which suggests that they have evolved through gene duplications. Similar conclusions were reached concerning the two main forms of attacin.

To understand the selective induction of the immune genes, a gene cloning program has been initiated. cDNA clones and sequences for the two main forms of attacin were first obtained (7, 8). We now report the isolation and sequence of two cDNA clones together containing the complete information for the precursor of cecropin B. The amino acid sequence deduced for the mature form of cecropin B differs at two residues in the COOH terminus from the tentative structure previously obtained from sequencing the peptide (3). To settle the discrepancy we synthesized the peptide predicted from the DNA sequence. A detailed comparison of the natural and the synthetic materials indicates that the two products are identical.

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.

MATERIALS AND METHODS

Molecular Cloning. The methods used in the cloning work were in general those given by Maniatis *et al.* (9). The construction of the cDNA bank and other methods used were recently described in detail (7, 8). The oligonucleotide probe was kindly synthesized by Staffan Josephson (KabiGen, Stockholm).

Solid-Phase Peptide Synthesis. *p*-Methylbenzhydrylamine resin·HCl was purchased from U.S. Biochemical (Cleveland, OH). Reagents and solvents for peptide synthesis were as previously described for the synthesis of cecropin A (10). The synthesis was carried out in a Beckman 990 automatic peptide synthesizer that had been modified for computer-assisted operation. Complete details on this computer-controlled synthesizer will be published elsewhere.

Purity Test and Bioassay of Cecropin B. Polyacrylamide gel electrophoresis of native cecropins at pH 4 with overlaying of *Escherichia coli* D31 and the antibacterial assay used for obtaining the lethal concentrations were as described earlier (5, 11).

Amino Acid Sequencing. The equipment used for automatic Edman degradation, the procedures for cleavage with *Staphylococcus aureus* V8 protease, amino acid analysis, and the HPLC separation of peptides were recently given in detail (6).

RESULTS

Our cDNA bank was screened for cecropin clones by using a synthetic probe (5′-C-C-C-A-T-$\frac{T}{C}$-T-T-$\frac{T}{C}$-T-C-$\frac{A}{G}$-A-T-3′) corresponding to residues 8–12 of cecropin B (3), the only region of the peptide that can be translated to a limited number of oligomers. This screening revealed three positive clones, two of which were sequenced and found to correspond to cecropin B. The cecropins were originally referred to as immune proteins P9 and the clones were therefore named pCP901 and pCP902, respectively (with C for Cecropia). The third clone was not sequenced, but judging from the restriction analysis it is most likely also a cecropin B clone.

Characterization and Sequence of Two cDNA Clones for Cecropin B. Fig. 1 shows the restriction map obtained for pCP901 and pCP902. Two enzymes, *Hin*cII and *Pvu* I, were used to cleave the clones into fragments suitable for sequencing, using the strategy indicated in the lower part of Fig. 1. Other restriction sites were obtained by computer analysis. The shorter clone (pCP901) is 173 bases, the longer one (pCP902) is 242 bases, and they have 152 bases in common, including the entire coding region for the mature cecropin B.

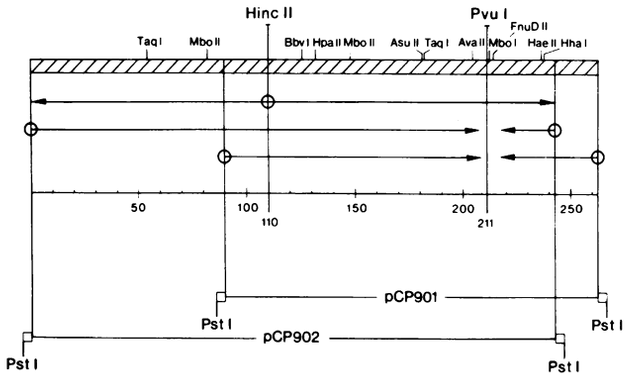

FIG. 1. Restriction analysis of the inserts in clones pCP901 and pCP902. The arrows in the lower part indicate the length and the direction of the sequencing experiments. The restriction sites printed in small letters were obtained by computer analysis. The boxes linked to the *Pst* I sites represent the G·C tails from the cloning procedure. The scale gives the number of nucleotides.

Fig. 2 gives the results obtained from the DNA sequencing performed according to Maxam and Gilbert (12). There are three points in the inferred protein structure that deserve comments: (*i*) Clone pCP902 contains a precursor sequence corresponding to 26 amino acid residues in the NH_2 terminus that is not present in the mature cecropin B. (*ii*) The mature cecropin B sequence encoded by the cDNA clones is lacking two amino acid residues present in the tentative structure previously published, namely Ile-35 and Ser-37. (*iii*) The COOH-terminal residue translated from the DNA sequence is Gly. Since cecropin A has been shown to carry a COOH-terminal residue blocked by an amide group (10) and since amide groups can be derived from glycine residues (see *Discussion*), cecropin B is believed to end with an amidated Leu at position 35. If so, the correct molecular weight for cecropin B is 3835.

The wobble base frequencies obtained from the sequence for the cecropin B precursor are as follows (in %): A, 29; T, 23; G, 27; C, 21. Thus, there is no evident preference for any of the bases, a result different from that found for the two attacins (7).

Solid-Phase Synthesis of Cecropin B. To confirm the cecropin B sequence given in Fig. 2, solid-phase methods (13) were used to synthesize this molecule with an amidated Leu-

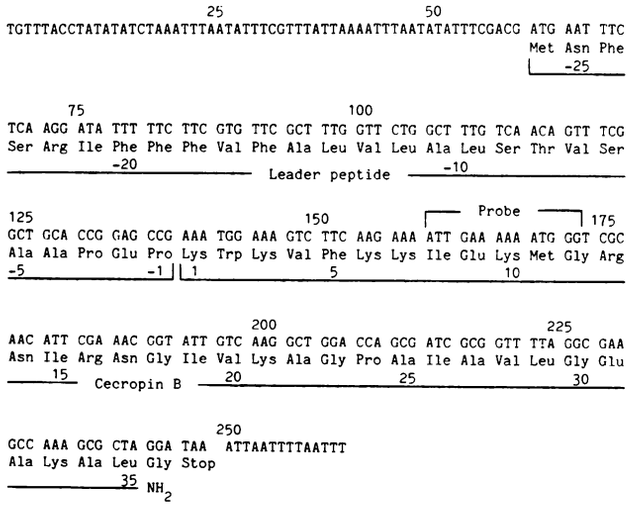

FIG. 2. Nucleotide sequence of the inserts in clones pCP901 and pCP902. The structures of the leader peptide and the mature cecropin B molecule are indicated by lines under the respective parts of the amino acid sequence. The principal structure of the probe (without alternative wobble bases) is indicated above the base sequence.

35 at the COOH terminus. *p*-Methylbenzhydrylamine resin was chosen as the solid support because of its comparable stability but better cleavage yields relative to benzhydrylamine resin (14). The synthesis was performed in a computer-controlled synthesizer using a standard dicyclohexylcarbodiimide double coupling protocol for all residues except Asn at positions 17 and 14, which was coupled as its 1-hydroxybenzotriazole active ester. As before (10), a low level of radioactivity was introduced in the COOH-terminal region, in this case by labeling of Leu-35 with ^3H (0.25 mCi/mmol; 1 Ci = 37 GBq). The structure of the completed peptide-resin is as follows:

Boc-Lys(ClZ)-Trp(For)-Lys(ClZ)-Val-Phe-Lys(ClZ)-Lys(ClZ)-Ile-Glu(OBzl)-Lys(ClZ)-Met(O)-Gly-Arg(Tos)-Asn-Ile-Arg(Tos)-Asn-Gly-Ile-Val-Lys(ClZ)-Ala-Gly-Pro-Ala-Ile-Ala-Val-Leu-Gly-Glu(OBzl)-Ala-Lys(ClZ)-Ala-Leu-NH-CH(*p*-CH_3-C_6H_5)-C_6H_4-resin,

in which ClZ is 2-chlorobenzyloxycarbonyl, For is formyl, OBzl indicates a benzyl ester, Met(O) indicates methionine sulfoxide, Tos is *p*-toluenesulfonyl, and the resin is copoly-(styrene/1% divinylbenzene).

Complete deprotection and cleavage of cecropin B from the resin was done by the recently developed low–high HF procedure (15). The peptide was purified to homogeneity by preparative reversed-phase liquid chromatography on C_{18}-silica (Fig. 3). The overall yield was 23% based on initial Leu-35. Amino acid analysis after hydrolysis with either 6 M HCl or 4 M methanesulfonic acid gave the molar ratios expected for theoretical cecropin B.

Identity of Natural and Synthetic Cecropin B. The electrophoretic mobilities of natural and synthetic cecropin B at pH 4 were compared (Fig. 4). Both samples showed similar properties, although the synthetic material was slightly purer.

The antibacterial activities of the two samples were also examined against seven test bacteria (Table 1). The differ-

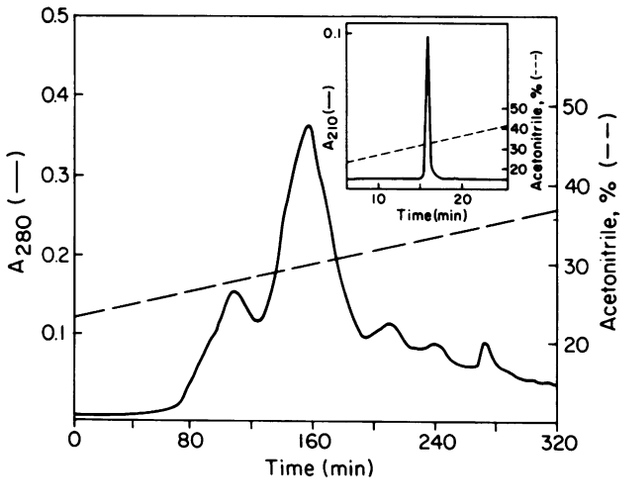

FIG. 3. Chromatographic purification of synthetic cecropin B. Crude peptide (24 mg) was dissolved in 2 ml of water and applied to a column (2.2 × 13 cm) of C_{18}-silica (10 μm, Waters Associates). Elution was with a linear gradient of acetonitrile in 0.5% trifluoroacetic acid in water at a flow rate of 2 ml/min. The mean peak was collected and rerun under the same conditions, yielding 9.5 mg of purified cecropin B. (*Inset*) Analytical HPLC of purified cecropin B on a Vydac 218 TP column (5 μm, 4.6 × 250 mm, The Separations Group, Hesperia, CA). Elution was with a linear gradient of acetonitrile in 0.45% trifluoroacetic acid in water at a flow rate of 1.5 ml/min.

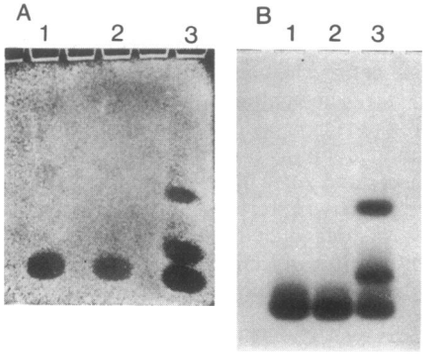

FIG. 4. Acid polyacrylamide gel electrophoresis of the cecropins. Samples were run towards the cathode in a 15% gel in β-alanine acetate buffer, pH 4. Gel *A* was overlayed with live *E. coli* D31, and gel *B* was stained for proteins with Coomassie brilliant blue. The samples and the amounts applied were as follows: gel *A*, lane 1, natural cecropin B (0.5 μg); lane 2, synthetic cecropin B (0.5 μg); lane 3, a mixture of natural cecropin A (middle spot) and cecropins B and D (0.5 μg of the B and D forms and 0.3 μg of the A form). Gel *B*, lane 1, natural cecropin B (12 μg); lane 2, synthetic cecropin B (12 μg), and lane 3, a mixture of cecropin A (middle spot) and cecropins B and D (6 μg of each). The difference in size between gels *A* and *B* is due to swelling of the latter during the staining procedure.

ences found are all within the experimental errors of the assay method. The results were compared with corresponding data for cecropin A (unpublished data). For six of the organisms, cecropin B was only slightly more potent than the A form. However, against *Staphylococcus capitis* cecropin B was about 3 times as active as cecropin A.

The differences between the cecropin B structure proposed earlier (3) and the one deduced from the cDNA sequence (Fig. 2) are confined to residues 35–37. Since residues 31 and 9 are Glu, digestion with V8 protease was used to liberate the COOH-terminal fragments from both natural and synthetic cecropin B. Both reaction mixtures were fractionated by HPLC on a μBondapak C_{18} column. The terminal tetrapeptide was identified by amino acid analysis in the case of the natural material (2 Ala, 1 Lys, and 1 Leu) and radioactivity in the case of the synthetic substance. When the two peptides were rerun under isocratic conditions, the retention times were identical (510 ± 5 sec, Fig. 5).

The natural tetrapeptide was also analyzed by ^{252}Cf plasma desorption time-of-flight mass spectroscopy (16). In the interval 120–800 daltons only one major peak was obtained, with a mass of 423.6 ± 0.2 daltons. The calculated value for Ala-Lys-Ala-Leu-NH_2·Na^+ is 423.51 daltons. The result is therefore consistent with an α-amidated COOH terminus in cecropin B. Finally, the natural tetrapeptide was deamidated and sequenced in the automatic Edman machine. The result confirmed the structure Ala-Lys-Ala-Leu.

Table 1. Lethal concentration for synthetic and natural cecropin B on different bacteria

Organism and strain	Lethal conc., μM	
	Synthetic cecropin B	Natural cecropin B
Escherichia coli D21	0.54	0.52
Acinetobacter calcoaceticus Ac11	0.67	0.72
Bacillus megaterium Bm11	0.75	0.95
Pseudomonas aeruginosa OT97	1.4	1.6
Micrococcus luteus Ml11	1.3	1.5
Staphylococcus capitis	6	7
Streptococcus pyogenes	5	5

Lethal concentration was computed from the inhibition zones obtained with a dilution series of cecropin B placed in small wells in thin agar plates seeded with the respective bacteria (5).

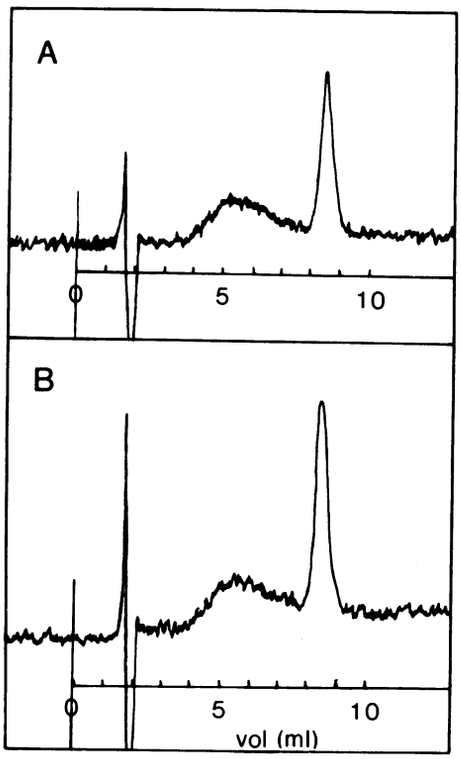

FIG. 5. Rechromatography of the COOH-terminal tetrapeptide from synthetic (*A*) and natural (*B*) cecropin B. The column was Waters μBondapak C_{18} eluted with 1% acetonitrile in 0.1% trifluoroacetic acid at a flow rate of 1 ml/min. The broad peak in front of the peptide is an impurity also appearing after an injection of water.

DISCUSSION

The sequencing of clones pCP901 and pCP902 yielded a 35-residue structure for cecropin B (Fig. 2) that in the COOH-terminal region differed from the 37-residue sequence previously proposed (3). Since the determination of the COOH-terminal structure by carboxypeptidase could have been disturbed by impurities in the cecropin B preparation we had more confidence in the cDNA sequence. It was therefore decided to synthesize the 35-residue structure with an amidated COOH terminus. The natural and the synthetic cecropin B were found to be indistinguishable with respect to electrophoretic mobility (Fig. 4) and antibacterial activity against seven different bacteria (Table 1). Moreover, the COOH-terminal tetrapeptides isolated from natural and synthetic cecropin B were also indistinguishable with respect to chromatographic behavior in HPLC (Fig. 5). Finally, the sequence of the tetrapeptide from the natural material was found to be Ala-Lys-Ala-Leu, while the mass spectroscopic analysis gave data consistent with an amidated Leu as the COOH-terminal residue. We therefore conclude that the COOH-terminal sequence previously reported is wrong and that the correct structure for cecropin B is that given in Fig. 2.

The precursor form of cecropin B shown in Fig. 2 suggests that the primary translation product is processed in two or possibly three steps to give the mature cecropin B. The leader peptide, which is 26 amino acid residues long, will have to be removed by one or more proteolytic enzymes. One possibility is that a membrane-bound signal peptidase recognizes the bond between Ala-(−5) and Ala-(−4) (refs. 17 and 18 and G. van Heijne, personal communication). If so, a dipeptidyl-peptidase similar to the one activating melittin (19) could recognize the two dipeptide units in Ala-Pro-Glu-Pro that preceed the NH_2 terminus of the mature cecropin B. An alternative mechanism with a signal peptidase that directly

recognizes a Pro-Lys bond may be less likely in view of present data (17).

Melittin and several hormones have been found to have amidated COOH-terminal residues (20, 21). In a number of these cases DNA sequencing has demonstrated that the COOH-terminal residue is followed by a Gly residue (21, 22) and a corresponding *in vitro* amidation has been demonstrated (23, 24). It is therefore most likely that amide formation from the terminal Gly residue is also a processing step in the biosynthesis of cecropin B.

We thank Dr. Donald Whitney for advice on the use of the new peptide synthesizer, Dr. Ivan Kamenski for the mass spectrum, and Gun Hellström for technical assistance. The work in Stockholm was supported by The Swedish Natural Science Research Council (Grant BU2453 to H.G.B.), in Uppsala by The Swedish Medical Research Council (Grant 13X-03556 to Å.E.) and in New York by the U.S. Public Health Service (Grant AM 01260 to R.B.M.). K.G.X. was a holder of a fellowship from the Alexander S. Onassis Public Benefit Foundation and D.A. was a fellow of the Fundación Juan March, Madrid.

1. Boman, H. G. & Steiner, H. (1981) *Curr. Top. Microbiol. Immunol.* **94/95,** 75–91.
2. Boman, H. G. (1982) *Zbl. Bakt. Suppl.* **12,** 211–222.
3. Steiner, H., Hultmark, D., Engström, Å., Bennich, H. & Boman, H. G. (1981) *Nature (London)* **292,** 246–248.
4. Hultmark, D., Engström, Å., Bennich, H., Kapur, R. & Boman, H. G. (1982) *Eur. J. Biochem.* **127,** 207–217.
5. Hultmark, D., Engström, Å., Andersson, K., Steiner, H., Bennich, H. & Boman, H. G. (1983) *EMBO J.* **2,** 571–576.
6. Engström, Å., Engström, P., Tao, Z.-J., Carlsson, A. & Bennich, H. (1984) *EMBO J.* **3,** 2065–2070.
7. Kockum, K., Faye, I., van Hofsten, P., Lee, J.-Y., Xanthopoulos, K. G. & Boman, H. G. (1984) *EMBO J.* **3,** 2071–2075.
8. Lee, J.-Y., Edlund, T., Ny, T., Faye, I. & Boman, H. G. (1983) *EMBO J.* **2,** 577–581.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
10. Andreu, D., Merrifield, R. B., Steiner, H. & Boman, H. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6475–6479.
11. Hultmark, D., Steiner, H., Rasmuson, T. & Boman, H. G. (1980) *Eur. J. Biochem.* **106,** 7–16.
12. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
13. Merrifield, R. B. (1963) *J. Am. Chem. Soc.* **85,** 2149–2154.
14. Matsueda, G. R. & Stewart, J. M. (1981) *Peptides* **2,** 45–50.
15. Tam, J. P., Heath, W. F. & Merrifield, R. B. (1983) *J. Am. Chem. Soc.* **105,** 6442–6454.
16. Sundqvist, B., Kamenski, I., Håkansson, P., Kjellberg, J., Salehpour, M., Widdiyasekera, S., Fohlman, J., Peterson, P. A. & Roepstorff, P. (1984) *Biomed. Mass Spectrom.* **11,** 242–257.
17. van Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21.
18. van Heijne, G. (1984) *J. Mol. Biol.* **173,** 243–251.
19. Kreil, G., Haiml, L. & Suchanek, G. (1980) *Eur. J. Biochem.* **111,** 49–58.
20. Kreil, G. (1973) *FEBS Lett.* **33,** 241–244.
21. Richter, D. (1983) *Trends Biochem. Sci.* **8,** 278–281.
22. Vlasak, R., Unger-Ullmann, C., Kreil, G. & Frischauf, A.-M. (1983) *Eur. J. Biochem.* **135,** 123–126.
23. Bradbury, A. F., Finnie, M. D. A. & Smyth, D. G. (1982) *Nature (London)* **296,** 686–688.
24. Eipper, B. A., Mains, R. E. & Glembotski, C. C. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5144–5148.