(1983) *Biophys. Struct. Mech. 9*, 235–244.
Hofmann, K. P., Emeis, D., & Schnetkamp, P. P. M. (1983) *Biochim. Biophys. Acta 725*, 60–70.
Kühn, H., Bennett, N., Michel-Villaz, M., & Chabre, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6873–6877.
Kühn, H., Hall, S. W., & Wilden, U. (1984) *FEBS Lett. 176*, 473–478.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Liebman, P. A., & Pugh, E. N., Jr. (1980) *Nature (London) 287*, 734–736.
Liebman, P. A., & Evanczuk, A. T. (1982) *Methods Enzymol. 81*, 532–542.
Liebman, P. A., & Pugh, E. N., Jr. (1982) *Vision Res. 22*, 1475–1480.
Miller, J. A., Paulsen, R., & Bownds, M. D. (1977) *Biochemistry 16*, 2633–2639.
Miller, J. L., Fox, D. A., & Litman, B. J. (1986) *Biochemistry 25*, 4983–4988.
Pfister, C., Chabre, M., Plouet, J., Tuyen, V. V., DeKozak, Y., Faure, J. P., & Kühn, H. (1985) *Science (Washington, D.C.) 228*, 891–893.
Pugh, E. N., Jr., & Cobbs, W. H. (1986) *Vision Res. 26*, 1613–1643.
Schnetkamp, P. P. M., Klompmakers, A. A., & Daemen, F. J. M. (1979) *Biochim. Biophys. Acta 552*, 379–389.
Sibley, D. R., Peters, J. R., Nambi, P., Caron, M. G., & Lefkowitz, R. J. (1984) *J. Biol. Chem. 259*, 9124–9149.
Sitaramayya, A. (1986a) *Biochemistry 25*, 5460–5468.
Sitaramayya, A. (1986b) *Invest. Ophthalmol. Visual Sci., Suppl. 27*, 217.
Sitaramayya, A., & Liebman, P. A. (1983) *J. Biol. Chem. 258*, 1205–1209.
Sitaramayya, A., & Casadevall, C. (1987) *Invest. Ophthalmol. Visual Sci. 28*, 94.
Sitaramayya, A., Harkness, J., & Liebman, P. A. (1985) *Invest. Ophthalmol. Visual Sci. 26*, 293.
Stadel, J. M., De Lean, A., Mullikan-Kilpatrick, B., Sawyer, D. D., & Lefkowitz, R. J. (1981) *J. Cyclic Nucleotide Res. 7*, 37–47.
Uhl, R., Borys, T., & Abrahamson, E. W. (1979) *FEBS Lett. 107*, 317–322.
Wilden, U., & Kühn, H. (1986) *FEBS Lett. 207*, 292–294.
Wilden, U., Hall, S. W., & Kühn, H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1174–1178.
Zuckerman, R., & Cheasty, J. E. (1986) *FEBS Lett. 207*, 35–41.

Molecular Cloning, Sequencing, and Characterization of cDNA for Sarcotoxin IIA, an Inducible Antibacterial Protein of *Sarcophaga peregrina* (Flesh Fly)†

Keiichi Ando and Shunji Natori*
Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
Received June 1, 1987; Revised Manuscript Received October 5, 1987

ABSTRACT: A cDNA clone for sarcotoxin IIA, an antibacterial protein of *Sarcophaga peregrina* (flesh fly) larvae [Ando, K., Okada, M., & Natori, S. (1987) *Biochemistry 26*, 226–230], was isolated and characterized. Sarcotoxin IIA was found to consist of 270 amino acid residues. Northern blot analysis showed that the sarcotoxin IIA gene was activated in response to injury of the body wall of the larvae. The gene was activated for much longer after injection of *Escherichia coli* into the abdominal cavity of larvae than after injection of saline alone. A common nucleotide sequence for mammalian inflammatory mediator protein cDNAs, TTATTTAT, was found in the 3′-untranslated region of sarcotoxin IIA cDNA, suggesting that this protein plays a role in the inflammatory response of this insect.

We are all familiar with the immune system in vertebrates. However, there are many unknown defense mechanisms that kill bacteria and viruses nonspecifically. The prototype of such nonspecific defense mechanisms can be found in invertebrates like insects. Therefore, it is important to investigate the defense mechanisms of insects from the viewpoint of comparative immunology.

Holometabolous insects have both humoral and cellular defense systems to protect themselves from various pathogens and to scavenge own-tissue fragments produced during metamorphosis (Boman et al., 1974; Pye & Boman, 1977; Scott, 1971; Gagen & Ratcliffe, 1976; Schmit & Ratcliffe, 1977). Of the humoral antibacterial proteins of *Lepidoptera*, cecropins and attacins, a series of antibacterial proteins of *Hyalophora cecropia*, have been purified (Steiner et al., 1981; Hultmark et al., 1982, 1983; Lee et al., 1983). Their primary structures were determined both by chemical sequencing of the proteins and from their cDNAs (engstrom et al., 1984; Kockum et al., 1984; von Hofsten et al., 1985). Cecropins were also isolated from *Antheraea pernyi*, and their primary structures were found to be almost identical with those of cecropins from *H. cecropia* (Qu et al., 1982).

In *Sarcophaga peregrina* (flesh fly), humoral antibacterial proteins are induced when the body wall of third instar larve is injured by a hypodermic needle (Natori, 1977). Previously we reported purification of antibacterial proteins termed sarcotoxin I and II (Okada & Natori, 1983; Ando et al., 1987). Sarcotoxin I consists of three structurally related proteins termed sarcotoxin IA, IB, and IC. Each protein has been purified to homogeneity and sequenced. These proteins each consist of 39 amino acid residues and differ only in 2 or 3 amino acid residues. From the sequences of sarcotoxin I and

† This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Science and Culture of Japan.
* Author to whom correspondence should be addressed.

0006-2960/88/0427-1715$01.50/0

cecropins, we concluded that these proteins belong to a similar protein family with antibacterial activity, although the former proteins are present in *Diptera* (Okada & Natori, 1985).

Sarcotoxin II consists of three structurally related proteins named sarcotoxin IIA, IIB, and IIC, each with a molecular mass of 24 000 Da (Ando et al., 1987). This paper reports molecular cloning and sequencing of cDNA for sarcotoxin IIA. We conclude that this protein consists of 270 amino acid residues with no Cys residue. Its overall amino acid sequence was not significantly homologous with those of attacins. However, significant homology was found between these proteins in a certain region near the carboxyl terminal, suggesting that the prototype gene of these antibacterial proteins existed before branching of *Diptera* and *Lepidoptera*.

Recently, we have obtained evidence that some defense proteins of *Sarcophaga* have dual functions both in the defense system and in development (Natori, 1987). Thus, studies on insect defense proteins are also important to understand the molecular mechanism of embryogenesis and adult development of holometabolous insects.

Materials and Methods

Sarcotoxin IIA and Its Antibody. Sarcotoxin IIA was purified from the hemolymph of third instar larvae of *Sarcophaga peregrina* as described previously (Ando et al., 1987). About 500 μg of pure sarcotoxin IIA was obtained from 260 mL of hemolymph. Antibody was raised against sarcotoxin IIA by immunizing a male albino rabbit with 75 μg of sarcotoxin IIA mixed with complete Freund's adjuvant and giving a booster injection of 50 μg of sarcotoxin IIA in incomplete Freund's adjuvant 21 days later. The animal was bled 8 days after the booster injection.

Isolation of Poly(A)$^+$ RNA and Its Translation in Vitro. Since sarcotoxin IIA is a hemolymph protein, it was likely to be synthesized in the fat body. So, poly(A)$^+$ RNA was isolated from the fat body of larvae that had been pricked with a hypodermic needle 6 h previously, according to the procedure described by Aviv and Leder (1972). For translation of poly(A)$^+$ RNA, the incubation mixture (final volume 12 μL) contained 7 μL of nuclease-treated reticulocyte lysate prepared by the method of Pelham and Jackson (1976), 20 μg of hemin/mL, 10 mM phosphocreatine, 1 mg of creatine kinase, 1 mM ATP, 0.2 mM GTP, 0.1 M KCl, 2 mM dithiothreitol, 1 mM magnesium acetate, 20 mM Hepes[1] (pH 7.0), a mixture of 19 amino acids (50 μM each) not including Met, 10 μCi of [^{35}S]Met (1100 Ci/mmol), and 230 ng of poly(A)$^+$ RNA. Incubation was done at 27 °C for 120 min.

Analysis of Translation Products. Translation products were treated with antibody against sarcotoxin IIA, and the resulting immunoprecipitate was analyzed by the method of Kessler (1975) with the Cowan I strain of *Staphylococcus aureus* as an adsorbent of the immunocomplex. For this, 12 μL of the translation reaction mixture was first treated with 10 μL of normal serum for 30 min at 4 °C. Then 300 μL of Cowan I cell suspension was added, and incubation was continued for 20 min. After centrifugation for 10 min at 1000*g*, 20 μL of antiserum was added to 280 μL of the resulting supernatant, and the mixture was incubated for 60 min at 4 °C. An equal volume of Cowan I cell suspension was then added, and incubation was continued for 20 min. The resulting immunocomplex was washed well by the procedure of Takahashi et al. (1985), and the final pellet was suspended in 40 μL of elution buffer [60 mM Tris-HCl buffer, pH 6.8, containing 1% (w/v) SDS, 10% (v/v) glycerol, and 5% (v/v) 2-mercaptoethanol], and heated at 100 °C for 5 min. The mixture was then centrifuged for 10 min at 10000*g*, and the resulting supernatant was subjected to SDS–polyacrylamide gel electrophoresis followed by fluorography.

Tryptic Digestion of Sarcotoxin IIA and Determination of the Amino Acid Sequences of the Resulting Peptides. To obtain partial amino acid sequences of sarcotoxin IIA suitable for the synthesis of DNA probes, sarcotoxin IIA (75 μg) was digested with tosylphenylalanyl chloromethyl ketone treated trypsin in 0.2 M $(NH_4)_2CO_3$ solution (pH 8.0) for 13 h at 37 °C at a molar enzyme/protein ratio of 1:50. The resulting peptides were analyzed on a reverse-phase HPLC column of Synchropak RP-P (C18) (250 × 4.1 mm, Synchrom Inc., Liden, IA) connected to a Gilson HPLC system and eluted with a linear gradient of 5–60% solution B [0.05% (v/v) trifluoroacetic acid in acetonitorile] in solution A [0.05% trifluoroacetic acid in H_2O] at a flow rate of 1 mL/min. The absorbance at 280 nm was monitored.

Fractions containing each tryptic fragment were collected, lyophilized, and subjected to automated sequence analysis in an Applied Biosystem 470A protein sequencer using the program described by Hunkapiller et al. (1983). Phenylthiohydantoin derivatives of amino acids were identified semiquantitatively in a Shimazu HPLC system equipped with a column of Unicilpack QC18.

Determination of the Partial Amino Acid Sequence from the Amino Terminal of Sarcotoxin IIA. The amino-terminal amino acid residue of sarcotoxin IIA was found to be blocked in some way, so we tested the effect of calf liver pyroglutamyl aminopeptidase by the method of Podell and Abraham (1978), assuming that the amino-terminal amino acid residue might be pyroglutamate. For this, 80 μg of sarcotoxin IIA was dissolved in 80 μL of 0.1 M phosphate buffer, pH 8.0, containing 5 mM dithiothreitol, 10 mM EDTA, 5% glycerol, and 8 μg of a lyophilized crude protein extract containing pyroglutamyl aminopeptidase (Boehringer Mannheim Biochemicals). After incubation for 9 h at 4 °C, the same amount of enzyme was added again and the mixture was incubated for 14 h at 20 °C. Then the sarcotoxin IIA was purified by HPLC and subjected to automated sequence analysis.

Cloning Procedure. A cDNA library was constructed by the method of Okayama and Berg (1982), using 10 μg of poly(A)$^+$ RNA extracted from the fat body and 3 μg of vector/primer DNA. *Escherichia coli* HB101 transformed by the method of Morrison (1979) was selected for ampicillin resistance as described by Hanahan and Meselson (1980) and screened by hybridization with a mixture of 32 synthetic oligodeoxyribonucleotides at 42 °C. These oligodeoxyribonucleotides containing three deoxyinosines (Ohtsuka et al., 1985) were designed to fit partial amino acid sequences of sarcotoxin IIA and were synthesized by a modification of the triester method (Hirose et al., 1978).

Other Analytical Procedures. Synthetic oligodeoxyribonucleotides were labeled as described by Kakidani et al. (1982). Nick translation with [α-^{32}P]dCTP was performed by the method of Weinstock et al. (1978). DNA sequences were determined by the dideoxyoligonucleotide chain termination method after each DNA fragment was subcloned into M13 phage (Sanger et al., 1977; Messing et al., 1981). DNA or RNA blot hybridization was performed in 50% (v/v) formamide/5 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate)/5 × Denhalt's solution [1 × Denhalt's solution =

[1] Abbreviations: Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; SDS, sodium dodecyl sulfate; HPLC, high-performance liquid chromatography; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid.

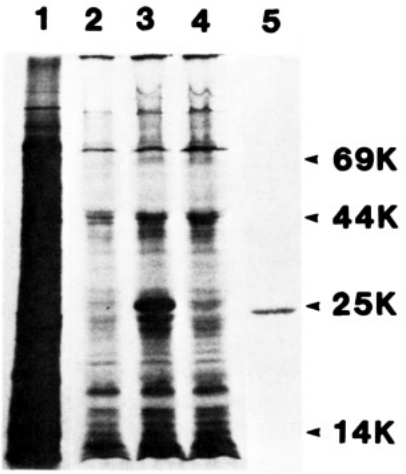


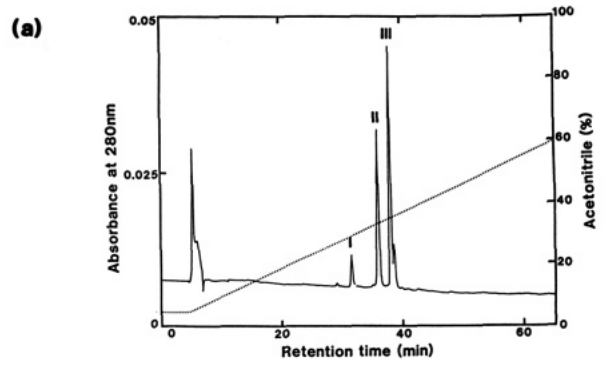
FIGURE 1: Analysis of proteins synthesized in vitro directed by poly(A)$^+$ RNA. Poly(A)$^+$ RNA was translated in vitro by using a rabbit reticulocytes lysate containing [^{35}S]methionine, and the immunoprecipitate formed with antibody against sarcotoxin IIA was analyzed by SDS–polyacrylamide gel electrophoresis followed by fluorography. (Lane 1) Total protein synthesized in the translation mixture (10^5 cpm); (lane 2) precipitate formed with normal serum (10^6 cpm); (lane 3) immunoprecipitate formed with antiserum (10^6 cpm); (lane 4) same as lane 3 but with 10 μg of cold sarcotoxin IIA added to the reaction mixture before addition of antiserum; (lane 5) sarcotoxin IIA electrophoresed and stained with Coomassie brilliant blue. Arrows indicate the positions of the following molecular mass markers: bovine serum albumin (69 kDa), ovalbumin (44 kDa), α-chymotrypsinogen (25 kDa), and egg-white lysozyme (14 kDa).

0.02% (w/v) Ficoll 400, 0.02% bovine serum albumin, 0.02% poly(vinylpyrrolidone) 40]/sonicated denatured salmon sperm DNA solution (50 μg/mL) for 18 h at 42 °C, and blots were washed successively by incubation in 2 × SSC for 15 min at room temperature and in 0.1 × SSC for 30 min at 42 °C (Thomas, 1980).

Results

In Vitro Translation of Fat Body Poly(A)$^+$ RNA. Since most hemolymph proteins are synthesized in the fat body (Takahashi et al., 1984; Komano et al., 1983), we checked whether poly(A)$^+$ RNA extracted from the fat body of injured larvae contains mRNA for sarcotoxin IIA. When the poly(A)$^+$ RNA was translated in vitro and the resulting translation products were precipitated with antibody against sarcotoxin IIA, a protein with a molecular mass of 26 kDa was specifically precipitated, as shown in Figure 1 (lane 3). This protein was not precipitated with normal serum (lane 2), and the intensity of the band decreased selectively when cold sarcotoxin IIA was added to the reaction mixture (lane 4). Therefore, we concluded that this band was due to sarcotoxin IIA and thus that fat body poly(A)$^+$ RNA contained mRNA for sarcotoxin IIA. The molecular mass of sarcotoxin IIA determined by SDS–polyacrylamide gel electrophoresis was 24 kDa (Ando et al., 1987). Thus, the 26-kDa protein in lane 3 may be a precursor of intact sarcotoxin IIA containing putative signal sequences.

Amino Acid Sequences of Tryptic Peptides and a Probe for cDNA Cloning. Since the above results showed that mRNA for sarcotoxin IIA is present in poly(A)$^+$ RNA extracted from the fat body of injured larvae, we next isolated its cDNA clone. For this, we first determined the amino acid sequences of tryptic fragments of sarcotoxin IIA. When intact sarcotoxin IIA was digested with trypsin and the resulting digestion products were analyzed by reverse-phase HPLC, four peaks with absorbance at 280 nm were detected, as shown in Figure 2a. Of these tryptic fragments, we partially sequenced those in peaks I, II, and III, and their sequences are shown in Figure



(b)

Peak I Asn-Asp-Asn-His-Asn-Leu-Asp-Ala-Ser-Val-Phe-

Peak II (a) Ala-Asn-Ala-Gly-Gly-Ser-Gln-Trp-Leu-Ser-Gly-Pro-Phe-Ala-Asn-Gln-

(b) Gly-Ile-Asp-Phe-Lys

Peak III Asp-Tyr-Ser-Phe-Gly-Leu-Gly-Leu-Ser-His-Asn-Ala-

5' CCIA$^{A}_{G}$ICC$^{A}_{G}$AAI$^{GAA}_{CTG}$TA$^{A}_{G}$TC 3'

FIGURE 2: Reverse-phase HPLC of tryptic peptides and their partial sequences. (a) Sarcotoxin IIA (75 μg) was digested with trypsin. The resulting peptides were applied to a reverse-phase column of Synchropak RP-P (C18) and analyzed as described under Materials and Methods. (—) Absorbance at 280 nm; (- - -) concentration of acetonitrile. (b) Partial or complete amino acid sequences of peptides in peaks I–III in (a).

2b. Peak II was found to contain two peptides. We then synthesized a probe DNA corresponding to Asp-Tyr-Ser-Phe-Gly-Leu-Gly in the peak III peptide. For this, we synthesized the following two series of mixtures of 16 eicosamers containing 3 deoxyinosines at the wobble positions: CCIA-(A/G)ICC(A/G)AAIGA(A/G)TA(A/G)TC and CCIA-(A/G)ICC(A/G)AAICT(A/G)TA(A/G)TC from the 5′- to the 3′-end, respectively (Ohtsuka et al., 1985). We used a mixture of these oligodeoxyribonucleotides as a probe. These oligodeoxyribonucleotides represent all possible complementary sequences corresponding to the above peptide except for the third nucleotide residues of the Gly codon of the carboxyl terminal.

Isolation and Characterization of a cDNA Clone. Using these synthetic oligodeoxyribonucleotides as probes, we isolated 20 hybridization-positive clones by screening about 30 000 transformants derived from a cDNA library for fat body poly(A)$^+$ RNA. Of these, three clones were found to contain inserts corresponding to full-length cDNA for sarcotoxin IIA mRNA. Since they gave identical restriction maps, the nucleotide sequence of only one of these clones, pTII20, was analyzed. This clone contained a cDNA insert of about 1.2 kilobase pairs (kb) including poly(dA)-poly(dT) and poly(dG)-poly(dC) tails. A restriction map and the strategy used to sequence the entire insert of pTII20 are shown in Figure 3, and the nucleotide sequence is shown in Figure 4 with the proposed amino acid sequence for sarcotoxin IIA. All four tryptic peptides shown in Figure 2b were included in this sequence, indicating that pTII20 is a cDNA clone for sarcotoxin IIA.

We could not determine the amino-terminal amino acid residue of sarcotoxin IIA because it seemed to be blocked in some way. Therefore, assuming that the amino-terminal amino acid residue might be pyroglutamate, we treated intact sarcotoxin IIA with calf liver pyroglutamyl aminopeptidase by the method of Podell and Abraham (1978) to hydrolyze the

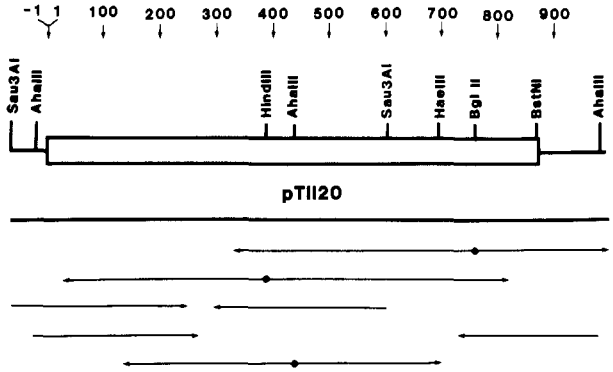


FIGURE 3: Strategy for sequencing cloned cDNA pTII20. The restriction map displays only relevant restriction endonuclease sites. Numbers indicate positions from the 5′-terminal nucleotide of the first Met codon. The poly(dA)-poly(dT) tract and poly(dG)-poly(dC) tails are not included in the restriction map. The protein coding region is indicated by an open box. The horizontal arrows show the directions and extents of sequencing.

pyrrolidone ring of pyroglutamate. After this treatment, partial sequencing of sarcotoxin IIA was possible, and the amino acid sequence obtained was Lys-Leu-Pro-Val-Pro-Ile-Pro-Pro-Pro-Thr-Asn-Pro-Pro-Val-Ala-Ala-Phe-His-Asn-Ser. Therefore, we concluded that the amino-terminal amino acid residue of sarcotoxin IIA is pyroglutamate. As the subsequent amino acid sequence was found to correspond to that at positions 26–45 in Figure 4, we assumed that the amino terminal of sarcotoxin IIA at position 25 was Gln. Upstream of this position, there are two candidates for an initiation codon giving putative signal peptides of 14 and 24 amino acid residues, respectively. At the moment it is uncertain from which of these initiation codons translation is initiated.

From the codon for amino-terminal Gln, there is an open reading frame for 270 amino acid residues before the first termination and poly(A) sequences start about 120 bases downstream from the termination codon. A consensus sequence for the poly(A) addition signal of AATAAA is located 19 bases upstream of the poly(A) addition site. The amino acid composition predicted from the nucleotide sequence was consistent with that obtained by analysis of intact sarcotoxin IIA, as shown in Table I. The content of basic amino acid residues was much higher than that of acidic amino acid residues, suggesting that sarcotoxin IIA is a basic protein. No Cys residue was found, indicating that the overall secondary structure of this protein is rather simple. It should also be noted that 7 of 10 Pro residues are located in the first 14 amino acid residues from the amino terminal. Especially noteworthy is the fact that there are tandem sequences of Pro-Pro-Pro and Pro-Pro, suggesting that this region has unique secondary structure.

A hydropathy plot for sarcotoxin IIA is shown in Figure 5. Except for a hydrophobic domain in the putative signal sequence, the protein seemed to be relatively hydrophilic, and the carboxyl-terminal half of the molecule tended to be rich in hydrophilic domains.

Identification of the mRNA and Gene for Sarcotoxin IIA. Using cloned pTII20 DNA as a probe, we identified sarcotoxin IIA mRNA. For this, total RNA was isolated from the fat body collected from larvae that had been injured with a hypodermic needle 6 h previously and was then denatured with formaldehyde and subjected to electrophoresis (Goldberg, 1980). After electrophoresis, the RNA was blotted onto a nitrocellulose filter (Thomas, 1980) and hybridized with a nick-translated *Hin*dIII–*Bst*NI fragment (480 bp) from the coding region of sarcotoxin IIA cDNA. As shown in Figure

Table I: Amino Acid Composition of Sarcotoxin IIA[a]

amino acid	predicted from nucleotide sequence		direct analysis[c]
	b	*c*	
Asp	14	13.7	13.1
Asn	23		
Thr	15	5.7	4.5
Ser	29	10.7	10.0
Glu	4	6.7	6.2
Gln	14		
Pro	10	3.7	3.1
Gly	35	13.0	14.3
Ala	24	8.9	10.1
Cys	0		
Val	18	6.7	6.4
Met	1	0.4	0.4
Ile	8	3.0	3.4
Leu	24	8.9	9.1
Tyr	3	1.1	1.2
Phe	12	4.4	4.6
His	7	2.6	2.7
Lys	14	5.2	5.5
Arg	13	4.8	4.8
Trp	2	0.7	0.7

[a] Predicted numbers of amino acids were calculated by assuming that sarcotoxin IIA is formed by removal of an N-terminal signal peptide of 24 amino acid residues. [b] Number of residues. [c] Mole percent.

6a, lane 2, a single band corresponding to about 1.2-kb RNA was detected with fat body RNA from injured larvae but not with that from normal larvae (lane 1). This length fitted that of the insert of pTII20, assuming that the mRNA contains a poly(A) tail of about 100 bases. No mRNA for sarcotoxin IIA was detected in the fat body RNA of normal larvae, indicating that the sarcotoxin IIA gene is usually dormant and that its transcription is activated by an injury to the body wall. Since sarcotoxin IIA is an antibacterial protein, we examined whether injection of bacteria into the abdominal cavity of larvae enhanced the transcription of sarcotoxin IIA genes. RNA was extracted from the whole body of larvae at various times after injection of a light suspension of *E. coli* and analyzed by Northern blot hybridization. As is evident from Figure 6b,c, transcription of the sarcotoxin IIA gene was activated by injection of either saline or *E. coli* suspension. However, a clear difference was observed in the persistence of mRNA: The apparent life of mRNA synthesized in response to *E. coli* was much longer than that synthesized in response to saline alone, and a high level of mRNA was maintained for more than 24 h after injection of *E. coli*. These results suggest that this gene is turned on in an emergency, such as on intrusion of a foreign substance through the damaged body wall, and that sarcotoxin IIA synthesized de novo is secreted into the hemolymph to kill invading bacteria. Probably, the presence of bacteria in the hemolymph keeps the gene turned on in some way, resulting in a longer life of the mRNA.

As reported before, sarcotoxin IIA is one of three sarcotoxin II proteins. The two other proteins, sarcotoxin IIB and IIC, give almost identical tryptic peptide maps (Ando et al., 1987). Therefore, we performed Southern blottin g analysis to obtain some information about the structures of the genes for sarcotoxin II, assuming that the coding sequences of pTII20 (*Hin*dIII–*Bst*NI fragment) also hybridize to the genes for sarcotoxin IIB and IIC. As shown in Figure 7, three to six hybridization-positive bands were detected, depending upon the restriction enzyme used, suggesting that multiple genes hybridize with this probe. Probably the sarcotoxin IIA gene belongs to a multigene family consisting of genes for sarcotoxin

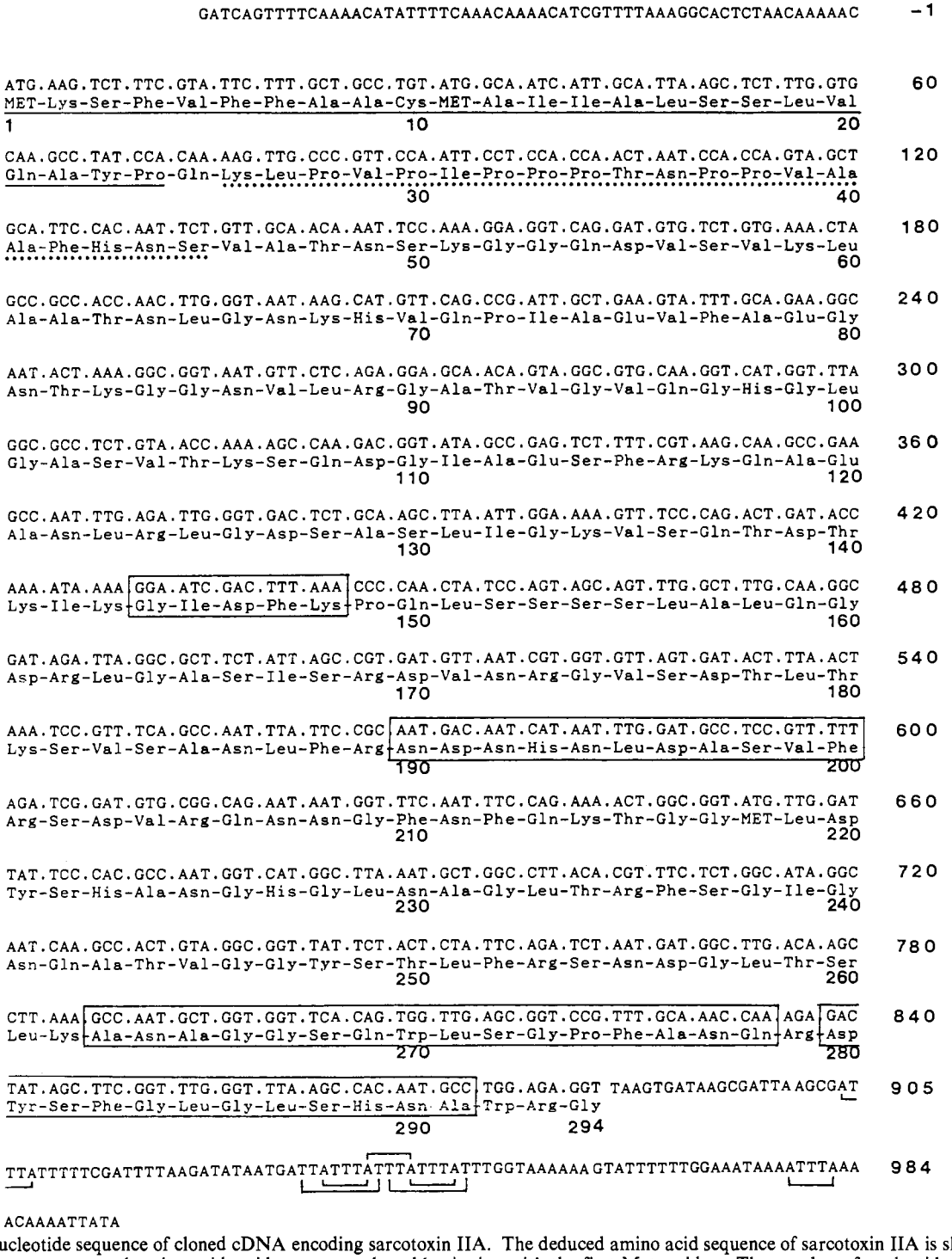


FIGURE 4: Nucleotide sequence of cloned cDNA encoding sarcotoxin IIA. The deduced amino acid sequence of sarcotoxin IIA is shown below the nucleotide sequence, and amino acid residues are numbered beginning with the first Met residue. The number of nucleotides is given at the right of each line. Tryptic fragments corresponding to peaks I, II, and III in Figure 2b are boxed. Amino acid sequences obtained after treatment with pyroglutamyl aminopeptidase are underlined by a dotted line and putative signal sequences by a solid line. AT-rich sequences in the 3′-untranslated region are shown in brackets (see Discussion).

IIB and IIC and other related genes.

Comparison of Sarcotoxin IIA and Attacins. Two cDNA clones for attacins of *Hyalophora cecropia* have been isolated and sequenced (Kockum et al., 1984). The molecular masses of sarcotoxin II and attacins are about the same, so it is possible that these two groups of proteins originated from the same ancestral gene. As shown in Figure 8, about 30% homology was found in the carboxyl-terminal halves of sarcotoxin IIA and basic attacin, but no significant homology was found in their amino-terminal halves. Probably, the carboxyl-terminal halves are important for the antibacterial actions of these proteins, and these portions have evolved from a common ancestral gene that existed before branching of *Diptera* and *Lepidoptera*.

Discussion

This paper describes the isolation and characterization of cDNA for sarcotoxin IIA, an antibacterial protein induced in the hemolymph of third instar larvae of *Sarcophaga peregrina* in response to injury of the body wall. Analysis of pTII20 showed that sarcotoxin IIA is a protein consisting of 270 amino acid residues with no Cys residue. Its molecular mass calculated from its amino acid sequence was about 28 200 Da, whereas that estimated from the mobility of purified

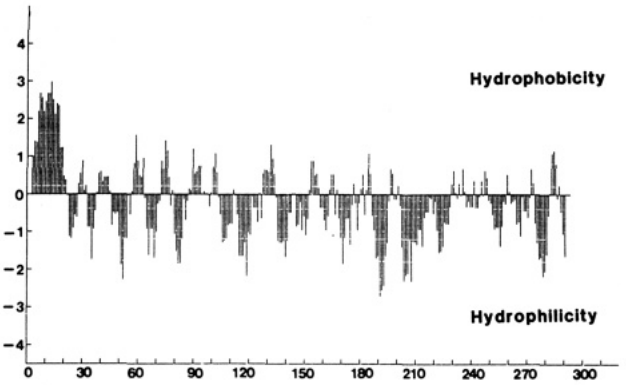


FIGURE 5: Hydropathy analysis of sarcotoxin IIA. The distribution of hydrophobic and hydrophilic domains of sarcotoxin IIA was analyzed by the method of Kyte and Doolittle (1982). Numbers of amino acid residues are shown at the bottom. Data presented as hydrophobic and hydrophilic portions are plotted above and below the vertical line.

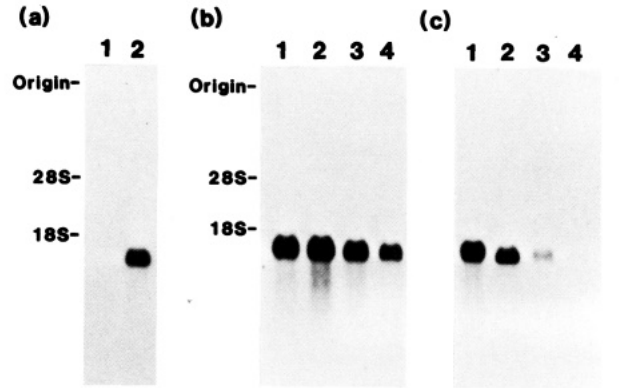


FIGURE 6: Expression of sarcotoxin IIA gene. Northern hybridization of sarcotoxin IIA cDNA. Total RNA (10 μg) isolated from the fat body of normal third instar larvae (lane 1) and those injured by pricking their body wall 6 h previously were subjected to electrophoresis on 1.2% agarose gel by the method of Goldberg (1980), blotted onto a nitrocellulose filter, and hybridization with the nick-translated *Hin*dIII–*Bst*NI (480 bp) fragment from the coding region of sarcotoxin IIA cDNA (a). RNA extracted at various times after injection of 5 μL of saline containing 2×10^6 *E. coli* (b) and saline alone (c). In (b) and (c), lanes 1–4 show RNA extracted 3, 6, 12, and 24 h after injections, respectively. The mouse ribosomal RNA was subjected to electrophoresis at the same time as a reference marker.

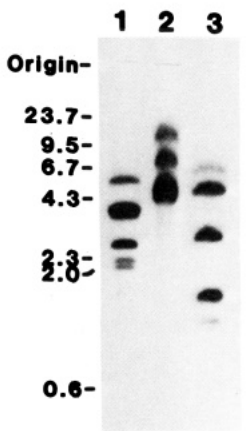


FIGURE 7: Southern hybridization of *Sarcophaga* genomic DNA. *Sarcophaga peregrina* DNA (10 μg) was digested with *Eco*RI (lane 1), *Pst*I (lane 2), or *Hin*dIII (lane 3). The digested samples were subjected to electrophoresis on 1% agarose gel, blotted onto a nitrocellulose filter, and hybridized with a nick-translated *Hin*dIII–*Bst*NI fragment of sarcotoxin IIA cDNA. Numbers indicate mobilities of *Hin*dIII-digested λ DNA fragments in kilobases.

sarcotoxin IIA on SDS–polyacrylamide gel electrophoresis was 24 000. Probably, intact sarcotoxin IIA migrates faster than

```
                 160         170         180         190         200
Sarcotoxin IIA   LALQG-DRLGASISRDVNRGVSDTLTKSVSANLFRNDNHNLDASVFRS-
                 ::      ::          :  :  :     ::: ::::   :  :
Basic attacin    LAYDNVNGHGATLTKTHIPGFGDKMTAAGKVNLFHNDNHDFSAKAFATK
                       60          70          80          90

                      210         220         230         240
                 DVRQNNGF-NFQKTGGMLDYSHANGHGLNAGLTRFSGIGNQA-TVGGYS
                       ::   :   ::      :  :        :        ::
                 NMPNIPQVPNFNTVGAGVDYMFKDKIGASANAAHTDFINRNDYSLGGKL
                       110         120         130         140

                 250       260       270
                 TLFRSNDGLTSLKANAGGSQWLSGPF
                  ::      :::  :::       ::
                 NLFKTPT--TSLDFNAG-WKKFDTPF
                 150         160         170
```

FIGURE 8: Comparison of the amino acid sequences of sarcotoxin IIA and basic attacin. Homology between sarcotoxin IIA and attacins (Kockum et al., 1984) was searched by computer, and the most homologous regions are aligned. Numbers indicate amino acid residues from the first Met shown in Figure 4.

usual proteins of the same size on SDS–polyacrylamide gel electrophoresis for some reason, and thus the molecular mass determined from its mobility may be an underestimate.

The sequence of the last two amino acid residues in the carboxyl terminal of sarcotoxin IIA was Arg-Gly, as deduced from the nucleotide sequence. This sequence was also found at the carboxyl terminal of sarcotoxin IA, another antibacterial protein, by analysis of cDNA for this protein (Matsumoto et al., 1986). However, the carboxyl-terminal amino acid residue of authentic sarcotoxin IA is amidated Arg, not Gly. Carboxyl-terminal Gly may be removed enzymatically, with formation of amidated Arg. Therefore, the carboxyl-terminal amino acid residue of sarcotoxin IIA may also be amidated Arg, like that of sarcotoxin IA. If so, the amino- and carboxyl-terminal amino acid residues of sarcotoxin IIA may both be modified, since the amino-terminal residue is pyroglutamate. It is known that the candidates for the last amino acid residue of the signal sequence are mainly Ala, Gly, Ser, Cys, Thr, and Gln in eukaryotic signal sequence (van Heijne, 1983). However, it was Pro in sarcotoxin IIA. It is possible that signal peptidase cleaves between Ala (+22) and Tyr (+23) first, producing sarcotoxin IIA containing extra dipeptide Tyr-Pro at its amino terminal, and then this dipeptide is removed to produce intact sarcotoxin IIA, as was suggested with cecropin B (van Hofsten, 1985).

Recent studies showed that the specific sequence TTATTTAT is conserved in the 3′-untranslated region of many cDNAs encoding mediator proteins related to the inflammatory response in mammals (Caput et al., 1986). Tandem repeats of this sequence were also found in the 3′-untranslated region of sarcotoxin IIA cDNA, as shown in Figure 4. If this sequence is specific for various inflammatory mediators of invertebrates as well as for those of vertebrates, sarcotoxin IIA may play a role in inflammatory responses, although it was originally found as an antibacterial protein. This sequence was not present in the mRNA for sarcotoxin IA. On the other hand, from studies on the effect of introducing AT-rich sequences of GM-CSF granulocyte macrophage colony stimulating factor (GM-CSF) cDNA into the rabbit β-globin gene, Show and Kamen (1986) suggested that AU-rich sequences in the 3′-untranslated region of mRNA of GM-CSF are the recognition signal for the specific degradation of this mRNA. If this is a general phenomenon, sarcotoxin IIA mRNA may be an unstable mRNA, since there are five AUUUA sequences in its 3′-untranslated region.

In *Sarcophaga peregrina* larvae, at least three groups of antibacterial proteins and one galactose-binding lectin appear in the hemolymph when the body wall is injured (Okada & Natori, 1983; Komano et al., 1980). The cDNAs for two of these proteins, sarcotoxin IA and *Sarcophaga* lectin, were cloned and characterized previously (Matsumoto et al., 1986;

Takahashi et al., 1985), and here we reported the cloning of sarcotoxin IIA cDNA. Since these genes are activated in the fat body in response to body injury, they may have a common regulatory sequence that responds to the stimulus of body injury. It would be interesting to find this common sequence and possibly also a common regulatory protein interacting with this sequence. We are now isolating genomic clones for these proteins. Analyses of the 5′-flanking sequences of these clones may give a clue to this regulatory sequence.

ACKNOWLEDGMENTS

We thank Dr. K. Miyoshi, Wakunaga Pharmaceutical Co. Ltd., for synthesis of oligodeoxyribonucleotides.

Registry No. DNA (*Sarcophaga peregrina* sarcotoxin IIA messenger RNA complementary), 112595-72-9; sarcotoxin IIA, 105430-42-0; sarcotoxin IIA (*Sarcophaga peregrina* precursor), 112595-74-1; sarcotoxin IIA (*Sarcophaga peregrina*), 112595-73-0.

REFERENCES

Ando, K., Okada, M., & Natori, S. (1987) *Biochemistry 26*, 226–230.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Boman, H. G., Nilsson-Faye, I., Paul, K., & Rasmuson, T., Jr. (1974) *Infect. Immun. 10*, 136–145.

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., & Cerami, A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1670–1674.

Engstrom, A., Engstrom, P., Tao, Z.-J., Carlsson, A., & Bennich, H. (1984) *EMBO J. 3*, 2065–2070.

Gagen, S. J., & Ratcliffe, N. A. (1976) *J. Invertebr. Pathol. 28*, 17–24.

Goldberg, D. A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5794–5798.

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63–67.

Hirose, T., Crea, R., & Itakura, K. (1978) *Tetrahedron Lett. 28*, 2449–2452.

Hultmark, D., Engstrom, A., Bennich, H., Kapur, R., & Boman, H. G. (1982) *Eur. J. Biochem. 127*, 207–217.

Hultmark, D., Engstrom, A., Andersson, K., Steiner, H., Bennich, H., & Boman, H. G. (1983) *EMBO J. 2*, 571–576.

Hunkapiller, M. W., Hewick, R. M., Dryer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399–413.

Kakidani, H., Furutani, Y., Takahashi, H., Noda, M., Morimoto, Y., Hirose, T., Asai, M., Nakanishi, S., & Numa, S. (1982) *Nature (London) 298*, 245–249.

Kessler, S. W. (1975) *J. Immunol. 115*, 1617–1624.

Kockum, K., Faye, I., van Hofsten, P., Lee, J.-Y., Xanthopoulos, K. G., & Boman, H. G. (1984) *EMBO J. 3*, 2071–2075.

Komano, H., Mizuno, D., & Natori, S. (1980) *J. Biol. Chem. 255*, 2919–2924.

Komano, H., Nozawa, R., Mizuno, D., & Natori, S. (1983) *J. Biol. Chem. 258*, 2143–2147.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Lee, J.-Y., Edlund, T., Ny, T., Faye, I., & Boman, H. G. (1983) *EMBO J. 2*, 577–581.

Matsumoto, N., Okada, M., Takahashi, H., Qu, X.-M., Nakajima, Y., Nakanishi, Y., Komano, H., & Natori, S. (1986) *Biochem. J. 239*, 717–722.

Messing, J. (1983) *Methods Enzymol. 101*, 20–39.

Morrison, D. A. (1979) *Methods Enzymol. 68*, 326–331.

Natori, S. (1977) *J. Insect Physiol. 23*, 1169–1171.

Natori, S. (1987) in *Molecular Entomology* (Law, J. H., Ed.) pp 369–378, Alan R. Liss, New York.

Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., & Matsubara, K. (1985) *J. Biol. Chem. 260*, 2605–2608.

Okada, M., & Natori, S. (1983) *Biochem. J. 211*, 727–734.

Okada, M., & Natori, S. (1985) *J. Biol. Chem. 260*, 7174–7177.

Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161–170.

Pelham, H. R. B., & Jackson, R. J. (1976) *Eur. J. Biochem. 67*, 247–256.

Podell, D. N., & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176–185.

Pye, A. E., & Boman, H. G. (1977) *Infect. Immun. 17*, 408–414.

Qu, X.-M., Steiner, H., Engstrom, A., Bennich, H., & Boman, H. G. (1982) *Eur. J. Biochem. 127*, 219–224.

Sanger, F., Nickien, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schmit, A. R., & Ratcliffe, N. A. (1977) *J. Insect Physiol. 23*, 175–184.

Scott, M. T. (1971) *Immunology 21*, 817–828.

Show, G., & Kamen, R. (1986) *Cell (Cambridge, Mass.) 46*, 659–667.

Steiner, H., Hultmark, D., Engstrom, A., Bennich, H., & Boman, H. G. (1981) *Nature (London) 292*, 246–248.

Takahashi, H., Komano, H., Kawaguchi, N., Kitamura, N., Nakanishi, S., & Natori, S. (1985) *J. Biol. Chem. 260*, 12228–12233.

Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.

van Heijne, G. (1983) *Eur. J. Biochem. 133*, 17–21.

van Hofsten, P., Faye, I., Kockum, K., Lee, J.-Y., Xanthopoulos, K. G., Boman, I. A., Boman, H. G., Engstrom, A., Andreu, D., & Merrifield, R. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2240–2243.

Weinstock, R., Sweet, R., Weiss, M., Cedar, H., & Axel, R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1299–1303.