Volume 258, number 2, 199-202 FEB 07885 December 1989

Cloning of gene cluster for sarcotoxin I, antibacterial proteins of *Sarcophaga peregrina*

Akio Kanai and Shunji Natori

Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Received 29 September 1989, revised version received 14 October 1989

A genomic clone of sarcotoxin I was isolated This clone contained four genes of structurally related proteins belonging to the sarcotoxin I family present in tandem array One of these genes was sequenced and found to be the sarcotoxin IB gene This gene contained a single intron of 95 bases

Gene structure, Multi-gene family, Sarcotoxin I, Antibacterial protein, (*Sarcophaga peregrina*)

1. INTRODUCTION

In *Sarcophaga peregrina* (flesh fly), various antibacterial proteins are induced in the hemolymph when the body wall of third-instar larvae is injured with a hypodermic needle [1-3]. Sarcotoxin I is a mixture of at least three cecropin-type antibacterial proteins termed sarcotoxin IA, IB and IC, which have almost identical primary structures [4]. We purified these proteins to homogeneity, determined their amino acid sequences, and isolated a cDNA for sarcotoxin IA [4,5]. As these proteins are synthesized simultaneously in the fat body in response to body injury, the genes for these proteins in the fat body of injured larvea must be activated in the same way.

Northern blotting analysis indicated that the sarcotoxin IA gene is also transiently expressed in the early embryonic stage and the early pupal stage of undamaged insects [6]. This expression pattern is exactly the same as that of *Sarcophaga* lectin, another defense protein synthesized by the fat body in response to body injury [7,8]. Thus, these two genes are normally activated twice during the life cycle of *Sarcophaga* as well as in response to body injury. We are interested in the mchanism of activation of a given gene under two different developmental and physiological conditions. To study this problem, it is necessary to clone the genes for these proteins and study their expression in vitro.

This paper describes the cloning of the sarcotoxin I gene. The genes for proteins of the sarcotoxin I family were found to exist as a cluster.

Correspondence address. S Natori, Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

2. MATERIALS AND METHODS

2 1 *Construction of genomic library of* Sarcophaga

A library was constructed essentially as described by Kaiser and Murray [9] Briefly, DNA isolated from adult flies was partially digested with *Sau*3AI, and the resulting DNA fragments were fractionated by a 10-40% sucrose density gradient centrifugation. Fragments with molecular lengths of 15-20 kb were collected, and ligated to *Bam*HI-cleaved λ EMBL3 arms (Amersham) The ligated DNA was packed into λ phage particles in vitro with a packaging extract (Stratagene) The infectious particles obtained (about 2×10^6 plaque forming units/μg *Sarcophaga* DNA) were used to screen the sarcotoxin I genes

2 2 *Screening of the sarcotoxin IA gene*

Escherichia coli Q358 was infected with the recombinant phages, and a total of 3×10^5 plaques were screened with pTO19, a cDNA clone for the sarcotoxin IA gene [5] Hybridization was carried out at 42°C for 12-16 h The filters were washed successively with 0 1 × SSC (1 × SSC = 0 15 M NaCl, 0.015 M sodium citrate) and 0 1% SDS solution at 42°C Then the blot was exposed to a Kodak XAR-5 film at −70°C

2 3 *DNA blot hybridization*

About 5 μg of DNA was digested with various restriction enzymes The digests were electrophoresed in 0.8% agarose gel and then transferred to a nitrocellulose filter as described by Thomas [10] The hybridization mixture consisted of 50% (v/v) formamide, 5 × SSC, 1 × Denhaldt's solution, 20 mM sodium phosphate (pH 6 5), and 100 μg/ml of denatured and sonicated salmon sperm DNA Hybridization was carried out with a nick-translated *Mbo*II-*Taq*I fragment of pTO19 at 42°C for 12-18 h The filter was washed for successive 15 min periods with 2 × SSC containing 0 1% SDS at room temperature and at 42°C, and then autoradiographed at −70°C

2 4 *DNA sequencing*

To sequence DNA, we first prepared various deletion derivatives of the DNA fragment using exonuclease III and mung bean nuclease, or site-specific restriction enzymes. Then each deletion derivative was inserted into a pUC118 vector Single strand DNA was obtained using a helper phage M13K07, and sequenced by the dideoxy chain termination method using Sequenase (United State Biochemical Corp) [11] Both strands of the DNA were sequenced, and the final sequence was inferred from overlaps between consecutive deletion derivatives

Published by Elsevier Science Publishers B V (Biomedical Division)
00145793/89/$3.50

3. RESULTS AND DISCUSSION

By screening a genomic library of *Sarcophaga*, we isolated one clone that hybridized with cDNA for sarcotoxin IA (pTO19) [5]. This colone λ IV32, contained a 16 kb insert. This insert was split into two with *Sal*I, and one of the resulting fragments was found to hybridize with pTO19. We subcloned this 11 kb fragment into a pUC118 vector, and named the resulting plasmid pTOGI11. From analysis of pTOGI11, this 11 kb fragment was concluded to contain four tandem repeats of a 2.4 kb *Bgl*II-*Eco*RI fragment, as illustrated in fig.1. This conclusion was based on the following lines of evidence: (1) When pTOGI11 was digested with *Hin*dIII, *Eco*RI, *Bgl*II, *Mlu*I and *Pst*I, a dense band of a 2.4 kb fragment that hybridized with pTO19 was always detected; and (2) As this pTOGI11 contained only one *Eco*RV site, restriction sites for *Hin*dIII, *Eco*RI, *Bgl*II and *Mlu*I could be mapped from the position of the *Eco*RV site. Since we could not map the *Pst*I sites completely, they are not included in fig.1. From these results, we speculated that pTOGI11 contains four genes for proteins belonging to the sarcotoxin I family, and that these genes are present in tandem array.

An example of DNA blotting hybridization obtained by digesting λ IV32 DNA with the three restriction enzymes is shown in fig.2. In all three digests, a 2.4 kb fragment was detected. All the other hybridization-positive bands except that of a 2.8 kb fragment in lane 3 are those of partially digested products or joint molecules of *Sarcophaga* DNA and vector DNA. The 2.8 kb fragment in lane 3 was found to be that of a DNA fragment containing unit 1 and the *Sal*I site in *Sarcophaga* DNA shown in fig.1.

Assuming that each *Bgl*II-*Eco*RI unit contains one gene for a sarcotoxin I protein, this gene structure is unique. Recently, four cecropin B genes were isolated [12]. It is not known whether they form a gene cluster, but judging from the present results, it is likely that the genes for these antibacterial proteins are present as a gene family, and that their expression is regulated in the same way.

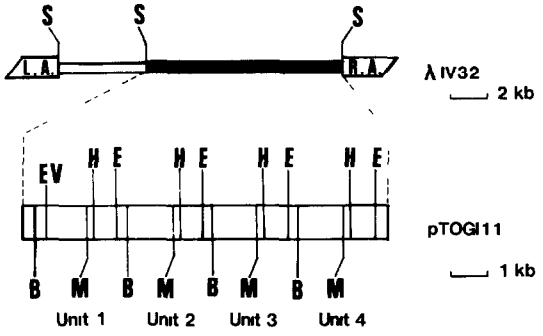

Fig 1 Multi-gene structure of the sarcotoxin I genomic clone An 11 kb *Sal*I-*Sal*I fragment of λ IV32 was subcloned (pTOGI11), and its precise restriction map was determined This clone contained four repeated units with almost identical restriction sites Restriction enzymes used S, *Sal*I; B, *Bgl*II; EV, *Eco*RV, M, *Mlu*I, H, *Hin*dIII, E, *Eco*RI L A and R A indicate the left arm and right arm of the vector used

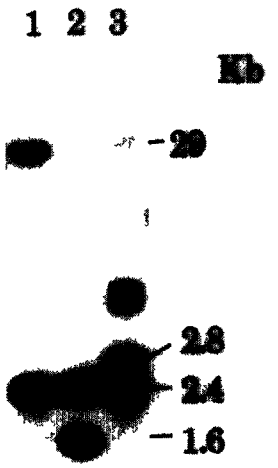

Fig 2 DNA blot hybridization of λ IV32 Recombinant phage DNA from λ IV32 was digested with each of *Hin*dIII, *Pst*I and *Eco*RI, and electrophoresed The digests were blotted onto a nitrocellulose filter and probed with sarcotoxin IA cDNA (pTO19) Figures indicate molecular lengths in kilobases Lane 1, *Hin*dIII digest, lane 2, *Pst*I digest, lane 3, *Eco*RI digest

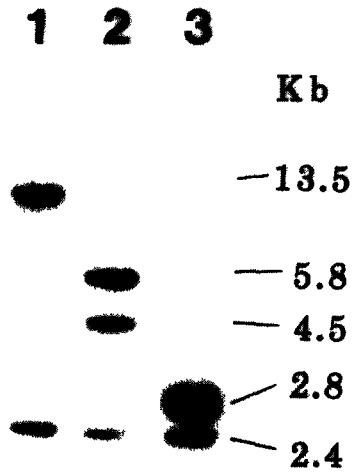

Fig 3 Genomic DNA blot hybridization of *Sarcophaga* DNA *Sarcophaga* DNA was digested with each of *Pst*I, *Hin*dIII and *Eco*RI, and electrophoresed The digests were blotted onto a nitrocellulose filter and probed with sarcotoxin IA cDNA (pTO19) Figures indicate molecular lengths in kilobases Lane 1, *Pst*I digest, lane 2, *Hin*dIII digest; lane 3, *Eco*RI digest

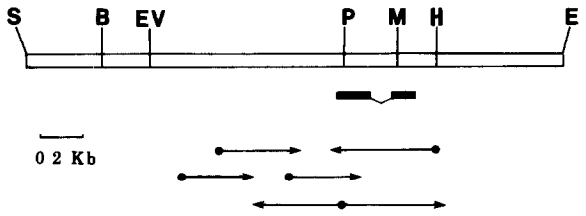

Fig 4 Diagram of the 2 6 kb *Sal*I-*Eco*RI fragment and its sequencing strategy The direction and extent of sequencing are shown with arrows The closed box shows the location of the sarcotoxin IB gene with an intron Restriction enzymes: S, *Sal*I, B, *Bgl*II, EV, *Eco*RV, P, *Pst*I; M, *Mlu*I, H, *Hin*dIII, E, *Eco*RI

An alternative, but unlikely possibility, was that λ IV32 is a cloning artifact. To examine this possibility, we analyzed *Sarcophaga* DNA by DNA blot hybridization to see if 2.4 kb fragments are produced. *Sarcophaga* DNA was digested with *Pst*I, *Hin*dIII and *Eco*RI, respectively, and probed with pTO19. As shown in fig.3, a 2.4 kb fragment was always detected irrespective of the restriction enzyme used. Particularly noteworthy is the fact that the digest with *Eco*RI gave bands of 2.4 kb and 2.8 kb fragments like the digestion of λ IV32 with *Eco*RI (cf. fig.2, lane 3). These results indicate that λ IV32 is not a cloning artifact, and that the sarcotoxin I gene in *Sarcophaga* DNA actually has a repeating structure. Then does each *Bgl*II-*Eco*RI fragment contain one gene for one protein of sarcotoxin I? To answer this question, we tried to sequence unit 1 shown in fig.1. For this purpose, we subcloned the *Sal*I-*Eco*RI 2.6 kb fragment, and its sequencing strategy was shown in fig.4. As pTO19 hybridized only with the *Eco*RV-*Hin*dIII fragment, we sequenced this fragment. As shown in fig.5, this fragment contained one gene for a protein of the sarcotoxin I family, and from its deduced sequence this protein was identified as sarcotoxin IB [4].

The structures of the sarcotoxin IB gene and the cecropin B gene are quite similar [12]. The sarcotoxin IB gene contains one intron as in the case of the cecropin B gene, but this intron in the sarcotoxin IB gene is much shorter, being only 95 bases. The intron of the sarcotoxin IB gene is present between Ile-10 and Glu-11, whereas that of the cecropin B gene is between Ile-8 and Glu-9. This Ile-Gle sequence is conserved among sarcotoxin I and cecropins A and B [4].

The cap site for the sarcotoxin IB gene was assigned as A in the insect-specific consensus sequence for the

```
   1 AAATACTAGCAAATTTTAATAGTTTTTTAAGTTGTTGACTATAAAATATCTTTAGAAAAA   60
  61 AAATGTATAAGAGAAGGTTTTGAACTTTTATTTTAACTGTTTGTTTCATCAAATAAAATG  120
 121 AATATCCAATGCAAATTTGATAGGATAATTAGAAATGATAAATAAATAAGTCAAGAAAAA  180
 181 AGTTTATTAATAACAAAATGTATTACAATATGACTACAATTTAAAATATTCGATGGAAAA  240
 241 TAGTTCAAATGAAATTACGATTTCATTTCATAATTTTTAAAATCATACTCACTTTCTTAT  300
 301 CACTTGAAATACTATGTTAAAAGCCACAATTTCTGTTAAATATAAACCCCCCTTTAAATC  360
 361 ACTTAAATTGTATGCGATCTATTTAAACACTCAGGGTATGTATGTATATACATACAATGT  420
 421 ATGTTGAAGATTCATTGAAAGCAAGTCCTATATAAACTAAGAACAATGCCAAGAATCAGC  480
 481 ATCATTTGCTTTTGAGCTGCAGTGTGAAAACTTAAAATTTAAATAATAAAAATTAATAAA  540
     . Cap    . Pst I .
 541 AAAAAATTAAAAAAACCATTATGAATTTCAATAAAGTATTCATTTTTGTTGCATTGATTT  600
                          MetAsnPheAsnLysValPheIlePheValAlaLeuIleLeu
 601 TGGCTGTTTTTGCTGGACAAAGTCAAGCTGGTTGGTTGAAAAAAATTGGCAAAAAAATTG  660
       AlaValPheAlaGlyGlnSerGlnAlaGlyTrpLeuLysLysIleGlyLysLysIle
                                  ▲ Exon 1
 661 TAAGTATTGTTATCAAGAGAAAACATTTTTATATTTTATATTAAAATGTATTAAAAATAT  720

 721 TAATTATTATTTAATTATTTTTAATTATTAAAAGGAACGCGTTGGCCAACATACACGTGA  780
                                    GluArgValGlyGlnHisThrArgAsp
 781 TGCCACCATTCAGGTTATTGGTGTTGCCCAACAGGCTGCAAATGTAGCTGCCACAGCCAG  840
       AlaThrIleGlnValIleGlyValAlaGlnGlnAlaAlaAsnValAlaAlaThrAlaArg
                                  . Exon 2 .
 841 AGGTTAAAAATTCTTCCTTTCTTTAGAATTAAATATTTTTAAGTGCCATATTAATATTAA  900
       Gly***
 901 GGATATAGTTTAAGATGAATAAATTATTTGAAAATTAAAAAAAGAATTTTAAGCTTAACA  960
                                                    Hind III
```

Fig 5 Nucleotide sequence of the sarcotoxin IB gene Coding regions are underlined and the corresponding amino acid sequences are shown A possible CAAT-box, TATA-box, Cap site and poly(A) addition signal are also underlined The arrowhead shows the cleavage site of a putative signal peptide

cap site ATCATTT [13], by comparison with the sequence of the sarcotoxin IA cDNA [5]. A TATA-box, a CAAT-box and a poly(A) addition signal were present in reasonable locations, indicating that the sarcotoxin IB gene is an independent compact gene. Probably, units 2, 3 and 4 contain similar genes corresponding to other proteins of the sarcotoxin I family.

Acknowledgements This study was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, and in part by a Grant-in-Aid (Bio Media Programs) from the Ministery of Agriculture, Forestry and Fisheries of Japan (89-IV-2-7)

REFERENCES

[1] Okada, M and Natori, S (1983) Biochem J. 211, 727–734
[2] Ando, K , Okada, M and Natori, S (1987) Biochemistry 26, 226–230
[3] Baba, K , Okada, M , Kawano, T , Komano, H and Natori, S (1987) J Biochem. 102, 69–74
[4] Okada, M. and Natori, S (1985) J. Biol. Chem 260, 7174–7177
[5] Matsumoto, N., Okada, M , Takahashi, H., Qu, X -M , Nakajima, Y., Nakanishi, Y , Komano, H and Natori, S (1986) Biochem. J 239, 717–722
[6] Nanbu, R , Nakajima, Y. and Natori, S (1988) Biochem Biophys Res Commun 150, 540–544.
[7] Komano, H., Mizuno, D and Natori, S (1980) J Biol Chem. 255, 2919–2924.
[8] Takahashi, H , Komano, H and Natori, S (1986) J Insect Physiol. 32, 771–779
[9] Kaiser, K and Murray, N E (1985) in: DNA Cloning, A Practical Approach (Glover, D M ed) pp 1–47, IRL Press, Oxford
[10] Thomas, P S (1980) Proc. Natl. Acad Sci USA 77, 5201–5205
[11] Sanger, F , Nicklen, S and Coulson, A R (1977) Proc. Natl Acad. Sci USA 74, 5463–5467
[12] Xanthopolos, K G , Lee, J.-Y , Gan, R , Kockum, K., Faye, I and Boman, H G (1988) Eur J. Biochem 172, 371–376
[13] Hultmark, D , Klemenz, R and Gehring, W J (1986) Cell 44, 429–438