Gentry-Weeks, C. R., Cookson, B. T., Goldman, W. E., Rimler, R. B., Porter, S., & Curtiss, R., III *Infect. Immun. 56*, 1698–1707.
Glauner, B., & Schwarz, U. (1983) in *The Target of Penicillin. Proceedings of the International FEMS Symposium* (Hakenbeck, R., Holtje, J. V., & Labischinski, H., Eds.) pp 29–34, de Gruyter, Berlin.
Goldman, W. E., Klapper, D. G., & Baseman, J. B. (1982) *Infect. Immun. 36*, 782–794.
Goodell, E. W., & Schwarz, U. (1985) *J. Bacteriol. 162*, 391–397.
Hebeler, B. H., & Young, F. E. (1976) *J. Bacteriol. 126*, 1180–1185.
Holtje, J. V., Mirelman, D., Sharon, N., & Schwarz, U. (1975) *J. Bacteriol. 124*, 1067–1076.
Karnovsky, M. L. (1986) *Fed. Proc. 45*, 2556–2560.
Kitano, K., Tuomanen, E., & Tomasz, A. (1986) *J. Bacteriol. 167*, 759–765.
Kotani, S., Tsujimoto, M., Koga, T., Nagao, S., Tanaka, A., & Kawata, S. (1986) *Fed. Proc. 45*, 2534–2540.
Krueger, J. M., Pappenheimer, J. R., & Karnovsky, M. L. (1982) *J. Biol. Chem. 257*, 1664–1699.
Krueger, J. M., Karnovsky, M. L., Martin, S. A., Pappenheimer, J. R., Walter, J., & Biemann, K. (1984) *J. Biol. Chem. 259*, 12659–12662.
Krueger, J. M., Rosenthal, R. S., Martin, S. A., Walter, J., Davenne, D., Shoham, S., Kubillus, S. L., & Biemann, K. (1987) *Brain Res. 403*, 249–257.
Mallory, R. B., & Hornor, A. A. (1912) *J. Med. Res. 27*, 115–123.
Martin, S. A., Karnovsky, M. L., Krueger, J. M., Pappenheimer, J. R., & Biemann, K. (1984) *J. Biol. Chem. 259*, 12652–12658.
Martin, S. A., Rosenthal, R. S., & Biemann, K. (1987) *J. Biol. Chem. 262*, 7514–7522.
Melly, M. A., McGee, Z. A., & Rosenthal, R. S. (1984) *J. Infect. Dis. 149*, 378–386.
Mett, H., Keck, W., Funk, A., & Schwarz, U. (1980) *J. Bacteriol. 144*, 45–52.
Miller, D. L., Ross, E. M., Alderslade, R., Bellman, M. H., & Rawson, N. S. B. (1981) *Br. Med. J. 282*, 1595–1599.
Olson, L. C. (1975) *Medicine (Baltimore) 54*, 427–469.
Pappenheimer, J. R., Koski, G., Fencl, V., Karnovsky, M. L., & Krueger, J. (1975) *J. Neurophysiol. 38*, 1299–1311.
Pisabarro, A. G., De Pedro, M. A., & Vazquez, D. (1985) *J. Bacteriol. 161*, 238–242.
Rosenthal, R. S. (1979) *Infect. Immun. 24*, 869–878.
Rosenthal, R. S., Nogami, W., Cookson, B. T., Goldman, W. E., & Folkening, W. J. (1987) *Infect. Immun. 55*, 2117–2120.
Schleifer, K. H., & Kandler, O. (1972) *Bacteriol. Rev. 36*, 407–477.
Sinha, R. K., & Rosenthal, R. S. (1980) *Infect. Immun. 29*, 914–925.
Taylor, A., Das, B. C., & Heijenoort, J. V. (1975) *Eur. J. Biochem. 53*, 47–54.
Weiss, A. A., & Hewlett, E. L. (1986) *Annu. Rev. Microbiol. 40*, 661–686.
Westall, F. C., & Root-Bernstein, R. S. (1983) *Mol. Immunol. 20*, 169–177.
Westall, F. C., & Root-Bernstein, R. S. (1986) *Lancet*, 251–252.

Molecular Cloning of cDNA for Sarcocystatin A and Analysis of the Expression of the Sarcocystatin A Gene during Development of *Sarcophaga peregrina*[†,‡]

Hiromi Saito, Toshiharu Suzuki, Kohji Ueno, Takeo Kubo, and Shunji Natori*
Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan
Received September 7, 1988; Revised Manuscript Received October 14, 1988

ABSTRACT: Sarcocystatin A is a cysteine proteinase inhibitor purified from the hemolymph of *Sarcophaga peregrina* larvae [Suzuki, T., & Natori, S. (1985) *J. Biol. Chem. 260*, 5115–5120]. We isolated a cDNA clone for sarcocystatin A and analyzed the structure and expression of the sarcocystatin A gene. Sarcocystatin A consists of 102 amino acid residues. Significant homology was found between amino acid sequences of sarcocystatin A and other mammalian cystatins, and highly conserved sequences among mammalian cystatins were also found in sarcocystatin A. Using cloned cDNA as a probe, we investigated expression of the sarcocystatin A gene during the development of *Sarcophaga*. Results showed that this gene was transiently activated in the very early embryonic stage and in the pupal stage, suggesting that sarcocystatin A participates in morphogenesis of larval and adult structures of *Sarcophaga*.

Low molecular mass cysteine proteinase inhibitors, termed cystatins, have been found in various mammalian tissues and sera (Brzin et al., 1983, 1984; Green et al., 1984; Isemura et al., 1984; Kominami et al., 1982). The physiological roles of cystatins are not clear, but these inhibitors are expected to regulate the activity of cysteine proteinases in several intracellular and extracellular biological reactions. Chicken egg white was also shown to contain a cystatin (Barrett, 1981), suggesting that cystatin plays a role in embryogenesis during animal development.

Previously we reported the purification and characterization of a low molecular mass cysteine proteinase inhibitor named sarcocystatin A from the hemolymph of *Sarcophaga peregrina* (flesh fly) larvae (Suzuki & Natori, 1985). Sarcocystatin A was found to be a mixture of sarcocystatin A_α and A_β in a

[†]This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.
[‡]The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02847.

0006-2960/89/0428-1749$01.50/0

molar ratio of 2:1. These two proteins have essentially the same primary structure and the same antigenicity and are thought to be derived from the same gene. The molecular weight of sarcocystatin A_β (M_r = 9500) was slightly smaller than that of A_α (M_r = 10000), possibly due to posttranslational processing or modification of A_α.

The amount of sarcocystatin A increases with time from the third instar and continues to increase after pupation, reaching a maximum level on day 5 after puparium formation. Then it decreases rapidly, and its level is very low in newly emerged flies (Suzuki & Natori, 1986). Sarcocystatin A selectively inhibits the activity of cysteine proteinase in the hemocytes, which ingest most larval tissues during metamorphosis, and so it is thought to protect developing adult tissues in pupae from attack by cysteine proteinase of hemocytes.

This paper reports the molecular cloning of cDNA for sarcocystatin A. This cDNA is actually for sarcocystatin A_α, but for simplicity we refer to it as cDNA for sarcocystatin A except where a distinction is necessary. We found significant homology between the amino acid sequence of sarcocystatin A and those of mammalian cystatins, in particular human kininogen segments (Salvesen et al., 1986). But, unlike kininogen, protein encoded by this cDNA was a single sarcocystatin A molecule, not a precursor protein containing multiple molecules of sarcocystatin A.

Materials and Methods

Sarcocystatin A and Its Tryptic Fragments. Sarcocystatin A was purified from the hemolymph of third instar *Sarcophaga* larvae as described before (Suzuki & Natori, 1985). Sarcocystatin A_α and A_β were separated by HPLC. To obtain tryptic fragments of sarcocystatin A containing amino acid sequence suitable for the synthesis of DNA probes, purified sarcocystatin A_α was digested with tosylphenylalanine chloromethyl ketone treated trypsin in 1 M triethylamine/bicarbonate buffer (pH 8.0) for 12 h at 37 °C at an enzyme/protein ratio of 1:25. The resulting peptides were applied to a reverse-phase HPLC column of Synchropak RP-P (C_{18}) connected to a Gilson HPLC system and eluted with a linear gradient of 5–60% solution B [0.05% (v/v) trifluoroacetic acid in acetonitrile] in solution A [0.05% (v/v) trifluoroacetic acid in H_2O] at a flow rate of 1 mL/min. The absorbance at 215 nm was monitored. Fractions containing each peptide were lyophilized and subjected to automated sequence analysis in an Applied Biosystem 470A protein sequencer.

Cloning Procedure and Sequencing of cDNA. A cDNA library for fat body mRNA of *Sarcophaga* larvae was constructed by the method of Okayama and Berg (1982) as described before (Takahashi et al., 1985). Colonies of *Escherichia coli* HB101 carrying recombinant plasmid were transferred onto nitrocellulose filters. The filters were baked at 80 °C for 2 h and then immersed in 3 × SSC (1 × SSC = 0.15 M NaCl and 0.015 M sodium citrate) containing 0.1% SDS kept at 60 °C for 24 h. Prehybridization was done in 3 × SSC/10 × Denhardt's solution [1 × Denhardt's solution = 0.02% (w/v) each of Ficoll 400, bovine serum albumin, and poly(vinylpyrolidone)-40] containing 50 μg/mL denatured and sonicated salmon sperm DNA. After prehybridization, the filters were hybridized with 5′-end-labeled synthetic probes in 4 × SSC/10 × Denhardt's solution containing 25 μg/mL denatured and sonicated salmon sperm DNA for 20 h at 42 °C. Then the filters were washed with 4 × SSC containing 0.1% SDS at 42 °C and autoradiographed.

For nucleotide sequencing, cloned cDNA was digested with various restriction enzymes, and the resulting fragments were subcloned into phage vectors M13 mp 18 and 19 (Messing, 1983). Recombinant phages were amplified in *E. coli* JM105 as described by Hanahan (1983), and the fragments of DNA were sequenced by the dideoxy method of Sanger et al. (1980).

RNA and DNA Blot Hybridizations. RNA was extracted from *Sarcophaga* at various developmental stages as described before (Takahashi et al., 1986). About 10 μg of RNA was denatured and electrophoresed in 1.2% agarose gel containing formaldehyde by the method of Goldberg (1980). After electrophoresis, RNA was transferred from the gel to a nitrocellulose filter as described by Thomas (1980). The hybridization mixture consisted of 50% (v/v) formamide, 5 × SSC, 1 × Denhardt's solution, 50 mM sodium phosphate (pH 6.5), 0.1% SDS, and 200 μg/mL denatured and sonicated salmon sperm DNA. Hybridization was carried out with a nick-translated cDNA fragment at 42 °C for 12–18 h. The filter was washed successively with 2 × SSC containing 0.1% SDS for 15 min at room temperature, and with 0.1 × SSC containing 0.1% SDS for 15 min at room temperature and then for 15 min at 42 °C, and then autoradiographed at −80 °C.

For DNA blot hybridization, 5 μg of DNA isolated from adult flies was digested completely with *Eco*RI or *Hin*dIII, and subjected to 1% agarose gel electrophoresis. The blotting, hybridization, and washing procedures were essentially the same as those for RNA blot hybridization. The probe used was the *Bal*I–*Asu*II fragment (positions −50 to +430) of pSC1 (see Figure 2).

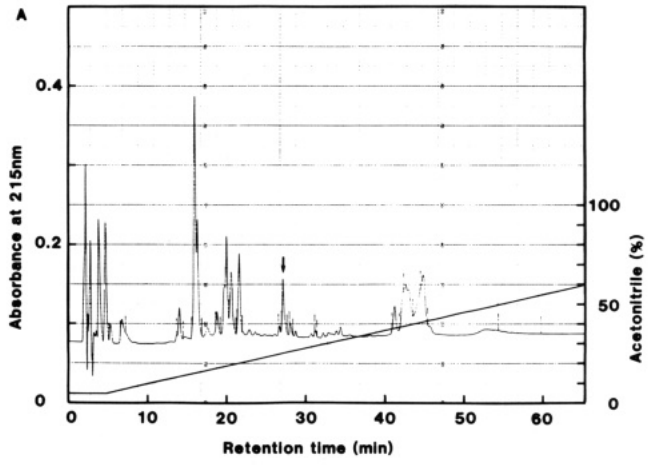
Other Methods. Oligodeoxyribonucleotides were synthesized by a modification of the triester method (Hirose et al., 1978), and were 5′-end labeled as described by Kakidani et al. (1982). Nick-translation with [α-^{32}P]dCTP was performed by the method of Weinstock et al. (1978). Pyroglutamate at the amino terminus of sarcocystatin A was hydrolyzed by the method of Podell and Abraham (1978).

Results

Partial Amino Acid Sequences of Sarcocystatin A_α and the Synthesis of Oligodeoxyribonucleotides. To isolate a cDNA clone for sarcocystatin A, we first determined the amino acid sequences of tryptic fragments of sarcocystatin A_α, since sarcocystatin A_β is thought to be a part or a modified form of sarcocystatin A_α (Suzuki & Natori, 1985, 1986). Sarcocystatin A_α was digested with trypsin, and the resulting digestion products were analyzed by reverse-phase HPLC. The pattern shown in Figure 1A was obtained reproducibly. Of these peptides, we determined the amino acid sequence of a peptide in the peak shown by an arrow. This peak contained a peptide consisting of 26 amino acid residues, the sequence of which is shown in Figure 1B. This peptide contained three Lys residues, one of which was the carboxyl terminus. Probably, trypsin cannot digest two tryptic sites inside this peptide completely because of steric hindrance.

We then synthesized oligodeoxyribonucleotides corresponding to Lys-Asp-Val-Ile-Asn-Ala-Asp (20-mer, mixture of 24) as shown in Figure 1C. These oligodeoxyribonucleotides represent all possible complementary sequences corresponding to this heptapeptide except for the third nucleotide residue of carboxyl-terminal Asp. In these oligodeoxyribonucleotides, the third nucleotide residues of the Val and Ala codons were replaced by deoxyinosinic acid (Ohtsuka et al., 1985).

Isolation and Characterization of a cDNA Clone for Sarcocystatin A. Using these oligodeoxyribonucleotides as probes, we screened about 30000 transformants derived from a cDNA library for fat body poly(A)+ RNA of *Sarcophaga* larvae and isolated eight hybridization-positive clones. Of these, five clones seemed to contain inserts of cDNA corresponding to



B

Ile-Asn-Ser-Ala-Thr-Thr-Gln-Val-Val-Ser-Gly-Ser-Lys

-Asp-Val-Ile-Asn-Ala-Asp-Leu-Lys-Asp-Glu-Asn-Asp-Lys

C

-Lys-Asp-Val-Ile-Asn-Ala-Asp-

3'— TT$^{T}_{C}$CT$^{A}_{G}$CAITAG$^{A}_{T}$TT$^{A}_{G}$CG ICT — 5'

FIGURE 1: Reverse-phase HPLC of tryptic peptide, the sequence of a tryptic fragment, and the synthetic oligodeoxyribonucleotides used as probes. (A) Sarcocystatin A_α (5 nmol) was digested with trypsin in 0.2 M ammonium carbonate solution (pH 8.0) for 12 h at 37 °C. The resulting peptides were applied to a HPLC column of Synchropak RP-P (C_{18}) and analyzed. (B) A peptide in the peak shown by an arrow in (A) was sequenced. (C) Mixture of 24 possible oligodeoxyribonucleotides used as probes which correspond to Lys-Asp-Val-Ile-Asn-Ala-Asp in the sequenced peptide.

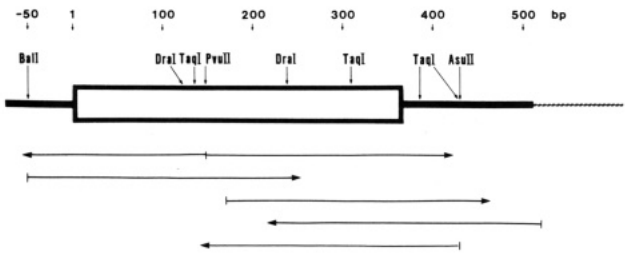


FIGURE 2: Restriction map and sequencing strategy of cloned cDNA pSC1. The restriction map displays only relevant restriction endonuclease sites. Numbers indicate positions from the 5′-terminal nucleotide of the first Met codon. The poly(dG)·poly(dC) tail is not indicated in the restriction map. The protein-encoding region is indicated by an open box. The horizontal arrows show the direction and extent of sequencing.

full-length sarcocystatin A mRNA and gave identical restriction maps. The inserts of the other three clones were much shorter. We determined the nucleotide sequence of one of these full-length clones, termed pSC1. This clone contained a cDNA insert of about 0.7 kb including the poly(dA)·poly(dT) and poly(dG)·poly(dC) tails. A restriction map and the strategy used to sequence the entire insert of pSC1 are shown in Figure 2, and its nucleotide sequence is shown in Figure 3 with the proposed amino acid sequence for sarcocystatin A.

As is evident from Figure 3, the amino acid sequence corresponding to the tryptic fragment shown in Figure 1B was present from position 61 to 86, indicating that this is a cDNA for sarcocystatin A. From the first Met codon, there is an open reading frame for 122 amino acid residues before the first termination codon. A consensus sequence for a poly(A) addition signal, AATAAA, is located about 150 bases downstream from this termination codon, and a poly(A) sequence

```
                                                    5'-----GGGGGGA -61

AAAATACTGGCCAGTCGTTATTTGAATCTTCATTGATTAACAATATCTTCATCGGACATT  -1

ATGAAATACGTTTTGATTTTGTGCGTTATCACTTTAGCTACTGTGGCTTATGCCCAGCCA  60
METLysTyrValLeuIleLeuCysValIleThrLeuAlaThrValAlaTyrAlaGlnPro
1                   10                  20

CAGTGTGTCGGTTGTCCCAGTGAAGTAAAGGGCGACAAACTTAAGCAATCGGAAGAAACT 120
GlnCysValGlyCysProSerGluValLysGlyAspLysLeuLysGlnSerGluGluThr
                    30                  40

TTAAACAAGTCACTATCGAAGTTAGCAGCTGGTGATGGACCCACTTACAAACTGGTTAAA 180
LeuAsnLysSerLeuSerLysLeuAlaAlaGlyAspGlyProThrTyrLysLeuValLys
                    50                  60

ATCAACTCAGCCACCACACAAGTTGTTTCTGGCAGTAAGGACGTAATAAATGCTGATTTA 240
IleAsnSerAlaThrThrGlnValValSerGlySerLysAspValIleAsnAlaAspLeu
                                        80

AAAGATGAAAACGATAAGACTAAGACCTGCGACATAACCATTTGGTCTCAACCTTGGTTG 300
LysAspGluAsnAspLysThrLysThrCysAspIleThrIleTrpSerGlnProTrpLeu
                    90                  100

GAAAATGGCATCGAAGTTACCTTCAATTGTCCCGGTGAACCGAAAGTGGTTAAGAAGCAC 360
GluAsnGlyIleGluValThrPheAsnCysProGlyGluProLysValValLysLysHis
                    110                 120

AGTGCCTAAAATCAATATAAAACAGTCGATTGCATTGGCTGACTGTTAATGCTCACTGAT 420
SerAla***

ATTCTACGTTCGAATACATAATTTTATAAATCAAATTGAATATTTTGTTTGTAATTTTTT 480

ATTTGCATAAAAAAGTGAAATAAAAAAAGAAAAACAAAAAAAAAAAA-----3'
```

FIGURE 3: Nucleotide sequence of cloned cDNA encoding sarcocystatin A. The deduced amino acid sequence of sarcocystatin A is shown below the nucleotide sequence, and amino acid residues are numbered beginning with the first Met residue. Numbers of nucleotides are given at the right of each line. The amino acid sequence corresponding to the tryptic fragment shown in Figure 1B is boxed. The putative signal sequence is underlined.

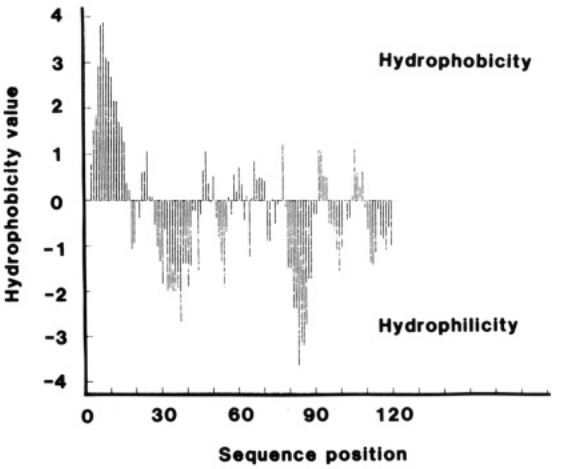


FIGURE 4: Hydropathy analysis of sarcocystatin A. The distribution of hydrophobic and hydrophilic domains of sarcocystatin A was analyzed by the method of Kyte and Doolittle (1982). Numbers of amino acid residues are shown at the bottom. Data presented as hydrophobic and hydrophilic portions are plotted above and below the vertical line, respectively.

starts from 12 bases downstream of the poly(A) addition signal.

A hydropathy profile of sarcocystatin A is shown in Figure 4. About 20 amino acid residues from the first Met were shown to be hydrophobic, but the residual region was relatively hydrophilic. Therefore, this hydrophobic region is likely to be a putative signal sequence. To confirm this possibility, we tried to determine the amino-terminal amino acid residue of sarcocystatin A. However, we were unsuccessful because this residue seemed to be blocked in some way. From the amino acid sequence and hydropathy profile, the amino-terminal amino acid residue seemed likely to be Gln at position 19 or 21 that had been converted to pyroglutamate. Therefore, we treated sarcocystatin A with calf liver pyroglutamyl aminopeptidase by the method of Podell and Abraham (1978) to hydrolyze the pyrrolidone ring of pyroglutamate. After this

FIGURE 5: Comparison of the amino acid sequence of sarcocystatin A with those of other proteins of the cystatin superfamily. Identical amino acid residues to those in sarcocystatin A are shaded. Numbering is from the amino terminus of sarcocystatin A. Gaps were introduced to obtain maximal sequence homology. (a) Human kininogen segment (residues 128–249) (Ohkubo et al., 1984); (b) human kininogen segment (residues 250–373) (Ohkubo et al., 1984); (c) chicken egg white cystatin (Shwabe et al., 1984); (d) bovine colostrum cystatin (Hirado et al., 1985); (e) human cystatin S (Isemura et al., 1984); (f) human cystatin C (Grubb & Lofberg, 1982); (g) human cystatin A (Machleidt et al., 1983); (h) rat cystatin α (Takio et al., 1984); (i) human cystatin B (Ritonja et al., 1985); (j) rat cystatin β (Takio et al., 1983).

treatment, sequencing of five amino acid residues of sarcocystatin A was possible, and the sequence obtained coincided with that from position 22 to 27. Therefore, the amino-terminal residue of sarcocystatin A was concluded to be pyroglutamate formed by modification of Gln at position 21.

Starting from this Gln, sarcocystatin A was found to consist of 102 amino acid residues, and its molecular weight was calculated to be 11 000, which was consistent with that obtained by electrophoretic analysis of sarcocystatin A_α. The amino acid composition predicted from the nucleotide sequence was consistent with that obtained by analysis of intact sarcocystatin A_α, as shown in Table I. This protein did not contain Met or Arg, but its Lys content was very high, 14 of 102 amino acid residues being Lys.

Comparison of the Amino Acid Sequence of Sarcocystatin A with Those of Other Cystatins. Sarcocystatin A is the first cystatin of invertebrate origin that had been purified. Comparison of its primary structure with those of other cystatins showed that a sequence of 58 amino acid residues coincided with the sequences of some of the 10 cystatins whose primary structures are known, including human kininogen segments, as seen in Figure 5. In particular, Gly at position 4 and Gln-Val-Val-(Ser)-Gly from position 47 to 51 are highly conserved sequences in various cystatins (Barrett et al., 1986; Barrett, 1987), and these sequences were also found in sarcocystatin A. It is interesting that sarcocystatin A has higher homology with family 2 cystatins and human kininogen segments than with family 1 cystatins (Barrett et al., 1986). The sequence of Cys-(Val)-Gly-Cys-(Pro) from position 2 to 6 is also found in human kininogen segments, and the sequence in the carboxyl-terminal region from position 70 to 90 shows significant homology with the carboxyl-terminal regions of family 2 cystatins and one of the human kininogen segments (Figure 5b–f). Therefore, it is conceivable that sarcocystatin A and these vertebrate cystatins have evolved from a common ancestral gene.

Table I: Amino Acid Composition of Sarcocystatin A

amino acid	predicted from nucleotide sequence: no. of residues	predicted from nucleotide sequence: molar %	molar % from direct analysis
Asn	6	5.9	12.7 (Asn + Asp)
Asp	7	6.9	
Thr	8	7.8	6.9
Ser	9	8.8	8.1
Gln	4	3.9	11.6 (Gln + Glu)
Glu	7	6.9	
Pro	5	4.9	5.2
Gly	7	6.9	6.9
Ala	5	4.9	5.9
Val	9	8.8	8.1
Met	0	0.0	0.0
Ile	5	4.9	4.6
Leu	7	6.9	6.9
Tyr	1	1.0	1.2
Phe	1	1.0	1.2
Lys	14	13.7	13.9
His	1	1.0	1.2
Arg	0	0.0	0.0
Trp	2	2.0	2.3
Cys	4	3.9	3.5

Expression of the Sarcocystatin A Gene during Development of Sarcophaga. In a previous paper, we discussed the possibility that this proteinase inhibitor might protect developing adult tissues in pupae from attack by a proteinase present in hemocytes (Suzuki & Natori, 1986). To obtain more insight into the function of sarcocystatin A during development of *Sarcophaga*, we investigated the expression of the sarcocystatin A gene at various developmental stages of *Sarcophaga*, by RNA blot hybridization. Since sarcocystatin A is present in

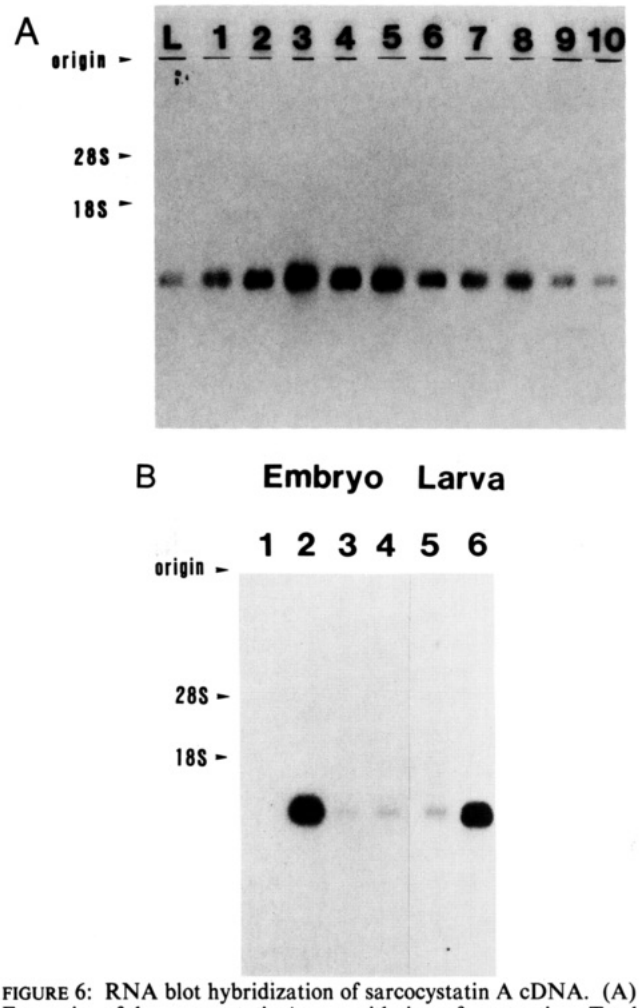


FIGURE 6: RNA blot hybridization of sarcocystatin A cDNA. (A) Expression of the sarcocystatin A gene with time after pupation. Total RNA was extracted from third instar larvae, pupae at various developmental stages, and newly emerged flies, and 10 μg of each RNA was analyzed by Northern blot hybridization by using the nick-translated *Bal*I–*Asu*II fragment of pSC1 as a probe. The mouse ribosomal RNA was subjected to electrophoresis as a reference marker. Lane L, third instar larvae; lanes 1–9, pupae harvested on days 1–9 after puparium formation; lane 10, newly emerged flies. (B) Expression of the sarcocystatin gene in embryos and larvae. Experimental conditions were essentially the same as for (A). Lane 1, eggs collected from 8-day-old virgin flies; lane 2, embryos collected from 8-day-old flies; lane 3, 10-day-old flies; lane 4, 12-day-old flies; lane 5, second instar larvae; lane 6, third instar larvae.

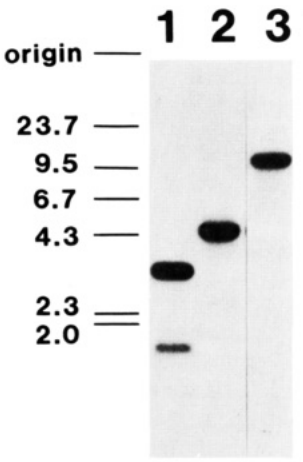


FIGURE 7: Southern blot hybridization analysis of *Sarcophaga* genomic DNA. *Sarcophaga* DNA (5 μg) was digested completely with *Eco*RI (lane 2) and *Hin*dIII (lane 3). The cloned cDNA pSC1 (170 pg) was cleaved with *Hin*dIII (lane 1). The digested DNA was subjected to electrophoresis, blotted onto a nitrocellulose filter, and hybridized with the nick-translated *Bal*I–*Asu*II fragment of pSC1 containing the coding region of sarcocystatin A cDNA. Numbers on the left indicate mobilities of *Hin*dIII-digested λ DNA fragments in kilobase pairs.

the hemolymph of third instar larvae, RNA extracted from these larvae should contain sarcocystatin A mRNA. In fact, a single band corresponding to 0.8–0.9-kilobase RNA corresponding to sarcocystatin A mRNA was detected by probing total RNA from third instar larvae with the *Bal*I–*Asu*II fragment of pSC1, as shown in Figure 6A, lane L. The amount of sarcocystatin A mRNA increased markedly with time after puparium formation, reaching a maximum on days 3–5 and then rapidly decreasing (lanes 1–9), but a significant level of mRNA was maintained throughout the pupal stage and mRNA was also detected in newly emerged flies (lane 10). These results indicate that the sarcocystatin A gene is activated significantly more after puparium formation than in the larval stage.

Previously, we demonstrated by radioimmunoassay that the amount of sarcocystatin A increased after puparium formation (Suzuki & Natori, 1986). At that time it was not certain whether de novo synthesis of sarcocystatin A was in fact enhanced or whether the antigenicity of sarcocystatin A increased by change in its quaternary structure (conversion of trimeric structures to free forms) at this stage (Suzuki & Natori, 1985). But these previous results together with our present results clearly show that de novo synthesis of sarcocystatin A increases in the pupal stage.

We also found that the sarcocystatin A gene was expressed transiently at a very early stage of embryogenesis. Since this insect is ovoviviparous, embryonic development usually starts in the uterus of 8-day-old female flies, and the first instar larvae are laid 4 days later. We collected embryos from the uterus of female flies on days 8, 10, and 12, respectively, after emergence. Their embryos are mature oocytes just after fertilization on day 8, whereas they are at a stage just before emergence on day 12 (Kurahashi & Ohtaki, 1979). We extracted RNA from these embryos and examined whether this RNA contained sarcocystatin A mRNA. As is evident from Figure 6B, a significant amount of sarcocystatin A mRNA was detected in embryos from 8-day-old flies (lane 2), the amount being comparable to that in third instar larvae. Levels of the mRNA at other developmental stages were very much lower (lanes 3 and 4), suggesting that the sarcocystatin A gene is activated soon after fertilization. This mRNA might have been maternal mRNA. To exclude this possibility, we separated male and female flies immediately after emergence, collected eggs from 8-day-old virgin flies, and compared their content of sarcocystatin A mRNA with that in embryos collected from 8-day-old flies. As shown in Figure 6B, lane 1, no mRNA of sarcocystatin A was detected in eggs from virgin flies, indicating that the sarcocystatin A gene is expressed after fertilization. Some mRNA was detected in second instar larvae, but much less than in third instar larvae (lanes 5 and 6).

Sarcocystatin A was originally found in the hemolymph of third instar larvae. However, the present results suggest that this proteinase inhibitor plays a role in the morphogenesis of both larval and adult structures.

Identification of the Gene for Sarcocystatin A. We used Southern blot hybridization analysis to obtain information about the organization of the sarcocystatin A gene. When 5 μg of restriction fragments of total *Sarcophaga* DNA was hybridized to radiolabeled pSC1 insert, single bands were detected in DNA fragments obtained by digestions with *Eco*RI

and *Hin*dIII, respectively, as shown in Figure 7. Moreover, the intensities of these bands were almost the same as that of the band of *Hin*dIII-digested pSC1 (170 pg) which is roughly equivalent to a single copy per haploid genome. From these results, we conclude that the copy number of the sarcocystatin A gene was probably one per haploid genome.

Discussion

This paper describes the isolation of cDNA for sarcocystatin A, a cystatin purified from the hemolymph of *Sarcophaga* larvae. Cystatins have been purified from various sources, and their primary structures are known (Barrett et al., 1986; Barrett, 1987). But only a few cystatin cDNA clones have been isolated, including human cystatin C (Abrahamson et al., 1987) and rice cystatin (Abe et al., 1987). pSC1 is the first cDNA clone of cystatin of invertebrate origin obtained. Analysis of pSC1 showed that sarcocystatin A is a protein consisting of 102 amino acid residues, with pyroglutamate as its amino terminus. A putative signal sequence of 20 amino acid residues is located upstream of the amino terminus. Since pSC1 contains full-length sarcocystatin A, this cDNA clone probably encodes sarcocystatin A_α, and sarcocystatin A_β may be derived from sarcocystatin A_α posttranslationally by some unknown molecular mechanism.

Two major points require discussion. One is the organization of the gene of sarcocystatin A. Significant homology was found between the sequence of sarcocystatin A and those of cystatins of vertebrate origin (Barrett et al., 1986; Barrett, 1987). In particular, about 30% homology was found between sarcocystatin A and human kininogen segments. Since human kininogen contains three possible cystatin sequences, cystatins may be produced by posttranslational processing (Salvesen et al., 1986). However, since pSC1 encodes only one molecule of sarcocystatin A, there is no precursor protein such as human kininogen. If sarcocystatin A and human kininogen segments, and possibly other cystatins, are derived from a common ancestral gene, this gene must have been present before the branching of Protostomia and Deuterostomia. In insects like *Sarcophaga*, this gene may have remained as a single gene, whereas in vertebrates like man it may have evolved into the kininogen gene by repeated duplication and fusion. Therefore, cystatin may be an important proteinase inhibitor that is present in most animals.

The other point that requires discussion is the unique expression of the sarcocystatin A gene. We found that this gene is activated transiently in both the early embryonic stage and the pupal stage, although it is continuously active during the second and third larval instars. These times of its expression provide a hint about the biological role of sarcocystatin A. Larval and adult structures are formed in the embryonic and pupal stages. In the pupal stage, hemocytes degrade and ingest unnecessary larval tissues, and in these reactions, proteinase secreted from hemocytes probably plays a role in decomposing larval tissues (Bodenstein, 1950; Richards & Davies, 1977). Thus, one possible function of sarcocystatin A in the pupal stage may be to protect developing adult tissues such as imaginal discs from attack by hemocyte proteinase in situ. Sarcocystatin A may have a similar role in embryos, since unnecessary cells such as certain nerve cells, which undergo programmed death, are probably digested by proteinase during embryogenesis (Truman, 1987).

It would be interesting to investigate the distribution of cystatin genes phylogenetically in the animal kingdom. The sarcocystatin A cDNA may be useful for such a study.

Registry No. pSC1, 118334-25-1; pSC1 deduced precursor amino acid sequence, 118334-30-8; pSC1 deduced amino acid sequence, 118334-31-9; sarcocystatin A_α, 96881-03-7.

References

Abe, K., Emori, Y., Kondo, H., Suzuki, K., & Arai, S. (1987) *J. Biol. Chem. 262*, 16793–16797.

Abrahamson, A., Grubb, A., Olafsson, I., & Lundwall, A. (1987) *FEBS Lett. 216*, 229–233.

Barrett, A. J. (1981) *Methods Enzymol. 80*, 771–778.

Barrett, A. J. (1987) *Trends Biochem. Sci. 12*, 193–196.

Barrett, A. J., Rawlings, N. D., Davies, M. E., Machleidt, W., Salvesen, G., & Turk, V. (1986) in *Proteinase Inhibitors* (Barrett, A. J., & Salvesen, G., Eds.) pp 515–569, Elsevier Science Publishers, Amsterdam.

Bodenstein, D. (1950) in *Biology of Drosophila* (Demerce, M., Ed.) pp 275–367, Hafner, New York.

Brzin, J., Kopitar, M., Turk, V., & Machleidt, W. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 1475–1480.

Brzin, J., Popovic, T., Turk, V., Borchart, U., & Machleidt, W. (1984) *Biochem. Biophys. Res. Commun. 118*, 103–109.

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

Green, G. D. J., Kembhavi, A. A., Davies, M. E., & Barrett, A. J. (1984) *Biochem. J. 218*, 939–946.

Grubb, A., & Lofberg, H. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3024–3027.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–580.

Hirado, M., Tsunasawa, S., Sakiyama, F., Niinobe, M., & Fujii, S. (1985) *FEBS Lett. 186*, 41–45.

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

Isemura, S., Saitoh, E., Ito, S., Isemura, M., & Sanada, K. (1984) *J. Biochem. 96*, 1311–1314.

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

Kominami, E., Wakamatsu, N., & Katsunuma, N. (1982) *J. Biol. Chem. 257*, 14648–14652.

Kurahashi, H., & Ohtaki, T. (1979) *Jap. J. Med. Sci. Biol. 32*, 77–82.

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

Machleidt, W., Borchart, U., Fritz, H., Brzin, J., Ritonja, A., & Turk, V. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 1481–1486.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., & Sasaki, M. (1984) *Biochemistry 23*, 5691–5697.

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

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

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

Richards, O. W., & Davies, R. G. (1977) in *IMMS' General Textbook of Entomology* (Richards, O. W., & Davies, R. G., Eds.) 10th ed., Vol. 1, p 240, John Wiley and Sons, New York.

Ritonja, A., Machleidt, W., & Barrett, A. J. (1985) *Biochem. Biophys. Res. Commun. 131*, 1187–1192.

Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A., & Barrett, A. J. (1986) *Biochem. J. 234*, 429–434.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., & Roe, B. (1980) *J. Mol. Biol. 143*, 161–178.

Shwabe, C., Anastasi, A., Crow, H., McDonald, J. K., & Barrett, A. J. (1984) *Biochem. J. 217*, 813–817.

Suzuki, T., & Natori, S. (1985) *J. Biol. Chem. 260*, 5115–5120.

Suzuki, T., & Natori, S. (1986) *Insect Biochem. 16*, 589–595.

Takahashi, H., Komano, H., Kawaguchi, N., Kitamura, N., Nakanishi, S., & Natori, S. (1985) *J. Biol. Chem. 260*, 12228–12233.
Takahashi, H., Komano, H., & Natori, S. (1986) *J. Insect Physiol. 32*, 771–779.
Takio, K., Kominami, E., Wakamatsu, N., Katsunuma, N., & Titani, K. (1983) *Biochem. Biophys. Res. Commun. 115*, 902–908.
Takio, K., Kominami, E., Bando, Y., Katsunuma, N., & Titani, K. (1984) *Biochem. Biophys. Res. Commun. 121*, 149–154.
Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5201–5205.
Truman, J. W. (1987) in *Current Topics in Developmental Biology* (Moscona, A. A., & Monroy, A., Eds.) Vol. 21, pp 99–116, Academic Press, New York.
Weinstock, R., Sweet, R., Weiss, M., Cedar, H., & Axel, R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1299–1303.

Complete Sequence of the Human Tissue Factor Gene, a Highly Regulated Cellular Receptor That Initiates the Coagulation Protease Cascade†,‡

Nigel Mackman, James H. Morrissey,* Bruce Fowler, and Thomas S. Edgington
Department of Immunology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037
Received September 6, 1988; Revised Manuscript Received October 12, 1988

ABSTRACT: Tissue factor (TF) is the high-affinity receptor for plasma factors VII and VIIa. TF plays a role in normal hemostasis by initiating the cell-surface assembly and propagation of the coagulation protease cascade. Outside the vasculature, TF expression is highly dependent upon cell type. TF can also be induced by inflammatory mediators to appear on monocytes and vascular endothelial cells as a component of cellular immune responses. As an initial step toward elucidating the regulatory regions involved in control of TF gene expression, we have established the organization of the 12.4 kbp human TF gene and its complete DNA sequence. There are six exons separated by five introns. Within intron 5, we have mapped the single nucleotide difference which leads to the previously described *Msp*I polymorphism; the same intron also contains an apparently polymorphic *Pst*I site. The TF gene also contains three full-length *Alu* repeats and one partial *Alu* repeat. A single major transcription start site was identified 26 bp downstream from a TATA consensus promoter element. The putative promoter and first exon are located within a 1.2 kbp region of very high G + C content which fits the criteria of an HTF island. A cluster of predicted binding sites for a number of known transcription factors was found to coincide with this putative promoter region. These factors included AP-1 and AP-2 which can mediate the effects of phorbol esters, agonists known to induce TF expression in monocytes and vascular endothelial cells.

Tissue factor (TF), a 47-kDa transmembrane glycoprotein, is the major cellular initiator of the coagulation protease cascade. TF functions in normal hemostasis as the high-affinity, cell-surface receptor and essential cofactor for factors VII and VIIa (Broze, 1982; Bach et al., 1986), thereby triggering the cell-surface assembly of a cascade of highly specific serine proteases. TF belongs to a unique class of highly regulated receptor proteins, including thrombomodulin (Esmon, 1987), that act as effector molecules to modulate the extracellular environment.

As expected, TF is not normally expressed at significant levels by cells within the vasculature. However, in a variety of inflammatory settings, TF can be induced to appear on cells of monocyte lineage, the only circulating cell type observed to express TF (Edwards et al., 1979; Levy et al., 1981; Gregory & Edgington, 1985). TF expression can also be induced in vascular endothelial cells by the inflammatory cytokines interleukin 1 (IL-1) and tumor necrosis factor (TNFα), as well as by other inducers such as endotoxin (Bevilacqua et al., 1984; Stern et al., 1985). Induction of TF in monocytes and vascular endothelial cells is thought to be an important component of cellular immune responses [reviewed by Ryan and Geczy (1987)] and is also implicated in the pathogenesis of disseminated intravascular coagulation and septic shock (Edwards & Rickles, 1984; Taylor et al., 1987). In the monocyte, the primary level of regulation of TF expression appears to be transcriptional; endotoxin stimulation of human monocytes results in transient, de novo transcription of the TF gene (Gregory et al., 1988) in a manner similar to the regulation of TNFα (Sariban et al., 1988).

In nonvascular cells, TF expression appears to be regulated by other mechanisms. In cell culture, for instance, TF activity ranges from undetectable levels in some cell types to relatively high levels in other cell types such as fibroblasts (Dvorak et al., 1983; Rodgers et al., 1984). Immunohistochemical localization of TF in tissues has confirmed the cell type specificity of TF expression in vivo (Drake et al., 1988).

The relatively complex pattern of TF expression in different cell types is likely to involve multiple tissue-specific and agonist-responsive cis-acting regulatory elements linked to the TF gene. Detailed analysis of these elements requires

† This is Publication No. 5394-IMM and was supported in part by NIH Grants HL16411 and CA41085. This work was done during the tenure of an Established Investigatorship (J.H.M.) of the American Heart Association and during the tenure of a research fellowship (N.M.) from the American Heart Association, California Affiliate, and with funds contributed by the San Diego County Chapter.
‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02846.
* To whom correspondence should be addressed.

0006-2960/89/0428-1755$01.50/0