Vol. 185, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 29, 1992 Pages 240-245

Primary Structure of the Human Elafin Precursor Preproelafin Deduced from the Nucleotide Sequence of Its Gene and the Presence of Unique Repetitive Sequences in the Prosegment

Toshihiko Saheki, Futoshi Ito, Hiromi Hagiwara, Yasuyoshi Saito, Jun Kuroki*, Shinro Tachibana*, and Shigehisa Hirose

Department of Biological Sciences, Tokyo Institute of Technology, Ookayama, Meguroku, Tokyo 152, Japan

**Tsukuba Research Laboratories, Eisai Co., Ltd., 5-1-3 Tokodai, Tsukuba 300-26, Japan*

Received April 7, 1992

SUMMARY: The human elafin gene was cloned and its entire nucleotide sequence was determined to deduce the amino acid sequence for the precursor of elafin, an elastase-specific inhibitor. The gene spans approximately 1.7 kb and is divided into 3 exons. The gene product preproelafin consists of 117 amino acids: the initiator Met, a putative 25-amino acid signal peptide, a pro-sequence of about 34 amino acids, and the C-terminal 57 amino acids for mature elafin. Possible covalent clotting of the prosegment and its physiological significance have been pointed out based on a remarkable sequence similarity between the pro-sequence and the guinea pig seminal clotting protein SVP-1.

Elafin is an elastase-specific inhibitor (1) which has recently been identified and found to bear significant amino acid sequence similarities to mucous proteinase inhibitor (HUSI-I) (2) and Na^+, K^+-ATPase inhibitor (SPAI) (3). These inhibitor molecules are relatively small, consisting of about 50-60 amino acids, and appear to be grouped together as a new protein superfamily that has an unusually high content of cysteine and proline and therefore is expected to have a compact structure tightly held by multiple disulfide bonds; HUSI-I has two such domains (4, 5). This cysteine-rich compact structure is termed "four-disulfide core" (6) or "WAP" motif (7).

While cloning the human SPAI gene using a porcine SPAI cDNA clone as a probe, we happened to isolate the human elafin gene. Here we report its complete nucleotide sequence and exon-intron organization and the amino acid sequence for the precursor preproelafin deduced from the exon sequences. The most interesting structural feature of the elafin precursor is the presence of multiple

0006-291X/92 $4.00

repetitive sequences in the prosegment. The repetitive sequences are rich in lysine and glutamine residues and very similar to the guinea pig seminal clotting protein SVP-1 (8, 9) which is covalently cross-linked by a transglutaminase through γ-glutamyl-ε-lysine isopeptide bonds. The similarity leads us to suggest that the precursor contains two distinct but related proteins: elafin and a clotting protein; both provide a powerful protection for extracellular matrices by inactivating destructive elastase and by forming extracellular-matrix-like meshwork, respectively.

MATERIALS and METHODS

Materials--A human genomic library, constructed in EMBL3 (*Bam*HI site) from *Sau*3A partial digests by S. Tomatsu and Y. Sakaki, was obtained from Japanese Cancer Research Resources Bank (JCRB)-Gene (Registration No.: LI 018). A full length cDNA clone (737 bp) encoding porcine SPAI was obtained by screening a porcine duodenum cDNA library (λgt10) using synthetic oligonucleotide probes designed according to the published amino acid sequence of SPAI (3); details including the nucleotide sequence will be published elsewhere. Restriction and modifying enzymes were purchased from Toyobo, Takara, Boehringer, and Pharmacia; pBluescript II was from Stratagene; Sequenase version 2.0 sequencing kit was from U.S. Biochemical Corporation; nitrocellulose filters were from Schleicher & Schuell; ^{32}P-labeled nucleotides were from Amersham. Oligonucleotides were synthesized with a Milligen/Biosearch Cyclone Plus DNA synthesizer.

Isolation and Characterization of Human Elafin Gene--The human genomic library was plated on *E. coli* NM538 and screened following transfer to nitrocellulose membranes using the 737-bp porcine SPAI cDNA probe. Hybridization was conducted at 42°C in 20% formamide containing 5 X SSPE, 5 X Denhardt's, 0.1% SDS, and 100 μg/ml herring sperm DNA. Filters were washed at 53°C in 1 X SSC containing 0.1% SDS. Twenty positive clones were identified by screening 5 X 10^5 plaques. Two clones (λhI-G1 and λhI-G2) with relatively strong hybridization signals were isolated by two cycles of replating at lower density and characterized with various restriction enzymes. Both clones had a Southern-hybridization-positive 4.4-kb *Eco*RI fragment (Fig. 1).

Sequencing--The 4.4-kb *Eco*RI fragment was subcloned into the pBluescript II plasmid vector and subjected to nested deletions employing a Pharmacia kit. Sequencing was performed by the double-stranded dideoxy chain termination technique (10).

RESULTS

A human genomic DNA library, obtained from the JCRB-Gene bank, was screened with ^{32}P-labeled porcine SPAI cDNA which encodes the complete preproSPAI sequence of 187 amino acids (to be published elsewhere) under moderate stringency conditions. Several positive clones were identified, plaque-purified, and characterized by restriction enzyme mapping. One typical clone, λhI-G1, was selected for further analysis and sequencing. Fig. 1 illustrates the restriction map of λhI-G1. The fragments carrying the coding exons were identified by Southern analysis, subcloned into the plasmid vector pBluescript II, and

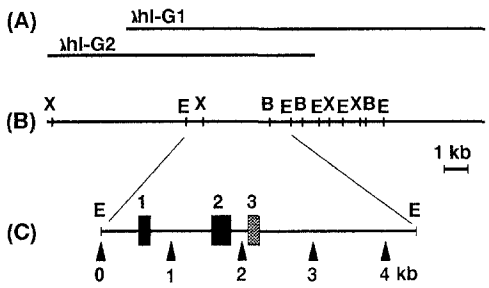

Fig. 1. **Clones analyzed (A), restriction map (B) and exon-intron organization (C) of the human elafin gene.** X, *Xba*I; E, *Eco*RI; and B, *Bam*HI. Exons are indicated by boxes: black boxes, coding exons; grey box, 3'-noncoding exon.

sequenced. The nucleotide sequence determined is shown in Fig. 2 together with the deduced amino acid sequence, which we first considered as the sequence for human SPAI but later turned out to be the human elafin precursor preproelafin since the COOH-terminal 57-amino-acid sequence coincided with the elafin sequence reported by Wiedow *et al.* (1); they have also isolated a partial cDNA clone (11) whose sequence is in complete agreement with the corresponding exon sequences in Fig. 2. Their cDNA clone, however, lacked a portion of the 5' region coding for the initiator methionine and the following hydrophobic signal sequence. As predicted from the exon sequences (Fig. 2), the human elafin precursor consists of 117 amino acids. The 60 amino acids preceding the mature sequence constitute a putative "pre" sequence of about 25 amino acids followed by a "pro" sequence of 34 amino acids.

The human elafin gene spanned approximately 1.7 kb and was comprised of 3 exons (Fig. 1). The first exon contained the 5'-noncoding region and 79 bases of the coding sequence to which the signal peptide sequence is confined. The "pro" and mature sequences were encoded by the second exon. The 3'-untranslated region was encoded entirely by exon 3. All splice site junctions conformed to the GT/AG rule (12) (Fig. 2).

DISCUSSION

Proteinase inhibitors that specifically inhibit the enzymes capable of degrading macromolecular components of the extracellular matrix has attracted considerable attention because of their roles in the control of extracellular matrix remodeling that is essential for morphogenesis, angiogenesis, tissue repair, and tumor metastasis. Neutrophil elastase, a serine proteinase contained in the azurophil granules of

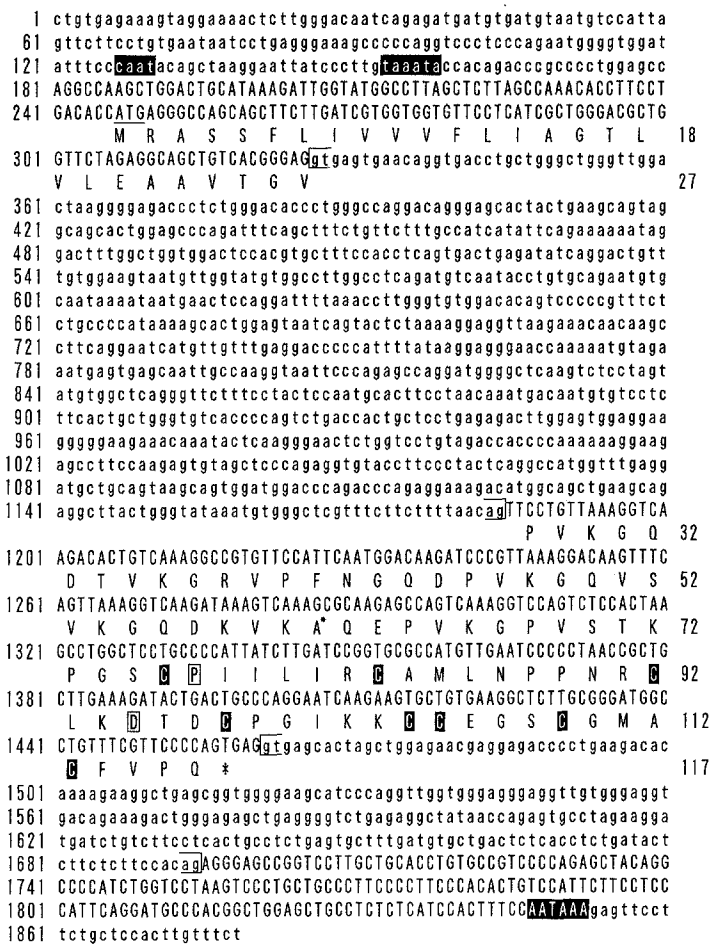

Fig. 2. **Nucleotide sequence of the human elafin gene**. The translated protein sequence is shown below the nucleotide sequence. The putative "CAAT" and "TATA" boxes and polyadenylation signal are marked by black boxes. The amino acids defining the "four-disulfide core" (6) or "WAP" motif (7) are indicated by filled boxes: black, cysteine; grey, aspartic acid and proline. The translation initiation site ATG is underlined. The exact Cap site and 3' end of the gene were not determined. The N-terminal amino acid of mature elafin is indicated by asterisk.

polymorphonuclear leukocytes, is released at sites of inflammation and initiates the breakdown of elastin, a major extracellular component of elastic tissues such as ligaments and blood vessels. Elafin is believed to play an important protective role against destructive degradation, by excessive elastase, of the structural integrity of elastin-containing tissues. In the present study, we established its secretory nature by determining the complete amino acid sequence of preproelafin through gene cloning. For clarification of the physiological relevance of elafin as an inhibitor of elastolytic enzymes, it is essential to determine its localization. The genomic DNA clone isolated here and its restriction enzyme fragments will serve as useful probes

for identifying the cells that synthesize, process, and secrete elafin. The following interesting question also remains to be answered: Is elafin secreted as the active form to exert its inhibitory effects directly or is it secreted as an inactive proform and then activated extracellularly?

Elafin belongs to the new family of proteins with the characteristic "four-disulfide core" (6). The amino acids defining this core are indicated by black boxes in Fig. 2 including 8 Cys, 1 Pro, and 1 Asp residues. The core sequence is also called "WAP" motif (7). The other members of the family so far known are: SPAI (13), HUSI-I (2), the ADMLX gene product (adhesion molecule-like from the X chromosome) (14), the WDNM1 gene product, a protein associated with the loss of metastatic potential (15), whey phosphoprotein (16), red sea turtle protease inhibitor (17), wheat germ agglutinin (6), neurophysin (18), scorpion and snake venom toxins (6, 19), and ragweed pollen allergen Ra5 (6).

The pro-sequence of the elafin precursor has a unique structural feature; it contains 5 homologous repeats of a 6-amino acid unit. The amino acid sequences of these repeats (Val-Lys-Gly-Gln-X-X) are very similar to those of the homologous repeats found in the major clotting protein SVP-1 from guinea pig seminal vesicle (8). Covalent clotting of SVP-1 is catalyzed by a transglutaminase which cross-links Lys and Gln side chains contained in the repeats. It seems, therefore, reasonable to speculate that the elafin profragment may be cross-linked to generate a protective meshwork at the sites of inflammation and injury where extracellular matrices tend to be damaged. If this is the case, the elafin precursor produces two proteins with different but coordinated functions.

Acknowledgments--We thank Setsuko Satoh and Kazuko Tanaka for their secretarial and technical assistance.

REFERENCES

1. Wiedow, O., Schröder, J., Gregory, H., Young, J.A., and Christophers, E. (1990) *J. Biol. Chem.* **265**, 14791-14795.
2. Seemüller, U., Arnhold, M., Fritz, H., Wiedenmann, K., Machleidt, W., Heinzel, R., Appelhans, H., Gassen, H., and Lottspeich, F. (1986) *FEBS Lett.* **199**, 43-48.
3. Araki, K., Kuroki, J., Ito, O., Kuwada, M., and Tachibana, S. (1989) *Biochem. Biophys. Res. Commun.* **164**, 496-502.
4. Heinzel, R., Appelhans, H., Gassen, G., Seemüller, U., Machleidt, W., Fritz, H., and Steffens, G. (1986) *Eur. J. Biochem.* **160**, 61-67.
5. Stetler, G., Brewer, M.T., and Thompson, R.C. (1986) *Nucleic Acids Res.* **14**, 7883-7896.
6. Drenth, J., Low, B.W., Richardson, J.S., and Wright, C.S. (1980) *J. Biol. Chem.* **255**, 2652-2655.
7. Bairoch, A. (1991) *Nucleic Acids Res.* **19**, 2241-2245.
8. Moore, J.T., Hagstrom, J., McCormick, D.J., Harvey, S., Madden, B., Holicky, E., Stanford, D.R., and Wieben, E.D. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6712-6714.

9. Hagstrom, J.E., Harvey, S., Madden, B., McCormick, J., and Wieben, E.D. (1989) *Mol. Endocrinol.* **3**, 1797-1806.
10. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.
11. Christophers, E., Schröder, J., Pioli, D., Wiedow, O., and Edge, M.D. (1990) Eur. Pat. Appl. EP 402068 A1.
12. Mount, S.M. (1982) *Nucleic Acids Res.* **10**, 459-471.
13. Araki, K., Kuwada, M., Ito, O., Kuroki, J., and Tachibana, S. (1990) *Biochem. Biophys. Res. Commun.* **172**, 42-46.
14. Legouis, R., Hardelin, J., Levilliers, J., Claverie, J., Compain, S., Wunderle, V., Millasseau, P., Paslier, D.L., Cohen, D., Caterina, D., Bougueleret, L., Delemarre-Van de Waal, H., Lutfalla, G., Weissenbach, J., and Petit, C. (1991) *Cell* **67**, 423-435.
15. Dear, T.N. and Kefford, R.F. (1991) *Biochem. Biophys. Res. Commun.* **176**, 247-254.
16. Dandekar, A.M., Robinson, E.A., Appella, E., and Qasba, P.K. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3987-3991.
17. Kato, I. and Tominaga, N. (1979) *Fed. Proc.* **38**, 832 (Abstract #3168).
18. Drenth, J. (1981) *J. Biol. Chem.* **256**, 2601-2602.
19. Fazal, A., Beg, O.U., Shafqat, J., Zaidi, Z.H., and Jörnvall, H. (1989) *FEBS Lett.* **257**, 260-262.