Volume 278, number 1, 87–90 FEBS 09311 January 1991

ADONIS 001457939100028I

Gene organization of oryzacystatin-II, a new cystatin superfamily member of plant origin, is closely related to that of oryzacystatin-I but different from those of animal cystatins

Hiroto Kondo[1], Keiko Abe[1], Yasufumi Emori[2] and Soichi Arai[1]

[1]*Department of Agricultural Chemistry, Faculty of Agriculture, University of Tokyo, 1-1-1 Yayoi-cho, Bunkyo-ku, Tokyo 113, Japan* and [2]*Department of Biophysics and Biochemistry, Faculty of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan*

Received 16 October 1990

The gene structure of oryzacystatin-II, a new cystatin superfamily member of rice seed origin, was determined. It spans approximately 2.5 kbp and comprises 3 exons. The number of exons and the intron-breakpoints coincide with those of oryzacystatin-I, the first well-defined plant cystatin. However, no similar sequences were observed between the two oryzacystatin genes in 5'-upstream regulatory regions, even though both are expressed specifically during the ripening stage of rice seeds. The gene organization of these two plant cystatins is generally different from that of animal cystatins.

Cysteine proteinase inhibitor; Rice seed; Genomic DNA; Intron breakpoint; Cystatin superfamily

1. INTRODUCTION

Cystatins refer to the protein members of a superfamily that specifically inhibits cysteine proteinases. Animal cystatins [1,2] have been well defined and are classified into 3 families on the basis of their molecular structures [3]. These include family-I for cystatins with no disulfide bond, family-II for those with two disulfide bonds, and family-III for those with a multitudinous structure, such as kininogens.

However, no detailed information has been obtained so far for cystatins of plant origin. Oryzacystatin-I (OC-I) [4–6] and oryzacystatin-II (OC-II) [7] are the first well-defined cystatins of plant origin. These proteins could be classified as members of a unique family, because while they contain no disulfide bonds, as in the case of family-I cystatins, overall they have a greater similarity to family-II members. Although these two oryzacystatins have significant amino acid sequence homology to each other (55%) [6,7], they show different inhibitory specificities against cysteine proteinases [7]; OC-I inhibits papain more effectively than cathepsin H whereas OC-II inhibits cathepsin H rather than papain. On the other hand, the expression of the mRNAs for OC-I and OC-II occurs only during the rice seed ripening stage and shows an almost identical time-course pattern [7]. This suggests that the same regulatory mechanism, such as a *cis*-element, may exist in their genes.

Oryzacystatins and animal cystatins are thought to derive from a cognate ancestral gene [8]. To clarify this notion, the elucidation of their genes is essential. In this study we report on the cloning and sequencing of a genomic DNA for OC-II to compare with that of OC-I. We also discuss the nature of the gene organization of plant and animal cystatins.

Correspondence address: S. Arai, Department of Agricultural Chemistry, Faculty of Agriculture, University of Tokyo, 1-1-1 Yayoi-cho, Bunkyo-ku, Tokyo 113, Japan

2. MATERIALS AND METHODS

2.1. *Materials*

Materials used in this study were obtained from the following sources: restriction enzymes from Takara Shuzo Co., modifying enzymes from Toyobo Co., and [α-^{32}P]dCTP, [γ-^{32}P]ATP, the multiprime labeling kit and nylon membrane Hybond N from Amersham Co. All other chemicals used were of reagent grade. Ears of the rice cultivar Nipponbare, *Oryza sativa* L. *japonica*, were picked two weeks after flowering and the mature seeds were harvested at the Experimental Farm, University of Tokyo. The ears and seeds were stored at −80°C until used.

2.2. *Methods*

We constructed a rice seed λEMBL3 genomic library as described previously [8]. All procedures from screening to sequencing were carried out using the same previously described method [8]. S_1 mapping analysis was conducted according to standard procedures [9]. Genomic DNA blot hybridization was carried out as described previously [10] using ^{32}P-labeled OC-II cDNA as a probe.

3. RESULTS AND DISCUSSION

3.1. *Gene structure of OC-II*

Restriction maps of 7 positive clones obtained from

the 1.8×10^6 plaques library revealed that all contained a common region. One of the clones, designated as λnOCg1, was chosen for further analysis. Nucleotide sequence analysis showed that λnOCg1 as a gene spans approximately 2.5 kbp comprising 3 exons (Fig. 1). In the 5′-region, typical TATA box and CAAT box sequences are detected at 30 bp and 75 bp, respectively, upstream from the transcription start point. Exon 1 contains the whole of the 5′-non-coding region and a part of the coding region corresponding to the first 43 amino acids of the protein as a gene expression product. Exon 2, which follows the first intron (408 bp), comprises 3 elements, the rest of the coding region, the stop codon TAA, and the first two nucleotides, G and C, of the 3′-non-coding region. Exon 3, which follows the second intron (132 bp), contains the 3′-non-coding region exclusively. The sequences of the two introns match the general consensus for intron sequences in that they follow the GT/AG rule [11]. Genomic Southern blotting (data not shown) revealed that the OC-II gene is a single copy gene, as is the case for the OC-I gene [8].

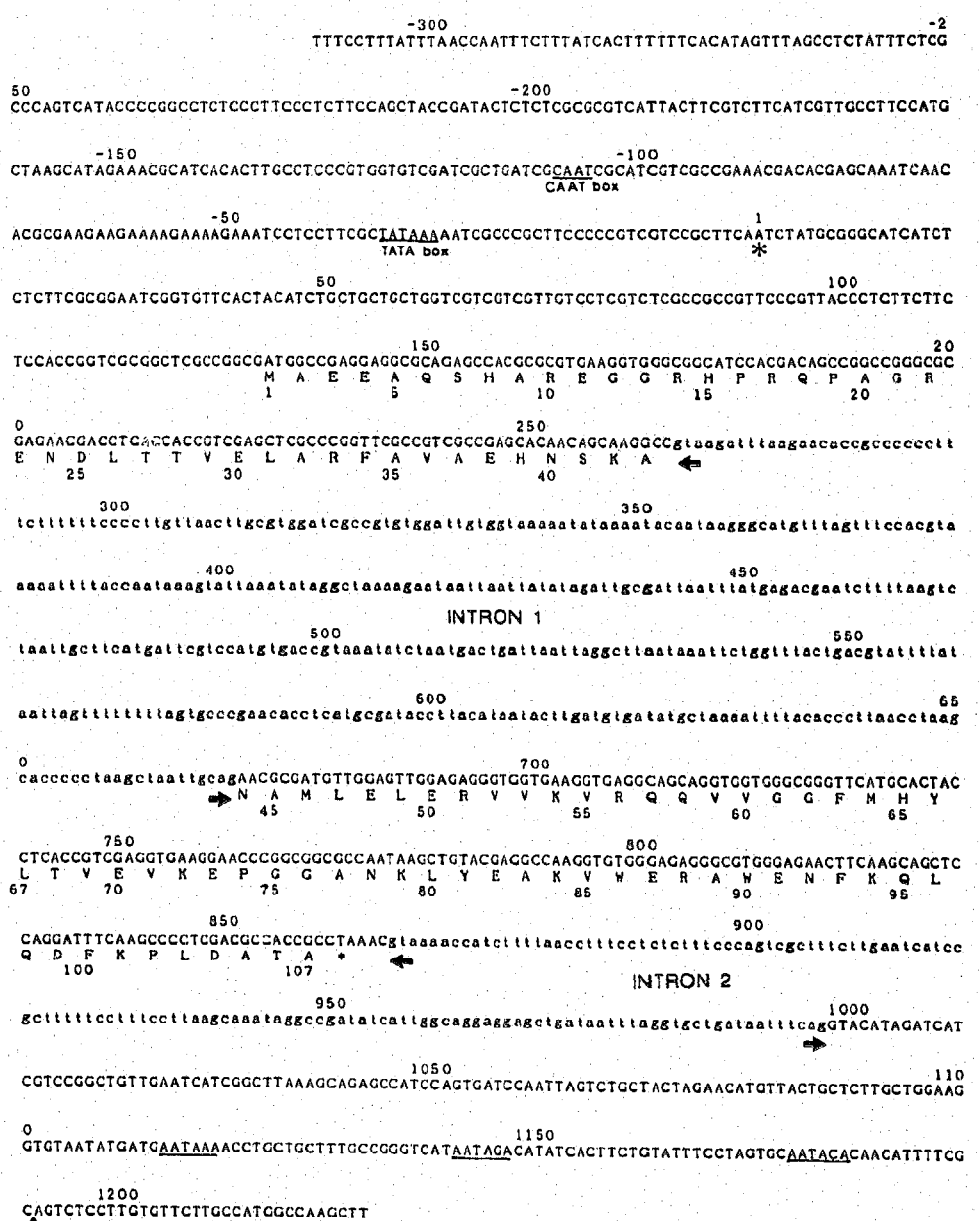

Fig. 1. Nucleotide sequence of the DNA fragment encoding oryzacystatin-II and its deduced amino acid sequence. The sequence is numbered from the transcription start point (denoted by asterisk) as determined by S1 mapping (data not shown). The last digits of the numbers are aligned to the corresponding nucleotide. The TATA box, CAAT box, polyadenylation signal sequence AATAAA and its analogous sequences are underlined. The arrowhead indicates the poly(A) addition site. The sequences of introns are shown in lower-case letters and the intron boundaries are represented by horizontal arrows. The exons are denoted by capital letters.

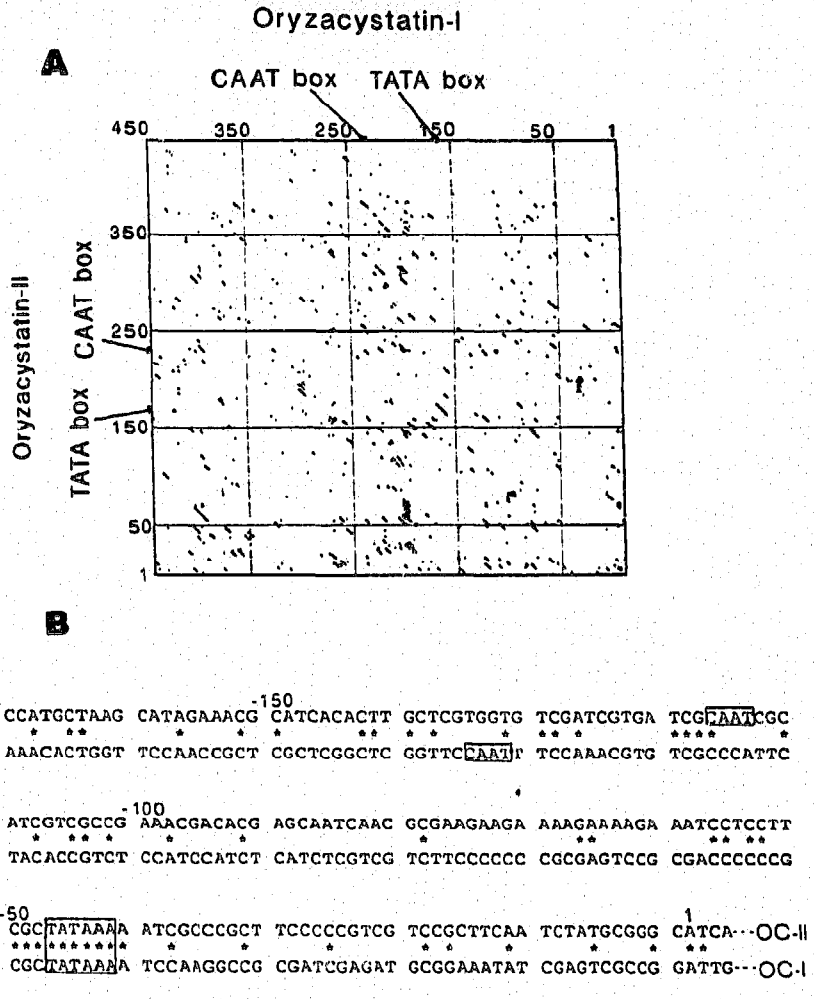

Fig. 2. A: Harr plot analysis of the 5'-upstream regions of OC-I (abscissa) and OC-II (ordinate). Numbers begin from the initiation Met (numbering 1). Homologous regions are shown as dots when the sequences match at 6 out of 8 nucleotides. Translation start points are located at positions 100 (OC-I) and 120 (OC-II). TATA boxes are located at positions 130 (OC-I) and 150 (OC-II), CAAT boxes are at position 230 (OC-I and OC-II). B: Comparison of the nucleotide sequences of OC-I and OC-II around the TATA and CAAT boxes. Identical nucleotides are asterisked. Positions are arranged to maximize the sequence homologies. The sequence is numbered from the transcription start point of OC-II (denoted by 1).

3.2. *Comparison of the gene structures of OC-I and OC-II*

The gene structure of OC-II (Fig. 1) is essentially similar to that of OC-I [8] with respect to the number and positions of its introns. In the case of OC-II, the first intron intervenes between Ala[43] and Asn[44], which corresponds to the position between Ala[37] and Asn[38] in OC-I [8], while in both genes the second intron is located next to the stop codon. A comparison of the 5'-non-coding regions, which might be important for the regulation of expression, shows that, except for the TATA box region, there is no significant homology as seen from the Harr plot analysis (Fig. 2A), although the expression of both the OC-I and OC-II mRNAs is limited to the rice seed ripening stage [6,7]. Little sequence homology is also observed between the two genes for OC-I and OC-II in the first and second introns and 3'-non-coding regions. Thus, no apparent sequence as *cis*-element for transcriptional regulation existed, although similar sequences around TATA box (Fig. 2B) might be considered to function as a common *cis*-element for their stage-specific expression.

3.3. *Comparison with animal cystatin genes*

We compared oryzacystatins and animal cystatins for gene structure. The comparison of nucleotide sequences and exon-intron breakpoints is interesting especially from an evolutionary point of view, although the meaning of the intron breakpoint position has not yet been clarified. Fig. 3 shows the gene structures of various cystatins with special emphasis on exon-intron boundaries.

All these animal cystatin genes are composed of three exons and two introns, and in this respect, the oryzacystatin genes are the same as those for the animal cystatins. However, with respect to the intron boun-

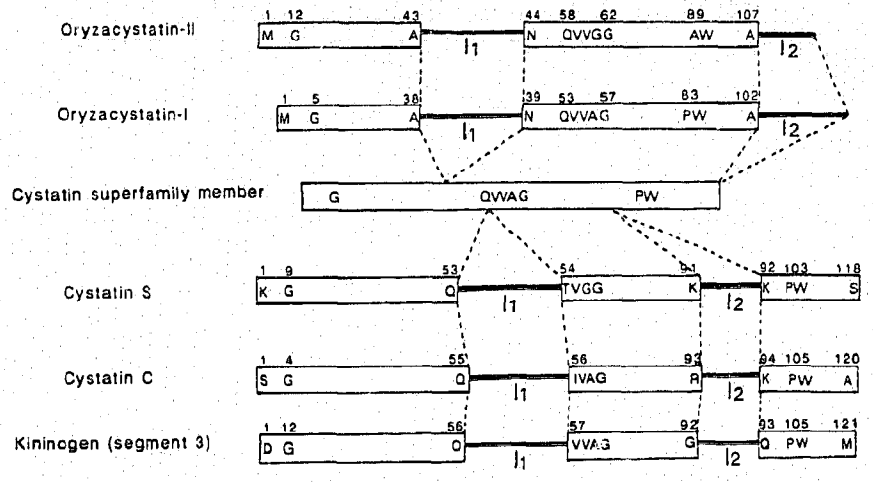

Fig. 3. Comparison of the gene organization of various cystatins. Open boxes denote the coding regions. Solid lines denote introns 1 (I_1) and 2 (I_2). The amino acid residues are indicated by the one-letter notation. The numbering starts at the initiation Met residue. The glycine residue (G) near the NH_2-terminal, the central part of the Gln-XXa-Val-XXa-Gly sequence (Q-V-G), and the Pro-Trp sequence (PW) near the COOH-terminal, 3 sequences homologous among cystatins, are noted for comparison.

daries and the length of each intron, there are clear differences between animal cystatins and oryzacystatins. Both introns in animal cystatins exist in the coding region, and the intron breakpoints of cystatin C (family-II) [12], cystatin S (family-II) [13], and human kininogen (family-III) [14] are equally conserved. The first intron intervenes between Q and V of the conserved QVVAG sequence and the second intron is located in the coding region near the COOH-terminal. As opposed to this, in the case of oryzacystatins the first intron exists upstream from the conserved QVVAG sequence, and the second intron is situated in the noncoding region.

Taking into consideration the differences in gene structure between plant and animal cystatins, besides the unique amino acid sequences, we propose that a new category, for example a phytocystatin family, should be organized to include oryzacystatins, in addition to 3 conventional cystatin families that comprise animal cystatins. Further data, especially about the gene structures of family-I cystatins, are essential for a more precise comparison of gene structures of plant and animal cystatins. However, it can be assumed that plant cystatin genes diverged from a cognate ancestral gene before the divergence of the animal cystatin genes.

REFERENCES

[1] Grubb, A. and Lofberg, H. (1982) Proc. Natl. Acad. Sci. USA 79, 3024–3027.
[2] Hirado, M., Tsunasawa, S.S., Sakiyama, F., Ninobe, M. and Fujii, S. (1985) FEBS Lett. 186, 41–45.
[3] Barrett, A.J., Rawlings, N.D., Davies, M.E., Machleidt, W., Salvesen, G. and Turk, V. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G. eds) pp. 515–569, Elsevier, Amsterdam.
[4] Abe, K., Kondo, H. and Arai, S. (1985) Agric. Biol. Chem. 49, 3349–3350.
[5] Abe, K., Kondo, H. and Arai, S. (1987) Agric. Biol. Chem. 51, 2763–2768.
[6] Abe, K., Emori, Y., Kondo, H., Suzuki, K. and Arai, S. (1987) J. Biol. Chem. 262, 16793–16797.
[7] Kondo, H., Abe, K., Nishimura, I., Watanabe, H., Emori, Y. and Arai, S. (1990) J. Biol. Chem. 265, 15832–15837.
[8] Kondo, H., Emori, Y., Abe, K., Suzuki, K. and Arai, S. (1989) Gene 81, 259–265.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[10] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[11] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[12] Saitoh, E., Sabatini, L., Eddy, R., Shows, T., Azen, E., Isemura, S. and Sanada, K. (1989) Biochem. Biophys. Res. Commun. 162, 1324–1331.
[13] Saitoh, E., Kim, H.-S., Smithies, O. and Maeda, N. (1987) Gene 61, 329–338.
[14] Müller-Esterl, W., Fritz, H., Kellerman, J., Lottspeich, F., Machleidt, W. and Turk, V. (1986) in: Cysteine Proteinases and Their Inhibitors (Turk, V. ed.) pp. 369–392, Walter de Gruyter, Berlin.