Volume 264, number 2, 279–282 FEBS 08453 May 1990

cDNA cloning and sequencing of component C9 of proteasomes from rat hepatoma cells

Atsushi Kumatori[1], Keiji Tanaka[1], Tomohiro Tamura[1], Tsutomu Fujiwara[2], Akira Ichihara[1], Fuminori Tokunaga[3], Aya Onikura[4] and Sadaaki Iwanaga[4]

[1]*Institute for Enzyme Research, The University of Tokushima, Tokushima 770,* [2]*Otsuka Pharmaceutical Company Ltd., Tokushima 771-01* and [3]*Department of Molecular Biology, Graduate School of Medical Science and* [4]*Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan*

Received 2 April 1990

The nucleotide sequence of component C9 of rat proteasomes (multicatalytic proteinase complexes) has been determined from a recombinant cDNA clone isolated by screening a Reuber H4TG hepatoma cell cDNA library using synthetic oligodeoxynucleotide probes corresponding to partial amino acid sequences of the protein. The predicted sequence of C9 consists of 261 amino acid residues with a calculated molecular weight of 29 496. The C9 component is a novel protein, differing from known proteins, but its primary structure resembles those of other proteasome components, including C2, C3 and C5, although its similarity to C5 is relatively low, suggesting that proteasomes consist of a family of proteins that have evolved from a common ancestor.

cDNA cloning; Multicatalytic proteinase; Proteasome; Subunit C9

1. INTRODUCTION

Proteasomes are unusually large multicatalytic proteinase complexes with a molecular mass of approximately 750 kDa, which are composed of over 10 subunits of non-identical polypeptides with similar molecular masses of 21–31 kDa and different p*I* values of 3–10 [1–5]. For determining the functions of these proteasomal multisubunit complexes, it is essential to obtain information about the structure–function relationships of individual subunits and the molecular mechanism of subunit organization. For this purpose, we are attempting to determine the primary structures of all the subunits of proteasomes by recombinant DNA techniques. So far we have isolated and sequenced cDNAs for components C2 [6], C3 [7] and C5 [8] of rat liver proteasomes. In this paper, we report the cloning and sequencing of a full-length cDNA for another component, C9, of rat proteasomes.

2. MATERIALS AND METHODS

Rat liver proteasomes were purified as reported previously [2,4]. Component C9 was isolated from the purified proteasomes by reversed-phase HPLC on a Cosmosil $5C_4$-300 column [5]. Fragments of C9 protein were obtained by digestion with lysyl endopeptidase, and their amino acid sequences were determined with a gas-phase sequencer (Applied Biosystems, model 477A), and 120A phenylthiohydantoin analyzer on-line system [6,7]. A cDNA library of Reuber H4TG hepatoma cells was constructed in a phagemid expression vector, Bluescript KS^+ (Stratagene) as described previously [7]. For isolation of cDNAs for component C9, about 30000 transformants were screened by hybridization with oligodeoxyribonucleotide probes (see section 3.2) that had been synthesized in an Applied Biosystems Model 380A DNA Synthesizer and labeled at their 5'-end with [γ-^{32}P]ATP. Colony hybridization was carried out by a reported method [6]. DNA sequencing was carried out by the dideoxy chain-termination method with a 7-DEAZA Sequencing Kit (Takara Shuzo) and Sequenase (Toyobo). Northern blot analysis was performed as described before [6] using 10 μg of poly(A)$^+$ mRNA extracted from various tissues of rats and a probe (600 bp *Xho*I-*Pst*I fragment of C9 cDNA) labeled by the multiprime DNA labeling method with [α-^{32}P]dCTP.

Correspondence address: K. Tanaka, Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan

Abbreviations: SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; bp, base pairs

3. RESULTS AND DISCUSSION

3.1. *Protein sequencing of component C9*

Previously, we reported the separation of multiple components of proteasomes from rat liver by reversed-phase HPLC [5]. First, 10 major components were separated on a Cosmosil $5C_4$-300 column and named component 1 (C1) to component 10 (C10) in order of their elution. Component C9 was recovered at 53% acetonitrile concentration. Fig.1 shows the profiles on SDS-PAGE of all the components of rat liver proteasomes and the isolated component C9. On SDS-PAGE analysis, the molecular mass of C9 was determined to be 28700 ± 700 [5,6].

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

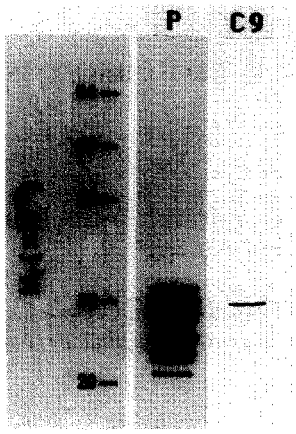

Fig. 1. Profiles of rat liver proteasomes and isolated component C9 on SDS-PAGE. Proteasomes (25 μg) and component C9 (approximately 1 μg) were subjected to SDS-PAGE. M_r, molecular mass markers; P, purified proteasome complexes; C9, isolated C9. Proteins were stained with Coomassie brilliant blue.

The N-terminus of C9 seemed to be blocked, because it was not reactive with phenylisothiocyanate. For determination of the primary structure of its internal region, samples of C9 were reduced, *S*-pyridylethylated and digested with lysyl endopeptidase. The resulting fragments were resolved by reversed-phase HPLC on a Chemcosorb 7-ODS-H column (data not shown) and their partial primary structures were analyzed by automated Edman degradation (solid lines in Fig. 2).

3.2. *Isolation of cDNA clones encoding component C9*

To isolate proteasome cDNA clones more efficiently, we constructed a new cDNA library with the phagemid Bluescript KS$^+$ vector using poly(A)$^+$ RNAs extracted from the H4TG cells [8], because the mRNA levels of components C2, and C3 in Reuber H4TG hepatoma cells were very much higher than those in normal rat liver (Kumatori et al., submitted). We screened this library by hybridization with synthetic deoxyribonucleotides as probes. For use as probes, we selected parts of the sequences of two of the proteolytic fragments of C9 with minimum codon ambiguity (Tyr-Ile-Gly-Trp-Asp-Lys and Asn-Glu-Asp-Met-Ala-Cys) and synthesized their antisense oligonucleotide sequences, 5′-TTRTCCCANCCDATRTA-3′ and 5′-CANGCCATRTCYTCRTT-3′ (N=AGTC, D=AGT, Y=TC, R=AG). We first screened about 30000 colonies of the H4TG cell cDNA library with 48 mixtures of the former 17mer oligonucleotides. For second screening, we used 32 mixtures of the latter 17mer oligonucleotides. Ten cDNA clones that gave strongly positive signals with both probes were isolated from the library by colony hybridization techniques. Of these, the clone (about 1.2 kb length including a poly(A) tail) that carried the largest cDNA insert was subjected to cDNA sequencing.

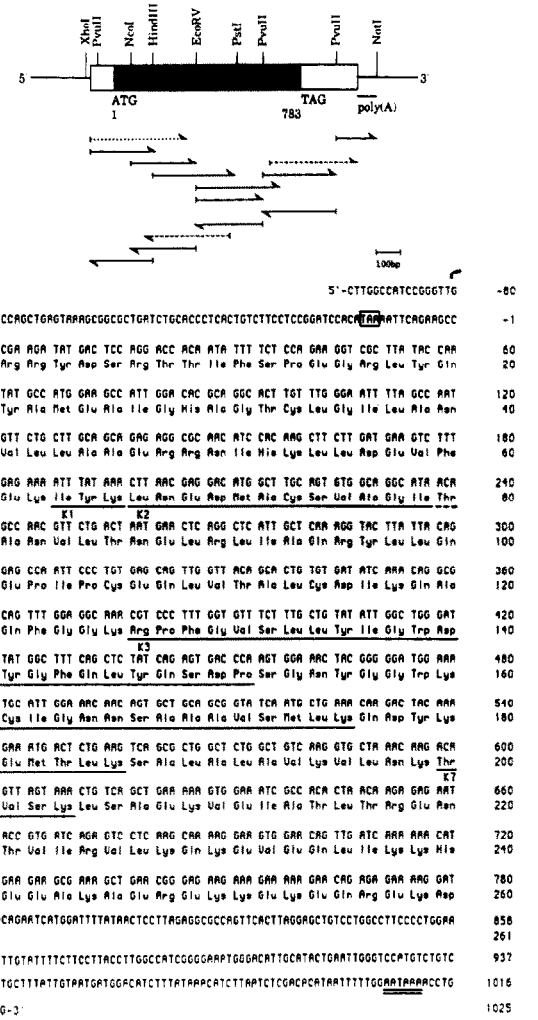

Fig. 2. Structure of cDNA for component C9 of rat proteasomes. Upper panel: Restriction endonuclease map of cloned cDNA for component C9 and sequencing strategy. The solid and open boxes show the coding region and 5′- and 3′-non-coding regions, respectively. Continuous lines indicate the sequence of the vector, Bluescript KS$^+$. The numbers below boxes indicate the nucleotide numbers of the first nucleotide of the initiation codon, ATG, and the nucleotide before the termination codon, TAG. Sequenced regions are shown by horizontal arrows. The sequence of the 5′-end of cDNA was determined by direct sequencing of an isolated clone, using a primer of the T3 promoter of Bluescript KS$^+$ (dotted arrow). Two 17mer oligonucleotides were synthesized as reported previously [6] and used as primers for sequencing some regions (dashed arrows). Some fragments were obtained by subcloning after cleavages with *Hin*dIII, *Nco*I, *Eco*RV and *Pvu*II and were sequenced with T3 and T7 primers (solid arrows). The bar represents 100 bp. Lower panel: Nucleotide sequence of the cDNA encoding component C9 and the amino acid sequence deduced from its largest open reading frame. Nucleotides are numbered in the 5′ to 3′ direction, beginning with the first residue of the putative initiation methionine codon ^{1}ATG^3. The nucleotides on the 5′ side are indicated by negative numbers. The predicted amino acid sequence of C9 is shown below the nucleotide sequence. Amino acid residues are numbered from the N-terminus. Continuous lines show the amino acid sequences corresponding to those obtained by Edman degradation of the fragments cleaved with lysyl endopeptidase. Amino acids shown by broken lines in the sequences of K2, and K4 were not identified by chemical analysis. The termination codon TAG is doubly underlined with broken lines. The possible polyadenylation signal (AATAAA) is doubly underlined with continuous lines.

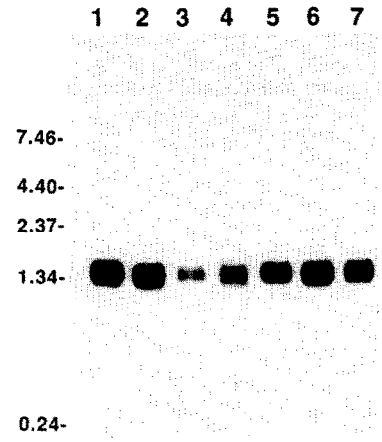

Fig. 3. RNA blot hybridization of rat liver C9 cDNA with poly(A)$^+$ RNAs from various rat tissues. Poly(A)$^+$ RNAs (10 μg) were used for Northern blot analysis. The *XhoI-PstI* fragment (600 bp) of cloned C9 cDNA was used as a probe (Fig. 2). Rat liver (lane 1), kidney (lane 2), skeletal muscle (lane 3), heart (lane 4), lung (lane 5), spleen (lane 6) and brain (lane 7) were examined. The positions of RNA standards (from Bethesda Research Laboratories) are shown (in kilobases) on the left.

3.3. *Primary structure deduced from the cDNA sequence*

The nucleotide sequence of the C9 cDNA clone was determined by the strategy illustrated in Fig. 2 (upper panel). The resulting nucleotide sequence of the cDNA and the primary structure of the C9 deduced from the cDNA sequence are shown in Fig. 2 (lower panel). The sequence of 1121 nucleotides included the entire coding region and 5′- and 3′-non-coding regions. The 3′-non-coding region consisted of 242 nucleotides, excluding the poly(A) tail. A putative polyadenylation signal (AATAAA), which is common to eukaryotic mRNAs, was located 20 nucleotides upstream from the poly(A) addition site. Part of the poly(A) sequence (approximately 90 nucleotides) is seen at the 3′-end of this cDNA sequence. The longest open reading frame (ORF) is 783 nucleotides long, which corresponds to a protein of 261 amino acids. From this deduced sequence, the molecular mass of C9 was calculated to be 29496. We concluded that ATG, located at nucleotides 1 to 3, is the initiation codon, because the longest ORF is usually thought to be the coding sequence of a protein, because a stop codon is present upstream of the predicted initiation codon ^{1}ATG^3 for C9 at nucleotides -13 to -15 (boxed in Fig. 2), and because this ^{1}ATG^3 is surrounded by a sequence that is similar to the consensus sequence for translation initiation [9]. Several lines of evidence indicated that the amino acid sequence shown in Fig. 2 is actually that of C9 of proteasomes: (i) the partial sequences of several fragments determined chemically were found to be in excellent accordance with those deduced from the nucleotide sequence of cDNA (solid lines in Fig. 2); (ii) the molecular mass of C9 estimated from its amino acid composition (29496) was similar to that estimated by SDS-PAGE (28700); and (iii) the size of mRNA for C9 was about 1.34 kb (Fig. 3), which was approximately the same as that of the insert of the isolated C9 cDNA clone (1.10 kb) plus poly(A) tail.

On computer-assisted homology analysis, no obvious overall structural similarity of component C9 with previously reported proteins could be found, suggesting that C9 is a novel protein. However, the primary structure of C9 showed marked similarity with those of other proteasome components, including C2 [6], C3 [7] and C5 [8] of rat proteasomes, and a PROS-35 protein of *Drosophila* proteasomes [10]. The identities of the amino acid residues of C9 with those of C2, C3, C5 and PROS-35 were 28.2%, 38.2%, 18.0% and 27.8%, respectively. The overall amino acid sequence of C9 showed low similarity with that of C5, but this was statistically significant. This structural similarity of different proteasomal components suggests that these components are members of a family of proteins and that they may have originated from a common ancestor.

3.4. *Expression of the gene for component C9 in various rat tissues*

To determine the tissue specificity of C9 gene expression, we examined the levels of C9-specific mRNA in various rat tissues. On Northern blot analysis, the mRNAs extracted from various rat tissues gave two hybridization bands of about 1.34 kb (Fig. 3). The level of mRNA for component C9 varied significantly in the different tissues examined, but was very similar to those of the mRNAs for components C2 [6], C3 [7] and C5 [8]. Moreover, these mRNA levels were fairly similar to the absolute tissue contents of proteasomes measured by quantitative enzyme immunoassay [2]. These results suggest that proteasomes with similar subunit compositions are present as large complexes in different cells and tissues.

Acknowledgements: This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan. We thank S. Kajiyama and Y. Nishina for performing amino acid analyses.

REFERENCES

[1] Rivett, A.J. (1989) Arch. Biochem. Biophys. 268, 1–8.
[2] Tanaka, K., Ii, K., Ichihara, A., Waxman, L. and Goldberg, A.L. (1986) J. Biol. Chem. 261, 15197–15203.
[3] Tanaka, K., Yoshimura, T., Ichihara, A., Kameyama, K. and Takagi, T. (1986) J. Biol. Chem. 261, 15204–15207.
[4] Tanaka, K., Yoshimura, T., Kumatori, A., Ichihara, A., Ikai, A., Nishigai, M., Kameyama, K. and Takagi, T. (1988) J. Biol. Chem. 263, 16209–16217.
[5] Tanaka, K., Yoshimura, T., Ichihara, A., Ikai, A., Nishigai, M., Morimoto, M., Sato, M., Tanaka, N., Katsube, Y., Kameyama, K. and Takagi, T. (1988) J. Mol. Biol. 203, 985–996.

[6] Fujiwara, T., Tanaka, K., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A. and Nakanishi, S. (1989) Biochemistry 28, 7332–7340.
[7] Tanaka, K., Fujiwara, T., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A. and Nakanishi, S. (1990) Biochemistry, in press.
[8] Tamura, T., Tanaka, K., Kumatori, A., Yamada, F., Tsurumi, C., Fujiwara, T., Ichihara, A., Tokunaga, F., Aruga, R. and Iwanaga, S. (1990) FEBS Lett., in press.
[9] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[10] Haass, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K. and Kloetzel, P.-M. (1989) EMBO J. 8, 2373–2379.