Volume 301, number 1, 65–68 FEBS 10917 April 1992

cDNA cloning of rat proteasome subunit RC1, a homologue of RING10 located in the human MHC class II region

Masashi Aki[a], Tomohiro Tamura[b], Fuminori Tokunaga[c,*], Sadaaki Iwanaga[c], Yoshihiro Kawamura[b], Naoki Shimbara[d], Susumu Kagawa[a], Keiji Tanaka[b] and Akira Ichihara[b]

[a]*Department of Urology, School of Medicine, and* [b]*Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan,* [c]*Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan* and [d]*Biomedical Research Institute, Yokohama 244, Japan*

Received 31 January 1992

The nucleotide sequence of a cDNA that encodes a new subunit, named RC1, of rat proteasomes (multicatalytic proteinase complexes) has been determined. The polypeptide predicted from the open reading frame consisted of 208 amino acid residues with a calculated molecular mass of 23,130, which is consistent with the size obtained by electrophoretic analysis of purified RC1. The partial amino acid sequences of several fragments of RC1, obtained by protein chemical analyses, were found to be in excellent accordance with those deduced from the cDNA sequence. Surprisingly, the overall structure of RC1 was found to be almost identical to that of recently isolated RING10, whose gene is located in the class II region of the human MHC gene cluster. This finding suggests that RC1 is a homologue of human RING10, supporting the proposal that proteasomes are involved in the antigen processing pathway.

Antigen processing; Class II MHC gene; cDNA cloning; Multicatalytic proteinase; Proteasome; RING10; subunit RC1

1. INTRODUCTION

Proteasomes are a novel type of intracellular proteinase complex with the unusually large size of 20S, and consist of approximately 15 non-identical subunits [1]. For determining the functions of this proteasomal multi-subunit complex, we are attempting to clarify the entire structure of the rat proteasome by recombinant DNA techniques, and so far we have isolated and sequenced 5 cDNAs encoding subunits RC2, RC3, RC5, RC8 and RC9 [2–6]. During structural analyses of the rat proteasome, we recently isolated a cDNA encoding a new subunit named RC1. Interestingly, we found by computer analysis that the primary structure of the RC1 showed marked similarity to that of RING10, found very recently as the product of a proteasome-related gene located in the human MHC class II region [7], suggesting that RC1 is a homologue of human RING10. In this paper, we report the cloning and sequencing of a full-length cDNA for RC1 and discuss the possible involvement of proteasomes in the antigen processing pathway, which has been proposed based on recent findings that certain proteasome-related genes are present within the MHC gene cluster [8].

Abbreviations: SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; MHC, major histocompatibility complex.

**Present address:* Himeji Institute for Technology, Faculty of Science, Harima Science Park City, Hyogo 678-12, Japan.

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

2. MATERIALS AND METHODS

Subunit RC1 was isolated from purified rat liver proteasomes by reversed-phase HPLC on a Cosmosil $5C_4$-300 column [9]. Fragments of RC1 protein were obtained by digestion with lysylendopeptidase, and their amino acid sequences were determined with a gas-phase sequencer (Applied Biosystems, model 477A), and 120A phenylthiohydantoin analyzer on-line system [2,3]. A cDNA library of Reuber H4TG hepatoma cells was constructed in a phagemid expression vector, Bluescript II KS$^-$ (Stratagene) [4]. For isolation of cDNA for RC1, about 120,000 transformants were screened by hybridization with oligodeoxyribonucleotide probes 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 [2]. DNA sequencing was carried out by the dideoxy chain-termination method with a Model 373A automatic DNA sequencer (Applied Biosystems).

3. RESULTS AND DISCUSSION

3.1. *Protein sequencing of RC1*

Previously, we reported the separation of multiple components of proteasomes from rat liver by reversed-phase HPLC [9]. First, 10 major components were separated on a Cosmosil $5C_4$-300 column and named component 1 (C1) to component 10 (C10) in the order of their elution. These components from rats were recently renamed RC1 to RC10 to distinguish them from those

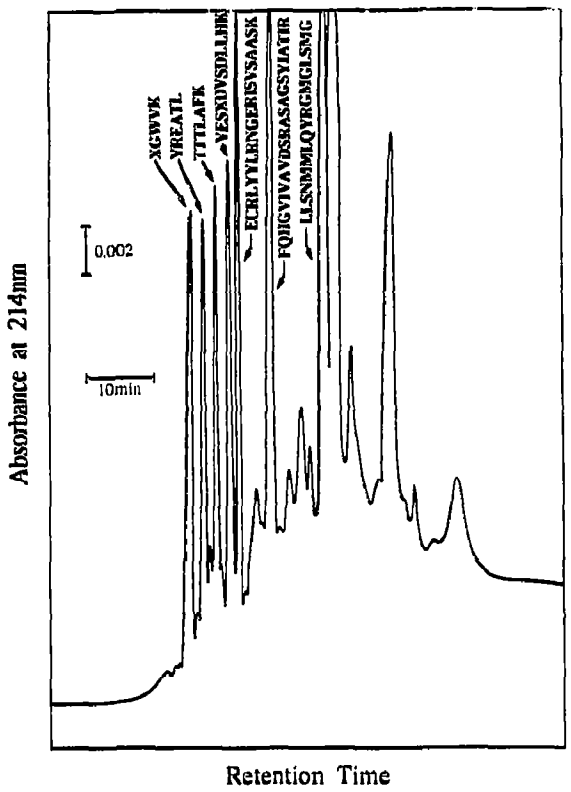
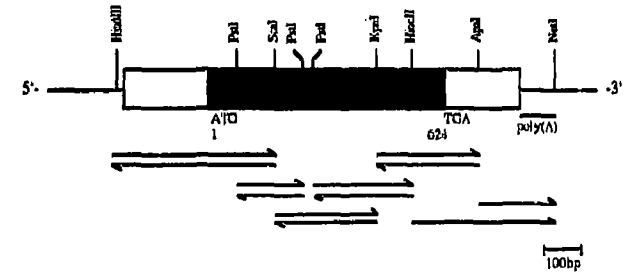


Fig. 1. Separation of fragments of subunit RC1 cleaved with lysylendopeptidase by reversed-phase HPLC. *S*-Pyridylethylated RC1 was digested with lysylendopeptidase and the fragments were resolved by HPLC on a Chemcosorb 7-ODS-H column with a linear gradient of acetonitrile (0 to 80%) in the presence of 0.1% trifluoroacetic acid at a flow rate of 0.5 ml/min. Partial amino acid sequences determined with a protein sequencer are also shown. X denotes an unidentified residue.



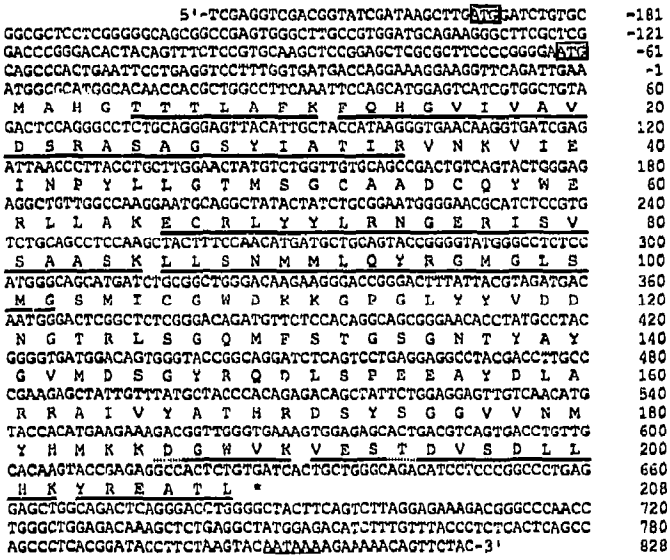


Fig. 2. Structre of cDNA for subunit RC1 of rat proteasomes. Upper panel: restriction endonuclease map of cloned cDNA for RC1 and sequencing strategy. The solid and open boxes show the coding region and 5'- and 3'-non-coding regions, respectively. Continuous lines indicate the sequence of the vector, Bluescript II 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, TGA. Sequenced regions are shown by horizontal arrows. The bar represents 100 bp. Lower panel: nucleotide sequence of the cDNA encoding component RC1 and the amino acid sequence deduced from its 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. Two other possible translation initiation sites are boxed. The nucleotides on the 5' side are indicated by negative numbers. The predicted amino acid sequence of RC1 is shown below the nucleotide sequence. Amino acid residues are numbered from the N-terminus. Continuous thick lines show the amino acid sequences corresponding to those obtained by Edman degradation of fragments cleaved with lysylendopeptidase. The amino acids shown by dotted lines were not identical with those found by chemical analysis. The termination codon TGA is marked with an asterisk. The possible polyadenylation signal (AATAAA) is underlined with a continuous thin line.

of other species. RC1 was recovered at 51% acetonitrile concentration. On SDS-PAGE analysis, the molecular weight of RC1 was determined to be 24,800 ± 800 [2]. The N-terminus of RC1 seemed to be blocked, because it was not reactive with phenyl*iso*thiocyanate [2]. For determination of the primary structure of its internal region, samples of RC1 were reduced, *S*-pyridylethylated and digested with lysylendopeptidase. The resulting fragments were resolved by reversed-phase HPLC on a Chemcosorb 7-ODS-H column and their partial primary structures were analyzed by automated Edman degradation (Fig. 1).

3.2. *Isolation of a cDNA clone encoding RC1*

To isolate cDNA encoding RC1, we screened a cDNA library with the phagemid Bluescript II KS$^+$ vector using poly(A)$^+$RNAs extracted from H4TG cells by hybridization with synthetic deoxyribonucleotides as probes [2,4]. For use as probes, we selected parts of the sequences of two of the proteolytic fragments of RC1 with minimum codon ambiguity (Asn-Met-Met-Leu-Gln-Tyr and Lys-Phe-Gln-His-Gly-Val) and synthesized their antisense oligonucleotide sequences 5'-TAYTGNAGCATCATRTT-3', 5'-TAYTGYAACATCATRTT-3', 5'-ACRCCRTGYTGRAAYTT-3' and 5'-ACYCCRTGYTGRAAYTT-3' (N=ACGT, R=AG Y=CT). We first screened about 120,000 colonies of the H4TG cell cDNA library with 24 mixtures of the former two 17-mer oligonucleotides. For second screening, we used 64 mixtures of the latter two 17-mer oligonucleo-

```
RC1        1 MAHGTTTLAFKFQHGVIVAVDSRASAGSYIATIRVNKVIEINPYLLGTMSGCAADCQYWE
             ****************** ************  ****************************
RING10     1 MAHGTTTLAFKFQHGVIAAVDSRASAGSYISALRVNKVIEINPYLLGTMSGCAADCQYWE

RC1       61 RLLAKECRLYYLRNGERISVSAASKLLSNMMLQYRGMGLSMGSMICGWDKKGPGLYYVDD
             ******************************** ************************** 
RING10    61 RLLAKECRLYYLRNGERISVSAASKLLSNMMCQYRGMGLSMGSMICGWDKKGPGLYYVDE

RC1      121 SGTRLSGQMFSTGSGNTYAYGVMSSGYRQDLSPEEAYDLARRAIVYATHRDSYSGGVVNM
              ****** ********************  ********* **** ***************
RING10   121 SGTRLSGNMFSTGSGNTYAYGVMSSGYRPNLSPEEAYDLGRRAIAYATHRDSYSGGVVNM

RC1      181 YHMKKDGWVKVESTDVSDLLHKYREATL
             **** ***************** ****
RING10   181 YHMKEDGWVKVESTDVSDLLHQYREANQ
```

Fig. 3. Comparison of the protein sequences of RC1 of rat liver proteasomes and human RING 10. Identical amino acid residues are marked with an asterisk. Numbers are residue numbers. The serine, histidine and aspartate residues, which form the consensus catalytic triad in the subtilisin family of serine proteases and are conserved in RING10 [7], are boxed in black. Note that the histidine residue in the consensus catalytic site was replaced by asparagine in RC1. Sequence data for RING 10 are taken from Glynne et al. [7].

tides. A single cDNA clone that gave a strongly positive signal with both probes was isolated from the library by colony hybridization techniques. The clone carried a cDNA insert of about 1.1 kb length including a poly(A) tail, and was subjected to cDNA sequencing.

3.3. *Primary structure of RC1*

The nucleotide sequence of the RC1 cDNA clone and the primary structure of the RC1 protein deduced from the cDNA sequence are shown in Fig. 2. The sequence of 1047 nucleotides included the entire coding region and 5′- and 3′-non-coding regions. The 3′-non-coding region consisted of 204 nucleotides, excluding the poly(A) tail. A putative polyadenylation signal (AATAAA), which is common to eukaryotic mRNAs, was located 22 nucleotides upstream from the poly(A) addition site. We concluded that ATG, located at nucleotides 1 to 3, is the initiation codon, because it is surrounded by a sequence that is similar to the consensus sequence for translation initiation [10], and because the molecular weight of RC1 estimated from the deduced amino acid sequence coded from this position (23,130) was similar to that of RC1 estimated by SDS-PAGE (24,800). In this nucleotide sequence, two ATG-codons were located in the 5′-upstream region of 'ATG' (boxed in Fig. 2), but similarity of amino acid sequence deduced from the isolated RC1 cDNA with its human homologue named RING10 was observed in the amino acid sequence from 'ATG', whereas no homology of RING 10 was found in the upstream regions of the two other large open reading frames (see Fig. 3), strongly suggesting that 'ATG' is the initiation codon of RC1. Subunit RC1 corresponds to a protein of 208 amino acids with a calculated molecular weight of 23,130. The amino acid sequence shown in Fig. 2 was confirmed to be that of RC1 of proteasomes by showing that the partial sequences of several fragments determined chemically were in excellent accordance with those deduced from the nucleotide sequence of cDNA (solid lines, Fig. 2).

3.4. *Similarity to Human RING10*

Based on structural information on proteasomes from various eukaryotes, the concept that the subunits of proteasomes consist of a family of conserved proteins with the same evolutional origin has been proposed [1]. However, the similarity of the primary structure of RC1 is less than those of other subunits, suggesting that it is a different type of subunit in the proteasomal complex. Moreover, the 5′-noncoding region of RC1 is considerably longer than those of other subunits. Thus, the RC1 gene seems to have diverged from the ancestral gene for proteasome subunits at an early stage of evolution.

Computer-assisted homology analysis showed no obvious overall structural similarity of component RC1 with most previously reported proteins, suggesting that RC1 is a novel protein. Surprisingly, however, the primary structure of RC1 showed remarkable similarity with that of the product of the recently cloned RING10 gene [7], which is located in the class II region of the human MHC. These two proteins have the same total numbers of amino acid residues and 93% identities of amino acid residues. In fact, as shown in Fig. 3, RC1 differs from RING10 in only 16 amino acid residues. Glynne et al. [7] classified RING10 in the subtilisin family of serine proteases, because the serine, histidine and aspartate residues, which are a consensus catalytic triad in subtilisin-like proteases, are conserved in RING10 (Fig. 3, residues boxed in black). However, we found that in RC1, the histidine residue in this active site was replaced by an asparagine residue. Therefore, it seems likely that proteasomes do not belong to the serine- protease family, and that they are a novel type of protease, as we initially proposed [1].

Very recently, another proteasome-related human protein, RING12 [11] and its mouse homologue, named LMP-2 [12], were cloned and both found to be located in the class II MHC region. Interestingly, the RING12 gene is linked to the RING10 gene [7,11]. Moreover, RC1 shows 34% and 31% similarities to these subunits RING12 and LMP-2, respectively, suggesting that these

proteins compose a sub-family of proteasome subunits, differing from other known subunits, such as RC2, RC3, RC5, RC8 and RC9.

3.5. *Possible functions of proteasomes*

In considering the functions of proteasomes, it is of interest that RING10, RING12 and LMP-2 are proteasome-related genes. However, as their similarities to most proteasome subunits are low, it is uncertain whether these genes actually encode proteasomal subunits. In the present study, we demonstrated that the rat proteasome subunit RC1 is a homologue of human RING10. Thus, it is now clear that two genes encoding proteasome subunits are actually located in the MHC class II region, suggesting possible involvement of proteasomes in the antigen processing pathway. This possibility is supported by the findings that certain subunit genes of proteasomes including LMP-2 show polymorphism [13,14] and are strongly induced by interferon-γ [7,12,14], because genes showing polymorphism and inducibility by interferon-γ are known to be important for the functions of MHC-linked genes. Moreover, these proteasome-related genes were found to be located between two putative peptide transporter loci in the class II region of the human MHC [8]. Thus, proteasomes have been proposed to be involved in processing in the cytosolic pathway for antigen presentation. That is, intracellular antigens are proposed to be processed by the proteasome and the peptides generated to be transported into the endoplasmic reticulum through two putative transporters to be assembled with class I MHC molecules for presentation on the cell surface [8]. However, there is no direct evidence that proteasomes are actually involved in the antigen processing pathway. As proteasomes have multicatalytic proteolytic functions [1,15], they are assumed to have the potential to generate antigenic peptides from various types of proteins. But it is still unknown whether the proteasome can produce peptides from known intracellular antigens in vitro or if so, whether these peptides are identical with those associated with the MHC class I molecule in the cell surface. Genetic studies are required for the clarification of the role of proteasomes in the pathway of antigen presentation. The 20S proteasome is known to associate with many other proteins to form a large multi-component proteolytic complex with an apparent sedimentation coefficient of 26S that catalyzes ATP-dependent selective breakdown of ubiquitin-ligated proteins [1]. Therefore, it will also be interesting to examine whether proteasome subunits encoded in the MHC region, such as RC1 (=RING10) and RING12 (=LMP-2) are assembled into the 20S proteasome or the 26S big complex, and whether the ubiquitin-mediated pathway is actually responsible for the processing of intracellular antigens.

Acknowledgements: This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan and the Sapporo Bioscience Foundation, Tokyo.

REFERENCES

[1] Tanaka, K., Tamura, T., Yoshimura, T. and Ichihara, A. (1992) The New Biologist, in press.
[2] 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.
[3] 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 29, 3777–3785.
[4] Tamura, T., Tanaka, K., Kumatori, A., Yamada, F., Tsurumi, C., Fujiwara, T., Ichihara, A., Tokunaga, F., Aruga, R. and Iwanaga, S. (1990) FEBS Lett. 264, 91–94.
[5] Tanaka, K., Kanayama, H., Tamura, T., Lee, D.H., Kumatori, A., Fujiwara, T., Ichihara, A., Tokunaga, F. and Iwanaga, S. (1990) Biochem. Biophys. Res. Commun. 171, 676–683.
[6] Kumatori, A., Tanaka, K., Tamura, T., Fujiwara, T., Ichihara, A., Tokunaga, F., Onikura, A. and Iwanaga, S. (1990) FEBS Lett. 264, 279–282.
[7] Glynne, R., Powis, S.H., Beck, S., Kelly, A., Kerr, L.-A. and Trowsdale, J. (1991) Nature 353, 357–360.
[8] Robertson, M. (1991) Nature 353, 300–301.
[9] Tanaka, K., Yoshimura, T., Ichihara, A., Iakai, A., Nishigai, M., Morimoto, M., Sato, M., Tanaka, N., Katsube, Y., Kameyama, K. and Takagi, T. (1988) J. Mol. Biol. 203, 985–996.
[10] Kozak, M. (1984) Nucleic Acids Res. 12, 857–872.
[11] Kelly, A., Powis, S.H., Glynne, R., Radley, E., Beck, S. and Trowsdale, J. (1991) Nature 353, 667–668.
[12] Martinez, C.K. and Monaco, J.J. (1991) Nature 353, 664–667.
[13] Brown, M.G., Driscoll, J. and Monaco, J.J. (1991) Nature 353, 355–367.
[14] Ortiz-Navarrete, V., Seeling, A., Gernold, M., Frentzel, S., Kloetzel, P.-M. and Hammerling, G.J. (1991) Nature 353, 662–664.
[15] Orlowski, M. (1990) Biochemistry 29, 10289–10297.