Volume 332, number 1,2, 52–56 FEBS 13071 October 1993

cDNA cloning of rat proteasome subunit RC7-I, a homologue of yeast PRE1 essential for chymotrypsin-like activity

Chihiro Nishimura[a], Tomohiro Tamura[b], Fuminori Tokunaga[c], Keiji Tanaka[b,*], Akira Ichihara[b]

[a]*Laboratory of Applied Toxicology, Nippon Kayaku Co, Ltd., 239 Iwahana-machi, Takasaki-city, Gunma 370-12, Japan*
[b]*Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan*
[c]*Himeji Institute of Technology, Faculty of Science, Harima Science Park City, Hyogo 678-12, Japan*

Received 5 August 1993

The nucleotide sequence of a cDNA that encodes a new subunit, named RC7-I, of the 20 S proteasome of rat hepatoma cells has been determined. The polypeptide predicted from the open reading frame consists of 201 amino acid residues with a calculated molecular weight of 22,912 and isoelectric point of 7.25. Approximately 80% of the partial amino acid sequences of several fragments of RC7-I, determined by protein chemical analyses, were found to be in excellent accordance with those deduced from the cDNA sequence. Computer analysis showed that RC7-I belongs to the β-type subgroup of proteasomes with similarity to the β-subunit of the archaebacterial proteasome, differing clearly from α-type subunits of the proteasome gene family. The overall structure of RC7-I was found to be homologous to that of yeast PRE1, which is necessary for chymotryptic activity.

Chymotrypsin-like activity; cDNA cloning; Multicatalytic proteinase; Proteasome; Subunit RC7-I

1. INTRODUCTION

The proteasome is a 20 S multi-molecular protease complex with a molecular weight of 750 kDa, consisting of 14–16 non-identical subunits of 21–32 kDa [1]. It is an essential component of the ATP-dependent 26 S proteolytic complex responsible for selective removal by extra-lysosomal proteolysis of abnormal and naturally occurring short-lived proteins [2]. All the genes of the 20 S proteasome examined so far were found to encode previously unidentified proteins that have been highly conserved during evolution. The proteasomal subunits, which have considerably high inter-subunit homology, can be classified into two subgroups, α and β, judging from their high similarities to the α- and β-subunit, respectively, of the archaebacterial proteasome [3]. 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 cDNAs for 10 subunits [4–10]. During structural analyses of the rat proteasome, we recently isolated a cDNA encoding a new subunit named RC7-I which belongs to be the β-type subgroup of the proteasome gene family. By computer analysis, the primary structure of the RC7-I was shown to have marked similarity to that of yeast PRE1, which is necessary for chymotrypsin-like activity of the yeast proteasome [11].

*Corresponding author.

2. MATERIALS AND METHODS

Subunit RC7-I was isolated from purified rat liver proteasomes, and its fragments were obtained by digestion with lysyl-endopeptidase by a similar method to that used for RC2 [4]. The amino acid sequences of the fragments were determined with a gas-phase sequencer (Applied Biosystems, model 477A), and 120A phenylthio-hydantoin analyzer on-line system [4]. A cDNA library of Reuber H4TG hepatoma cells was constructed in a phagemid expression vector, Bluescript II KS$^+$ (Stratagene) [5]. For isolation of cDNA for RC7-I, about 80,000 transformants were screened by hybridization with a cDNA fragment that had been synthesized by the polymerase chain reaction (PCR; for details, see text) and labeled with [α-^{32}P]dCTP as a probe. Colony hybridization was carried out by a reported method [4]. DNA sequencing was carried out by the dideoxy chain-termination method with an A.L.F. automatic DNA sequencer (Pharmacia LKB Biotechnology Inc.).

3. RESULTS AND DISCUSSION

3.1. *Protein sequencing of RC7-I*

Previously, we reported the separation of multiple components of proteasomes from rat liver by reversed-phase high-performance liquid chromatography [12]. 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. These components from rats were recently renamed RC1 to RC10 to distinguish them from those of other species. RC7 was eluted from the column with 55% acetonitrile. This RC7 fraction gave bands of two components of different sizes (I and II) on SDS-PAGE and these were separated by rechromatography on a TSK-GEL phenyl-5PW column. For determination of the primary

 Published by Elsevier Science Publishers B.V.

structure of its internal region, samples of RC7-I 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 and their partial primary structures were analyzed by automated Edman degradation (Table I).

3.2. *Isolation of a cDNA clone encoding RC7-I*

To isolate cDNA encoding RC7-I, we screened a cDNA library with the phagemid Bluescript II KS$^+$ vector using poly(A)$^+$ RNAs extracted from H4TG cells by hybridization with a cDNA fragment synthesized by PCR as a probe [4,5]. For use as PCR primers, we selected parts of the sequences of two of the proteolytic fragments RC7-I-42 and RC7-I-50 of RC7-I with DTVQFAEYIQK and AGYDEHEGPA, respectively (Table I). The following oligonucleotides corresponding to the protein sequences were synthesized:
Forward primer: 5′-GATACTGTTCAGTTTGCTGAGTATATTCAGAA-3′.
Reverse primer: 5′-GCAGGACCCTCATGCTCATCATAACCAGC-3′.

Using these primers, a fragment of approximately 190 bp was synthesized by PCR against first strand cDNA complementary to mRNA from fetal rats as a template. We then screened about 80,000 colonies of the H4TG cell cDNA library with the cDNA fragment. A single cDNA clone that gave a strongly positive signal with the probe was isolated from the library by colony hybridization techniques. The clone carried a cDNA insert of about 0.8 kb length including a poly(A) tail, and was subjected to cDNA sequencing.

3.3. *Primary structure of RC7-I*

The nucleotide sequence of the RC7-I cDNA clone and the primary structure of the RC7-I protein deduced from the cDNA sequence are shown in Fig. 1. The sequence of 792 nucleotides included the entire coding region and 5′- and 3′-non-coding regions. The 3′-non-coding region consisted of 131 nucleotides, excluding the poly(A) tail. A putative polyadenylation signal (AATAAA), which is common to eukaryotic mRNAs, was located 25 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 [13], and because the size of this open reading frame is very similar to that of the yeast homologue named PRE1 [11] (see Fig. 2). Subunit RC7-I corresponds to a protein of 201 amino acids with a calculated molecular weight of 22,912 Da. The amino acid sequence shown in Fig. 1 was confirmed to be that of RC7-I of proteasomes by showing that the partial sequences of 5 fragments determined chemically were in excellent accordance with those deduced from the nucleotide sequence of the cDNA (solid underlines, in Fig. 1). The isoelectric point (pI) of RC7-I was calculated to be 7.25 by the method of Skoog and Wichman [14].

3.4. *Inter-subunit homology of RC7-I in rat proteasomes*

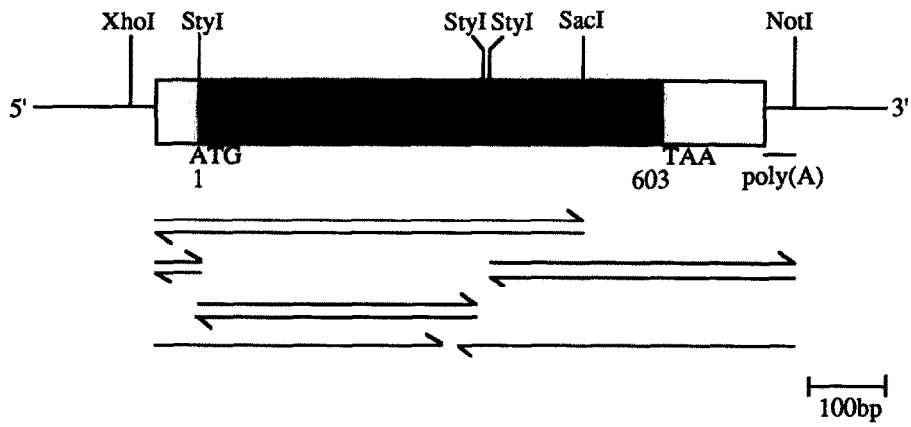
Proteasomal subunits, which have considerably high inter-subunit homology, can be classified into two subfamilies with high similarities to the α- and β-subunit, respectively, of the archaebacterial proteasome [3]. Table II shows computer-assisted homology analysis of the rat proteasomal subunits sequenced so far. Interestingly, members of the α-subunit family are highly homologous, whereas, except for RC1, rLMP2 and rDELTA, those of the β-subunit family show less similarity, suggesting that the latter subunit family may have diverged at an earlier stage during evolution. The α- and β-subgroups probably have distinct functions. Most α-type subunits contain a consensus signal sequence for nuclear translocation (NLS) and an NLS complementary sequence (cNLS), which consist of basic and acidic amino acid residues, respectively [1,15]. Thus one role of the α-subunit family may be in control of the intracellular distribution of the proteasome. In contrast, several subunits related with proteolytic functions belong to the β-subunit family, suggesting that most β-type subunits in the proteasome complex have catalytic functions [3]. For example, RN3 is a homologue of yeast PRE4, which is necessary for peptidylglutamylpeptide hydrolyzing activity [16]. Moreover, RC1 (rLMP7) and rLMP12 are suggested to be involved in

Table I

Automated amino acid sequence analysis of fragments in subunit RC7-I cleaved with lysyl-endopeptidase

No.	Fragment	
1	RC7-I-26	DGIHNLENITFTK
2	RC7-I-42	ILLLCVGEAGDTVQFAEYIQK
3	RC7-I-47	FILNLPTFSVRVIDK
4	RC7-I-49	APFAAHGYGAFLTLSILDRYYTPTISRERAVELLRK
5	RC7-I-50	MLNGYELSPTAAVNFTRRNLADPLRSRTPYGVNLLLAGYDEHEGPALYYMDXLAXL

S-Pyridylethylated RC7-I was digested with lysyl-endopeptidase 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 were determined with a protein sequencer. X denotes an unidentified residue. The amino acids underlined were not identical with those deduced from the nucleotide sequence of the cDNA encoding RC7-I.




```
5'-CGCGTACCCGGGAAGTCCCCGTGTGTGTGTACCTTGCAGCCTCCCGCTTCCCGGAACC   -58
ATGGAGTACCTCATTGGCATCCAGGGTCCTGACTATGTCCTGGTCGCTTCTGATCGGGTG    60
 M  E  Y  L  I  G  I  Q  G  P  D  Y  V  L  V  A  S  D  R  V     20

GCCGCCAGCAATATTGTCCAGATGAAGGACGATCATGACAAGATGTTTAAGATGAGTGAA   120
 A  A  S  N  I  V  Q  M  K  D  D  H  D  K  M  F  K  M  S  E     40

AAAATCTTACTCCTATGTGTTGGAGAGGCTGGAGACACTGTACAGTTTGCAGAATATATC   180
 K  I  L  L  L  C  V  G  E  A  G  D  T  V  Q  F  A  E  Y  I     60

CAGAAAAACGTGCAGCTTTATAAAATGCGGAATGGATATGAATTGTCCCCCACAGCAGCA   240
 Q  K  N  V  Q  L  Y  K  M  R  N  G  Y  E  L  S  P  T  A  A     80

GCTAATTTCACACGCCGAAACCTGGCCGACTGTCTTCGGAGTCGGACCCCTTATCATGTG   300
 A  N  F  T  R  R  N  L  A  D  C  L  R  S  R  T  P  Y  H  V    100

AACCTCCTCCTGGCTGGCTATGACGAGCATGAGGGGCCAGCACTTTACTACATGGACTAC   360
 N  L  L  L  A  G  Y  D  E  H  E  G  P  A  L  Y  Y  M  D  Y    120

CTAGCAGCCTTGGCCAAGGCTCCTTTTGCAGCTCATGGCTATGGTGCCTTCCTGACCCTC   420
 L  A  A  L  A  K  A  P  F  A  A  H  G  Y  G  A  F  L  T  L    140

AGCATCCTTGACCGATACTACACACCAACTATCTCACGTGAGAGGGCAGTGGAGCTTCTT   480
 S  I  L  D  R  Y  Y  T  P  T  I  S  R  E  R  A  V  E  L  L    160

AGGAAGTGTCTGGAGGAGCTCCAGAAGCGCTTCATCTTGAATCTGCCCACCTTCAGCGTT   540
 R  K  C  L  E  E  L  Q  K  R  F  I  L  N  L  P  T  F  S  V    180

CGGGTCATTGACAAAGATGGCATCCACAATCTGGAGAACATTACCTTCACTAAGCGGAGT   600
 R  V  I  D  K  D  G  I  H  N  L  E  N  I  T  F  T  K  R  S    200

TCCTAACATCATGTTCTTCCCTCCCACTTACCAGGGACCTTTTTTTGATGGGTTCCTTTT   660
 S  *                                                           201

ATTTTTTTCTACTCTTTTGACGTTCACTCTCAATAGATGGTTAATTCAAAATAAAGCTGA   720
ACACAGCTAAATTG-3'                                              734
```

Fig. 1. Structure of cDNA for subunit RC7-I of rat proteasomes. Upper panel: restriction endonuclease map of cloned cDNA for RC7-I 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, TAA. Sequenced regions are shown by horizontal arrows. The bar represents 100 bp. Lower panel: nucleotide sequence of the cDNA encoding component RC7-I 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. The nucleotides on the 5′ side are indicated by negative numbers. The predicted amino acid sequence of RC7-I is shown below the nucleotide sequence. Amino acid residues are numbered from the N-terminus. Continuous underlines show the amino acid sequences corresponding to those obtained by Edman degradation of fragments cleaved with lysyl-endopeptidase (see Table I). The amino acids shown by dotted lines were not identical with those found by chemical analysis. The termination codon TAA is marked with an asterisk. The possible polyadenylation signal (AATAAA) is underlined with a continuous thin line.

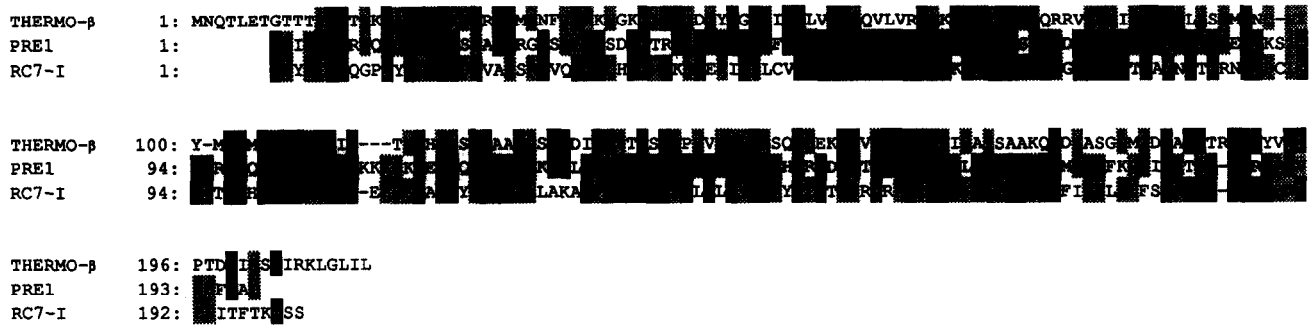


Fig. 2. Comparison of the protein sequences of RC7-I of rat liver proteasomes, yeast PRE1 and the β-subunit of the archaebacterial proteasome. Identical and conserved amino acid residues are shown by black and shaded boxes, respectively. Numbers are residue numbers. Conserved amino acids are defined as reported by Needleman and Wusch [20]: A, S, T, P and G; N, D, E and Q; R, K and H; I, V, L and M; F, Y and W.

antigen processing [2,17]. As shown in Table II, the primary structure of RC7-I does not show high homology to those of other subunits, suggesting that it is a somewhat unique subunit in the proteasomal complex. However, this subunit is a homologue of yeast PRE1, which is necessary for chymotryptic activity (see next section). Thus, the divergence of β-type subunits including RC7-I may have been associated with the acquirment of specific functions of proteasomes.

3.5. *Similarity of the RC7-I subunit to yeast PRE1 and the β-subunit of* T. acidophilum *and its possible functions*

Computer analysis showed no obvious overall structural similarity of component RC7-I with most previously reported proteins. Interestingly, however, the primary structure of RC7-I showed remarkable similarity with that of the product of the yeast PRE1 gene [11]. In addition, the protein sequence of RC7-I has considerable similarity to that of the β-subunit of the archaebacterial proteasome [3]. Fig. 2 shows its alignments with the yeast PRE1 and the β-subunit of *T. acidophilum*. Similarities are observed throughout almost their entire sequences, suggesting functional similarity. The β-subunit of the archaebacterial proteasome is responsible for the expression of chymotrypsin-like activity, because the archaebacterial proteasome has only chymotryptic activity unlike eukaryotic proteasomes, which have multiple proteolytic activities [3]. On the other hand, the PRE1 gene was separated by functional complementation of mutant yeast cells lacking the chymotrypsin-like activity of the yeast proteasome [11]. In the mutant cells, protein degradation was decreased, resulting in accumulation of ubiquitinated proteins [11] and loss of activity for protein degradation mediated by the N-end rule pathway [18,19], indicating the importance of this activity in protein breakdown in yeast cells. As shown in Fig. 2, the entire

Table II

Similarities of primary structures of various rat proteasome subunits

	RC2	RC3	RC8	RC9	rIOTA	rZETA	RC5	RC1	rLPM2	rDELTA	RN3	RC7-I
α-Type subunits												
RC2	1279											
RC3	336	1148										
RC8	262	299	1238									
RC9	263	405	298	1271								
rIOTA	292	355	336	355	1221							
rZETA	333	362	280	338	317	1115						
β-Type subunits												
RC5	56	61	33	90	85	50	1169					
RC1	40	139	38	98	ns	102	103	1056				
rLMP2	44	36	42	93	60	95	101	249	1047			
rDELTA	ns	ns	54	115	50	85	83	256	659	977		
RN3	36	46	39	38	35	40	76	91	83	59	1150	
RC7-I	72	62	100	76	68	64	157	106	93	102	120	1004

The sources of sequence data are as follows: RC1 [10]; RC2 [4]; RC3 [5]; RC5 [6]; RC8 [7]; RC9 [8]; rIOTA, rDELTA, rZETA and rLMP12 [9]; RN3 [15]. The sequences of the subunits were aligned to achieve maximal homology, and scores of the resulting pairs were determined by computer analysis according to Dayhoff et al. [21]. The extent of similarity between pairs of subunits was deduced by comparing the scores of the pairs with those of the two subunits. ns indicates no significant homology (Dayhoff' score <28). High scores (>200) for similarities of β-type subunits are underlined.

amino acid sequences of RC7-I and these two proteins show considerable similarity. Therefore, the subunit RC7-I may be involved in expression of the chymotrypsin-like activity in the rat proteasome complex.

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

REFERENCES

[1] Tanaka, K., Tamura, T., Yoshimura, T. and Ichihara, A. (1992) The New Biologist 4, 173–187.
[2] Goldberg, A. and Rock, K.L. (1992) Nature 357, 375–379.
[3] Zwickl, P., Grziwa, G., Puhler, G., Dahlmann, B., Lottspeich, F. and Baumeister, W. (1992) Biochemistry 31, 964–972.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] Kumatori, A., Tanaka, K., Tamura, T., Fujiwara, T., Ichihara, A., Tokunaga, F., Onikura, A. and Iwanaga, S. (1990) FEBS Lett. 264, 279–282.
[9] Tamura, T., Shimbara, N., Aki, M., Ishida, N., Bey, F., Scherrer, K., Tanaka, K. and Ichihara, A. (1992) J. Biochem. 112, 530–534.
[10] Aki, M., Tamura, T., Tokunaga, F., Iwanaga, S., Kawamura, Y, Shimbara, N., Kagawa, S., Tanaka, K. and Ichihara, A. (1992) FEBS Lett. 301, 65–68.
[11] Heinemeyer, V., Kleinschmidt, J.A., Saidowsky, J., Escher, C. and Wolf, D.H. (1991) EMBO J. 10, 555–562.
[12] 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.
[13] Kozak, M. (1984) Nucleic Acid Res. 12, 857–872.
[14] Skoog, B. and Wichman, A. (1986) Trends Anal. Chem. 5, 82–83.
[15] Tanaka, K., Yoshimura, T., Tamura, T., Fujiwara, T., Kumatori, A. and Ichihara, A. (1990) FEBS Lett. 271, 41–46.
[16] Thompson, S., Balson, D.F. and Rivett, A.J. (1993) FEBS Lett. 322, 135–138.
[17] Monaco, J.J. (1992) Immunol. Today 13, 173–179.
[18] Seufert, W. and Jentsch, S. (1992) EMBO J. 8, 3077–3080.
[19] Richter-Ruoff, B., Heinemeyer, W. and Wolf, D.H. (1992) FEBS Lett. 302, 192–196.
[20] Needleman S.B. and Wusch C.D. (1970) J. Mol. Biol. 48, 443–453.
[21] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O., Ed.) Vol. 5, Suppl. 3, National Biomedical Research Foundation, Silver Spring, MD.