Russell, D. W., Smith, M., Cox, D., Williamson, V. M., & Young, E. W. (1983) *Nature 304*, 652-654.
Sakata, Y., & Aoki, N. (1980) *J. Clin. Invest. 65*, 290-297.
Sakata, Y., & Aoki, N. (1982) *J. Clin. Invest. 69*, 536-542.
Schwartz, M. L., Pizzo, S. V., Hill, R. L., & McKee, P. A. (1971) *J. Biol. Chem. 246*, 5851-5854.
Schwartz, M. L., Pizzo, S. V., Hill, R. L., & McKee, P. A. (1973) *J. Biol. Chem. 248*, 1395-1407.
Seelig, G. F., & Folk, J. E. (1980) *J. Biol. Chem. 255*, 9589-9593.
Seydoux, F., Malhotra, O. P., & Bernhard, S. A. (1974) *CRC Crit. Rev. Biochem. 2*, 227-257.
Shafer, J. A., & Higgins, D. L. (1988) *CRC Crit. Rev. Clin. Lab. Sci. 26*, 1-41.
Takagi, T., & Doolittle, R. F. (1974) *Biochemistry 13*, 750-756.
Takahashi, N., Takahashi, Y., & Putnam, F. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8019-8023.
Tamaki, T., & Aoki, N. (1981) *Biochim. Biophys. Acta 661*, 280-286.

Molecular Cloning of cDNA for Proteasomes (Multicatalytic Proteinase Complexes) from Rat Liver: Primary Structure of the Largest Component (C2)†

Tsutomu Fujiwara,‡ Keiji Tanaka,*,§ Atsushi Kumatori,§ Sadahito Shin,‡ Tetsuro Yoshimura,§ Akira Ichihara,§ Fuminori Tokunaga,∥ Rie Aruga,⊥ Sadaaki Iwanaga,⊥ Akira Kakizuka,# and Shigetada Nakanishi#

Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan, Otsuka Pharmaceutical Company Ltd., Tokushima 770, Japan, Department of Molecular Biology, Graduate School of Medical Science, and Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan, and Institute for Immunology, Kyoto University Faculty of Medicine, Kyoto 606, Japan

Received February 24, 1989; Revised Manuscript Received May 12, 1989

ABSTRACT: Proteasomes (multicatalytic proteinase complexes) from rat liver are composed of at least 13 nonidentical components [Tanaka, K., Yoshimura, T., Ichihara, A., Ikai, A., Nishigai, M., Morimoto, M., Sato, M., Tanaka, N., Katsube, Y., Kameyama, K., & Takagi, T. (1988) *J. Mol. Biol. 203*, 985-996]. The nucleotide sequence of one major component (C2) of the proteasomes has been determined from a recombinant cDNA clone isolated by screening a rat liver cDNA library with a mixture of synthetic deoxyribonucleotides as a probe. The sequence was composed of 1174 nucleotides including a coding region for the entire protein and noncoding regions of both the 5′- and 3′-sides. The polypeptide deduced from the open reading frame consisted of 263 amino acid residues, and its molecular weight was calculated to be 29 516. The partial amino acid sequences of several fragments (approximately 45% of the total residues), which were obtained by cleavage of C2 with lysyl endopeptidase and cyanogen bromide, were determined by automated Edman degradation and found to be in complete accordance with those deduced from the cDNA sequence. The amino acid composition of C2, determined by chemical analysis, was also consistent with that deduced from the cDNA sequence, indicating that the cloned cDNA actually encoded component C2. Computer analysis revealed little structural similarity of C2 to other proteins reported so far. Northern blot hybridization analyses showed that the mRNA encoding this novel protein C2 was expressesd in all the rat tissues examined and in a variety of eukaryotic organisms such as amphibia, birds, and mammals with slight species-specific differences in size. This finding suggests that the gene encoding proteasomes has been conserved in eukaryotes during evolution.

During investigations on the intracellular proteolytic system, very large proteases have been found in mammalian cells [for recent review, see Rivett (1989)]. This enzyme was recently named "multicatalytic proteinase", because of its unique catalytic properties, that is, cleavage of peptide bonds on the carboxyl side of basic, hydrophobic, or acidic amino acid residues (Wilk & Orlowski, 1983; Dahlmann et al., 1985; Tanaka et al., 1986a). These multiple proteolytic activities seem to be manifested at independent catalytic sites within the single large enzyme complex. We demonstrated that these multicatalytic proteinases are widely distributed in a variety of eukaryotic cells ranging from human to yeast cells (Tanaka et al., 1988a). Recently, these proteinases have been reported to be responsible for ATP-stimulated breakdown of intracellular proteins in mammalian cells (McGuire et al., 1988; Tanaka & Ichihara, 1988; Driscoll & Goldberg, 1989; Matthews et al., 1989). In addition, we showed that these enzymes are present in the nucleus as well as in the cytoplasm of various cells (Arrigo et al., 1988; Tanaka et al., 1989).

On the other hand, ring-shaped 19S-22S particles have been found in a wide variety of eukaryotic organisms and shown to have common structural characteristics, such as an unusually large size, a unique symmetrical shape, and a multisubunit structure (Martins de Sa et al., 1986, Arrigo et al.,

† This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan and the Foundation for Application of Enzymes, Osaka.
* To whom correspondence should be addressed.
‡ Otsuka Pharmaceutical Co. Ltd.
§ Institute for Enzyme Research, The University of Tokushima.
∥ Department of Molecular Biology, Graduate School of Medical Science, Kyushu University 33.
⊥ Department of Biology, Faculty of Science, Kyushu University 33.
Institute for Immunology, Kyoto University Faculty of Medicine.

0006-2960/89/0428-7332$01.50/0

1987). The functions of most of these particles are still unknown, but some particles have been reported to have functions such as in repression of mRNA translation (Martins de Sa et al., 1986; Kloetzel et al., 1987) and tRNA processing (Castano et al., 1986). In addition, some have been considered to be ribonucleoprotein particles associated with specific small RNA species (Martins de Sa et al., 1986; Arrigo et al., 1987) or with certain heat-shock proteins (Schuldt & Kloetzel, 1985; Arrigo et al., 1985). Thus, these 20S particles seem to have essential roles in a variety of cellular processes.

Interestingly, the size, shape, and subunit multiplicity of these 20S ring-shaped particles are remarkably similar to those of the multicatalytic proteinases mentioned above. These two types of molecules are also similar in being ubiquitously distributed in eukaryotes and in being colocalized in the nucleus and cytoplasm. Very recently, the identity of 19S–22S ring-shaped particles with mammalian or yeast multicatalytic proteinases was reported independently by us (Arrigo et al., 1988), Falkenburg et al. (1988), and Kleinschmidt et al. (1988). We named these 20S particles with multiple protease activities *proteasomes* (Arrigo et al., 1988; Tanaka et al., 1988a). We also showed that these enzymes are symmetrical ring-shaped particles with a sedimentation coefficient of approximately 20S and a molecular mass of 700–900 kDa and are multisubunit complexes consisting of at least 13 nonidentical polypeptides with molecular masses of 21–31 kDa and isoelectric points of 3–10 (Tanaka & Ichihara, 1988; Tanaka et al., 1988a,b). Judging from these protein structural analyses together with the gene analysis reported in this paper, proteasomes seem to be multicatalytic proteinase *complexes*. Interestingly, Haass and Kloetzel (1989) reported that the *Drosophila* proteasome undergoes change in its subunit pattern during development, indicating diversification of the subunit multiplicity in different states of the cells.

As proteasomes are large multisubunit complexes with several possible functions, the individual subunits may each have different functions, including as yet unknown functions. For clarification of the various functions of these 20S particles, it is essential to obtain information about the structure–function relationship of the individual subunits. But it is difficult to isolate individual subunits in an active state. Therefore, our strategy is to determine the primary structures of these subunits by recombinant DNA techniques. Studies on homology of their primary structures may provide information about their structure–function relationships, and genetic studies should throw light on their functions. In this paper, we report cloning of cDNA for the largest subunit of rat liver proteasomal particles and Northern blot hybridization analysis of its distribution in various tissues and eukaryotic cells. To our knowledge, this is the first report of the cloning of a subunit of a multicatalytic proteinase complex or 20S ring-shaped particle.

Experimental Procedures

Materials

The following reagents were used: lysyl endopeptidase from Wako Pure Chemical Industries; [α-^{32}P]dCTP (3000 Ci/mmol) and [γ-^{32}P]ATP (5000 Ci/mmol) from Amersham Corp.; *Escherichia coli* DNA polymerase I and *E. coli* DNA ligase from New England Biolabs; the Klenow fragment of *E. coli* DNA polymerase I from Boehringer Mannheim; avian myeloblatosis virus reverse transcriptase from Life Sciences Inc.; RNase H, T4 polynucleotide kinase, T4 DNA ligase, and bacterial alkaline phosphatase from Takara Shuzo; various restriction endonucleases from Takara Shuzo, Bethesda Research Laboratories, and New England Biolabs.

Methods

Purification of Proteasomes and Subunit C2. Proteasomes were purified to homogeneity from rat liver by conventional chromatographic techniques as described (Tanaka et al., 1986a, 1988a). Component C2 was isolated directly from the purified proteasomes by reversed-phase high-performance liquid chromatography (HPLC)[1] on a Cosmosil 5C_4-300 column (10 × 250 mm, Nacalai Tesque) developed with a gradient of acetonitrile in aqueous 0.05% trifluoroacetic acid (v/v) as described previously (Tanaka et al., 1988b). The preparation of C2 was more than 95% pure as judged by SDS–PAGE (Laemmli, 1970).

Lysyl Endopeptidase Digestion. The protein was reduced and S-pyridylethylated as described previously (Aketagawa et al., 1986). The S-pyridylethylated protein was digested with lysyl endopeptidase at 37 °C for 12 h in 50 mM Tris-HCl buffer (pH 8) containing 2 M urea at a molecular ratio of enzyme to substrate 1:40. The digest was separated by reversed-phase HPLC on a Chemcosorb 7-ODS-H column (2.1 × 150 mm, Chemco) developed with a linear gradient of acetonitrile containing 0.1% trifluoroacetic acid.

Cyanogen Bromide Cleavage. Cleavage at methionyl bonds was achieved by treatment with 10% cyanogen bromide (w/w) in 70% formic acid (v/v) at room temperature for 24 h. The reaction was terminated by lyophilization, and the peptides obtained were purified by reversed-phase HPLC on a Cosmosil 5C_4-300 column (4.6 × 150 mm).

Amino Acid Analysis. The protein and the peptides obtained by enzymatic or chemical cleavage were hydrolyzed in vapor of 6 M HCl at 110 °C for 20 h. The hydrolysates were analyzed by the phenylthiocarbamoyl method (Bidlingmeyer et al., 1984) with a PICO-TAG system (Waters, Millipore Corp.) according to the instruction manual.

Protein Sequencing. Amino acid sequences were analyzed with a gas-phase sequencer (Applied Biosystems, Model 477A), and phenylthiohydantoin derivatives were identified with an Applied Biosystems 120A phenylthiohydantoin analyzer on-line system.

Syntheses of Oligonucleotide Probes and Primers for DNA Sequencing. Oligonucleotide hybridization probes and primers for DNA sequencing were synthesized in an Applied Biosystems Model 380B DNA synthesizer. 5′-End labeling of a mixture of oligodeoxyribonucleotides with [γ-^{32}P]ATP and T4 polynucleotide kinase was carried out as described (Maxam & Gilbert, 1980).

Preparation of Poly(A)+ RNA. Total RNA was extracted from rat liver by the guanidinium thiocyanate method (Chirgwin et al., 1979), and poly(A)+ RNA was isolated by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972).

Construction of a cDNA Library. A rat liver cDNA library was constructed in a phagemid expression vector, Bluescript KS M13$^+$ (Stratagene). Double-stranded DNA complementary to rat liver poly(A)+ RNA was synthesized by the method of Gubler and Hoffman (1983), except that 5′-AGGCCATGGCGGCCGCTTTTTTTTTTTTTTT-3′ was used as a primer for synthesis of the first strand of cDNA (Takumi et al., 1988). This primer includes unique restriction endonuclease sites for *Not*I (5′-GCGGCCGC-3′) and *Sfi*I (5′-GGCCNNNNNGGCC-3′) at the 5′-side of oligo(dT)$_{15}$.

[1] Abbreviations: SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; CNBr, cyanogen bromide; MOPS, 3-(*N*-morpholino)propanesulfonic acid; SSC, 150 mM NaCl, 15 mM sodium citrate; SSPE, 0.15 M NaCl, 10 mM NaH_2PO_4 (pH 7.4), 1 mM EDTA.

The double-stranded DNA mixtures synthesized were treated with Klenow fragment to blunt their ends. The resulting DNA mixtures were digested with *Not*I. Excess linkers and *Not*I-digested linker fragments were removed by fractionation on a Sepharose CL-4B column (0.8 × 4 cm, Pharmacia LKB Biotechnology Inc.). The DNA mixtures were then ligated with vector Bluescript. The vector had been digested with *Not*I and *Eco*RV at the multicloning site and then dephosphorylated by treatment with bacterial alkaline phosphatase to block self-ligation. *E. coli* DH5α competent cells (Bethesda Research Laboratories) were transformed with the ligated DNA.

Screening of the cDNA Library by Oligonucleotide Hybridization. For isolation of cDNAs for component C2, about 50000 transformants were screened by hybridization with ^{32}P-labeled oligodeoxyribonucleotide probes synthesized as described above. Colony hybridization was carried out by the method of Hanahan and Meselson (1980). Briefly, transformants that had been replicated onto duplicate nylon filters were prehybridized at 37 °C for 6 h in a solution of 50 mM Tris-HCl (pH 7.5), 1 M NaCl, 1% SDS, and 200 μg/mL yeast RNA and hybridized overnight in the same solution containing 100 μg/mL sonicated salmon sperm DNA and labeled probes (approximately 10^7 cpm/mL), which had a specific activity of about 5×10^6 cpm/pmol. The filters were washed sequentially with 6× SSC containing 0.1% SDS at room temperature for 5 min, at 37 °C for 10 min, and then at 42 °C for 2 min. The filters were then exposed to X-ray film at −70 °C with an intensifying screen.

DNA Sequencing. DNA sequencing was carried out by the dideoxy chain-termination method (Sanger et al., 1977) with a 7-DEAZA sequencing kit from Takara Shuzo.

Northern Blot Analysis. Electrophoresis and RNA blotting were carried out as described by Lehrach et al. (1977) and Thomas (1983), respectively. Poly(A)+ RNA (10 μg) was denatured in a solution of 20 mM MOPS buffer (pH 7.0) containing 6% formaldehyde, 50% formamide, 5 mM sodium acetate, and 1 mM EDTA at 65 °C for 5 min and then electrophoresed on 1.2% agarose gel containing 6% formaldehyde. The separated RNA was transferred onto a Hybond-N nylon membrane (Amersham) and hybridized with a ^{32}P-labeled probe. The probe used was a *Hin*dIII–*Pvu*II fragment (about 1-kb length) of cDNA for C2 protein labeled by the multiprime DNA labeling method (Feinberg & Vogelstein, 1984). Before hybridization, the nylon membrane was prehybridized in 50% formamide, 5× SSPE, 200 μg/mL sonicated salmon sperm DNA, 0.1% SDS, and 5× Denhardt's solution at 42 °C for 6 h. Hybridization was performed at 42 °C for 14 h in the same solution containing 10% dextran sulfate and labeled cDNA (10^6 cpm/mL) which had a specific activity of about 10^9 cpm/μg. After hybridization, the nylon membrane was washed four times with 2× SSPE containing 0.1% SDS for 15 min each time at 42 °C and then with 1× SSPE containing 0.1% SDS for 30 min at 42 °C. The membrane was autoradiographed with Kodak XAR-5 film at −70 °C with an intensifying screen.

Results

Subunit Compositions of Proteasomes. Previously, we reported the separation of multiple components of proteasomes from rat liver by reversed-phase HPLC (Tanaka et al., 1988b). First, 10 major components were separated on a Cosmosil $5C_4$-300 column and were named component 1 (C1) to component 10 (C10) in order of their elution. The fractions of C6, C7, and C10 each contained two components of different sizes (I and II) that could be separated by rechromatography on a TSK gel phenyl-5PW RP column. Table I shows the molecular weights of these components determined by SDS–PAGE analysis. When these purified materials were subjected to automated Edman degradation to determine their N-terminal amino acids, no appreciable signals were obtained except in the case of component C5, suggesting that the N-termini of most of these components were blocked.

Table I: Multiple Components of Proteasomes from Rat Liver

component[a]	molecular weight[b]	component[a]	molecular weight[b]
C1	24800 ± 800[c]	C7-I	27100 ± 500
C2	30800 ± 1200	C7-II	23500 ± 700
C3	25800 ± 700	C8	28800 ± 500
C4	26400 ± 700	C9	28700 ± 700
C5	25100 ± 900	C10-I	27900 ± 800
C6-I	27800 ± 700	C10-II	25700 ± 800
C6-II	21500 ± 500		

[a] For nomenclature of components, see text. [b] Molecular weights were determined by SDS–PAGE. [c] Values are means ± SD for four experiments.

Component C2 is the largest of the subunits of proteasomes, and it was obtained as a homogeneous preparation in good yield by reversed-phase HPLC. Therefore, we first attempted to isolate cDNA for component C2 of proteasomes.

Isolation and Protein Sequencing of Fragments of Component C2. As the N-terminus of C2 was blocked, as described above, we attempted to obtain information on the primary structure of its internal region. Samples of C2 were reduced and S-pyridylethylated to protect cysteine residues, and part (1.73 nmol) was then digested with lysyl endopeptidase, and the rest (39.7 nmol) was cleaved with cyanogen bromide. The resulting fragments were resolved by reversed-phase HPLC on a Chemcosorb 7-ODS-H column and a Cosmosil $5C_4$-300 column, respectively. When the elution profile was monitored at 214 nm, one large component (C2-CN-A) was obtained from the CNBr-treated preparation (Figure 1A), and nine major components (named C2-A to C2-I in order to their elution) were separated from the lysyl endopeptidase digest (Figure 1B). The amino acid compositions of these isolated peptides were determined (data not shown), and their partial primary structures were analyzed by automated Edman degradation (Table II). Peptide C2-A was considered to be the C-terminal fragment of C2 because it did not contain a lysine residue (lysyl endopeptidase cleaves the carboxy side of lysine residues in a polypeptide chain). Indeed, its primary structure was consistent with that in the C-terminal region deduced from the nucleotide sequence of cDNA for C2 as described later. Peptide C2-G was concluded to be the N-terminal fragment, because its N-terminal amino acid was blocked. This conclusion was supported by the finding that the amino acid composition of the C2-G fragment corresponded to that of the 30 amino acids in the N-terminal region deduced from the sequence of the C2 cDNA (data not shown). Peptide C2-CN-A was also thought to be an N-terminal fragment of C2, because its primary structure was identical with that of the N-terminal region (24 amino acids) after the N-terminal methionine (cleaved by CNBr treatment) deduced from the nucleotide sequence of C2-cDNA.

Isolation of cDNA Clones Encoding Component C2. A rat cDNA library was constructed with the phagemid Bluescript vector using poly(A)+ RNAs extracted from rat liver as described under Experimental Procedures. For isolation of cDNA for component C2, this library was screened by hybridization with synthetic deoxyribonucleotides as probes. These probes were synthesized as follows. Of the C2 fragments obtained by proteolytic digestion (Figure 1B), parts of the sequences of two fragments (Ala-Asp-Glu-Pro-Met-Glu of

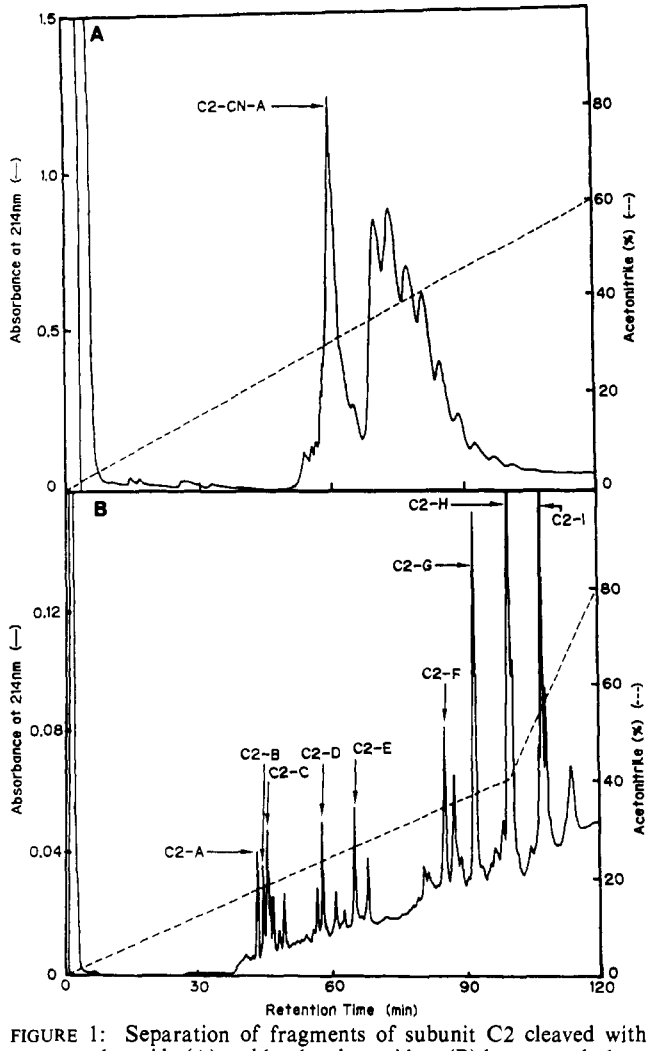

FIGURE 1: Separation of fragments of subunit C2 cleaved with cyanogen bromide (A) and lysyl endopeptidase (B) by reversed-phase HPLC. Fragments of C2 were resolved by HPLC on a Cosmosil $5C_4$-300 column (A) or a Chemcosorb 7-ODS-H column (B) with a linear gradient of acetonitrile as described under Experimental Procedures. The fragment obtained by cyanogen bromide cleavage was named C2-CN-A. The major proteolytic fragments were named C2-A to C2-I in order of their elution.

C2-A and Asp-Glu-Pro-Ala-Glu-Lys of C2-B in Table II) with minimum codon ambiguity were selected, and the oligonucleotide probes 5′-TCCATNGGYTCRTCRGC-3′ and 5′-TTYTCRGCRGGYTCRTC-3′ (N = AGTC, Y = TC, R = AG) were synthesized. These probes had a deletion of the third codon for the carboxyl-terminal amino acid of these peptides and were designed as antisense nucleotide sequences for these peptides according to the coding frequencies of eukaryotic proteins (Lathe, 1985). We screened about 50 000 colonies of a rat liver cDNA library with two sets of 32 mixtures of these 17-mer oligonucleotides. Three cDNA clones that gave strongly positive signals for both probes were isolated from the library by colony hybridization techniques. As the cDNAs of all three clones (pC2-α, -β, and -γ) gave similar cleavage maps with several restriction endonucleases, clone pC2-α [about 1.2-kb length plus a poly(A) tail], which carried the largest cDNA insert of the three, was subjected to cDNA sequence determination.

Primary Structure Deduced from the cDNA Sequence. The nucleotide sequence of the C2 cDNA clone pC2-α was determined by the strategy illustrated in Figure 2. The resulting nucleotide sequence of the cDNA and the primary structure of the C2 deduced from the cDNA sequence are shown in Figure 3. The sequence of 1174 nucleotides included the

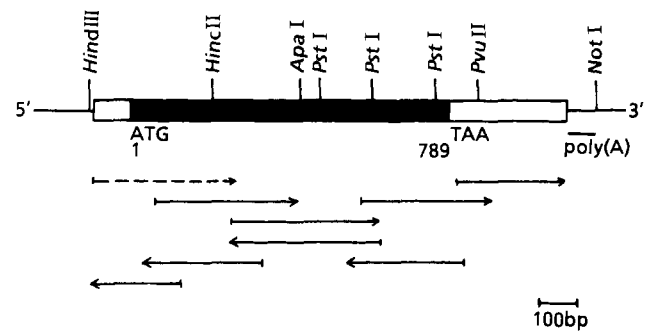

FIGURE 2: Restriction endonuclease map of cloned cDNA for subunit C2 and the sequencing strategy. The solid and open boxes show the coding and 5′- and 3′-noncoding regions, respectively. The solid lines indicate the sequence of the vector Bluescript KS^+. The numbers below the boxes indicate the nucleotide 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 (base pairs). The sequence of the 5′-end of cDNA (dashed arrow) was determined by direct sequencing of an isolated clone, pC2-α, with the primer of T3 promoter of Bluescript KS^+. Various oligonucleotides (17–20 nucleotides in length) were synthesized as described under Experimental Procedures and used as primers for sequencing other regions (solid arrows).

entire coding region and 5′- and 3′-noncoding regions. The 3′-noncoding region consisted of 297 nucleotides, excluding the poly(A) tail. A polyadenylation signal (AATAAA) that is common in eukaryotic mRNAs (Proudfoot & Brownlee, 1978; Kozak, 1984) was located 17 nucleotides upstream from the poly(A) addition site. Part of the poly(A) sequence (approximately 70 nucleotides) is seen at the 3′-end of this cDNA clone. The protein coding region is 789 nucleotides long, is capable of encoding 263 amino acids, and extends from ATG at nucleotide position 1–3 to nucleotide position 789. This coding sequence is followed by the termination codon TAA. Although the N-terminal amino acid sequence of C2 is unknown, we concluded that ATG, located at nucleotides 1–3, is the initiation codon for the following four reasons. First, the open reading frame starting from this ATG at nucleotide position 1–3 is the longest. Second, ^{1}ATG^3 is the first ATG appearing in the sequence. Third, there is a termination codon TGA at nucleotides −54 to −52 in the 5′-noncoding region, and ^{1}ATG^3 followed by this termination codon in the same frame should be an initiation codon. Fourth, a CNBr fragment (C2-CN-A) of 24 amino acids with the sequence from Phe to Ala (Table II) corresponds to nucleotide positions 4–75 (indicated by a broken line in Figure 3), indicating that the N-terminal amino acid of the C2 protein is Met, corresponding to nucleotides 1–3, and that its amino group is blocked.

Several lines of evidence indicated that the amino acid sequence shown in Figure 3 is actually that of C2 of proteasomes: (1) amino acid sequences determined by the Edman degradation method (Table II) were found in the amino acid sequence deduced from the nucleotide sequence (solid lines and broken line in Figure 3), (2) the amino acid composition (Table III) of the whole C2 molecule determined by chemical analysis coincided well with that deduced from the sequence shown in Figure 3, and (3) the molecular weight of C2 estimated from its amino acid composition was similar to that estimated by SDS–PAGE (Table I). Thus, it is reasonable to conclude tht pC2-α is a clone of subunit C2 cDNA.

Component C2 had one potential N-glycosylation site (^{209}Asn-Val-Ser^{211} in Figure 3) although it does not seem to be a glycoprotein, because no appreciable glucosamine or galactosamine was detected during amino acid analysis (data not shown). Of the amino acids in component C2, 35 are

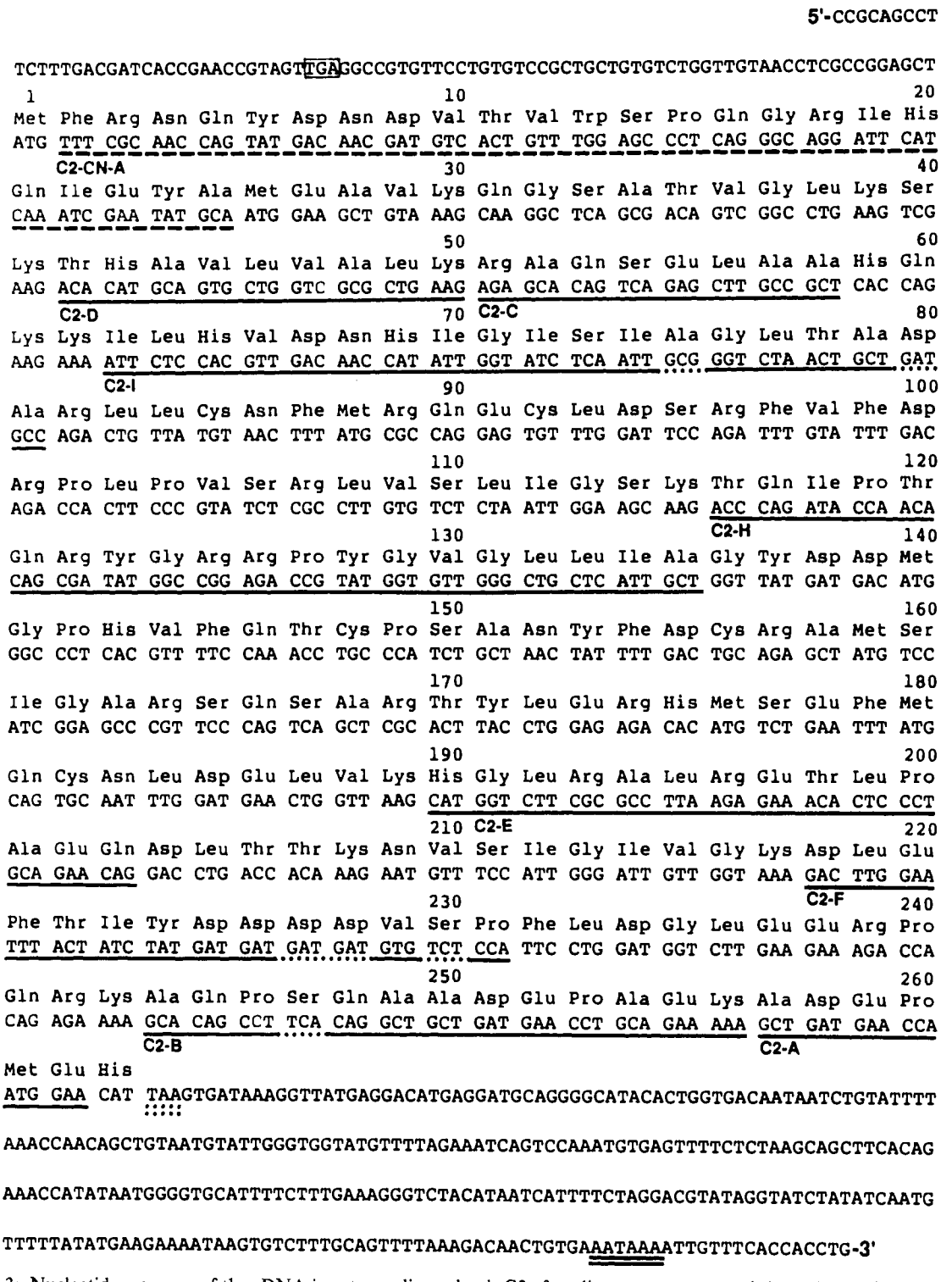

FIGURE 3: Nucleotide sequence of the cDNA insert encoding subunit C2 of rat liver proteasomes and the amino acid sequence deduced from its open reading frame. Nucleotides are numbered in the 5′- to 3′-direction, beginning with the first residue of the initiation methionine codon ATG. The nucleotides on the 5′-side are indicated by negative numbers. The predicted amino acid sequence of C2 is shown above the nucleotide sequence. Amino acid residues are numbered from the N-terminus. Solid (C2-A to C2-I) and broken (C2-CN-A) lines show the amino acid sequences corresponding to those obtained by Edman degradation of fragments cleaved with lysyl endopeptidase and CNBr (Table II), respectively. Amino acids shown by dotted lines in the sequences of C2-B, C2-F, and C2-I were not identified by chemical analysis. The termination codon TGA in the 5′-noncoding region is boxed. The termination codon TAA is doubly underlined with dotted lines. The possible polyadenylation signal AATAAA is doubly underlined.

acidic and 40 are basic amino acid residues. Interestingly, an odd number of five cysteine residues was found in component C2 (Table II), indicating the presence of at least one free sulfhydryl residue. Studies are required on whether this is important for the function of C2.

To obtain information about the structure–function relationship of C2, we searched for structural homologies of the nucleotide and protein sequences using the data bases of EMBL/GDB (European Molecular Biology Laboratory), NBRF/PDB (National Biomedical Research Foundation), and PRF/SEQDB (Protein Research Foundation). No proteins with overall similarity could be found by computer analyses, suggesting that component C2 is distinct from known proteins. However, some proteins with partially identical amino acid

Table II: Automated Amino Acid Sequence Analysis of Fragments Derived from Subunit C2 of Proteasomes Cleaved by Lysyl Endopeptidase and Cyanogen Bromide

	C2-A		C2-B		C2-C		C2-D		C2-E		C2-F		C2-H		C2-I		C2-CN-A	
cycle no.	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)	amino acid	yield (pmol)
1	Ala	348	Ala	nd[a]	Arg	319	Thr	132	His	173	Asp	nd	Thr	64	Ile	123	Phe	437
2	Asp	144	Gln	86	Ala	nd	His	60	Gly	136	Leu	144	Gln	74	Leu	135	Arg	47
3	Glu	95	Pro	195	Gln	52	Ala	98	Leu	213	Glu	112	Ile	94	His	70	Asn	177
4	Pro	109	X[b]		Ser	22	Val	44	Arg	539	Phe	137	Pro	57	Val	86	Gln	242
5	Met	117	Gln	69	Glu	29	Leu	42	Ala	140	Thr	69	Thr	24	Asp	60	Tyr	301
6	Glu	41	Ala	127	Leu	24	Val	22	Leu	130	Ile	126	Gln	34	Asn	47	Asp	244
7			Ala	nd	Ala	nd	Ala	27	Arg	nd	Tyr	86	Arg	75	His	30	Asn	244
8			Asp	57	Ala	nd	Leu	31	Glu	81	Asp	67	Tyr	31	Ile	61	Asp	337
9			Glu	55			Lys	6	Thr	38	Asp	nd	Gly	54	Gly	64	Val	298
10			Pro	20					Leu	63	X		Arg	89	Ile	63	Thr	97
11			Ala	nd					Pro	49	X		Arg	144	Ser	14	Val	258
12			Glu	27					Ala	42	Val	17	Pro	36	Ile	46	Trp	96
13			Lys	6					Glu	32	X		Tyr	27	X		Ser	53
14									Gln	22	Pro	nd	Gly	42	Gly	46	Pro	474
15													Val	18	Leu	34	Gln	108
16													Gly	41	Thr	9	Gly	95
17													Leu	32	Ala	24	Arg	41
18													Leu	43	X		Ile	72
19													Ile	21	Ala	26	His	27
20													Ala	19			Gln	69
21																	Ile	70
22																	Glu	67
23																	Tyr	59
24																	Ala	35
25																		
position	257–262		244–256		51–58		42–50		190–203		218–231		116–135		63–81		2–25	
repetitive	Glu-3–Glu-6		Gln-2–Gln-5				Ala-3–Ala-7		Leu-3–Leu-10				Gly-9–Gly-14		Ile-1–Ile-8		Gln-4–Gln-15	
yield (%)	76		93		nd		72		84		nd		95		90		93	

[a] Not determined quantitatively. [b] Not identified.

Table III: Amino Acid Composition of Subunit C2 of Proteasomes[a]

	amino acid composition (mol %)				amino acid composition (mol %)		
amino acid	chemical analysis	cDNA sequence analysis	no. of residues[b]	amino acid	chemical analysis	cDNA sequence analysis	no. of residues[b]
Gly	6.25	6.08	16	Glu		6.08	16
Ala	8.39	8.75	23	Gln		6.08	16
Val	6.36	6.08	16	Arg	7.78	7.22	19
Leu	9.88	9.13	24	Lys	4.80	4.56	12
Ile	4.76	4.94	13	His	3.16	3.42	9
Ser	5.72	6.46	17	Phe	3.48	3.42	9
Thr	4.44	4.56	12	Tyr	3.16	3.04	8
Cys	1.53[c]	1.90	5	Trp	0.43[d]	0.38	1
Met	2.77	3.04	8	Pro	4.90	4.94	13
total Asx	10.20	9.88					
Asp		7.22	19	total	100	100	263
Asn		2.66	7				
total Glx	11.98	12.16			molecular weight: 29 516		

[a] The amino acid composition was determined by analysis of a hydrolysate of purified subunit C2 or deduced from the sequence of nucleotides in a recombinant cDNA. [b] Predicted from the nucleotide sequence. [c] Determined as S-pyridylethylated cysteine. [d] Determined after hydrolysis for 20 h with 3 M mercaptoethanesulfonic acid.

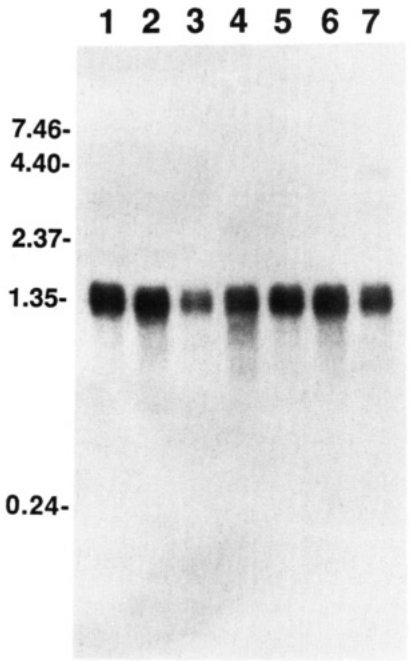

FIGURE 4: Expression of C2-specific mRNA in various tissues of rats determined by Northern blot analysis. Poly(A)+ RNAs (10 μg) from rat liver (lane 1), kidney (lane 2), skeletal muscle (lane 3), heart (lane 4), lung (lane 5), spleen (lane 6), and brain (lane 7) were analyzed as described under Experimental Procedures. The *Hin*dIII–*Pvu*II fragment of cloned C2 cDNA was used as a probe (see Figure 2). The positions of RNA standards are shown (in kilobases) on the left. These RNA size markers were obtained from Bethesda Research Laboratories.

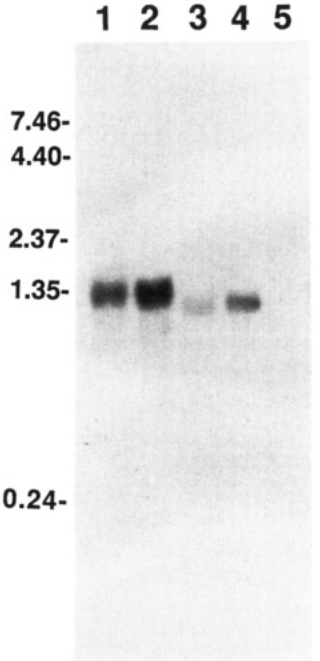

FIGURE 5: RNA blot hybridization with rat liver C2 cDNA of the poly(A)+ RNAs from various eukaryotic cells. Experimental conditions were as described for Figure 4, except that poly(A)+ RNAs from human liver (lane 1), rat liver (lane 2), chicken liver (lane 3), *Xenopus laevis* ovary (lane 4), and bakers' yeast (*Saccharomyces cerevisiae*) (lane 5) were used.

sequences were found. One of these is a large subunit of *chicken* calcium-dependent protease (calpain). A sequence of 48 amino acid residues in component C2 (residues 27–74) is apparently similar to part of this calcium-dependent protease (residues 195–244) (Ohno et al., 1984), the identity and structural analogy of amino acids in the sequences being 33% and 42%, respectively. Moreover, the homology of these sequences of the two proteins was shown to be highly significant by statistical analysis (Toh et al., 1983). In contrast, no significant homology was found between proteasome C2 and *human* calcium-dependent protease (Aoki et al., 1986). It is thus obscure whether this region of C2 is functionally important. Indeed, the corresponding sequence does not contain the active-site Cys-108 and His-265 residues of the calcium-dependent protease, although it is located in the thiol protease domain.

Expression of the Gene for Component C2 in Various Eukaryotic Organisms. To determine the tissue specificity of C2 gene expression, we examined the levels of C2-specific mRNAs in various rat tissues. On Northern blot analysis, one hybridization band of about 1.3–1.4 kb was observed in mRNAs extracted from various rat tissues (Figure 4). The levels of mRNA for component C2 were very similar in all tissues tested except skeletal muscle.

We next examined whether the C2 mRNA is expressed in other eukaryotic cells. As shown in Figure 5, a single band hybridizing to rat C2 cDNA was observed in various tissues of different species, such as human liver, chicken liver, and *Xenopus laevis* ovary, but was not found in bakers' yeast under the present experimental conditions. The size of the rat RNA hybridizing with the rat C2 cDNA was similar to that of human RNA, but somewhat different from those of chicken and *Xenopus* RNAs. These findings indicate that the mRNA encoding component C2 is widely expressed in a variety of eukaryotes including mammals, birds, and amphibia, but with minor species-specific variations in mRNA structure.

Discussion

In the present study, cDNA for component C2 of proteasomes (multicatalytic proteinase complexes) from rat liver was isolated and sequenced. This is the first report of the cloning of cDNA from eukaryotic proteasomes. The cloned cDNA was 1174 nucleotides long and included an open reading frame for the full-length coding region of C2 and 5′- and 3′-noncoding regions. The 3′-noncoding region contained a putative polyadenylation signal. Approximately 45% of the total amino acid sequence of C2 was determined by Edman degradation of fragments and found to be in complete accordance with that deduced from the cDNA sequence. The amino acid composition of component C2 determined by chemical analysis was also consistent with that deduced from the cDNA sequence. On computer analyses, no structural homology of component C2 with previously reported proteins could be found, and so no information could be obtained about the function of C2. These findings suggest that component C2 must be a novel protein, possibly with a new, unidentified function(s).

Previously, we found that proteasomes are composed of a set of polypeptide subunits ranging in size from 21 to 31 kDa with p*I* values from 3 to 10 (Tanaka et al., 1988a,b). Peptide mapping analysis showed that these multiple subunits were nonidentical (Tanaka et al., 1988b). However, the possibility that they were generated from larger polypeptides by post-transcriptional cleavage could not be ruled out. In the present study, however, we demonstrated that C2 is the unmodified product of a single gene, by the findings that its N- and C-terminal sequences, determined by Edman degradation, were coded in a single open reading frame of isolated C2 cDNA as an entire structure (Figure 3). This finding proves that component C2 is not produced from a precursor form. Thus, the multiple subunits of proteasomes are likely each to be coded by a different gene. To confirm this more directly, we are now determining the primary structures of other compo-

nents by their cDNA cloning.

By quantitative enzyme immunoassay, we showed that proteasomes are widely distributed in various rat cells and tissues (Tanaka et al., 1986a). Consistent with this finding, we showed by Northern blot analysis that mRNA encoding C2 is expressed in various rat tissues (Figure 4).

Recently, we demonstrated that proteasomes with similar proteolytic functions are present in a great variety of eukaryotic cells, such as human, rat, chicken, *Xenopus*, and yeast cells (Tanaka et al., 1988a). These proteasomes, which are similar to various eukaryotic 20S ring-shaped particles in size and shape (Martins de Sa et al., 1986; Arrigo et al., 1987), are complexes of 13–20 subunits. However, they showed species-specific differences in subunit multiplicity,and they differed immunologically, as shown by Ouchterlony immunodiffusion tests and immunoblotting analyses, although cross-immunoreactivities of some subunits or domains were observed. These results indicate that the sizes and shapes of these proteasomes have been highly conserved during evolution but that they show species-specific differences in conformation and subunit structure (Tanaka et al., 1988a). More recently, we isolated four monoclonal antibodies against component C2 of human proteasomes (Tanaka et al., unpublished data). These monoclonal antibodies did not show any cross-reactivity with other subunits, and all had different epitopes, while on immunoblotting analyses they strongly cross-reacted with proteasomes from human, rat, and chicken but not with those from *Xenopus* or yeast, suggesting that the primary structure of C2 has changed during evolution. Consistent with these observations, as shown by Northern blot analysis (Figure 5), most eukaryotic cells contained mRNAs that hybridized with rat liver C2 cDNA, but the sizes of these mRNAs in various species and their hybridization reactivities differed. In addition, no RNA species that hybridized with rat C2 cDNA was found in bakers' yeast (Figure 5, lane 5). It is unknown whether this finding indicates that the gene corresponding to component C2 of rats is not present in the yeast genome or that although a gene related to C2 is present in yeast, it has changed so greatly that its mRNA could not hybridize with rat C2 cDNA under the present experimental conditions. A species-specific difference in size of mRNA may also not actually reflect a difference in primary structure of the C2 protein, because the difference is likely to be due to a difference in the noncoding region of the mRNA. Judging from the present results, therefore, it seems that the gene encoding component C2 has probably been highly conserved during evolution, even though it shows minor species-specific variations in structure. For further elucidation of its conservation, the nucleotide sequences of C2 genes and/or the primary structures of C2 proteins in other species must be determined. Therefore, we are now attempting to isolate cDNA that hybridizes with rat liver C2 cDNA or an equivalent gene from other organisms such as humans and *Xenopus*.

In conclusion, in this work we isolated and sequenced a cDNA encoding component C2 of proteasomes from rat liver. The isolated C2 cDNA will be useful in further studies on proteasomes such as for elucidation of the mechanisms of coordinated regulation of the syntheses of these multiple components, possibly for measurement of its mRNA level in cells or the transcriptional activity of genes encoding proteasomes, and for isolation of the C2 gene from other species, which is of particular interest in elucidating the significance of conservation of proteasomes (multicatalytic proteinase complexes) during evolution of a wide variety of eukaryotes ranging from yeast to man.

Acknowledgments

We are grateful to Drs. Takashi Miyata and Keiichi Kuma of the Department of Biology, Faculty of Science, Kyushu University (Fukuoka 812, Japan), for statistical analysis of homologies.

References

Aketagawa, J., Miyata, T., Ohtsubo, S., Nakamura, T., Morita, T., Hayashida, H., Miyata, T., Iwanaga, S., Takao, T., & Shimonishi, Y. (1986) *J. Biol. Chem. 261*, 7357–7365.

Aoki, K., Imajoh, S., Ohno, S., Emori, Y., Koike, M., Kosaki, G., & Suzuki, K. (1986) *FEBS Lett. 205*, 313–317.

Arrigo, A.-P., Darlix, J.-L., Khandjian, E. W., Simon, M., & Spahr, P.-F. (1985) *EMBO J. 4*, 399–406.

Arrigo, A.-P., Simon, M., Darlix, J.-L., & Spahr, P.-F. (1987) *J. Mol. Evol. 25*, 141–150.

Arrigo, A.-P., Tanaka, K., Goldberg, A. L., & Welch, W. J. (1988) *Nature 331*, 192–194.

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93–104.

Castano, J. G., Ornberg, R., Koster, J. G., Tobian, J. A., & Zasloff, M. (1986) *Cell 46*, 377–387.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry 18*, 5294–5299.

Dahlmann, B., Rutschmann, M., Kuehn, L., & Reinauer, H. (1985) *Biochem. J. 228*, 171–177.

Driscoll, J., & Goldberg, A. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 787–791.

Falkenburg, P.-E., Haass, C., Kloetzel, P.-M., Niedel, B., Kopp, F., Kuehn, L., & Dahlmann, B. (1988) *Nature 331*, 190–192.

Feinberg, A. P., & Vogelstein, B. A. (1984) *Anal. Biochem. 137*, 266–267.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–268.

Haass, C., & Kloetzel, P.-M. (1989) *Exp. Cell Res. 180*, 243–252.

Hanahan, D., & Meselson, M. (1980) *Gene 10*, 63–67.

Kleinschmidt, J. A., Escher, C., & Wolf, D. H. (1988) *FEBS Lett. 239*, 35–42.

Kloetzel, P.-M., Falkenburg, P.-E., Hossl, P., & Glatzer, K. H. (1987) *Exp. Cell Res. 170*, 204–213.

Kozak, M. (1984) *Nucleic Acids Res. 12*, 8857–8872.

Laemmli, U. K. (1970) *Nature 227*, 680–685.

Lathe, R. (1985) *J. Mol. Biol. 183*, 1–12.

Lahrach, H. D., Diamond, J. M., Wozner, M., & Boedtker, H. (1977) *Biochemistry 16*, 4743–4751.

Martins de Sa, C., Grossi de Sa, M.-F., Akhayat, O., Broders, F., & Scherrer, K. (1986) *J. Mol. Biol. 187*, 479–493.

Matthews, W., Tanaka, K., Driscoll, J., Ichihara, A., & Goldberg, A. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2597–2601.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

McGuire, M. J., Croll, D. E., & DeMartino, G. N. (1988) *Arch. Biochim. Biophys. 262*, 273–285.

Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M., & Suzuki, K. (1984) *Nature 312*, 566–570.

Proudfoot, N. J., & Brownlee, G. G. (1978) *Nature 263*, 211–214.

Rivett, A. J. (1989) *Arch. Biochem. Biophys. 268*, 1–8.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Schuldt, C., & Kloetzel, P.-M. (1985) *Dev. Biol. 110*, 65–74.

Takumi, T., Ohkubo, H., & Nakanishi, S. (1988) *Science 242*, 1042-1045.
Tanaka, K., & Ichihara, A. (1988) *FEBS Lett. 236*, 159-162.
Tanaka, K., Ii, K., Ichihara, A., Waxman, L., & Goldberg, A. L. (1986a) *J. Biol. Chem. 261*, 15197-15203.
Tanaka, K., Yoshimura, T., Ichihara, A., Kameyama, K., & Takagi, T. (1986b) *J. Biol. Chem. 261*, 15204-15207.
Tanaka, K., Yoshimura, T., Kumatori, A., Ichihara, A., Ikai, A., Nishigai, M., Kameyama, K., & Takagi, T. (1988a) *J. Biol. Chem. 263*, 16209-16217.
Tanaka, K., Yoshimura, T., Ichihara, A., Ikai, A., Nishigai, M., Morimoto, M., Sato, M., Tanaka, N., Katsube, Y., Kameyama, K., & Takagi, T. (1988b) *J. Mol. Biol. 203*, 985-996.
Tanaka, K., Kumatori, A., Ii, K., & Ichihara, A., (1989) *J. Cell. Physiol. 139*, 34-41.
Thomas, P. S. (1983) *Methods Enzymol. 100*, 255-267.
Toh, H., Hayashida, H., & Miyata, T. (1983) *Nature 305*, 827-829.
Wilk, S., & Orlowski, M. (1983) *J. Neurochem. 40*, 842-849.

cDNA Cloning and Sequence and cDNA-Directed Expression of Human P450 IIB1: Identification of a Normal and Two Variant cDNAs Derived from the *CYP2B* Locus on Chromosome 19 and Differential Expression of the IIB mRNAs in Human Liver†

Shigeru Yamano,‡,§ Patson T. Nhamburo,‡ Toshifumi Aoyama,‡ Urs A. Meyer,‖ Tadanobu Inaba,⊥ Werner Kalow,⊥ Harry V. Gelboin,‡ O. Wesley McBride,# and Frank J. Gonzalez*,‡

Laboratory of Molecular Carcinogenesis and Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Department of Pharmacology, Biocenter, University of Basel, CH-4056 Basel, Switzerland, and Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Received January 17, 1989; Revised Manuscript Received April 17, 1989

ABSTRACT: A cDNA designated hIIB1, representing the entire coding sequence of a P450 in the IIB gene family, was isolated from a human liver λgt11 library by using the rat IIB1 cDNA as a probe. The hIIB1 protein, deduced from the cDNA sequence, contained 491 amino acids, had a calculated molecular weight of 56 286, and displayed 76% amino acid similarity with the rat IIB1 protein. Expression of this cDNA, using the vaccinia virus system, yielded a P450 that had a reduced CO-binding spectrum with an absorption maximum of 452 nm. The expressed human enzyme was able to catalyze the deethylation of 7-ethoxycoumarin. Total RNA from 13 livers was probed for levels of hIIB mRNA. Two livers had high levels, four contained moderate levels, and eight contained very low, or no detectable, mRNA. These data suggest either that defective hIIB1 genes exist in humans or that the hIIB1 gene is regulated and variably induced in our liver specimens. To search for mutant mRNA transcripts, libraries were constructed from livers expressing low levels of hIIB1 mRNA. A cDNA, designated hIIB2, was isolated that was identical with the hIIB1 cDNA except for the presence of an unusual alteration of the DNA near its 5′ end corresponding to the putative exon 4. This alteration was caused by a deletion of 29 bp and an insertion of 44 bp of nonhomologous DNA. This sequence replacement occurs at the junction of the third and fourth exons as predicted from the structure of the rat IIB1 gene, suggesting that a faulty splice might have given rise to the variant hIIb2 transcript. Due to the presence of an in-frame termination codon in the inserted DNA, this variant transcript can only produce a prematurely terminated protein. A third cDNA, designated hIIB3, was identified in two separate libraries that displayed 95% nucleotide and 93% cDNA-deduced amino acid sequence similarities to hIIB1. This transcript was found to possess a C → T change that resulted in a termination codon. The IIB genes (*CYP2B* locus) were localized to human chromosome 19 using the somatic cell hybrid mapping strategy. High-frequency restriction fragment length polymorphisms were detected in both *Bam*HI and *Bgl*II digests.

Hepatic P450s are the principal enzymes required for the metabolism and ultimate removal of foreign chemicals from the body. The great majority of foreign substances, including drugs and toxins, when ingested, pass through the liver and are converted to other chemical forms by P450. Hydrophobic compounds are frequently hydroxylated, and the hydroxy group can then be further metabolized by conjugating enzymes, the net result being that lipophilic material is converted to a more hydrophilic compound that can be easily eliminated through the bile or urine. P450 reactions on carcinogenic compounds sometimes yield highly reactive intermediates that can damage DNA and initiate tumorgenesis. Individual differences in P450 levels could determine susceptibility or resistance to the detrimental effects of drugs or carcinogens. In fact, genetically inherited P450-mediated drug oxidation polymorphisms have been described, including the debrisoquine/sparteine polymorphism (Mahgoub et al., 1977; Ei-

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02864.

‡ Laboratory of Molecular Carcinogenesis, National Institutes of Health.

§ Supported by a Japanese Overseas Cancer Fellowship of the Foundation for Promotion of Cancer Research.

‖ University of Basel.
⊥ University of Toronto.
Laboratory of Biochemistry, National Institutes of Health.