Kosk-Kosicka, D., Bzdega, T., & Johnson, J. D. (1990b) *Biochemistry 29*, 1875–1879.
Lanzetta, P., Alvarez, L. J., Reinach, P., & Candia, O. (1979) *Anal. Biochem. 100*, 95–97.
Lin, T.-I., & Morales, M. E. (1977) *Anal. Biochem. 77*, 10–17.
Lowry, O. H., Rosenbrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.
Mitchinson, C., Wilderspin, A. F., Trinnaman, B. I., & Green, N. M. (1982) *FEBS Lett. 146*, 87–92.
Rossi, J. P. F. C., Rega, A. F., & Garrahan, P. J. (1985) *Biochim. Biophys. Acta 816*, 379–386.
Schatzmann, H. J. (1966) *Experientia 22*, 364–368.
Schwartzenbach, G., Senn, H., & Andereff, G. (1957) *Helv. Chim. Acta 40*, 1886–1900.

Molecular Cloning of cDNA for Proteasomes from Rat Liver: Primary Structure of Component C3 with a Possible Tyrosine Phosphorylation Site†,‡

Keiji Tanaka,*,§ Tsutomu Fujiwara,∥ 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 771-01, 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 October 12, 1989; Revised Manuscript Received December 4, 1989

ABSTRACT: Proteasomes are multicatalytic proteinase complexes consisting of multiple components. Previously, we reported the cloning and sequencing of cDNA for the largest component, C2, of rat liver proteasomes [Fujiwara, T., Tanaka, K., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A., & Nakanishi, S. (1989) *Biochemistry 28*, 7332–7340]. In the present study, the nucleotide sequence of another component (C3) of proteasomes has been determined from a recombinant cDNA clone isolated by screening a rat liver cDNA library with synthetic oligodeoxynucleotide probes corresponding to partial amino acid sequences of the protein. The deduced sequence of component C3 consists of 234 amino acid residues with a calculated molecular weight of 25 925. These values are consistent with those obtained by protein chemical analyses. A single mRNA species hybridizing to the C3 cDNA of rat liver was expressed in all rat tissues examined and in a variety of other eukaryotic organisms, its distribution being similar to that of C2 mRNA. The wide distribution of the gene product, possibly C3, suggests that this protein functions similarly in most eukaryotes. C3 is an unmodified protein of a single gene and differs from any other known protein, but its overall amino acid sequence resembles those of other proteasomal components, including C2, suggesting that these components belong to a single family of proteins with the same evolutionary origin. Interestingly, a sequence of about 70 amino acid residues in the C3 protein is very similar to the amino acid sequences, including a conserved autophosphorylated tyrosine residue, present in various cellular tyrosine kinases such as *src* gene products and some receptor proteins; the amino acids in this sequence show about 30% identify with those of the sequences of tyrosine kinases. Conceivably, therefore, the C3 protein has a tyrosine phosphorylation site.

Proteasomes (multicatalytic proteinase complexes) are widely distributed in eukaryotes ranging from man to yeast (Tanaka et al., 1988a; Rivett, 1989). There is increasing evidence that they have ATP-dependent proteolytic activities (Waxman et al., 1985; Driscoll & Goldberg, 1989) and are involved in an ATP/ubiquitin-dependent nonlysosomal proteolytic pathway (Ganoth et al., 1988; McGuire et al., 1988; Tanaka & Ichihara, 1988; Matthews et al., 1989). Proteasomes were initially found in the cytoplasm of cells, but later they were demonstrated in the nucleus also at considerably high concentration (Arrigo et al., 1988; Tanaka et al., 1989). Interestingly, these proteasomes are similar in size, shape, and subunit structure to the 19S–22S ring-shaped particles that are ubiquitous in eukaryotes (Martins de Sa et al., 1986; Arrigo et al., 1987) and that are proposed to have critical cellular functions, such as in repressing mRNA translation (Martins de Sa et al., 1986) and tRNA processing (Castano et al., 1986). In fact, the identity of proteasomes as multicatalytic proteinases with these 20S particles was recently reported by us (Arrigo et al., 1988) and others (Falkenburg et al., 1988; Kleinschmidt et al., 1988). These 20S proteasomal particles appear to play an important role in development in the sea urchin (Akhayat et al., 1987), newt (Gounon et al., 1988), axolotl (Gautier et al., 1988), and *Drosophila* (Haass & Kloetzel, 1989; Haass et al., 1989), but their biological significance is still obscure.

Rat liver proteasomes are symmetrical ring-shaped particles with a sedimentation coefficient of approximately 20 S and

† 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, Japan.
‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02897.
* To whom correspondence should be addressed.
§ Institute for Enzyme Research, The University of Tokushima.
∥ Otsuka Pharmaceutical Co. Ltd.
⊥ 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.

0006-2960/90/0429-3777$02.50/0

a molecular mass of 750 kDa and are composed of a set of polypeptide subunits ranging in size from 21 to 31 kDa and in p*I* value from 3 to 10 (Tanaka et al., 1986b, 1988b). The reason why proteasomes have such an unusual, complex structure is unknown, but as the individual subunits may each have different functions, including multiple proteinase activities, the complex structure may facilitate cooperation of these functions in complex biological processes. For determining the functions of these proteasomal multisubunit particles, it is essential to obtain information about the structure–function relationships of individual subunits. For this purpose, our strategy is to determine the primary structures of all the subunits of proteasomes by recombinant DNA techniques. So far, we have cloned and sequenced cDNA for the largest component (C2) of rat liver proteasomes and demonstrated that C2 is the product of a single gene and is a novel protein, differing from all proteins reported so far (Fujiwara et al., 1989).

In this paper, we report the cloning and sequencing of a full-length cDNA for another component (C3) of proteasomes from rat liver. This component bears a possible tyrosine phosphorylation site like that in a family of tyrosine kinases. We also discuss the structural similarity of C2 with other components of proteasomes on the basis of the information available so far.

Experimental Procedures

Materials. The materials used were as follows: lysyl endopeptidase (Wako Pure Chemical Industries, Osaka, Japan) [α-^{32}P]dCTP (3000 Ci/mmol, Amersham Corp.), [γ-^{32}P]ATP (5000 Ci/mmol, Amersham Corp.), various restriction endonucleases (Takara Shuzo, Kyoto, Japan; Bethesda Research Laboratories; and New England Biolabs), a Cosmosil 5C$_4$-300 column (Nacalai Tesque, Kyoto, Japan) a TSK GEL phenyl-5PW column (Tosoh Co., Tokyo, Japan), and a Chemcosorb 7-ODS-H column (Chemco, Tokyo, Japan).

Purification of Proteasomes and Component C3. Proteasomes were purified to homogeneity from rat liver by conventional chromatographic techniques as reported (Tanaka et al., 1986a, 1988a). Component C3 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 C3 was more than 95% pure as judged by SDS–PAGE (Laemmli, 1970).

Protein Sequencing. The methods used were reported previously (Fujiwara et al., 1989). Briefly, the protein C3 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 molar ratio of enzyme to substrate of 1:40. The digest was separated by reversed-phase HPLC on a Chemcosorb 7-ODS-H column (2.1 × 150 mm) developed with a linear gradient of acetonitrile containing 0.1% trifluoroacetic acid. The protein and peptides obtained by enzymatic digestion were hydrolyzed in the vapor of 6 M HCl at 110 °C for 20 h, and the hydrolysates were analyzed by the phenylthiocarbamoyl method (Bidlingmeyer et al., 1984) in a PICO-TAG system (Waters, Millipore Corp.) according to the instruction manual. The amino acid sequences of the peptides were determined with a gas-phase sequencer (Applied Biosystems, Model 477A), and phenylthiohydantoin derivatives were identified with an Applied Biosystems 120A phenylthiohydantoin analyzer on-line system.

Construction of a cDNA Library. A cDNA library of fetal liver from Wistar rats was constructed in a phagemid expression vector, Bluescript KS M13$^+$ (Stratagene). Double-stranded DNA complementary to rat liver poly(A+) RNA was synthesized as described before (Fujiwara et al., 1989), except that 5′-TAGGTCGACGCGGCCGCTTTTTTTT-TTTTTTT-3′ was used as primer for synthesis of the first strand of cDNA. This primer includes unique restriction endonuclease sites for *Not*I (5′-GCGGCCGC-3′) and *Sal*I (5′-GTCGAC-3′) at the 5′-side of oligo(dT)$_{15}$. 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 QIAGEN-tip column (Funakoshi Co. Ltd., Tokyo, Japan). The DNA mixtures were then ligated with vector Bluescript. The vector had been digested with *Not*I and *Eco*RV at the multicloning site. *Escherichia coli* HB101 competent cells (Takara Shuzo) were transformed with the ligated DNA.

Isolation of C3 cDNA Clones. For isolation of cDNAs for component C3, about 180000 transformants were screened by hybridization with oligodeoxyribonucleotide probes that had been synthesized in an Applied Biosystems Model 380B DNA synthesizer and labeled at their 5′-end with [γ-^{32}P]ATP and T4 polynucleotide kinase (Maxam & Gilbert, 1980). Colony hybridization was carried out by a reported method (Hanahan & Meselson, 1980).

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

RNA Hybridization Analysis. Adult male Wistar strain rats and chickens were obtained from Shizuoka Laboratory Animal Center (Shizuoka, Japan) and Ishii Animal Center (Tokushima, Japan), respectively. *Xenopus laevis* toads were obtained from the Animal Bank of Seibu Department Store (Tokyo, Japan). The animals were decapitated, and various tissues were rapidly removed, frozen in liquid nitrogen, and stored at −70 °C until use. A noncancerous portion of the liver of a patient with a hepatoma was obtained at surgery. Bakers' yeast (*Saccharomyces cerevisiae*) was obtained from Oriental Yeast Co. (Osaka, Japan). Total RNA was extracted from various tissues of rats and other eukaryotes by the guanidinium thiocyanate/cesium chloride method (Chirgwin et al., 1979), and poly(A+) RNA was isolated by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). RNA samples [10 μg of poly(A+) RNA] were denatured and electrophoresed in agarose gel containing formaldehyde. After electrophoresis, the RNA was transferred to a Hybond-N nylon membrane (Amersham) and hybridized with a ^{32}P-labeled probe under conditions similar to those described previously (Fujiwara et al., 1989). The probe used was a *Hin*dIII fragment (about 600-bp length) of cDNA for C3 protein labeled by the multiprime DNA labeling method (Feinberg & Vogelstein, 1984). The membrane was washed as described (Fujiwara et al., 1989) and exposed to Kodak XAR-5 film at −70 °C with an intensifying screen.

Results and Discussion

Isolation of Subunit C3 of Proteasomes. Previously, we reported the separation of multiple components of proteasomes

[1] Abbreviations: SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; Src (pp60$^{c\text{-}src}$), cellular protein with homology to the oncogene product of Rous avian sarcoma virus; EGFR, human epidermal growth factor receptor; INSR, human insulin receptor; PDGFR, mouse platelet-derived growth factor receptor.

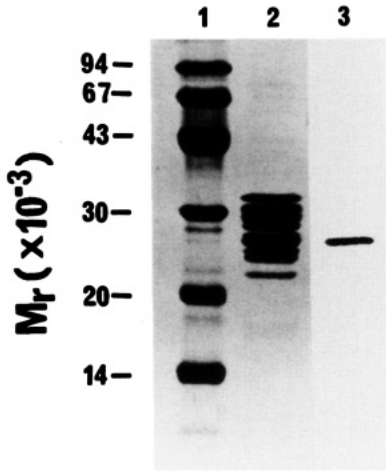

FIGURE 1: Profiles of rat liver proteasomes and isolated component C3 on SDS–PAGE. Proteasomes (25 μg) and component C3 (approximately 2 μg) were subjected to SDS–PAGE: (lane 1) molecular weight markers; (lane 2) purified proteasome complexes; (lane 3) isolated C3. Proteins were stained with Coomassie Brilliant Blue.

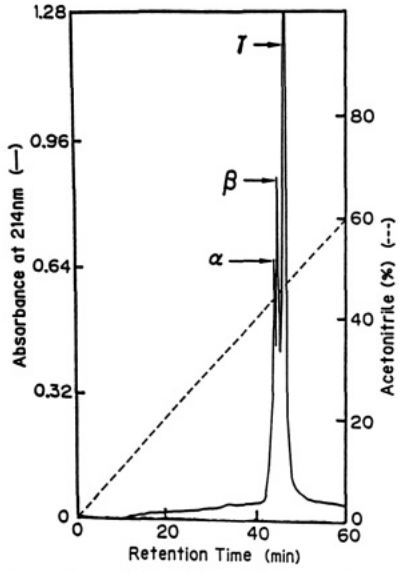

FIGURE 2: Separation of three forms of S-pyridylethylated C3 by reversed-phase HPLC. Purified C3 (approximately 2 mg of protein) was reduced and S-pyridylethylated and then subjected to reversed-phase HPLC on a TSK-GEL phenyl-5PW RP column with a linear gradient of acetonitrile (0–60%). The three components separated were named C3α, C3β, and C3γ in order of their elution.

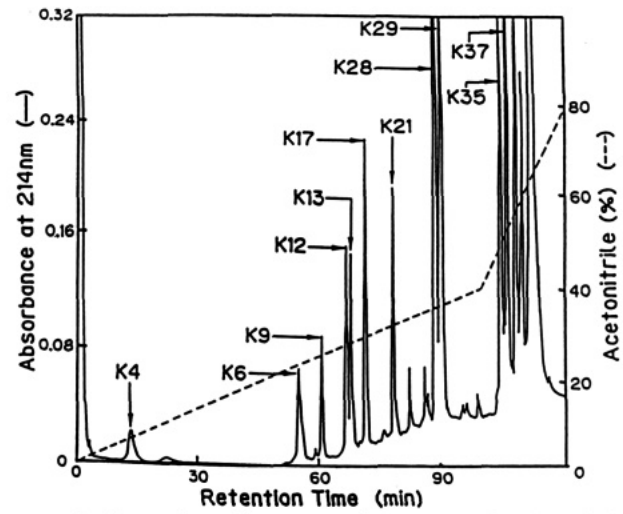

FIGURE 3: Separation of fragments of component C3 cleaved with lysyl endopeptidase by reversed-phase HPLC. S-Pyridylethylated C3 (C3γ) was digested with lysyl endopeptidase as described under Experimental Procedures, and the fragments were resolved by HPLC on a Chemcosorb 7-ODS-H column with a linear gradient of acetonitrile.

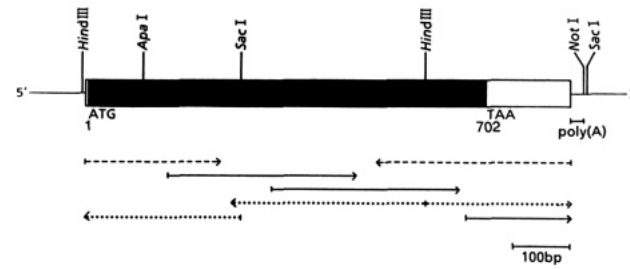

FIGURE 4: Restriction endonuclease map of cloned cDNA for component C3 and sequencing strategy. The solid and open boxes show the coding region and 5′- and 3′-noncoding regions, respectively. Solid lines indicate the sequence of the vector, Bluescript KS$^+$. The numbers below boxes indicate the nucleotide numbers of the first nucleotide of the initiation codon, ATG, and the nucleotide before the termination codon, TAA. Sequenced regions are shown by horizontal arrows. The bar represents 100 bp (base pairs). The sequences of the 5′- and 3′-ends of cDNA (dashed arrows) were determined by direct sequencing of an isolated clone, using primers of T3 to T7 promoter of Bluescript KS$^+$. Some fragments were obtained by subcloning after cleavages with *Sac*I and *Hin*dIII and sequences with T3 and T7 primers (dotted arrows). Various oligonucleotides (17–20 nucleotides long) were synthesized as described under Experimental Procedures and used as primers for sequencing some other regions (solid arrows).

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 named component 1 (C1) to component 10 (C10) in order of their elution. Component C3 was recovered at 48% acetonitrile concentration. Figure 1 shows the profiles on SDS–PAGE of all the components of rat liver proteasomes and the isolated component C3. The molecular weight of C3 was determined to be 25 800 ± 700 (Tanaka et al., 1988b; Fujiwara et al., 1989). When the purified C3 was subjected to automated Edman degradation to determine the amino acid sequence of its N-terminal region, no appreciable signal was obtained, suggesting that the N-terminus was blocked.

Isolation and Protein Sequencing of Fragments of Component C3. As the N-terminus of C3 seemed to be blocked, we attempted to obtain information on the primary structure of its internal region. For this, samples of C3 were reduced, S-pyridylethylated to protect cysteine residues, and subjected to reversed-phase HPLC on a TSK GEL phenyl-5PW RP column. Three components were separated and named C3α, C3β, and C3γ in order of their elution (Figure 2). The amino acid compositions of these three proteins were very similar (data not shown). The molecular ratio of C3α, C3β, and C3γ in the purified preparation was approximately 3:4:16, as judged by the yields on amino acid analysis. Parts of C3γ (6.45 nmol) and C3β (3.08 nmol) were then digested with lysyl endopeptidase, and the resulting fragments were resolved by reversed-phase HPLC on a Chemcosorb 7-ODS-H column. The elution profile monitored at 214 nm showed the presence of 11 major components in the lysyl endopeptidase digest of C3γ (Figure 3). The peptide map of C3γ was very similar to that of C3β (data not shown). Furthermore, the sequences of two fragments of approximately 40 amino acid residues obtained from C3β were identical with those of C3γ determined from the cDNA sequence, as described later. Thus C3β, and probably C3α also, is a modified form of C3γ. Therefore, the amino acid compositions of these peptides isolated from C3γ were determined (data not shown), and their partial primary structures were analyzed by automated Edman degradation

Table I: Automated Amino Acid Sequence Analysis of Fragments in Component C3 of Proteasomes Cleaved by Lysyl Endopeptidase

	C3K4		C3K6		C3K9		C3K12		C3K13		C3K17		C3K21		C3K28		C3K35		C3K37	
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)	amino acid	yield (pmol)
1	Ala	647	Asn	594	Val	855	Thr	239	Ala	1035	Ser	88	Ser	26	His	nd[a]	Leu	1076	Glu	902
2	Thr	146	Tyr	902	Glu	308	Phe	742	Ala	723	Ile	965	Ile	181	Ile	752	Val	1034	Ser	267
3	Ala	462	Val	959	Pro	319	Leu	548	Asn	260	Leu	862	Leu	163	Gly	542	Gln/Glu		Phe	1122
4	Met	604	Asn	555	Ile	484	Glu	287	Gly	570	Tyr	292	Tyr	61	Leu	746	Ile	839	Glu	686
5	Gly	463	Gly	736	Thr	78	Lys	301	Val	655	Asp	99	Asp	19	Val	531	Glu	436	Gly	1702
6	Lys	264	Lys	377	Lys	137			Val	745	Glu	182	Glu	41	Tyr	458	Tyr	418	Gln	1210
7									Leu	834	Arg	135	X[b]		Ser	87	Ala	873	Met	1241
8									Ala	679	Ser	60	Ser	10	Gly	494	Leu	768	Thr	367
9									Thr	126	Val	525	Val	83	Met	467	Ala	970	Glu	779
10									Glu	175	His	nd	(His)	nd	Gly	416	Ala	822	Asp	337
11									Lys	120	Lys	178	Lys	40	Pro	412	Val	581	Asn	681
12													Val	60	Asp	180	Ala	591	Ile	656
13													Glu	22	Tyr	317	Gly	495	Glu	686
14													Pro	22	Arg	490	Gly	434	Val	873
15															Val	371	Ala	602	Gly	763
16															Leu	295	Pro	309	Ile	501
17															Val	397	Ser	65	Cys	nd
18															(His)	nd	Val	279	Asn	317
19															Arg	395	Gly	205	Glu	260
20															Ala	337	Ile	226	Ala	525
21															Arg	952	Lys	84	Gly	589
22															Lys	86			Phe	380
23																			Arg	nd
24																			Arg	nd
25																			Leu	nd
26																			Thr	110
27																			Pro	276
28																			Thr	56
29																			Glu	117
30																			Val	221
repetitive yield (%)	Ala-1–Ala-3 85		Asn-1–Asn-4 98		nd		nd		Ala-1–Ala-8 94		Ser-1–Ser-8 95		Ser-1–Ser-8 87		Gly-3–Gly-8 98		Leu-1–Leu-8 95		Glu-1–Glu-4 91	

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

Table II: Amino Acid Composition of Component C3 of Proteasomes

	amino acid compositions (mol %)[a]				amino acid compositions (mol %)[a]				amino acid compositions (mol %)[a]		
amino acid	chemical analysis	cDNA sequence analysis	no. of residues[b]	amino acid	chemical analysis	cDNA sequence analysis	no. of residues[b]	amino acid	chemical analysis	cDNA sequence analysis	no. of residues[b]
Gly	8.36	8.55	20	Met	1.60	2.14	5	Lys	5.79	5.56	13
Ala	9.54	9.83	23	total Asx	6.93	5.98	14	His	1.56	1.71	4
Val	8.11	8.12	19	Asp		2.99	7	Phe	3.51	3.42	8
Leu	9.54	8.55	20	Asn		2.99	7	Tyr	5.79	5.98	14
Ile	5.03	5.13	12	total Glx	12.33	12.39	29	Trp	0.76[d]	0.85	2
Ser	5.28	5.98	14	Glu		7.69	18	Pro	4.39	4.27	10
Thr	5.19	5.56	13	Gln		4.70	11	total	100	100	234
Cys	0.59[c]	0.85	2	Arg	5.70	5.13	12	mol wt			25925

[a] The amino acid composition was determined by analysis of a hydrolysate of purified component C3 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.

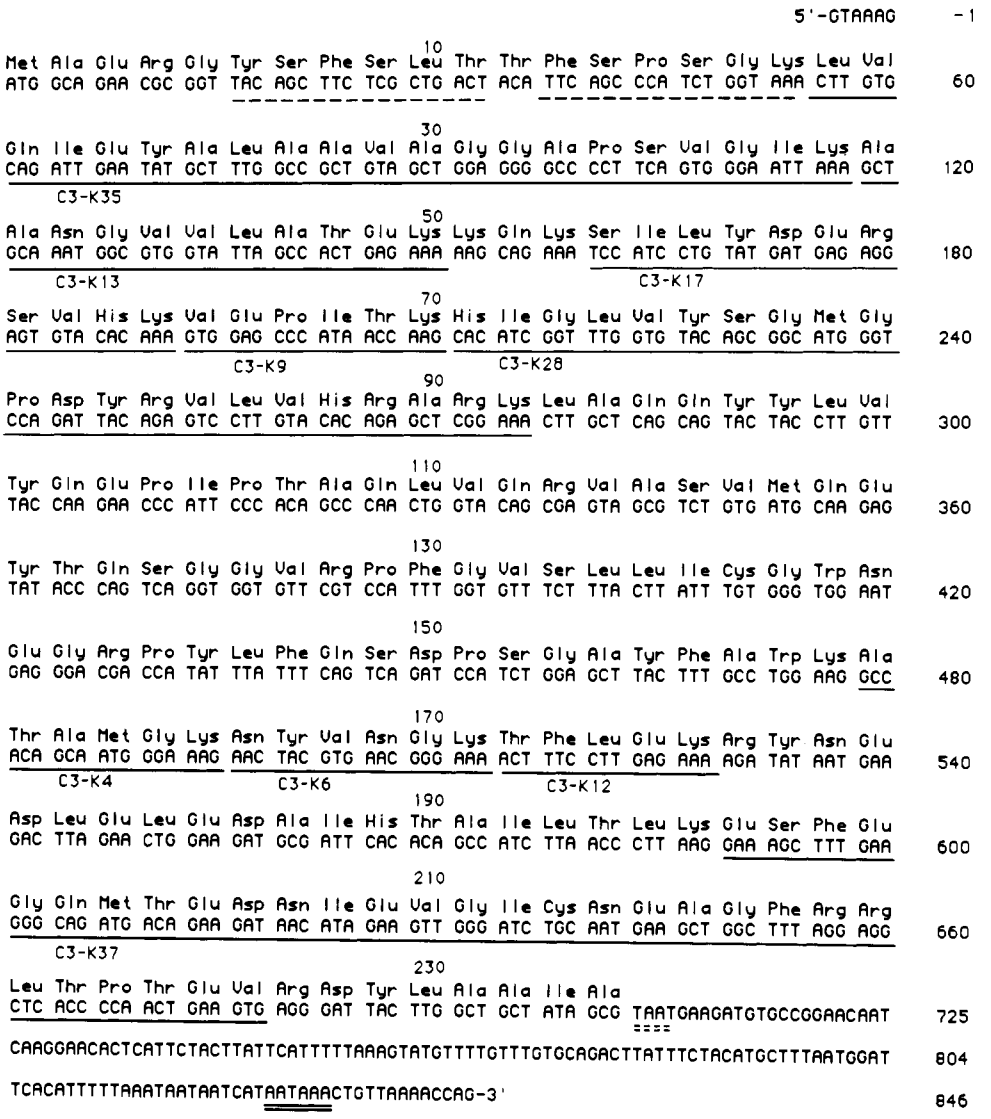

FIGURE 5: Nucleotide sequence of the cDNA insert encoding component C3 of rat liver proteasomes and the amino acid sequence deduced from its open reading frame. Nucleotides are numbered in the 5'- to 3'-direction, beginning with the first residue of the initiation methionine codon ATG. The nucleotides on the 5'-side are indicated by negative numbers. The predicted amino acid sequence of C3 is shown above the nucleotide sequence. Amino acid residues are numbered from the N-terminus. Solid and broken lines show the amino acid sequences corresponding to those obtained by Edman degradation of fragments cleaved with lysyl endopeptidase (Table I) and of a trypsin-cleaved fragment of the N-terminal C3-K29 fragment (Figure 3), respectively. The termination codon TAA is doubly underlined with broken lines. The possible polyadenylation signal AATAAA is doubly underlined with solid lines.

(Table I). In this paper, C3γ is called C3 for simplicity, unless otherwise specified. Peptide C3-K37 was considered to be the C-terminal fragment of C3, because it did not contain a lysine residue (lysyl endopeptidase cleaves the carboxy side of lysine residues in a polypeptide chain). Indeed, the primary structure of peptide C3-K37 was consistent with that of the C-terminal region of C3 deduced from the nucleotide sequence of the cDNA, as described later. Peptide C3-K29 was concluded to be the N-terminal fragment, because its N-terminal amino acid was blocked. Peptide C3-K21 was a larger fragment and was concluded to include C3-K17, because it contained a single lysine residue and a sequence of 10 amino acid residues that was identical with the sequence of C3-K17 (Table I).

Isolation of cDNA Clones Encoding Component C3. A rat cDNA library was constructed with the phagemid Bluescript vector using poly(A+) RNAs extracted from fetal rat liver as described under Experimental Procedures. For isolation of cDNA for component C3, this library was screened by hybridization with synthetic deoxyribonucleotides as probes. These probes were synthesized as follows. Of the C3 fragments obtained by proteolytic digestion (Figure 2), parts of the sequences of two fragments (Phe-Glu-Gly-Gln-Met-Thr of C3-K37 and Lys-Asn-Tyr-Val-Asn-Gly-Lys of C3-K6 in Table I) with minimum codon ambiguity were selected, and the oligonucleotide probes 5'-GTCATYTGNCCYTCRAA-3' and 5'-TTGCCRTTNACRTARTTTTT-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 180000 colonies of a fetal rat liver cDNA library with two sets of mixtures of 32 of each of these 17mer or 20mer oligonucleotides. Twenty-four cDNA clones that gave strongly positive signals for both probes were isolated from the library by colony hybridization techniques. As the cDNAs of these 24 clones gave similar cleavage maps with several restriction endonucleases, the clone [about 880-bp length including a poly(A) tail] that carried the largest cDNA insert was subjected to cDNA sequencing.

Primary Structure Deduced from the cDNA Sequence. The nucleotide sequence of the largest C3 cDNA clone was de-

termined by the strategy illustrated in Figure 4. The resulting nucleotide sequence of the cDNA and the primary structure of the C3 deduced from the cDNA sequence are shown in Figure 5. The sequence of 852 nucleotides included the entire coding region and 5′- and 3′-noncoding regions. The 3′-noncoding region consisted of 144 nucleotides, excluding the poly(A) tail. A polyadenylation signal (AAUAAA) that is common in eukaryotic mRNAs (Proudfoot & Brownlee, 1976) was located 14 nucleotides upstream from the poly(A) addition site. Part of the poly(A) sequence (approximately 25 nucleotides) is seen at the 3′-end of this cDNA clone. The coding region is 702 nucleotides long, which corresponds to a protein of 234 amino acids, and extends from ATG at nucleotide positions 1–3 to nucleotide 702. From this deduced sequence, the molecular weight was calculated to be 25 925. This coding sequence was followed by the termination codon TAA. Although the N-terminal amino acid sequence of C3 is unknown, we concluded that ATG, located at nucleotides 1–3, is the initiation codon for two reasons. (1) The open reading frame starting from this ATG at nucleotide positions 1–3 is the longest. (2) When the N-terminal peptide C3-K29 (Figure 3) was further digested with trypsin (because it contained 1 mol of Arg residue as determined by analysis of its amino acid composition) and the resulting fragment was resolved by reversed-phase HPLC on a Chemcosorb 7-ODS-H column, a 12 amino acid sequence (X-Tyr-Ser-Phe-Ser-Leu-Thr-X-Phe-Ser-Pro-Ser-Gly-Lys), determined by Edman degradation, was found to correspond to nucleotide positions 16–54 of C3 cDNA (indicated by a broken line in Figure 5). For these two reasons, the Met residue, corresponding to nucleotides 1–3, was concluded to be the N-terminal amino acid of the C3 protein.

The following observations indicate that the amino acid sequence shown in Figure 5 is actually that of C3 of proteasomes: (1) The partial amino acid sequences of several fragments (approximately 50% of the total residues) determined by Edman degradation (Table I) were found to be in complete accordance with those deduced from the nucleotide sequence of cDNA (solid and broken lines, in Figure 5). (2) The amino acid composition (Table II) of the whole C3 molecule determined by chemical analysis coincided well with that deduced from the sequence shown in Figure 5. (3) The molecular weight (25 925) of C3 estimated from its amino acid composition was similar to that estimated by SDS–PAGE (25 800) (Figure 1). Thus, it is reasonable to conclude that the isolated clone is that of component C3 cDNA.

Component C3 had no potential N-glycosylation site, which is in accordance with the observation that no appreciable glucosamine or galactosamine was detected during amino acid analysis (data not shown). Of the amino acid residues in component C3, 25 are acidic and 29 are basic residues.

Expression of the Gene for Component C3 in Various Tissues of Eukaryotic Organisms. To investigate the tissue specificity of C3 gene expression, we examined the levels of C3-specific mRNA in various rat tissues. On Northern blot analysis, the mRNAs extracted from various rat tissues gave a single hybridization band of about 1.0 kb (Figure 6, left panel). The level of mRNA for component C3 varied significantly in the different tissues examined but was very similar to that of mRNA for component C2 (Fujiwara et al., 1989), suggesting that proteasome complexes, presumably all the components, are expressed similarly in different cells and tissues.

We next examined whether the C3 mRNA is expressed in other eukaryotic cells (Figure 6, right panel). A single band

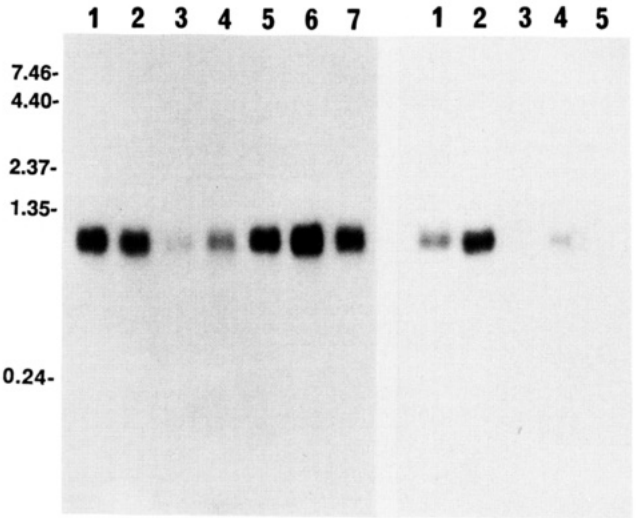

FIGURE 6: RNA blot hybridization of rat liver C3 cDNA with poly(A+) RNAs from various rat tissues and other eukaryotic cells. Poly(A+) RNAs (10 μg) were used for Northern blot analysis as described under Experimental Procedures. The *Hin*dIII fragment (600 bp) of cloned C3 cDNA was used as a probe (see Figure 4). Left panel: rat liver (lane 1), kidney (lane 2), skeletal muscle (lane 3), heart (lane 4), lung (lane 5), spleen (lane 6), and brain (lane 7). Right panel: human liver (lane 1), rat liver (lane 2), chicken liver (lane 3), *X. laevis* ovary (lane 4), and bakers' yeast (*S. cerevisiae*) (lane 5). The positions of RNA standards (from Bethesda Research Laboratories) are shown (in kilobases) on the left.

hybridizing to rat C3 cDNA was observed in mRNA fractions from human liver. It was the same size (about 1.0 kb) as that in rat liver, although its reactivity was somewhat weaker than that of rat liver mRNA, suggesting close similarity in the structures of mRNAs for C3 in various mammals. Rat C3 cDNA also hybridized to mRNAs from various tissues of other species, such as those of chicken liver and *Xenopus laevis* ovary, but the hybridizing efficiencies of these C3 mRNAs were low, and their sizes were somewhat different from those of mammalian tissues. No band hybridizing with rat C3 cDNA was found in bakers' yeast under the present experimental conditions. Thus, mRNA encoding component C3 is widely expressed in a variety of eukaryotes, including mammals, a bird, and an amphibian, but with minor species-specific variations in mRNA structure. This suggests that C3 functions similarly in almost all eukaryotes. The distributions of C3 mRNAs in different species were quite similar to those of C2 mRNAs reported before (Fujiwara et al., 1989).

Comparison of Primary Structures of C3 and Other Proteasome Subunits. We searched for structural homologies of the nucleotide and amino acid sequences of C3 with those of other proteins, using the data bases EMBL/GDB (European Molecular Biology Laboratory) and NBRF/PDB (National Biomedical Research Foundation). No proteins with overall similarity to C3 were found, suggesting that component C3 is distinct from any other known protein. However, a sequence of amino acid residues of this C3 protein showed marked overall similarity to that of component C2, which was recently deduced from the sequence of its cDNA clone isolated in this laboratory (Fujiwara et al., 1989). Figure 7 shows the primary structures of these proteins deduced from the nucleotide sequences of their cDNA clones. The amino acid sequence of C3 resembled that of C2: the identity of amino acids in their sequences was 31%. These findings suggest that these proteins belong to a family with the same evolutionary origin. This suggestion is supported by our recent findings that parts of the primary structures (approximately 50% of the total amino acid residues) of C8 and C9, other components of rat liver

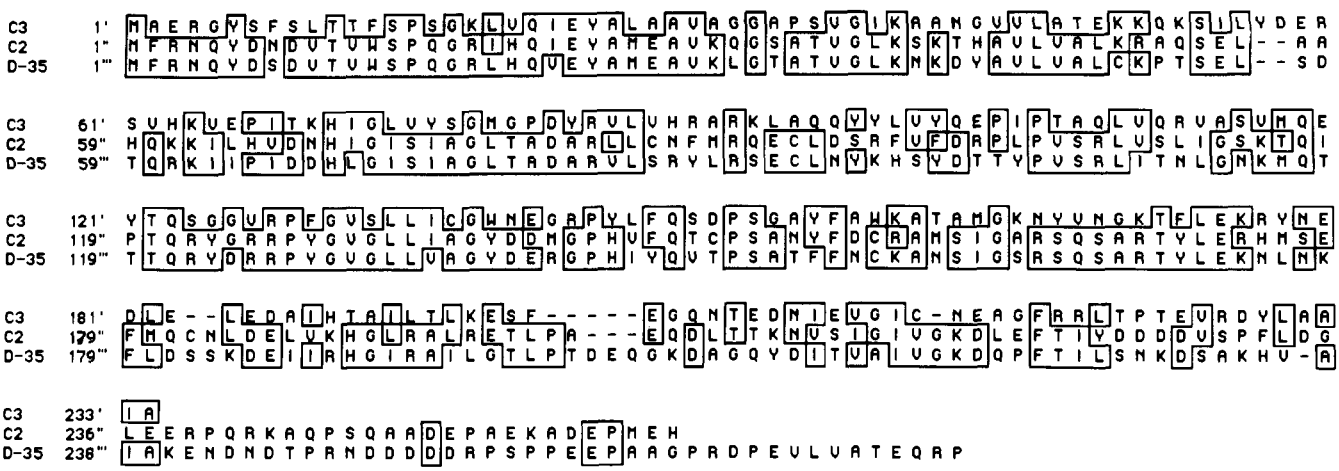

FIGURE 7: Comparison of the protein sequences of C3 and C2 of rat liver proteasomes and the 35000 component of *Drosophila* proteasomes. Identical amino acid residues are boxed. Numbers are residue numbers of component C3. Gaps (shown by bars) are inserted to achieve maximum sequence homologies of C3 and the 35000 component of *Drosophila* (D-35) with the sequence of C2. Sequence data for C2 and the *Drosophila* PROS-35 gene product are taken from Fujiwara et al. (1989) and Haass et al. (1989), respectively.

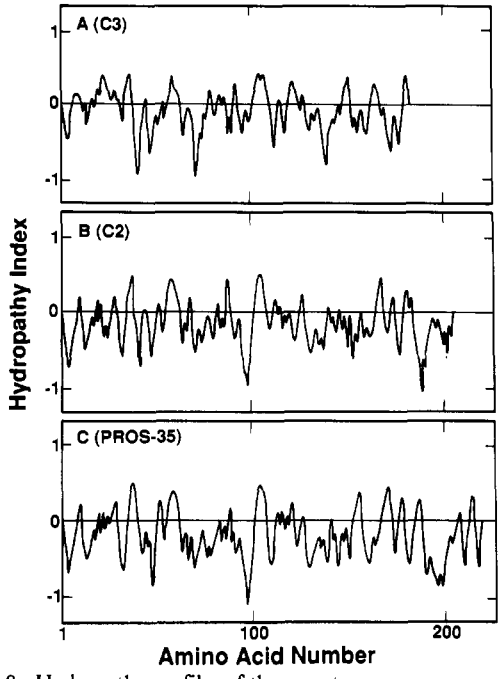

FIGURE 8: Hydropathy profiles of three proteasome components: (A) rat C3; (B) rat C2; (C) *Drosophila* PROS-35 gene product. Hydropathy was analyzed by the procedure of Eisenberg et al. (1984). The peaks above and below the zero line represent hydrophobic and hydrophilic domains, respectively.

proteasomes (Tanaka et al., 1988b), were also very similar to those of C2 and C3 (unpublished data).

Interestingly, the amino acid sequences of components C2 and C3 are remarkably similar to that of a component with a molecular weight of 35000 from *Drosophila* proteasomes (PROS-35 gene product) that was recently reported by Haass et al. (1989). The overall similarity of C3 of rat proteasomes with the PROS-35 protein is approximately 35%. Moreover, the amino acid sequence of C2 also resembles that of the *Drosophila* 35000 protein more closely than that of C3, the identity of amino acids in the sequences of the two being 52% (Figure 7). Furthermore, the hydropathy profiles of these three proteins shown in Figure 8 indicate closer structural similarities between rat C2 and the *Drosophila* protein than between C2 and C3 or between C3 and the *Drosophila* PROS-35 protein. From these profiles, the tertiary structures of C2 and the PROS-35 gene product could be very similar, although with minor differences in their C-terminal regions, suggesting their similar function(s) in cells. The structural conservation of proteasome components in these two different eukaryotic organisms suggests that they evolved from the same ancestral gene.

Identification of a Possible Tyrosine Phosphorylation Site in the C3 Protein. Computer analysis revealed little overall structural similarity of C3 to any other known protein but some partial identity of amino acid sequences: a sequence of 73 amino acid residues in component C3 (residues 78–150) closely

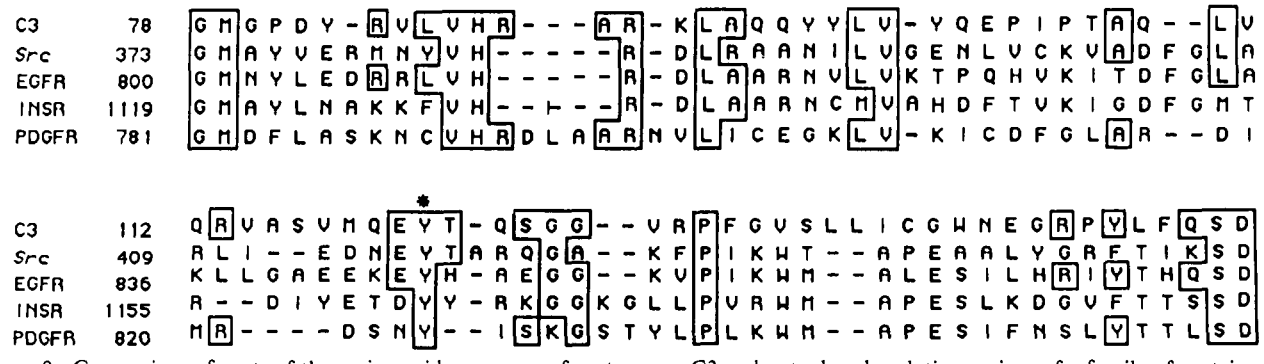

FIGURE 9: Comparison of parts of the amino acid sequences of proteasome C3 and autophosphorylation regions of a family of proteins with tyrosine kinase activity. Residues that are identical in rat proteasome C3 and other gene products are boxed. The numbers shown are the residue numbers of the respective proteins. Gaps (bars) are inserted for optimal alignment of the sequences. Chicken Src ($pp60^{c\text{-}src}$), cellular protein with homology to the oncogene product from Rous avian sarcoma virus (Takeya & Hanafusa, 1983); EGFR, human epidermal growth factor receptor (Ullrich et al., 1984); INSR , human insulin receptor (Ebina et al., 1985); PDGFR, mouse platelet-derived growth factor receptor (Yarden et al., 1986). The asterisk indicates the autophosphorylated tyrosine residue conserved in this family of tyrosine kinases.

resembled partial sequences, including a conserved possible autophosphorylation site, found in viral and receptor proteins with tyrosine kinase activity, such as $pp60^{c\text{-}src}$, a cellular protein with homology to the oncogene product of Rous avian sarcoma virus (Takeya & Hanafusa, 1983), and receptors of epidermal growth factor (EGFR; Ullrich et al., 1984), insulin (INSR; Ebina et al., 1985), and platelet-derived growth factor (PDGFR; Yarden et al., 1986). The amino acid sequences of these other proteins with similarity to part of C3 are shown in Figure 9. The identity of amino acids in these sequences was about 30%. Similar homology was also observed in the sequences of a variety of other tyrosine kinases [for recent review, see Hanks et al. (1988)]. Residue Tyr419 of $pp60^{v\text{-}src}$ (Smart et al., 1981) and Tyr1162 and Tyr1163 of INSR (Ellis et al., 1986) in these regions have been reported to be phosphorylated in vivo. Moreover, Tyr825 of PDGFR may also be an autophosphorylated residue (Yarden et al., 1986). Thus, Tyr121 of C3 may be a phosphorylation site, like the autophosphorylated tyrosine sites in these tyrosine kinases (Figure 9). However, in this connection it is noteworthy that Tyr845 of EGFR is not phosphorylated in vivo (Downward et al., 1984) although it is conserved. Moreover, very recently in preliminary studies we found that the partial amino acid sequence of another subunit, named C4, is very similar to or identical with that of C3 and that monoclonal antibody against phosphotyrosine did not react with C3 but reacted specifically with C4 (unpublished data). Thus, C4 could be a phosphorylated form of C3.

Recently, Haass and Kloetzel (1989) reported that in *Drosophila* proteasomes undergo changes in subunit pattern during development and that these developmental changes in subunit multiplicity are due to phosphorylation of some components. Moreover, they found that the PROS-35 protein of *Drosophila* proteasomes bears a similar tyrosine phosphorylation site to the viral and cellular enzymes. It is very interesting that both rat C3 and the PROS-35 protein contain a possible tyrosine phosphorylation site. However, its position in C3 (Tyr121) is clearly different from that in the PROS-35 subunit (Tyr103), as judged by the optimal alignment of primary structures of the two proteins for homology (Figure 7). Moreover, C2, which shows closer structural similarity to PROS-35 protein, does not contain a potential phosphorylated tyrosine residue. Thus, C3 may only be related to PROS-35 protein functionally: they may have a potential regulatory effect on another component(s) of the proteasome complex through tyrosine phosphorylation.

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.

Akhayat, O., Grossi de Sa, F., & Infante, A. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 1595–1599.

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.

Downward, J., Parker, P., & Waterfield, M. D. (1984) *Nature 311*, 483–485.

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

Ebina, Y., Ellis, L., Jarnagin, K., Edery, M., Graf, L., Clauser, E., Ou, J., Masiarz, F., Kan, Y. W., Goldfine, I. D., Roth, R. A., & Rutter, W. J. (1985) *Cell 40*, 747–758.

Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 140–144.

Ellis, L., Clauser, E., Morgan, D. O., Edery, M., Roth, R. A., & Rutter, W. J. (1986) *Cell 45*, 721–732.

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.

Fujiwara, T., Tanaka, K., Kumatori, A., Shin, S., Yoshimura, T., Ichihara, A., Tokunaga, F., Aruga, R., Iwanaga, S., Kakizuka, A., & Nakanishi, S. (1989) *Biochemistry 28*, 7332–7340.

Ganoth, D., Leshinsky, E., Eytan, E., & Hershko, A. (1988) *J. Biol. Chem. 263*, 12412–12419.

Gautier, J., Pal, J. K., Grossi de Sa, M.-F., Beetschen, J. C., & Scherrer, K. (1988) *J. Cell Sci. 90*, 543–553.

Gounon, J. K P., Grossi, M.-F., & Scherrer, K. (1988) *J. Cell Sci. 90*, 555–567.

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

Haass, C., Pesold-Hurt, B., Multhaup, G., Beyreuther, K., & Kloetzel, P.-M. (1989) *EMBO J. 8*, 2373–2379.

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

Hanks, S. K., Quinn, A. M., & Hunter, T. (1988) *Science 241*, 42–52.

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.

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

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

Martins de Sa, C., Grossi de Sa, M.-F., Akhayat, O., Broders, F., Scherrer, K., Horsch, A., & Schmid, H.-P. (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.

Proudfoot, N. J., & Brownlee, G. G. (1976) *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.

Smart, J. E., Oppermann, H., Czernilofsky, A. P., Purchio, A. F., Erikson, R. L., & Bishop, J. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6013–6017.

Takeya, T., & Hanafusa, H. (1983) *Cell 32*, 881–890.

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.
Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayers, E. L. V., Whittle, N., Waterfield, M. D., & Seeburg, P. H. (1984) *Nature 309*, 418–425.
Waxman, L., Fagan, J. M., Tanaka, K., & Goldberg, A. L. (1985) *J. Biol. Chem. 260*, 1194–1200.
Yarden, Y., Escobedo, J. A., Kuang, W.-J., Yang-Feng, T. L., Daniel, T. O., Tremble, P. M., Chen, E. Y., Ando, M. E., Harkins, R. N., Francke, U., Fried, V. A., Ullrich, A., & Williams, L. T. (1986) *Nature 323*, 226–232.
Yoneda, Y., Imamoto-Sonobe, N., Matsuoka, Y., Iwamoto, R., Kiho, Y., & Uchida, T. (1988) *Science 242*, 275–278.

^1H and ^{31}P Nuclear Magnetic Resonance Investigation of the Interaction between 2,3-Diphosphoglycerate and Human Normal Adult Hemoglobin†

Irina M. Russu,‡ Shing-Shing Wu, Keith A. Bupp,§ Nancy T. Ho, and Chien Ho*
Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
Received June 28, 1989; Revised Manuscript Received November 22, 1989

ABSTRACT: High-resolution ^1H and ^{31}P nuclear magnetic resonance spectroscopy has been used to investigate the binding of 2,3-diphosphoglycerate to human normal adult hemoglobin and the molecular interactions involved in the allosteric effect of the 2,3-diphosphoglycerate molecule on hemoglobin. Individual hydrogen ion NMR titration curves have been obtained for 22–26 histidyl residues of hemoglobin and for each phosphate group of 2,3-diphosphoglycerate with hemoglobin in both the deoxy and carbonmonoxy forms. The results indicate that 2,3-diphosphoglycerate binds to deoxyhemoglobin at the central cavity between the two β chains and the binding involves the β2-histidyl residues. Moreover, the results suggest that the binding site of 2,3-diphosphoglycerate to carbonmonoxyhemoglobin contains the same (or at least some of the same) amino acid residues responsible for binding in the deoxy form. As a result of the specific interactions with 2,3-diphosphoglycerate, the β2-histidyl residues make a significant contribution to the alkaline Bohr effect under these experimental conditions (up to 0.5 proton/Hb tetramer). 2,3-Diphosphoglycerate also affects the individual hydrogen ion equilibria of several histidyl residues located away from the binding site on the surface of the hemoglobin molecule, and, possibly, in the heme pockets. These results give the first experimental demonstration that long-range electrostatic and/or conformational effects of the binding could play an important role in the allosteric effect of 2,3-diphosphoglycerate on hemoglobin. The ^{31}P nuclear magnetic resonance titration data for each phosphate group of 2,3-diphosphoglycerate have been used to calculate the pK values of the phosphate groups in 2,3-diphosphoglycerate bound to deoxy- and carbonmonoxyhemoglobin and the proton uptake by 2,3-diphosphoglycerate upon ligand binding to hemoglobin.

2,3-Diphosphoglycerate (2,3-DPG)[1] is the predominant phosphorylated metabolite inside red blood cells. The allosteric effect of this compound on hemoglobin (Hb) leads to a dramatic decrease in the oxygen affinity of the Hb molecule and thus facilitates the unloading of oxygen to the tissues (Benesch & Benesch, 1969, 1974). 2,3-DPG also influences the pH dependence of the oxygen affinity of Hb by increasing both the alkaline and the acid Bohr effect (de Bruin et al., 1971, 1973, 1974; de Bruin & Janssen, 1973; Kilmartin, 1974). The allosteric effect of 2,3-DPG upon Hb function results from the higher binding affinity of 2,3-DPG for the deoxy form of the Hb molecule. The site of the binding of 2,3-DPG to human normal adult Hb (Hb A) in the deoxy form has been identified by X-ray crystallography to be at the central cavity between the two β chains on the 2-fold symmetry axis of the molecule (Arnone, 1972). The site of the binding of 2,3-DPG to the ligated form of Hb is not known, although it has been suggested that it may be close to that in the deoxy form (Gupta et al., 1979).

An understanding of the allosteric effect of 2,3-DPG on Hb requires a characterization of the roles played by individual groups of Hb and 2,3-DPG in the binding process. Nuclear magnetic resonance (NMR) spectroscopy is uniquely suited to obtain such a characterization, since it is the only technique

† Preliminary results were presented at the 27th Annual Meeting of the Biophysical Society, Feb 14–17, 1983, San Diego, CA. This research was supported by a research grant from the National Institutes of Health (HL-24525).
* Address correspondence for this author at the Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Ave., Pittsburgh, PA 15213.
‡ Present address: Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT 06457.
§ Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

[1] Abbreviations: 2,3-DPG, 2,3-diphosphoglycerate; Hb, hemoglobin; Hb A, human normal adult hemoglobin; HbCO, carbonmonoxyhemoglobin; HbO_2, oxyhemoglobin; deoxy-Hb, deoxyhemoglobin; NMR, nuclear magnetic resonance; DSS, 2,2-dimethyl-2-silapentane-5-sulfonate; Bis-Tris, [bis(2-hydroxyethyl)amino]tris(hydroxymethyl)-methane; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/90/0429-3785$02.50/0