cDNAs and Deduced Amino Acid Sequences of Subunits in the Binding Component of Mouse Bactericidal Factor, Ra-Reactive Factor: Similarity to Mannose-Binding Proteins†,‡

Shusuke Kuge,§,‖ Setsunosuke Ihara,§,⊥ Emiko Watanabe,§,# Mariko Watanabe,§ Kunio Takishima,△ Tatsuko Suga,△ Gunji Mamiya,△ and Masaya Kawakami*,§

Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara, Kanagawa-ken 228, Japan, and Department of Biochemistry, National Defense Medical College, Tokorozawa, Saitama 359, Japan

Received March 2, 1992; Revised Manuscript Received May 4, 1992

ABSTRACT: The complement-dependent bactericidal factor, Ra-reactive factor, binds specifically to Ra polysaccharide, which is common to some strains of Gram-negative enterobacteria, and it is a complex of proteins composed of a polysaccharide-binding component and a component that is presumably responsible for the complement activation. The former component consists of two different 28-kDa polypeptides, P28a and P28b. We determined the partial amino acid sequences of P28a and P28b, and the results indicated that these polypeptides were similar to two species of mannose-binding protein, MBP-C and MBP-A (alternative names, liver and serum mannan-binding proteins, respectively), which have been isolated from rat liver and/or serum [Drickamer, K., Dordal, M. S., & Reynolds, L. (1986) *J. Biol. Chem. 261*, 6878–6887; Oka, S., Itoh, N., Kawasaki, T., & Yamashina, I. (1987) *J. Biochem. 101*, 135–144]. Thus, we cloned the respective cDNAs, using as probes synthetic oligonucleotides for which the sequences had been deduced from the amino acid sequences of P28a and P28b and of rat MBP cDNAs. The primary structures of P28a and P28b deduced from the cloned cDNAs are homologous to one another. They have three domains, a short NH_2-terminal domain, a collagen-like domain, and a domain homologous to regions of some carbohydrate-binding proteins, as has been reported for rat MBPs. Southern and Northern blotting analyses using these cDNAs indicated that the P28a and P28b polypeptides are the products of two unique mouse genes which are expressed in hepatic cells.

Ra-reactive factor (RaRF)[1] is a complement-dependent bactericidal factor, that reacts with many strains of *Salmonella* and *Escherichia coli*. RaRF binds specifically to the Ra polysaccharide that is common to almost all strains of *Salmonella* and to the R2 polysaccharide present in certain strains of *E. coli*. The most important components of the determinant to which the factor binds are the L-glycero-D-mannoheptosyl and *N*-acetyl-D-glucosaminyl residues that are present at the nonreducing ends of the polysaccharide chains (Ihara et al., 1982a). After the binding, RaRF kills the bacteria via complement activation (Kawakami et al., 1982). Such anti-Ra bactericidal activities are present in sera of a wide variety of vertebrates, including mammals, birds, reptiles, amphibians, and bony and cartilaginous fish (Kawakami et al., 1984). From these facts it has been considered that RaRF may play an essential role in resistance of vertebrates to infection. Detailed analysis of the activation of complement that is triggered by mouse RaRF has indicated that it activates the C4 and C2 components of complement directly, without involvement of the C1 component (Ji et al., 1988). This unique complement-activation pathway differs from that triggered by the antigen–antibody complex.

Affinity-purified mouse RaRF has a molecular weight of about 350K and is composed of a polysaccharide-binding component (larger than 200 kDa) and components that are probably responsible for the complement activation (50–200 kDa). The polysaccharide-binding component is composed of equimolar amounts of two different 28-kDa polypeptides, designated P28a and P28b, and the putative complement-activating component contains 29- and 70-kDa polypeptides (Ihara et al., 1991). The ability to bind specifically to the Ra determinant was found to be associated with the polysaccharide-binding component composed of P28a and P28b, while the complement-activating ability was not.

Determination of the primary structures of P28 polypeptides should provide information about the mechanism of action of RaRF. In the present studies, the results of partial amino acid analysis of P28a and P28b indicated that their sequences exhibit significant homology to those of rat mannose-binding proteins, MBP-A and MBP-C, which have also been referred to as liver and serum mannan-binding protein, respectively (Drickamer et al., 1986; Kawasaki et al., 1987; Oka et al., 1987, 1988). It was noted that the polysaccharide-binding component of RaRF was similar to those of MBPs in certain respects. For example, the RaRF had the ability to bind to polysaccharide composed of *N*-acetyl-D-glucosamine and D-

† This work was supported in part by grants from the Ministry of Education, Science and Culture, Japan (58570197), and from the Terumo Life Science Foundation.

‡ The nucleic acid sequence in this paper has been submitted to GenBank under Accession Numbers D11440 for P28a and D11441 for P28b.

* Address correspondence and reprint requests to this author at the Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara, Kanagawa-ken 228, Japan.

§ Kitasato University.

‖ Present address: Gene Regulation Laboratory, Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX, England.

⊥ Present address: Developmental Biology Laboratory, Department of Plastic Surgery, School of Medicine, Kitasato University, Sagamihara, Kanagawa 228, Japan.

Present address: Graduate School of Science and Technology, Niigata University, Niigata 950-21.

△ National Defense Medical College.

[1] Abbreviations: RaRF, Ra-reactive factor; MBP, mannose-binding protein or mannan-binding protein; PAGE, polyacrylamide gel electrophoresis.

0006-2960/92/0431-6943$03.00/0

mannose; there was a requirement for Ca^{2+} for the binding; and it was unstable under acidic conditions (Ihara et al., 1982b; Kawakami et al., 1982; Kawasaki et al., 1987). The molecular masses of P28 and MBP polypeptides are also similar to one another. To clone the cDNA of P28a and P28b, therefore, we synthesized oligonucleotides with sequences deduced from the partial amino acid sequences of these polypeptides, by reference to the nucleotide sequences of rat MBP cDNAs. It was found that each of the polypeptide sequences deduced from the cloned cDNA sequences have three domains, a short NH_2-terminal domain, a collagen-like domain, and a domain with a sequence homologous to regions of other carbohydrate-binding proteins. These structural features are similar to those of the C1q subcomponent of complement (Reid, 1985; Wood et al., 1988; Sellar et al., 1991) and of serum lectins, such as rat and human mannan-binding proteins (Drickamer et al., 1986; Ikeda et al., 1987; Oka et al., 1987; Ezekowitz et al., 1988).

EXPERIMENTAL PROCEDURES

Materials. Restriction endonucleases, *Eco*RI linker nucleotide, GGAATTCC, and DNA-modification enzymes were purchased from Takara Shuzo Co., Tokyo, New England Biolabs, Beverly, MA, and Boehringer-Mannheim BmbH, Mannheim, Germany, respectively. Reverse transcriptase from avian myeloblastosis was purchased from Japan Bio-Rad Laboratories, Tokyo, and radiolabeled compounds were purchased from Amersham International Co., Bucks, U.K. The enzymes and kits were used according to the instructions from the manufacturers.

Purification of RaRF and Its Polypeptide Components. RaRF was purified from delipidized serum of ICR mice by absorption onto a formalin-killed Ra chemotype strain, rfb388, of *S. typhimurium* (Kawakami et al., 1982), with subsequent extraction with a solution of *N*-acetyl-D-glucosamine (Ihara et al., 1991). After reduction of RaRF with dithiothreitol and carboxymethylation with iodoacetamide (Crestfield et al., 1963), P28a and P28b polypeptides were separated by polyacrylamide gel electrophoresis (PAGE) on a 15% gel that contained 6.25 M urea and 0.9 M acetic acid, as described by Panyim and Chalkley (1969). Polypeptides were recovered electrophoretically from the gel as described previously (Ihara et al., 1987). The resolved P28a and P28b (approximately 1 mol of each) were dissolved separately in 100 μL of 0.1 M ammonium bicarbonate that contained 0.1 mM $CaCl_2$ and 2 M urea, and then each was digested with 4 μg of 1-chloro-4-phenyl-3-L-(tosylamino)butan-2-one-treated trypsin (Seikagaku Co., Tokyo) at 37 °C for 24 h. The peptide fragments obtained were separated by reverse-phase HPLC on a BQ603 column (Gasukuro Kogyo Co., Tokyo). The column was eluted with linear gradients of solvents made by mixing 20% acetonitrile in 0.1% trifluoroacetic acid and 80% acetonitrile in 0.08% trifluoroacetic acid.

Determination of Amino Acid Sequences. Automated Edman degradation was conducted in a gas-phase sequencer (Model 470A; Applied Biosystems Inc., Foster City, CA). Phenylthiohydantoin derivatives of amino acids were identified either by reverse-phase HPLC on an SSC-SEQ-4 column (Senshu Kagaku, Tokyo) with a single-wavelength (270 nm) detector or by on-line injection with a PTH analyzer (Model 120A, Applied Biosystems).

Screening of cDNA Library. Poly(A)$^+$ RNA was prepared from total RNA from the liver of a female BALB/c mouse by the standard method (Maniatis et al., 1982). Oligo(dT)-primed first-strand synthesis and RNase H- and DNA polymerase I-mediated second-strand synthesis of cDNA were carried out by the method of Gubler and Hoffman (1983). After treatment with *Eco*RI methylase and addition of *Eco*RI linker, double-stranded cDNA was ligated to *Eco*RI-digested λgt10, as described by Huynh et al. (1988). The DNA mixture was then packaged in vitro to generate phage particles by the standard method (Maniatis et al., 1982). The cloning efficiency was 3×10^7 recombinants per microgram of double-stranded cDNA. Approximately 3×10^5 recombinant phage particles were used to infect host bacteria, and propagated phages were used for cDNA cloning.

Two pairs of oligonucleotide probes with the following sequences were synthesized and used to screen P28a and P28b cDNAs:

```
pa-1  5'-GCTACTTTGAaCCCCTTCGGTTAAGGTCTC-3'

pa-2  5'-CTCATTCCAGTTAGTGTAGCGCACTCTGTTTCCTGTCAGGTCCTCAAA-3'

pb-1  5'-AGGGTgTCCTCACAGGTTTGTGACCCTG-3'

pb-2  5'-ATTGGGCTCATCCTTTTTCCAGTTGCTGTAGGTGAGCCTCCCCCCTG-3'
```

Probes pa-1 and pa-2 are complementary to nucleotides 253–264 and nucleotides 751–796 of rat MBP-C cDNA, respectively, while pb-1 and pb-2 are complementary to nucleotides 11–38 and nucleotides 524–570 of rat MBP-A cDNA, respectively, with the exception that the bases indicated by small letters were modified by reference to the amino acid sequences of tryptic peptides of P28a and P28b. Oligonucleotides were synthesized, using a DNA synthesizer (Model 380B; Applied Biosystems), purified by PAGE, and labeled with [γ-^{32}P]ATP, using T4 polynucleotide kinase (Maniatis et al., 1982). The specific activity was approximately 2×10^6 cpm/pmol of oligonucleotide.

Phage DNA, transferred to nitrocellulose paper (Schleicher and Schuell Co.) was allowed to hybridize with the oligonucleotide probes at 60 °C for 14 h in 180 mM Tris-HCl, pH 8.0, that contained 0.9 M NaCl, 6 mM EDTA, 0.5% SDS, 5 × Denhardt's solution, and 0.5 pmol/mL radiolabeled probe. The filters were then washed four times at room temperature with 90 mM trisodium citrate that contained 0.9 M NaCl for 5 min, and once at 60 °C with the same solution for 2 min.

DNA Sequence Determination. The cloned DNAs were subcloned into the pUC118 vector. The cDNA fragments were subjected to sequential deletions (Henikoff, 1984) with exonuclease III and mung bean nuclease, using the Deletion Kit for Kilo-sequence (Takara Shuzo Co.) or by digestion with restriction enzymes. Single-stranded DNAs were recovered from the M13 phage in association with the M13 helper virus in *E. coli* harboring pUC plasmids, as described elsewhere (Vieira & Messing, 1987), and then they were sequenced by the dideoxy method (Sanger et al., 1977). Sequence data were analyzed by the sequence-analysis software packages of the University of Wisconsin Genetics Computer Group, UWGCG (Devereux et al., 1984), and the Integrated Database and Extended Analysis System for Nucleic Acids and Protein, IDEAS (Kanehisa, 1982). Programs were run on a VAX-11/750 computer system (Digital Equipment Co., Tokyo) at the Institute of Medical Science, University of Tokyo.

Northern and Southern Blot Analyses. Poly(A)$^+$ RNAs were separated by electrophoresis on 1% agarose gel in the presence of glyoxal and transferred onto nitrocellulose membrane as described elsewhere (Maniatis et al., 1982; Kuge et al., 1986). After digestion with various restriction endonucleases, genomic DNA was fractionated by electrophoresis on 1% agarose gel and transferred to nitrocellulose paper as

Table I: Amino Acid Sequences of P28a and P28b of RaRF[a]

	P28a (a-N)		P28b (b-N)	
cycle		pmol		pmol
1	Glu	159	Ser	108
2			Gly	150
3	Leu	86	Ser[b]	ND[c]
4	Thr[b]	ND[c]	Gln	95
5	Glu	31	Thr	12
6	Gly	52	Cys	ND[c]
7	Val	73	Glu	80
8	Gln	40	Asp	87
9	Asn	16	Thr[b]	ND[c]
10	Ser[b]	ND[c]	Leu	42
11			Lys	32
12	Pro	21	Thr[b]	ND[c]
13	Val	28		
14	Val	25	Ser[b]	ND[c]
15			Val	35
16			Ile	28
17			Ala	33

[a] Yields of amino acids estimated by HPLC. [b] Identified at 313 nm. [c] Not determined.

described previously (Kuge et al., 1986). ^{32}P-Labeled hybridization probes (approximately 2×10^9 cpm/μg of DNA) were prepared from the purified restriction fragments of P28a and P28b cDNAs, using the Random Primed DNA Labeling kit (Boehringer-Mannheim). Hybridization was carried out at 65 °C for 14–16 h in 60 mM trisodium citrate containing 0.6 M NaCl, 2.5 × Denhardt's solution, 0.5% SDS, 40 μg/mL poly(A), 0.1 mg/mL denatured salmon sperm DNA, and denatured ^{32}P-labeled DNA probe, after prehybridization at 65 °C for 3 h in the same mixture without labeled DNA and poly(A). The paper was washed in 15 mM trisodium citrate containing 0.15 M NaCl at 65 °C for 45 min, as described previously (Kuge et al., 1986).

RESULTS

Determination of Partial Amino Acid Sequences of P28 Polypeptides. The P28a and P28b polypeptides were resolved by acid/urea–PAGE, and the amino acid sequences of their NH_2-terminal regions were determined (Table I). Three tryptic fragments each were obtained from P28a and P28b, and they were also sequenced after the isolation (Table II). Comparison of the amino acid sequences with those of other polypeptides of known primary structure indicated that the partial amino acid sequences of the P28a polypeptide exhibited the highest homology (65%) to the MBP-C polypeptide of rat mannan-binding protein, while the partial sequence of P28b exhibited the highest homology (81%) to MBP-A (Drickamer et al., 1986; Oka et al., 1987). These results prompted us to make DNA probes by reference to the sequences of cDNAs for MBPs, and to clone the cDNAs of the P28 polypeptides, using these probes.

The amino acid sequence of the NH_2 region of the P28a polypeptide (a-N) is homologous to that of MBP-C (residues 1–14). Therefore, oligonucleotide probe pa-1 was designed by reference to the a-N sequence and to the base sequence of the cDNA for MBP-C that corresponds to this region. The amino acid sequence of the P28a a-25 fragment exhibits homology to the COOH-terminal region of MBP-C (residues 169–180). Thus, the pa-2 probe was designed by reference to this region of MBP-C. In the same way, the pb-1 and pb-2 probes were prepared by reference to the sequences of the NH_2- and COOH-terminal regions of P28b (b-N and b-34) and MBP-A (residues 1–17 and 172–187). The sequences of these probes are complementary to the mRNA.

Isolation of cDNA Clones from a Mouse cDNA Library. cDNAs corresponding to P28a and P28b were cloned from a mouse liver cDNA library constructed in λ gt10 vectors, with the synthetic DNAs used as probes. When the mouse liver cDNA library was screened with probe pa-2, eight positive plaques were detected out of 250 000 screened. After purification of the plaques, five clones were found to hybridize with both probes pa-1 and pa-2. Of the clones purified from these plaques, a7 and a10 were found to contain longer inserts that hybridized with these probes, as determined by agarose gel electrophoresis after digestion with *Eco*RI. In the same way, six clones hybridizing with both the pb-1 and pb-2 probes were isolated from ten positive plaques obtained by the screening with probe pb-2. Finally, clones b51 and b60 carrying large inserts were selected.

Characterization of cDNA Sequences. The nucleotide sequences of these inserts were determined after subcloning of each cDNA fragment into pUC118 vector. Results of partial determination of nucleotide sequence suggested that b51 cDNA was simply a part of b60 cDNA, whereas there was a difference between the sequences of a7 and a10 (data not shown). Therefore, the complete nucleotide sequences of a7, a10, and b60 cDNA clones were determined. The restriction maps and strategies for sequencing of these cDNAs are shown in Figure 1. The complete nucleotide sequences of a10 and b60 cDNA are indicated in Figures 2 and 3, respectively. The lengths of the coding regions deduced from the open reading frames of a10 and b60 are 732 and 717 nucleotides, respectively. Sequence identity between the coding region of a10 cDNA and that of rat MBP-C cDNA (Drickamer et al., 1986; Oka et al., 1987) is 86.2% in terms of nucleotide sequence and 81.2% in terms of amino acid sequence. Between the coding region of b60 cDNA and that of rat MBP-A cDNA (Drickamer et al., 1986), there is 91.3% in terms of nucleotide sequence and 89.5% in terms of amino acid sequence. It is also possible to align the amino acid sequences encoded by a10 cDNA and b60 cDNA with each other to give 52.0% identity, which is consistent with the homology between rat MBP-C and MBP-A (52.2%). The sequence of a possible poly(A)-addition signal is found at nucleotide 1038 in the 3′-untranslated sequence prior to the oligo-d(A) sequence of a10 cDNA, but not in the b60 cDNA sequence (Figure 3). The sequence of a7 is the same as that of a10, except that heterogeneity in nucleotide sequence is found in their 5′-untranslated regions. As shown in Figure 4, the 5′-untranslated regions of a10 and a7 cDNAs can be aligned at three homologous sequences I–III. However, there are additional sequences between sequences II and III in a10, and between sequences I and II in a7. The possible involvement of alternative splicing that generates two species of mRNA with different untranslated sequences is discussed below.

Amino Acid Sequences Deduced from the cDNA Sequences. The amino acid sequences of the NH_2-terminal regions and tryptic fragments of P28a and P28b are shown in Figures 2 and 3, together with those deduced from the a10 and b60 cDNA sequences. There is a sequence of 18 amino acids, which is thought to be a signal peptide, in the predicted sequence of both P28a and P28b. The determined sequences of all P28b fragments can be accounted for in the amino acid sequence predicted from the b60 cDNA (Figure 3). However, the amino acid sequence predicted from the a10 cDNA differs from the determined sequence at residues 60, 158, and 172 in the P28a polypeptide (Figure 2). While the P28a polypeptide was isolated from ICR mice, the a10 cDNA was prepared from BALB/c liver cDNA library. As discussed

Table II: Amino Acid Sequences of Tryptic Peptides Derived from P28a and P28b of RaRF[a]

	a-15		a-35		a-25		b-24		b-34		b-44	
cycle		pmol		pmol		pmol		pmol		pmol		pmol
1	Val	93	Gly/Ser[b]	101/50	Val	115	Leu	370	Ser[b]	20	Lys	50
2	Gly	87	Ile	138	Glu	96	His	60	Leu	160	Asp	47
3	Pro	53	Ala	205	Gly	98	Ala	140	Cys	ND[c]	Glu	50
4	Thr[b]	ND[c]	Tyr	137	His	56	Phe	210	Thr	10	Pro	45
5	Gly	56	Leu	152	Phe	91	Ser	20	Glu	90	Asn	13
6	Pro	49	Gly	99	Glu	74	Met	110	Leu	70	Asn	35
7	His	31	Ile	97	Asp	53	Gly	80	Gln	100		
8	Gly	67	Thr[b]	17	Leu	65	Lys	70	Gly	90	Gly	19
9	Asn	36	Asp	32	Thr[b]	ND[c]						
10			Val	88	Gly	60			Val	70		
11					Asn	51			Ala	70	Glu	9
12					Arg	ND[c]			Ile	60		

[a] Yields of amino acids estimated by HPLC. [b] Identified at 313 nm. [c] Not determined.

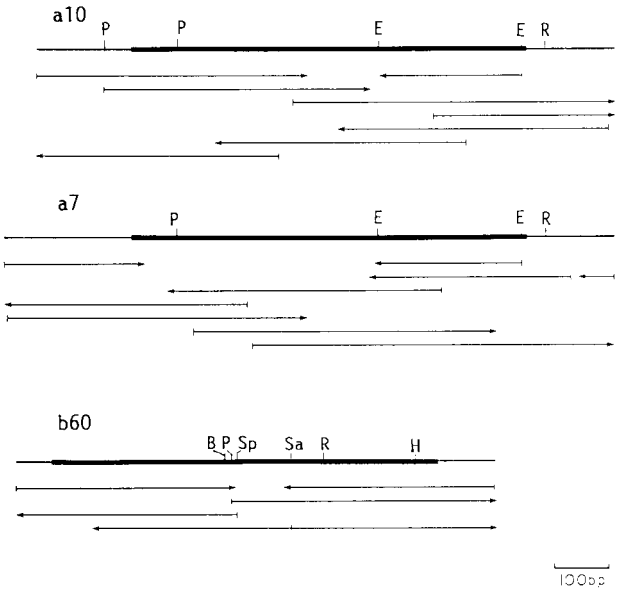

FIGURE 1: Restriction maps and sequencing strategies for a10, a7, and b60 cDNAs. The cDNA is indicated at the top of the each illustration with its coding (bold line) and noncoding regions (thin line). Relevant restriction sites are shown: P, *Pst*I; E, *Eco*RI; R, *Rsa*I; B, *Bam*HI; Sp, *Sph*I; Sa, *Sac*I; H, *Hin*dIII. The a10 and a7 cDNAs, each having two *Eco*RI cleavage sites, were subcloned using both *Hin*dII (map position 32.47 kilobase pairs) and *Bgl*II (map position 33.61 kilobase pairs) sites of vector phage. Poly(dA)/poly(dT) stretches of 90, 66, and 18 bases long (omitted from the figure) were present at the downstream ends of a10, a7, and b60 cDNAs, respectively. The *Eco*RI linker nucleotide sequence, GGAATTCC, used for construction of the cDNA library, is found at both ends of each insert. The horizontal arrows indicate the direction and extent of sequence analyses by the dideoxy chain-termination method.

below, this difference may be the result of allotypic differences between the P28a proteins in these strains. The amino acid compositions determined from both P28a and P28b polypeptides (Ihara et al., 1991) are very similar to those deduced from a10 and b60 cDNAs, respectively. The calculated molecular masses were 24 kDa for the polypeptide encoded by a10 and 23.5 kDa for that encoded by b60. The molecular masses of P28a and P28b (28 kDa) as estimated from gel electrophoresis exceed those calculated from the sequences. This discrepancy is thought to be the result of the glycosylation of polypeptides, since the bands of the polypeptides after PAGE can be stained by Schiff's reagent (data not shown).

Analysis of the Genomic DNA of Mouse and Other Mammals. Southern blot analysis of mouse genomic DNA was carried out (Figure 5). Mouse genomic DNA was digested with *Eco*RI or *Hin*dIII and probed with a 264-base-pair fragment of a10 or a 237-base-pair fragment of b60 cDNA, respectively. No *Hin*dIII sites are found in a10 cDNA and no *Eco*RI sites in b60 cDNA (Figure 1). A single band (2.2 kilobases) was observed when *Hin*dIII-digested mouse DNA was probed with the a10 fragment (panel A, lane 4). Probing *Eco*RI-digested mouse DNA with the b60 fragment, we also found a single band (1.8 kilobases; panel B, lane 10). The sequence homology between the a10 fragment and intact b60 cDNA is 66.3%, and that between the b60 fragment and intact a7 or a10 cDNA is 70.9%. Cross-hybridization of these probes with the cDNAs gave only weak bands (lanes 1 and 7), indicating the satisfactory stringency of this hybridization. The intensities of the bands of mouse genomic DNA (lanes 4 and 10) corresponded to that of a7 or b60 cDNA that were equivalent to a single copy per haploid genome (lanes 3 and 9). Single bands were also produced when the *Rsa*I or *Sal*I digest of mouse genomic DNA was probed with the a10 or b60 fragment, respectively (data not shown). The intensity of the bands again corresponded to a single copy per haploid genome. Therefore, each of the two genes corresponding to a10 and b60 cDNA is present as approximately one single copy per haploid genome.

RaRF activity has been found in the sera of various mammals (Kawakami et al., 1984), suggesting the presence in these animals of genes which are similar to the P28 genes of mice. In this study, genomic DNAs isolated from rat liver and a human cell line were examined by Southern blotting on the same gel as mentioned above (Figure 5). Rat DNA hybridizing with the b60 cDNA probe gave a single band (lane 11), whereas the same DNA hybridizing with the a10 probe gave several bands of considerable intensity (lane 5). It is suggested that the rat DNAs which hybridize with a10 and b60 are MBP-C and MBP-A, respectively. In the rat genome, there is a pseudogene for MBP-C in addition to the transcribed genes for MBP-A and MBP-C (Drickamer et al., 1987). Probably, the presence of MBP-C pseudogene may result in several bands in the blotting of rat DNA with the a10 probe. Human cell DNA hybridized with the b60 probe (lanes 12). Probably, the DNA hybridizing with b60 is the gene for human MBP since the b60 and human MBP cDNAs (Ezekowitz et al., 1988) have fairly homologous sequences. In contrast to b60 probe, the a10 probe did not hybridize with the human cell DNA (lane 6). This may be due to the absence of human genes of MBP that have sequence homology with the P28a DNA.

Tissue Expression of P28 RNAs. Northern blot hybridization analysis was undertaken to identify those tissues that express P28a and P28b mRNA. Poly(A)$^+$ RNAs prepared from mouse liver, lung, spleen, kidney, and brain were tested for their ability to hybridize with cDNA for P28a (Figure 6,

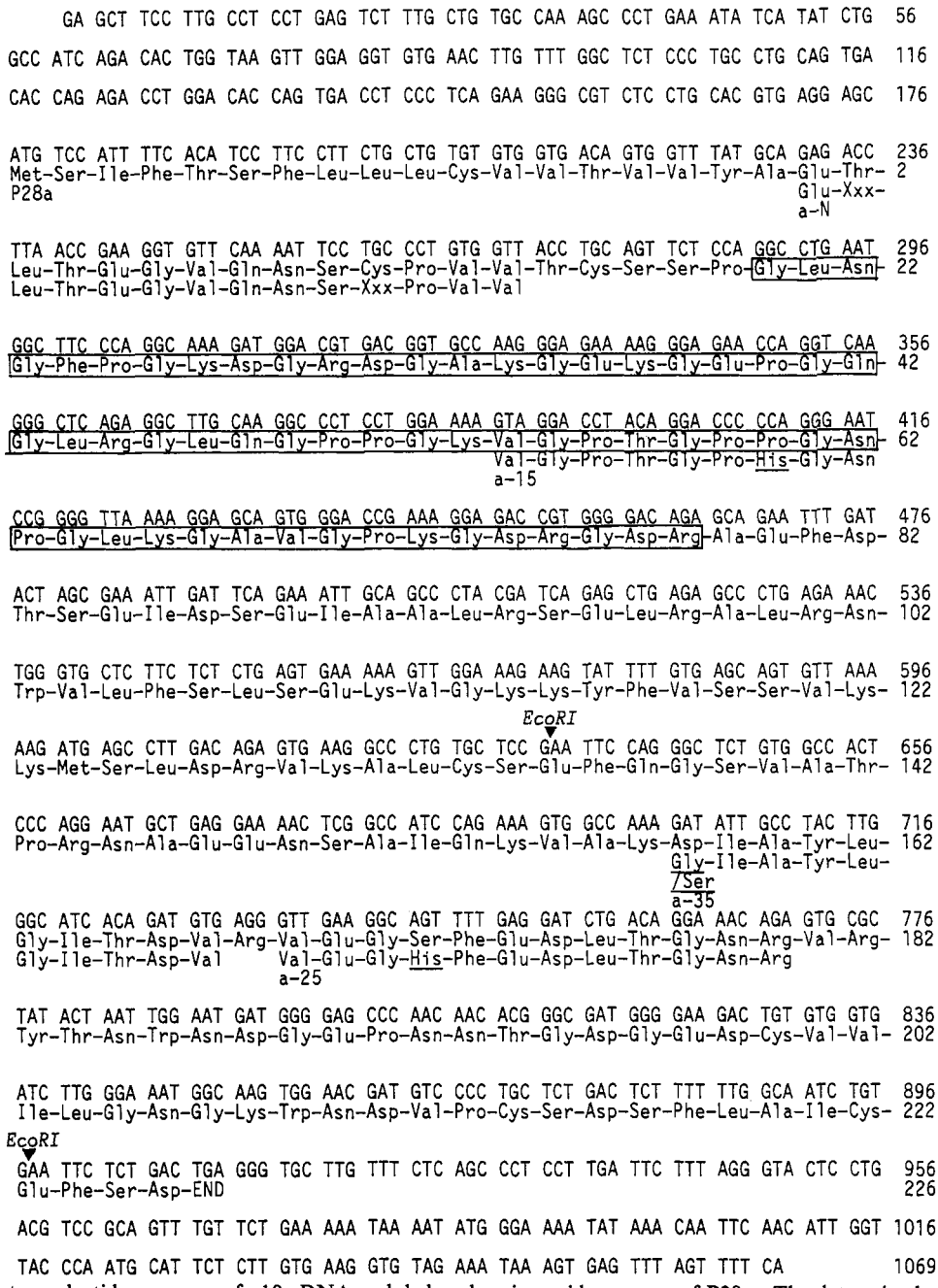

FIGURE 2: Complete nucleotide sequence of a10 cDNA and deduced amino acid sequence of P28a. The determined amino acid sequences of the NH_2-terminal region of the intact protein (a-N) and the tryptic peptides (a-15, a-35, and a-25) of P28a are aligned with the amino acid sequence predicted from the nucleotide sequence. The underlined amino acids differ between the predicted and determined sequences. Boxes indicate the repeats of the amino acid sequence, Gly-X-Y. The poly(dA) sequence corresponding to the poly(A) structure has been omitted from the sequence. Triangles indicate the sites of restriction enzymes used for the preparation of probe (a10 fragment).

panel A) and P28b (Figure 6, panel B). Single bands appeared in the case of poly(A)$^+$ RNA from liver. This result is consistent with the results of an immunohistological study which demonstrated that anti-RaRF antibodies reacted only with liver cells (Ohmomo et al., unpublished results). It is, thus, likely that the P28a and P28b polypeptides of mouse RaRF are both synthesized at least in liver cells. From the lanes of RNA from lung, spleen, kidney, and brain, no bands were detected. However, further analysis is necessary to obtain a conclusive result.

DISCUSSION

When we cloned the cDNA for P28a, we obtained two clones, a7 and a10, that have the same sequences at their coding regions. Differences between these DNAs occur only in the sequences of the 5′-untranslated regions (Figure 4). It is unclear whether this difference reflects the presence of different mRNAs or is an artifact that results from the cloning of cDNAs. S1 nuclease protection analysis of total mouse liver RNA indicated that the major population of transcripts consists of RNAs with untranslated sequences of the a10 type (unpublished results), suggesting the possible artifactual production of a7. However, part of the 5′-untranslated region characteristic of a7 cDNA (nucleotides 136–233) is homologous to that of rat MBP-C cDNA (nucleotides 82–178; Drickamer et al., 1986). Furthermore, a splice-junction sequence is present in the 5′-side sequence preceding the homologous region II of a7 cDNA (tagG), as well as in the 3′-side sequence next to region II of a10 cDNA (gtaa). Thus, possibly, different splicing events may give rise the 5′-untranslated sequences of the a7 and a10 types.

```
                                                                  C TGG  4

ACT CGA GAC ATA GTT TCT CTT CCA CTG CTC CTT TAC TCT AAA GAA ACC CTA GTA AGG ACC  64

ATG CTT CTG CTT CCA TTA CTC CCT GTC CTT CTG TGT GTG GTG AGT GTG TCC TCA TCA GGG 124
Met-Leu-Leu-Leu-Pro-Leu-Leu-Pro-Val-Leu-Leu-Cys-Val-Val-Ser-Val-Ser-Ser-Ser-Gly- 2
P28b                                                                 Ser-Gly-
                                                                     b-N

TCA CAA ACC TGT GAG GAC ACC CTG AAG ACT TGC TCT GTG ATA GCC TGT GGC AGA GAT GGG 184
Ser-Gln-Thr-Cys-Glu-Asp-Thr-Leu-Lys-Thr-Cys-Ser-Val-Ile-Ala-Cys-Gly-Arg-Asp-Gly- 22
Ser-Gln-Thr-Cys-Glu-Asp-Thr-Leu-Lys-Thr-Xxx-Ser-Val-Ile-Ala

AGA GAT GGA CCC AAA GGG GAG AAG GGA GAA CCA GGT CAA GGG CTC AGG GGC TTG CAG GGC 244
Arg-Asp-Gly-Pro-Lys-Gly-Glu-Lys-Gly-Glu-Pro-Gly-Gln-Gly-Leu-Arg-Gly-Leu-Gln-Gly- 42

CCT CCA GGG AAA TTG GGG CCT CCA GGA AGT GTT GGA AGC CCT GGA AGT CCA GGA CCA AAA 304
Pro-Pro-Gly-Lys-Leu-Gly-Pro-Pro-Gly-Ser-Val-Gly-Ser-Pro-Gly-Ser-Pro-Gly-Pro-Lys- 62

GGC CAA AAG GGG GAC CAT GGA GAC AAT AGA GCC ATT GAG GAG AAG CTG GCA AAT ATG GAG 364
Gly-Gln-Lys-Gly-Asp-His-Gly-Asp-Asn-Arg-Ala-Ile-Glu-Glu-Lys-Leu-Ala-Asn-Met-Glu- 82

GCA GAG ATA AGG ATC CTG AAA TCA AAA CTG CAG CTA ACC AAC AAG TTG CAT GCC TTC TCA 424
Ala-Glu-Ile-Arg-Ile-Leu-Lys-Ser-Lys-Leu-Gln-Leu-Thr-Asn-Lys-Leu-His-Ala-Phe-Ser- 102
                                                        Leu-His-Ala-Phe-Ser-
                                                        b-24

ATG GGC AAA AAG TCT GGG AAG AAG TTG TTT GTG ACC AAC CAT GAG AAG ATG CCC TTT TCC 484
Met-Gly-Lys-Lys-Ser-Gly-Lys-Lys-Leu-Phe-Val-Thr-Asn-His-Glu-Lys-Met-Pro-Phe-Ser- 122
Met-Gly-Lys
                          SacI
                           ▼
AAA GTG AAG TCT CTG TGC ACA GAG CTC CAA GGC ACT GTG GCT ATC CCC AGG AAT GCT GAA 544
Lys-Val-Lys-Ser-Leu-Cys-Thr-Glu-Leu-Gln-Gly-Thr-Val-Ala-Ile-Pro-Arg-Asn-Ala-Glu- 142
           Ser-Leu-Cys-Thr-Glu-Leu-Gln-Gly-Xxx-Val-Ala-Ile
           b-34

GAG AAC AAG GCC ATT CAA GAA GTG GCC ACA GGC ATT GCC TTC CTA GGC ATC ACG GAC GAG 604
Glu-Asn-Lys-Ala-Ile-Gln-Glu-Val-Ala-Thr-Gly-Ile-Ala-Phe-Leu-Gly-Ile-Thr-Asp-Glu- 162

GCG ACT GAA GGG CAG TTC ATG TAC GTG ACA GGG GGG AGG CTC ACC TAC AGC AAC TGG AAA 664
Ala-Thr-Glu-Gly-Gln-Phe-Met-Tyr-Val-Thr-Gly-Gly-Arg-Leu-Thr-Tyr-Ser-Asn-Trp-Lys- 182

AAG GAT GAG CCA AAT AAC CAT GGC TCT GGG GAA GAC TGT GTC ATT ATA TTA GAT AAT GGT 724
Lys-Asp-Glu-Pro-Asn-Asn-His-Gly-Ser-Gly-Glu-Asp-Cys-Val-Ile-Ile-Leu-Asp-Asn-Gly- 202
Lys-Asp-Glu-Pro(Asn)Asn-Xxx-Gly-Xxx-Xxx-Glu
b-44
                      HindIII
                        ▼
TTG TGG AAT GAC ATT TCC TGT CAA GCT TCC TTC AAG GCT GTC TGC GAG TTC CCA GCC TGA 784
Leu-Trp-Asn-Asp-Ile-Ser-Cys-Gln-Ala-Ser-Phe-Lys-Ala-Val-Cys-Glu-Phe-Pro-Ala-END 221

GGA AAC GAG TGC CTC CAT ATT CTC CTT GCC TCC TCT CTG GAC TCT CAC TTG CTT CCA AAG 844

AAA ATT CAG TAC TTG TTT CTC A                                              866
```

FIGURE 3: Complete nucleotide sequences of b60 cDNA and deduced amino acid sequence of P28b. The determined amino acid sequences of the NH_2-terminal region of the intact protein (b-N) and the tryptic peptides (b-24, b-34, and b-44) of P28b are aligned with the amino acid sequence predicted from the nucleotide sequence. Boxes indicate the repeats of the amino acid sequence, Gly-X-Y. The poly(dA) sequence corresponding to the poly(A) structure has been omitted from the sequence. Triangles indicate the sites of restriction enzymes used for the preparation of probe (b60 fragment).

The discrepancy between the amino acid sequence deduced from a10 cDNA and that determined by the direct analysis of the P28a polypeptide suggests that several similar proteins might be produced by multiple genes that are related to one another. However, the present Southern blotting analysis demonstrated the presence of two unique genes corresponding to a10 and b60 cDNAs, respectively. Therefore, such a possibility seems unlikely, and thus, the discrepancy in amino acid sequence is thought to arise from allotypic heterogeneity between BALB/c and ICR mice.

A short, specific NH_2-terminal domain, followed by a collagen-like domain and C-terminal domain, is present in all of these polypeptides. In the collagen-like region of the A and C chains of C1q, there is not only the Gly-X-Y repeating sequence but also a structural irregularity, which has been suggested to form a kink in the triple helix of the C1q molecule (Kilchherr et al., 1985). This irregularity is also found in the collagen-like region of the P28 polypeptides (Figures 2 and 3).

Like the C1q chains, the MBP chains of rats and humans are composed of a specific NH_2-terminal domain, a collagen-like domain, and a COOH-terminal domain (Drickamer et al., 1986; Oka et al., 1987; Ezekowitz et al., 1988). The P28 polypeptides are similar to the MBPs not only in the constitution of these domains but also in the overall sequences; there is 81.2% homology between P28a and MBP-C, and 89.5% between P28b and MBP-A. There are cysteine residues in the NH_2-terminal regions of both the C1q chains (Porter & Reid, 1979) and rat MBP-C (Drickamer et al., 1986; Oka et al., 1987) which are thought to be involved in the formation of interchain disulfide bonds. Such cysteine residues are also found in the NH_2-terminal domain of the P28 polypeptides. Therefore, it appears that the P28a polypeptide is a mouse protein corresponding to rat MBP-C, while P28b corresponds to MBP-A.

One of the unique characteristics of the COOH-terminal domain of MBP is the presence of a sequence common to certain animal lectins (Drickamer et al., 1986). The COOH-

```
                                   I
a10   GAGCTTCCTTGCCTCCTGAGTCTTTGCTGTGCCAAAGCCCTGAAATATCATATCTG                                  56
a7    GAGCTTCCTTGCCTCCTGAGTCTTTGCTGTGCCAAAGCCCTGAAATATCATATCTGcctatagtttattgctagaaacgg        80

a10                                                                                          56
a7    aggaaagcctatctaactggttttggagagaaaccatcttttccagtgtgaaaataggattggggctgctgatggtggca      160
MBP-C                                                                 ggatttgggctgcagaccgtggca 105
                                                                                II
a10                                                                 GCCATCAGACACTGgtaagttggaggtg  84
a7    cacatgtgagagccttctgaacatagtgtctccgctcgttctcaaccctagGCCATCAGACACTG                        226
MBP-C cacatgtgagatccttcggaacacagtgtctccggtcatcctcaaccctaagccatccgacactg                       171

a10   tgaacttgtttggctctccctgcctgcagtgacaccagagacctggacaccagtgacctccctcagaagggcgtctcctg       164
a7                                                                                           226
MBP-C                                                                                        171

          III
a10   cacGTGAGGAGCATG                                                                        179
a7      GTGAGGAGCATG                                                                         238
MBP-C   gtgaggaccatg                                                                         183
```

FIGURE 4: Nucleotide sequence heterogeneity between the 5′-noncoding sequences of a10 and a7 cDNAs. The 5′-noncoding sequences of a10 cDNA (176 bases long) and a7 cDNA (235 bases long), together with that of rat MBP-C cDNA, are aligned with respect to the three homologous regions, I–III. Splice-junction sequences are underlined. Initiation codons at the downstream ends are included in boldface.

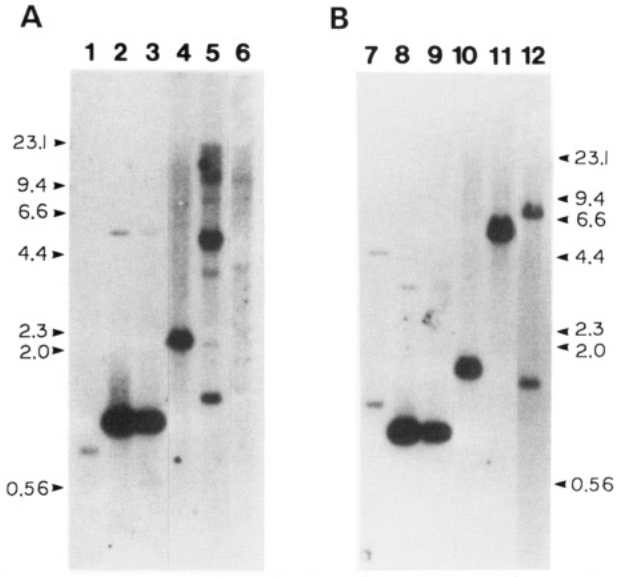

FIGURE 5: Southern blot hybridization of genomic DNAs from various mammals. DNA (10 μg) from mouse liver (lanes 4 and 10), rat liver (lanes 5 and 11), and HeLa cells (lanes 6 and 12) was digested with *Hin*dIII (A) or *Eco*RI (B), subjected to 1% agarose gel electrophoresis, transferred to nitrocellulose filter paper, and allowed to hybridize with a ^{32}P-labeled, 264-base-pair *Eco*RI–*Eco*RI fragment of a10 cDNA (nucleotides 634–897 in Figure 2) (A) or with a ^{32}P-labeled, 237-base-pair *Sac*I–*Hin*dIII fragment of b60 cDNA (nucleotides 511–747 in Figure 3) (B). The 1.1-kilobase-pair *Rsa*I fragment composed almost entirely of a7 cDNA (lanes 2, 3, and 7) and 0.87-kilobase-pair b60 cDNA (lanes 1, 8, and 9) were analyzed on the same gel. The amounts were equivalent to one copy per haploid genome (lanes 3 and 9) or to three copies per haploid genome (lanes 1, 2, 7, and 8). The stringency in this experiment refers to the patterns of hybridization with aliquots of mouse cDNAs (lanes 1 and 7). Positions of DNA molecular size markers are indicated in kilobase pairs on the left and right sides of the autographs.

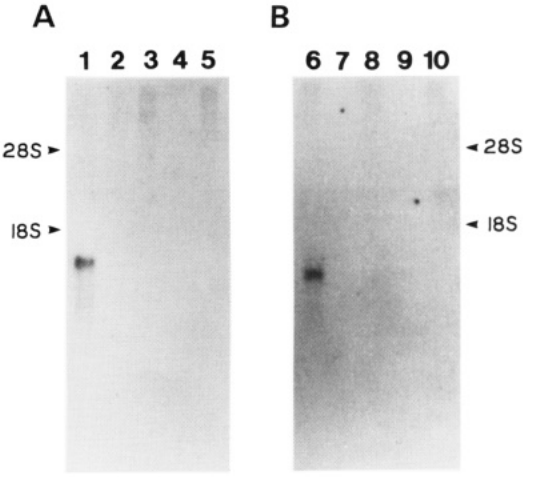

FIGURE 6: Northern blot analysis of poly(A)$^+$ RNA isolated from various mouse tissues. Poly(A)$^+$ RNA was prepared from mouse liver (lanes 1 and 6), lung (lanes 2 and 7), spleen (lanes 3 and 8), kidney (lanes 4 and 9), and brain (lanes 5 and 10), subjected to agarose gel electrophoresis, blotted on nitrocellulose filter paper, and allowed to hybridize with a ^{32}P-labeled, 264-base-pair fragment of a10 cDNA (A) or with a ^{32}P-labeled, 237-base-pair fragment of b60 cDNA (B) (see legend to Figure 5). Positions of mouse ribosomal RNAs are indicated on the left and right of the autographs.

terminal domains of P28a and P28b also exhibit homologies with the lectin-like region of rat MBPs. The capacity for specific binding to Ra polysaccharide of the RaRF complex is associated with the polysaccharide-binding component which is composed of the P28a and P28b polypeptides (Ihara et al., 1991). Therefore, it is reasonable to assume that the COOH-terminal domain of P28 polypeptides is a carbohydrate-recognition region and participates in the specific binding of the RaRF complex to Ra polysaccharide.

Rat serum MBP is an oligomer of homologous polypeptides corresponding to MBP-A (Oka et al., 1987). Ikeda et al. (1987) have demonstrated that MBP activates the C4-dependent pathway of complement. Lu et al. (1990) have reported that human MBP, which is also an oligomer of homologous polypeptides similar to MBP-A, activates C4, using C1r and C1s in serum. In contrast to MBP, the RaRF complex can activate C4 independently. Since the RaRF complex activates the C4 and C2 components of complement without participation of C1, the complex is thought to contain a C1s-like serine protease that is capable of cleaving C4 and C2 (Mackinnon et al., 1987). However, the C4-activating capacity is not associated with the polysaccharide-binding component (Ihara et al., 1991), and no sequences characteristic of serine proteases are found in P28a and P28b. Therefore, a polypeptide component(s) other than the P28 polypeptides plays (play) a role in the activation of C4 and C2. Probably, the 29- and 70-kDa polypeptides in the low-molecular-weight component are responsible for the C4/C2 activation (Ihara et al., 1988, 1991). Taking these results into consideration, we propose that RaRF is a complex composed of the C4-activating proteases, together with the P28a and P28b.

ACKNOWLEDGMENT

We are grateful to Masao Ono and Haruo Suzuki for many useful suggestions and discussions and to Masayuki Yamamoto and Akira Itoh for their help in sequence analysis using the VAX computer system.

REFERENCES

Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Drickamer, K., & McCreary, V. (1987) *J. Biol. Chem. 262*, 2582–2589.

Drickamer, K., Dordal, M. S., & Reynolds, L. (1986) *J. Biol. Chem. 261*, 6878–6887.

Ezekowitz, R. A. B., Day, L. E., & Herman, G. A. (1988) *J. Exp. Med. 167*, 1034–1046.

Gubler, U., & Hoffman, B. J. (1983) *Gene 25*, 263–269.

Henikoff, S. (1984) *Gene 28*, 351–359.

Huynh, T. V., Young, R. A., & Davis, R. W. (1988) in *DNA Cloning* (Glover, D. M., Ed.) Vol. 1, pp 49–78, IRL Press, Oxford.

Ihara, I., Harada, Y., Ihara, S., & Kawakami, M. A. (1982a) *J. Immunol. 128*, 1256–1260.

Ihara, I., Ueda, H., Suzuki, A., & Kawakami, M. (1982b) *Biochem. Biophys. Res. Commun. 107*, 1185–1190.

Ihara, I., Ihara, S., Nagashima, A., Ji, Y. H., & Kawakami, M. (1988) *Biochem. Biophys. Res. Commun. 152*, 636–641.

Ihara, S., Suzuki, H., & Kawakami, M. (1987) *Anal. Biochem. 166*, 349–352.

Ihara, S., Takahashi, A., Hatsuse, H., Sumitomo, Doi, K., & Kawakami, M. (1991) *J. Immunol. 146*, 1874–1879.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., & Yamashina, I. (1987) *J. Biol. Chem. 262*, 7451–7454.

Ji, Y. H., Matsushita, M., Okada, H., Fujita, T., & Kawakami, M. (1988) *J. Immunol. 141*, 4271–4275.

Kanehisa, M. (1982) *Nucleic Acids Res. 10*, 183–196.

Kawakami, M., Ihara, I., Suzuki, A., & Harada, Y. (1982) *J. Immunol. 129*, 2198–2201.

Kawakami, M., Ihara, I., Ihara, S., Suzuki, A., & Fukui, K. (1984) *J. Immunol. 132*, 2578–2580.

Kawasaki, T., Mori, K., Oka, S., Kawasaki, N., Kozutumi, Y., & Yamashina, I. (1987) in *Vertebrate Lectins* (Olden, K., & Parent, J. B., Eds.) pp 92–107, Van Nostrand Reinhold, New York.

Kilchherr, E., Hofman, H., Steigemann, W., & Engel, J. (1985) *J. Mol. Biol. 186*, 403–415.

Kuge, S., Saito, I., & Nomoto, A. (1986) *J. Mol. Biol. 192*, 473–487.

Lu, J., Thiel, S., Wiedemann, H., Timpl, R., & Reid, K. B. M. (1990) *J. Immunol. 144*, 2287–2294.

Mackinnon, C. M., Carter, P. E., Smyth, J., Dunbar, B., & Fothergill, J. E. (1987) *Eur. J. Biochem. 169*, 547–553.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Oka, S., Itoh, N., Kawasaki, T., & Yamashina, I. (1987) *J. Biochem. 101*, 135–144.

Oka, S., Ikeda, K., Kawasaki, T., & Yamashina, I. (1988) *Arch. Biochem. Biophys. 260*, 257–266.

Panyim, S., & Chalkley, R. (1969) *Arch. Biochem. Biophys. 130*, 337–346.

Porter, R. R., & Reid, K. B. M. (1979) *Adv. Protein Chem. 33*, 1–71.

Reid, K. B. M. (1985) *Biochem. J. 231*, 729–735.

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

Sellar, G. G., Blake, D. J., & Reid, K. B. M. (1991) *Biochem. J. 274*, 481–490.

Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.

Wood, L., Pulaski, S., & Vogeli, G. (1988) *Immunology 17*, 115–120.