THE HUMAN MANNOSE-BINDING PROTEIN GENE

Exon Structure Reveals its Evolutionary Relationship to a Human Pulmonary Surfactant Gene and Localization to Chromosome 10

By KEDARNATH SASTRY,* GARY A. HERMAN,* LINDA DAY,* EILEEN DEIGNAN,* GAIL BRUNS,‡ CYNTHIA C. MORTON,§ AND R. ALAN B. EZEKOWITZ*

*From the *Division of Hematology/Oncology, Children's Hospital and Dana Farber Cancer Institute, Harvard Department of Pediatrics; the ‡Division of Genetics, Children's Hospital; and the §Division of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115*

The human mannose-binding protein (MBP)[1] is an acute phase serum protein of ~300 kD comprised of multimers of a 32-kD subunit (1-3). It is a member of an ever-growing family of animal lectins that share at least 18 invariant residues in their carbohydrate recognition domain (CRD). The family can be divided into membrane proteins and soluble proteins, and all bind ligands optimally at neutral pH in the presence of calcium (reviewed in reference 4). For membrane proteins, the CRD is attached to a membrane anchor domain as found in the asialoglycoprotein receptors of rat (5) and human (6, 7), the chicken hepatic lectin (8), the lymphocyte IgE Fc receptor (9), and a rat Kupffer cell-binding protein (10). This group has recently been expanded to include a lymphocyte homing receptor (11), a granular membrane protein of platelets and endothelium, GMP-140 (12), and ELAM-1, a cytokine-inducible endothelial cell receptor (13). The function of the attachment domains of the soluble lectins, which include MBP (14, 15), the apoprotein of pulmonary surfactant SP-A (16, 17), cartilage proteoglycan core protein (18), a fly (19) and sea urchin lectin (20), and an acorn barnacle lectin (21) are not known. Only in the rat and human MBP, and dog, human, and rabbit (22) surfactant A protein, are the noncarbohydrate recognition domains similar. In these proteins there is a collagen-like region that is preceded by a cysteine-rich NH_2-terminal region that mediates interchain disulphide bonds. As our knowledge of the function of MBP and surfactant A protein expands, it appears that each domain of these proteins may have a defined purpose.

Until recently, the function of MBP was not known. However, a function in host defense is suggested by its ability to bind yeast mannans (1) and interact with the complement cascade (23), and by our finding that MBP prevents infection of H9 lymphoblasts by HIV by binding to the high mannose glycans expressed on the

This work was supported by National Institutes of Health grant R01 AI-23786 and A Grant-in-Aid from the Squibb Medical Institute (R. A. B. Ezekowitz). Address correspondence to R. A. B. Ezekowitz, Division of Hematology-Oncology, The Children's Hospital, Enders Building, 7th Floor, 300 Longwood Avenue, Boston, MA 02115.

[1] *Abbreviations used in this paper:* CRD, carbohydrate recognition domain; MBP, mannose-binding protein.

envelope glycoprotein of HIV (24). In addition, in other experiments, we have shown that native and recombinant MBP can serve an opsonic role in the serum and thereby enhance clearance of mannose-rich pathogens by phagocytes (25). Interaction of MBP with these organisms results in attachment, uptake, and killing of opsonized bacteria, as well as activation of complement via the alternative complement pathway (26). As SP-A has similar overall organization to MBP, it may serve as a local opsonin in the lung. A role for SP-A in local host defense is supported by its ability to bind carbohydrates (27) and enhance phagocytosis of opsonized particles (28), as well as its ability to opsonize certain serotypes of group B streptococcus (R. A. B. Ezekowitz, unpublished observation). This suggests that MBP and human SP-A may have a common evolutionary origin and that each domain of the molecule may have a distinct function.

To gain greater insight into the structure-function relationship of various domains of human MBP and to enhance our understanding of the origins of this chimeric protein, we undertook to examine its gene structure and chromosomal localization. Like the human SP-A and rat MBP-A (29), the human MBP gene appears to have evolved by recombination of an ancestral nonfibrillar collagen gene with a gene that encodes carbohydrate recognition. Interestingly, the gene locus is located on the long arm of chromosome 10 at 10q11.2-q21, a region that includes the locus for multiple endocrine neoplasia type 2 (30).

Materials and Methods

Screening of Human Genomic Library. A human genomic library, prepared by partial digestion of genomic DNA with Sau 3AI and ligation into the Bam HI site of phage λ EMBL 3A was plated in *Escherichia coli* strain LE392 at a titre of ~10,000 plaque-forming units/plate and transferred in duplicate to nitrocellulose filters. The filters were screened with a radiolabeled cDNA fragment of 550 bp corresponding to the carbohydrate-binding domain in 6× SSC/5× Denhardts/0.1% SDS/5 mM EDTA/50 μg/ml salmon sperm DNA at 68°C. After hybridization overnight, the filters were washed to 0.1× SSC at 68°C. Of ~500,000 plaques screened, two positives were obtained, MBPG26 and MBPG22, containing ~18-20 kb of insert DNA. MBPG26 was further characterized using Southern blot analysis (Fig. 1).

Southern Blot Analysis/Subcloning of Relevant Regions. As shown in Fig. 1, relevant portions of the phage λ (MBPG26) clone were identified by Southern blot analysis using probes constructed from different portions of the cDNA molecule. 1-2 μg of DNA was digested with either Eco RI or Pst I, size fractionated on a 1% agarose gel, and transferred (31) to Magnagraph nylon membranes (American Laboratory Supply Co., Natick, MA). The blots were then hybridized under stringent conditions specified by the manufacturer with ^{32}P oligo-labeled cDNA probes. Fig. 1 *D* shows the location of the probes used relative to the MBP cDNA: (*a*) a 150-bp Pst I-Bst EII fragment spanning a portion of the 5′ cDNA; (*b*) a 400-bp Pst I fragment spanning the middle region of the protein-coding portion of the cDNA; and (*c*) a 2.2-kb Xba I-Eco RI fragment encoding the 3′ untranslated region of the cDNA. Fig. 1, *A-C* show the relevant Eco RI and Pst I fragments that hybridized to the cDNA probes. The fragments were then subcloned into phosphatased pUC13 and pUC19 vectors and transformed into competent *E. coli* strain DK1 (recb, recc, sbcb). The respective fragments that hybridized to the 3′, middle, and 5′ cDNAs were identified and subcloned in M13 or pUC vectors for further analysis.

DNA Sequence Analysis. DNA sequence analysis was accomplished using the strategy outlined in Fig. 2. The Sanger dideoxy chain termination method (32) was employed using both single-stranded M13 vectors (33) and double-stranded pUC vectors. The single-stranded vectors were then primed with a 20-bp universal primer and extended in the presence of α-[^{35}S]dCTP with the Klenow fragment of DNA Polymerase I (Amersham Corp., Arlington Heights, IL) as previously described (33). The pUC clones containing fragments hybridizing

to the middle cDNA probe were sequenced directly using the double-stranded sequenase method (U. S. Biochemical, Cleveland, OH) described by manufacturer. The clones were primed in both directions with forward and reverse primers and extended in the presence of α-[^{35}S]dATP. The boundaries of the third exon were located by using two primers, 5′ ATG GTG ATA GTA GCC TGG C 3′ and 5′ ACT TTT TGA TAC GTG CC 3′, similar to the cDNA sequence. Two primers, 5′ TCT CGA GGG GTA GAG GGC TCC CCT AAT 3′ and 5′ AGG ATC CAT CAA AGC AAG GTT GAT 3′, derived by cDNA sequence at the beginning of the fourth exon and the end of the 3′ untranslated region, respectively, were used to amplify DNA using λ G26 genomic clone. The Taq polymerase-amplified product obtained after 30 cycles (94°C, 15 s denaturation; 55°C, 15 s annealing; and 72°C, 1 min extension) was ~2.3 kb. Restriction mapping and sequencing of the amplified product indicated that it is identical to the cDNA sequence in this region. The genomic λ clone was also sequenced directly using these primers. Together, these results confirm that the CRD and the 3′ UT are contained in exon 4. Sequencing reactions were analyzed on 8% polyacrylamide 7 M urea denaturing gels at high voltage. Sequence data was analyzed with the Microgenie sequencing program (Beckman Instruments, Inc., Palo Alto, CA).

Chromosomal Localization. Chromosome in situ hybridization was performed as previously described (34). The recombinant plasmid, pMBP68 (35), was labeled by nick translation using all four dNTPs to a specific activity of 1.09×10^7 cpm/μg. Chromosomes were stained with quinacrine mustard dihydrochloride. Data were collected at the microscope using a combination of incident ultraviolet and transmitted visible light to permit simultaneous visualization of chromosome bands and silver grains.

The human-rodent somatic cell hybrids were derived from fusions of hypoxanthine phosphoribosyltransferase-deficient Chinese hamster E36 or mouse RAG cells with white blood cells from two unrelated female carriers of different X;19 translocations (36); with fibroblasts from a female carrier of an X;13 translocation (37); or with white blood cells from a normal male.

Results and Discussion

The MBP Gene. As the human MBP appears to have distinct domains, we were interested in determining whether each domain of the protein was represented in a single exon. Comparison of the genomic nucleotide sequence with the cDNA sequence revealed that the protein coding region is represented by four exons, which are 251 bp, 117 bp, 69 bp, and 3.1 kb. Each intron exon boundary has a consensus donor acceptor site of GT/AG (38).

5′ Region. The first exon that represents the 5′UT region and the NH_2-terminal portion of the protein ends at the first G of a glycine codon, where there is an aberration in the collagen-like domain. Analysis of both genomic clones and a further three cDNA clones from a HepG2 cDNA library revealed three errors in the original cDNA sequence in the signal peptide. The inclusions of three Cs instead of two at position 70 and two Cs instead of a single C at positions 79 and 93 result in an alteration in the signal peptide. Although these changes result in the insertion of an extra amino acid, the signal peptide has a typical hydrophobic sequence as is expected in a secreted molecule. The NH_2 terminus of the native protein is a glutamic acid residue based on the NH_2-terminal amino acid sequence of MBP isolated from human serum. This residue is preceded by a serine that conforms to the general rule that the amino acid in position −1 preceding cleavage by the signal peptidase tends to be a residue with a small side chain like serine or alanine (39). The first in frame ATG is surrounded by ACCATGT, which is a strong Kozak consensus, as there is purine at the −3 position. The presence of an upstream out of frame ATG at position +8 is unusual, and in some cases has been shown to decrease translation efficiency of mRNA (reviewed in reference 40). The role, if any, of this out of frame upstream ATG in regulating MBP expression is not known.

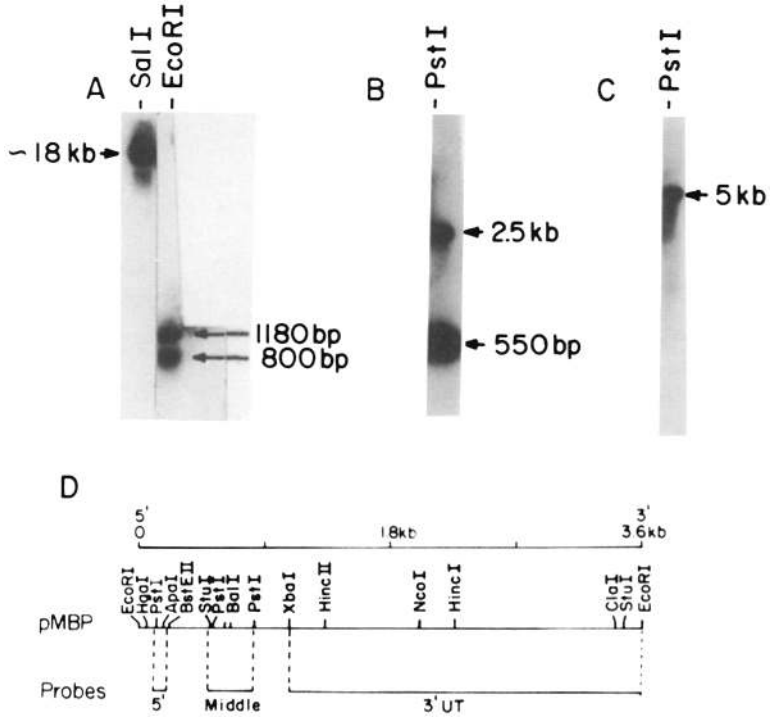

FIGURE 1. Southern blot analysis of MBPG26. 1-2 μg of phage λ MBPG26 DNA was digested with either Sal I, Eco RI, or Pst I, size fractionated on a 1% agarose gel, transferred to nylon membranes, and then hybridized with oligo ^{32}P-labeled cDNA probe encompassing different regions of the human cDNA (14). (*A*) (*Left hand lane*) Hybridization of Sal I digest (Sal I sites flank the insert in the EMBL 3A vector) with a labeled full-length cDNA probe. (*Right hand lane*) Hybridization of Eco RI digest with labeled 5' cDNA probe. (*B*) Hybridization of Pst I digest with a labeled cDNA probe encompassing the majority of the protein-coding portion of the MBP. (*C*) Hybridization of Pst I digest with labeled 2.5-kb cDNA probe spanning the 3' untranslated portion of the cDNA. (*D*) Restriction map of the full-length human MBP-C cDNA (14) is depicted as pMBP. Indicated below are regions of the molecule used to construct oligo-labeled probes encompassing the 5', middle, and 3' untranslated portions of the human MBP-C cDNA. Probes were then hybridized to blots as described in *A*, *B*, and *C*, and washed under stringent conditions.

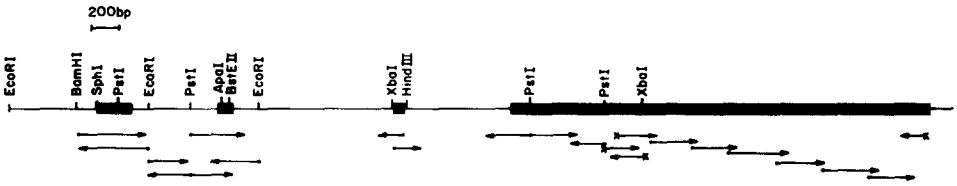

FIGURE 2. Restriction map of the human MBP gene (~7 kb in length). The thickened lines correspond to exons (14). The arrows indicate the direction and extent of sequence analysis by the dideoxy nucleotide chain termination method utilizing either single-stranded M13 vectors or double-stranded pUC vectors.

Primer extension and S1 nuclease protection experiments indicate that the transcription initiation site designated position 1 coincides with the end of the longest cDNA clone (not shown). There is a putative TATATA box −41 bp downstream from this site and a putative CCAAAT box at −82 (Fig. 3). Studies are in progress to assess whether this region can act as a functional promoter.

The 5′ upstream region contains a motif, taagaaatttccag (Fig. 3, *bold type*), that is similar to a heat shock element first described in Drosophila (41). Recent studies indicate that maximum transcription requires two of these elements (42), to which a heat shock transcription factor (43) binds and mediates enhanced transcription of cloned mammalian and Drosophila heat shock genes (reviewed in reference 44).

FIGURE 3. Nucleotide sequence of the human MBP gene. Exon regions are indicated by capital letters and the coding sequence is shown as the three-letter code for amino acids. The presumptive start site is indicated by an arrow. The "TATA-box," "CCAAT-box," and heat shock consensus sequence are in bold type, and the instability motif ATTTA is underscored. It should be noted that there are some differences in the coding region from the previously published cDNA sequence (14). These differences, which result in neutral changes in the encoded protein, are due to DNA sequencing errors of the original clones and do not represent polymorphisms. The corrected sequence has been submitted to EMBL/GenBank Data Libraries under the accession number 15422.

A similar element has been identified in the 5′ flanking region of the human C-reactive protein gene, whose product, like the human MBP, is an acute phase protein (45). Experiments are in progress to determine whether heat shock has a role in MBP gene expression, as well as to assess the relative role of various cytokines that have been shown to induce transcription of genes that encode other acute phase proteins.

The MBP 32-kD subunit forms a trimer that then multimerizes to form a protein of ~300 kD. The NH_2 cysteine-rich area, which mediates interchain disulphide bonds, is found in this exon. Interestingly, this region is followed by a stretch of alternating positively and negatively charged amino acids, interspersed with four glycine residues. These amino acids are largely conserved in the rat MBP-C (Fig. 4), implicating an exceptional structural importance of this peculiar sequence motif. The above amino acids, when arranged on a helical wheel (Fig. 4), form a structure with one face positively charged and the other negatively charged. Such a structure may greatly assist multimer formation not only within each trimeric unit but also between adjacent subunits, via opposite charge-charge neutralization. The abundance of helix-breaking glycines, however, argues against a canonical α-helical conformation of this segment. As much as an extended (β strand) conformation seems to be an equally unlikely possibility, we propose that this region folds into a series of four (approximately helical) turns, the conformation of which becomes stabilized by intimate intermolecular interactions among the subunits.

Collagen Region. Human MBP consists of 19 repeats of Gly-X-Y with the 10th repeat being Gly-Gln-Gly. The first exon ends at this site and the position of the intron strongly supports the contention that two regions of the molecule have evolved from nonfibrillar collagen genes (discussed below). The next exon encodes 11 repeats of Gly-X-Y, with four of the Y amino acids being hydroxyproline (46). A similar collagen stem is found in SP-A and the first complement component, C1q (47). The overall organization of MBP is similar to C1q, although no actual homology exists between the two molecules. The location of the intron in the B chain of the C1q gene, which interrupts a glycine codon after the first G, is similar to rat and human

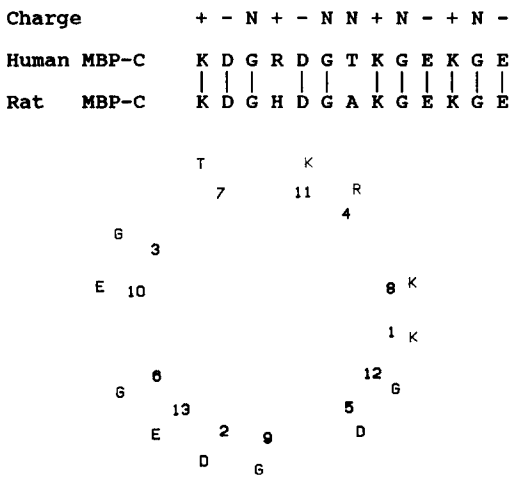

FIGURE 4. Alignment of like-charged amino acids of human MBP-C and rat MBP-C reveals that the positively charged amino acids can be aligned on one face of an α helix with neutral and negatively charged amino acids on the other face, as depicted by a helical wheel. We would like to thank Dr. Jiri Novotny for the helical wheel diagram.

MBP and SP-A genes (47). Initial reports suggested that MBP and C1q may be functionally equivalent, as it was determined that MBP is able to activate complement via the classical pathway (23). Subsequent experiments using purified complement components have shown that human serum MBP is a potent activator of the alternative complement cascade and not the classical pathway (26). As this collagen region is similar to C1q, we predict that this may represent the effector region that interacts with complement components. In vitro mutagenesis experiments are underway to investigate this possibility.

The Putative Cell-binding Domain. MBP appears to selectively recognize configurations of high mannose that are present on pathogens and does appear to bind to normal cells (24, 25). However, interaction of MBP with mannose-rich bacteria results in MBP-dependent attachment and ingestion by phagocytic cells (25). The engagement of the pathogen most likely involves a conformational change in the molecule that then unmasks a cell attachment domain of MBP. Inspection of the sequence at the end of exon 2 reveals a sequence of KSPD that is then followed by GDS in exon 3. Although this sequence does not match the exact consensus sequence recognized by the integrin family of molecules, it may form a loose consensus, which after a conformational change is able to bind to an integrin molecule (48). This family is well represented on the surface of phagocytes, and hence, they are likely candidate receptors for MBP-coated ligands.

The CRD and 3'UT Region. The CRD and the large 2.5-kb 3'UT region are encoded by a single exon. The previously published cDNAs were isolated from a GC-tailed human liver cDNA library and appeared to have a 310-bp 3'UT region followed by an adenylation tail. However, analysis of full-length cDNAs isolated from a λ GT10 HepG2 cDNA library revealed a much longer 3'UT region. The predominant mRNA species on Northern blot analysis is 3.5 kb, although a smear is sometimes observed at 1.2 kb. It is possible that the shorter clones previously described are truncated as a result of the use of an alternative polyadenylation site. Alternately, the poly A tract observed in the shorter clone may be a cloning artifact, as has been described for other cDNAs isolated from GC-tailed libraries (49). The 3'UT region contains seven motifs of AUUUA, which represent the so-called instability sequence. This motif is found in the 3'UT of mRNAs in many cytokines (49), oncogenes, and growth factors (50), and its presence can account for rapid degradation of mRNA. Its presence in the MBP mRNA may be of functional significance as MBP is an acute phase protein. In mice MBP, mRNA is induced between 16 and 18 h after an in vivo challenge of thioglycollate broth, a known stimulus of other active phase reactants. By 24 h, the levels return to baseline (R. A. B. Ezekowitz, unpublished observation). More detalied analysis is necessary to determine the precise contribution of the AUUUA motifs in this phenomena.

Comparison between the Human MBP-C and Rat A Genes. Previously, attention has been drawn to the similarity in both actual homology and organization of the human MBP-C and rat MBP-A and -C. All three molecules have a collagen-like domain that is similar to nonfibrillar collagen proteins. This group of proteins shares with the MBPs a large globular region as well as a triple-helical region in which there are distortions in the regular pattern as a result of an irregularity in a Gly-X-Y repeat. The human MBP gene, like the rat A gene, has features of nonfibrillar collagen genes (Fig. 3). There is an interruption after the first G of the glycine codon at the

site where there is an aberration in the collagen sequence. Both the disturbance of regular collagen α helix and the characteristic intron placement are distinctive features of nonfibrillar collagen genes (51).

The placement of introns is remarkably conserved between the human and rat genes (Fig. 5). The coding regions of both proteins are interrupted by three introns that are located at identical positions (Fig. 5). Like the rat A gene, the CRD and the 3'UT region appeared to be encoded by one exon. However, in the human gene there is a large 2.5-kb 3'UT region, as discussed above.

Drickamer and McCreary (29) pointed out that the functional domains of the rat MBP protein appear to have evolved by juxtaposition of exons by recombination within introns of genes that encode nonfibrillar collagens and genes that encode CRDs. The similarity between the MBP human C gene and rat A gene strengthen this hypothesis, which was proposed to explain the evolution of chimeric proteins (52). Like the rat MBP, the human MBP CRD is encoded by one exon, whereas the rat asialoglycoprotein CRD is encoded by three exons, suggesting that the progenitor gene for this portion of the molecule was divided into three regions and that deletion of introns through evolution led to exon fusion (53), thereby accounting for the genomic organization of the MBP genes. Analysis of the CRDs of a lectin-like family of proteins reveals 18 invariant residues found in a conserved 130-amino acid domain. These proteins bind ligands in a calcium-dependent manner, which have distinct and overlapping specificities. It appears likely that the conserved residues form a polypeptide framework that facilitates ligand binding, but certain residues involved in ligand binding have diverged to accommodate distinct saccharide structures, as proposed by Drickamer (4).

Comparison with the Gene for Surfactant Apoprotein SP-A. In addition to an overall organization homology, the 28–36-kD apoprotein of human surfactant (SP-A), shares 30% overall homology with the human MBP and rat MBPs. The noted homology with rat MBP led to the investigation and verification of the carbohydrate-binding

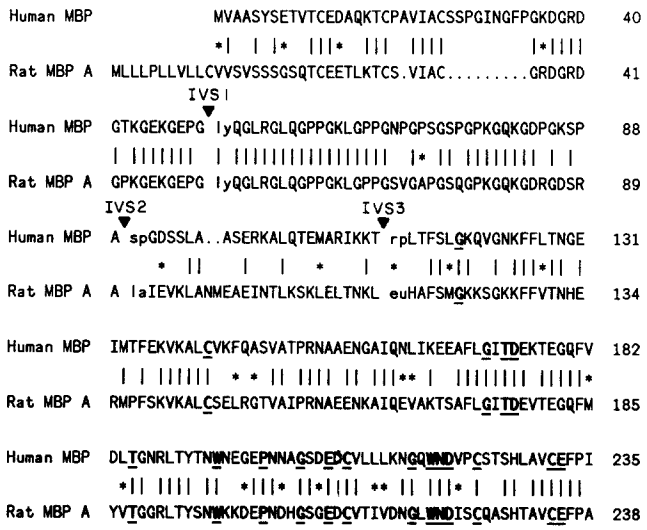

FIGURE 5. Comparison of intron locations and amino acid homology between human MBP-C and rat MBP-A genes. Areas of amino acid homology are indicated by a dash for identical amino acids, whereas conserving submutations are indicated by asterisks. Dots represent gaps introduced to maximize alignment. The locations of introns in human MBP-C and rat MBP-A are denoted by arrows (Bestifit Program, Genetics Computer Group, University of Wisconsin, Madison, WI). The 18 invariant residues characteristic of animal lectins (4) have been underscored.

properties of SP-A (27). The genomic organization of human SP-A was compared with human MBP-C gene (Fig. 6) to ascertain whether the similarities between the two proteins are mirrored at the level of the gene. Several similarities were noted. First, the intron that interrupts the collagen domain is located between the first and second positions of a glycine codon. In both proteins the intron interrupts the helix-encoding region of the gene near a single irregularity in the Gly-X-Y sequences. The nature of the irregularity is different; in H-MBP, one Gly-X-Y sequence is replaced by Gly-Gln-Gly, while in SP-A, the entire sequence Gly-Glu-Met-Pro-Cys-Pro-Pro is inserted. Although the site of the intron in these two collagen containing proteins is near an irregularity in the helix, the placement of the intron is not invariable in nonfibrillar collagen genes (51), and therefore, the relationship of intron placement and evolution of helical distortion remains unclear. Second, the site and size of the two exons and one intron that make up the rest of the noncarbohydrate domain are remarkably conserved in both genes. Third, the carbohydrate recognition domain in the coding region shows 30% homology between the two proteins and the 3'UT region is encoded by a single exon.

The SP-A 3'UT region does not contain the instability sequences that have been identified in the human MBP. MBP and SP-A are chimeric molecules that have apparently evolved either by similar pairing of two ancestral genes or duplication of a common gene. Their structural similarity suggests that these two proteins may share similar functions. Recent work from our laboratory, which showed that human MBP serves as an opsonin in serum, led us to investigate whether SP-A may have a similar role in the hyperphase of the lung. Work in progress confirms this hypothesis, as purified SP-A is able to bind specifically to certain serotypes of group b strep-

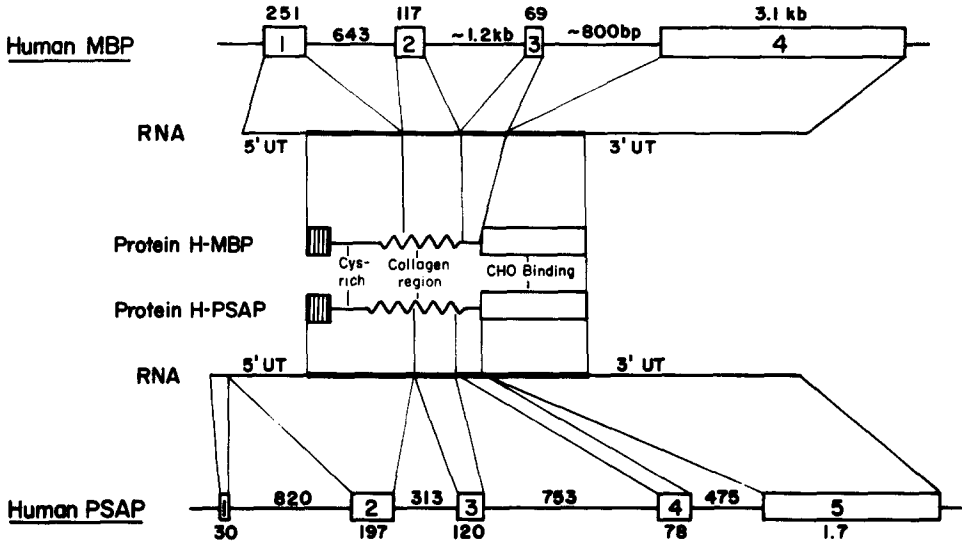

FIGURE 6. Summary of the functional regions in the human MBP-C gene, mRNA, and polypeptide, and comparison with human SP-A gene. The human MBP-C shares 30% overall homology with SP-A as well as similar genomic and functional organization. Most of the protein homology is contained in the exons encoding the carbohydrate-binding domain of both MBP-A and SP-A.

tococci and enhance uptake of the opsonized bacteria by phagocytes (Ezekowitz, R. A. B., S. Gould, M. Wessells, and D. Phelps, unpublished observation).

Chromosomal Localization. Southern blot analysis of DNAs from 23 human/rodent somatic cell hybrids (54) revealed a hybridization pattern of radiolabeled human MBP cDNA that was completely concordant with the segregation of human chromosome 10 (Table I). The pattern was independent of that of other autosomes and sex chromosomes, which had discordant fractions ranging from 0.16 to 0.6.

TABLE I

Segregation of MBP cDNAs with DNAs from Human-Rodent Somatic Cell Hybrids

Chromosome*	Hybridization pattern‡				Discordant fractions§
	+ / +	− / −	+ / −	− / +	
1	3	7	4	1	0.33
2	3	7	2	0	0.16
3	4	6	3	2	0.33
4	2	4	4	4	0.57
5	2	7	5	1	0.40
6	3	4	4	3	0.50
7	4	6	3	2	0.33
8	3	6	4	2	0.40
9	1	5	6	3	0.60
10	7	8	0	0	0.00
11	3	5	4	2	0.42
12	3	7	4	0	0.29
13	4	5	3	3	0.40
14	4	4	2	4	0.43
15	3	6	4	2	0.40
16	5	5	2	3	0.33
17	3	6	4	2	0.40
18	3	7	4	1	0.33
19 and $19q^{+}$‖	6	2	1	6	0.47
20	4	5	3	3	0.40
21	4	5	3	3	0.40
22	4	7	3	1	0.27
X and Xq^{-}¶	3	5	4	3	0.47
Y	0	8	7	0	0.47

* The human chromosome complements of the hybrids were determined by isozyme and cytogenetic techniques (56) and by analysis of hybrid DNAs with DNA probes for each autosome and the X chromosome.

‡ The column designations are: + / +, hybridization signal and chromosome both present; − / −, hybridization signal and chromosome both absent; + / −, hybridization present but chromosome absent; and − / +, hybridization absent with chromosome present.

§ Hybrids with a rearranged chromosome or in which the chromosome was present in <15% of cells were excluded for calculation of discordant fractions.

‖ For the 12 hybrid clones derived from fusions with leukocytes from two different X;19 translocation carriers, this category represents the (der)-19 chromosome.

¶ The X category includes hybrids with an intact X and those with der-(X) translocation chromosomes.

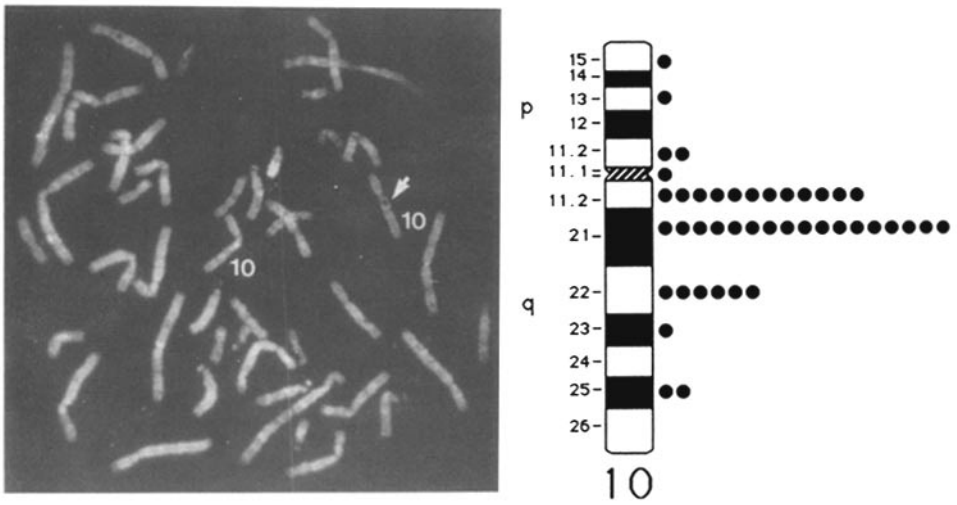

FIGURE 7. Photograph of a representative metaphase spread illuminated by a combination of ultraviolet and visible lights showing hybridization of the MBP gene at 10q11.2-q21. An ideogram of chromosome 10 with the distribution of silver grains scored either on or directly beside this chromosome is seen at the right. A significant peak of hybridization is seen at 10q11.2-q21.

To verify independently the somatic cell hybrid assignment of the human MBP gene and to determine the regionalization of hybridization on the chromosome, the radiolabeled MBP cDNA was utlilized for chromosomal in situ hybridization. A total of 170 metaphase spreads were analyzed for localization of the MBP gene. Of the 373 silver grains present either on or directly beside chromosomes, 29 grains (7.8% of the total number of grains) were located at 10q11.2-q21 (Fig. 7). This represented 67.4% of the total number of grains on chromosome 10. Approximately 14.1% of metaphases were labeled at 10q11.2-q21. No additional sites of hybridization were noted. It is interesting to note that the SP-A gene has been assigned a broad location between 10q21-24 (54). If the precise localization is closer to 10q21, it may indicate that these similar genes may have evolved by duplication of a common gene. However, if the distance between the loci is further away, a more likely explanation is that they evolved from a common ancestral gene and happen by chance to reside on the same chromosome. Surprisingly, human MBP appears to map to the region including the locus for multiple endocrine neoplasia type 2A. The closest marker for this disease locus is the interstitial retinal binding protein, which maps to 10q11.2-21 (30, 55). The polymorphic nature of the MBP gene (M. Kuhlman and R. A. B. Ezekowitz, unpublished observations) suggests that this gene may be a useful marker for the disease. Studies with informative families are in progress to test the possible linkage of MBP to this disease locus.

Summary

The human mannose-binding protein (MBP) plays a role in first line host defense against certain pathogens. It is an acute phase protein that exists in serum as a multimer of a 32-kD subunit. The NH_2 terminus is rich in cysteines that mediate in-

terchain disulphide bonds and stabilize the second collagen-like region. This is followed by a short intervening region, and the carbohydrate recognition domain is found in the COOH-terminal region. Analysis of the human MBP gene reveals that the coding region is interrupted by three introns, and all four exons appear to encode a distinct domain of the protein. It appears that the human MBP gene has evolved by recombination of an ancestral nonfibrillar collagen gene with a gene that encodes carbohydrate recognition, and is therefore similar to the human surfactant SP-A gene and the rat MBP gene. The gene for MBP is located on the long arm of chromosome 10 at 10q11.2–q21, a region that is included in the assignment for the gene for multiple endocrine neoplasia type 2A.

We thank Drs. Kurt Drickamer and Jiri Novotny for useful discussions and Lois Juergens for technical assistance in the chromosomal in situ hybridization analysis.

Received for publication 21 April 1989 and in revised form 22 June 1989.

Note added in proof: This is to acknowledge a manuscript by Taylor et al. (57). This work was brought to our attention after the submission of this manuscript. We would like to thank Dr. John Summerfield for communicating results before publication.

References

1. Kawasaki, M., T. Kawasaki, and I. Yamashura. 1983. Isolation and characterization of a mannose-binding protein from human serum. *J. Biochem. (Tokyo).* 94:937.
2. Wild, J., D. Robinson, and B. Winchester. 1983. Isolation of mannose-binding proteins from human liver. *Biochem. J.* 210:167.
3. Summerfield, J. A., and M. E. Taylor. 1986. Mannose-binding proteins human serum: identification of mannose-specific immunoglobulins and a calcium dependent lectin, of broader carbohydrate specificity secreted by hepatocytes. *Biochem. Biophys. Acta.* 883:197.
4. Drickamer, K. 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Biol. Chem.* 263:9557.
5. Leung, J. O., E. C. Holland, and K. Drickamer. 1985. Characterization of the gene encoding the major rat liver asialoglycoprotein receptor. *J. Biol. Chem.* 260:12523.
6. Halberg, D. F., H. E. Wager, D. C. Farrell, J. Hidreth, M. S. Queensberry, J. A. Loeb, E. C. Holland, and K. Drickamer. 1987. Major and minor forms of the rat liver asialoglycoprotein receptor are independent galactose-binding proteins. Primary structures and glycosylation heterogeneity of minor receptor forms. *J. Biol. Chem.* 262:9828.
7. Bischoff, J., and H. F. Lodish. 1987. Two asialoglycoprotein receptor polypeptides in human hepatoma cells. *J. Biol. Chem.* 262:11825.
8. Drickamer, K. 1981. Complete amino acid sequence of a membrane receptor for glycoproteins. Sequence of the chicken hepatic lectin. *J. Biol. Chem.* 256:5827.
9. Ikuta, K., M. Takami, C. W. Kim, T. Honjo, T. Miyoski, Y. Tagaya, T. Kawabe, and J. Yodou. 1987. Human lymphocyte Fc receptor for IgE. Sequence homology of its cloned cDNA with animal lectins. *Proc. Natl. Acad. Sci. USA.* 84:819.
10. Hoyle, G. W., and R. L. Hill. 1988. Molecular cloning and sequencing of a cDNA for a carbohydrate binding receptor unique to rat Kupffer cells. *J. Biol. Chem.* 263:7487.
11. Lasky, L. A., M. S. Singer, T. A. Yednock, D. Dowbenko, C. Fennie, H. Rodriguez, T. Nguyen, S. Stachel, and S. D. Rosen. 1989. Cloning of a lymphocyte homing receptor reveals a lectin domain. *Cell.* 56:1045.
12. Johnston, G. I., R. G. Cook, and R. P. McEver. 1989. Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sequence similarity to proteins involved

in cell adhesion and inflammation. *Cell.* 56:1033.
13. Bevilacqua, M. P., S. Stengelin, Jr., M. A. Gimbrone, and B. Seed. 1989. Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins. *Science (Wash. DC).* 243:1160.
14. Ezekowitz, R. A. B., L. Day, and G. Herman. 1988. A human mannose-binding protein is an acute phase reactant that shares sequence homology with other vertebrate lectins. *J. Exp. Med.* 167:1034.
15. Drickamer, K., M. S. Dordal, and L. Reynolds. 1986. Mannose-binding proteins isolated from rat liver contain carbohydrate-recognition domains linked to collagenous tail. *J. Biol. Chem.* 261:6878.
16. Benson, B., S. Hawgood, J. Schilling, J. Clements, P. Danim, B. Cordell, and R. T. White. 1985. Structure of canine pulmonary surfactant apoprotein: cDNA and complete amino acid sequence. *Proc. Natl. Acad. Sci. USA.* 82:6379.
17. Floros, J., R. Steinbrink, K. Jacobs, D. Phelps, R. Kriz, M. Recny, L. Sultzman, S. Jones, H. W. Taeusch, H. A. Frank, and E. F. Fritsch. 1986. Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfactant-associated protein. *J. Biol. Chem.* 261:9029.
18. Shigaku, S., T. Takaaki, R. A. Kosher, and M. L. Tanzier. 1986. Cloning and sequence analysis of a partial cDNA for chicken cartilage proteoglycan core protein. *Proc. Natl. Acad. Sci. USA.* 83:5081.
19. Takahashi, H., H. Komano, N. Kawaguchi, N. Kitamura, S. Nakanishi, and S. Naton. 1985. Cloning and sequencing of a cDNA of Sarcophaga perginina humoral lectin induced by injury of the body wall. *J. Biol. Chem.* 260:12228.
20. Giga, Y., I. Atsushi, and K. Takahashi. 1987. The complete amino acid sequence of echinoiden, a lectin from the coelomic fluid of the sea urchin Anthociadaris crassispina. *J. Biol. Chem.* 262:6197.
21. Muramoto, K., and H. Kamiya. 1986. The amino-acid sequence of a lectin from the acorn barnacle Megabalanus rosa. *Biochem. Biophys. Acta.* 874:285.
22. Boggaram, V., K. Qing, and C. R. Mendelson. 1988. The major apoprotein of rabbit pulmonary surfactant. Elucidation of primary sequence and cyclic AMP and developmental regulation. *J. Biol. Chem.* 263:2939.
23. Ikeda, K., T. Sannoh, N. Kawasaki, T. Kawasaki, and I. Yamashina. 1987. Serum lectin with known structure activates complement through the classical pathway. *J. Biol. Chem.* 262:7451.
24. Ezekowitz, R. A. B., M. Kuhlman, J. Groopman, and R. Byrn. 1989. A human serum mannose-binding protein inhibits in vitro infection by the human immunodeficiency virus. *J. Exp. Med.* 169:185.
25. Kuhlman, M., K. Joiner, and R. A. B. Ezekowitz. 1989. The human mannose-binding protein functions as an opsonin. *J. Exp. Med.* 169:1733.
26. Schweinle, J., R. A. B. Ezekowitz, A. Tenner, M. Kuhlman, and K. Joiner. 1989. Human mannose-binding protein activates the alternative complement pathway and enhances serum bactericidal activity on a mannose-rich isolate of Salmonella *J. Clin. Invest.* In press.
27. Haagsman, H. P., S. Hawgood, T. Sargeant, D. Buckley, R. T. White, K. Drickamer, and B. J. Benson. 1987. The major lung surfactant protein, SP28-36, is a calcium dependent, carbohydrate-binding protein. *J. Biol. Chem.* 262:13877.
28. Tenner, A. J., S. Robinson, J. D. Borchelt, and J. R. Wright. 1988. Human pulmonary surfactant protein (SP-A), a protein structurally homologous to C1q, can enhance FcR and CR1 mediated phagocytosis. *J. Cell Biol.* 107:642*a*. (Abstr.)
29. Drickamer, K., and V. McCreary. 1987. Exon structure of a mannose-binding protein gene reflects its evolutionary relationship to the asialoglycoprotein receptor nonfibrillar collagens. *J. Biol. Chem.* 262:2582.

30. Simpson, N. E., K. K. Kidd, P. J. Goodfellow, H. McDermid, S. Myers, J. R. Kidd, C. E. Jackson, A. M. V. Duncan, L. A. Farrer, K. Brasch, C. Castiglione, M. Genel, J. Gertner, C. R. Greenberg, J. F. Gusella, J. J. A. Holden, and B. N. White. 1987. Assignment of multiple endocrine neoplasia type 2A to chromosome 10 by linkage. *Nature (Lond.).* 328:528.
31. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503.
32. Sanger, F., A. S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
33. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* 101:20.
34. Morton, C. C., I. R. Kirsch, R. Taub, S. H. Orkin, and J. A. Brown. 1984. Localization of the beta-globin gene by chromosomal in situ hybridization. *Am. J. Hum. Genet.* 36:576.
35. Loos, M., and F. Clas. 1986. Antibody independent killing of gram-negative bacteria via the classical pathway of complement. *Immunol. Lett.* 14:203.
36. Brook, J. D., D. J. Shaw, L. Meredith, G. A. P. Bruns, and P. S. Harper. 1985. Localization of genetic markers and orientation of the linkage group on chromosome 19. *Hum. Genet.* 69:282.
37. Latt, S. A., H. F. Willard, and P. S. Gerald. 1976. Hoechst analysis of DNA replication in human lymphocytes with supernumerary or structurally abnormal X chromosomes. *Chromosoma (Berl.).* 57:135.
38. Breathnach, R., and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349.
39. Von Heijne, G. 1983. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem.* 133:17.
40. Kozak, M. 1989. The scanning model for translation. An update. *J. Cell Biol.* 108:229.
41. Pelham, H. R. B. 1982. A regulatory upstream promoter element in the Drosophilia Hsp 70 heat shock gene. *Cell.* 30:517.
42. Cohen, R. S., and M. Meselson. 1988. Periodic interactions of heat shock transcriptional elements. *Nature (Lond.).* 332:856.
43. Wu, C., S. Wilson, B. Walker, I. David, T. Paisley, V. Zimarino, and H. Ueda. 1987. Purification and properties of Drosophila heat shock activator protein. *Science (Wash. DC).* 238:1247.
44. Lindquist, S. 1986. The heat-shock response. *Annu. Rev. Biochem.* 55:1151.
45. Woo, P., J. R. Korenberg, and A. S. Whitehead. 1985. Characterization of genomic and complementary DNA sequence of human c-reactive protein and comparison with the complementary DNA sequence of serum amyloid Component. *J. Biol. Chem.* 260:13384.
46. Colley, K. M., C. Beranek, and J. U. Baenziger. 1988. Purification and characterization of human and rat core specific lectins. *Biochem. J.* 256:61.
47. Reid, K. B. M. 1983. Proteins involved in the activation and control of the two pathways of human complement. *Biochem. Soc. Trans.* 11:1.
48. Hynes, R. O. 1988. Integrins: a family of cell surface receptors. *Cell.* 48:549.
49. Caput, D., B. Beutler, K. Hartog, R. Thayer, S. Brown-Shiner, and A. Cerami. 1986. Identification of a common nucleotide sequence in the 3' untranslated region of mRNA molecules specifying inflammatory mediators. *Proc. Natl. Acad. Sci. USA.* 83:1670.
50. Shaw, G., and R. Kamen. 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA. *Cell.* 46:659.
51. Martin, G. R., R. Rimpl, P. K. Muller, and K. Kuhn. 1985. The genetically distinct collagens. *TIBS (Trends Biochem. Sci.).* 10:285.
52. Blake, C. C. 1985. Exons and the evolution of proteins. *Int. Rev. Cytol.* 93:149.
53. Gilbert, W. 1978. Why genes in pieces? *Nature (Lond.).* 271:501.
54. Bruns, G., G. M. Veldman, H. Stroh, S. A. Latt, and J. Floros. 1987. The 35 kd pulmo-

nary surfactant-associated protein is encoded in chromosome 10. *Hum. Genet.* 76:58.
55. Mathew, C. G. P., K. S. Chin, D. F. Easton, K. Thorpe, C. Carter, G. I. Liou, S.-L. Fong, C. D. B. Bridges, H. Haak, A. C. Nieuwenhuijzen Kruseman, S. Schifter, H. H. Hansen, H. Telenius, M. Telenius-Berg, and B. A. J. Ponder. 1987. A linked genetic marker for multiple endocrine neoplasia type 2A on chromosome 10. *Nature (Lond.).* 328:527.
56. Bruns, G. A. P., B. J. Mintz, A. C. Leary, V. M. Regina, and P. S. Gerald. 1979. Human lysosomal genes: arylsulfatase A and beta-galactosidase. *Biochem. Genet.* 17:1031.
57. Taylor, M. E., E. M. Brickell, R. K. Craig, and J. A. Summerfield. 1989. The structure and evolutionary origin of the gene encoding a human mannose-binding protein. *Biochem J.* In press.