Structure and evolutionary origin of the gene encoding a human serum mannose-binding protein

Maureen E. TAYLOR,*† Paul M. BRICKELL,† Roger K. CRAIG†§ and John A. SUMMERFIELD*||

*Department of Medicine, St. Mary's Hospital Medical School, Norfolk Place, London W2 1NY,

and †The Medical Molecular Biology Unit, Biochemistry Department, University College and Middlesex School of Medicine, London W1P 6DB, U.K.

The *N*-terminal sequence of the major human serum mannose-binding protein (MBP1) was shown to be identical at all positions determined with the amino acid sequence predicted from a cDNA clone of a human liver MBP mRNA. An oligonucleotide corresponding to part of the sequence of this cDNA clone was used to isolate a cosmid genomic clone containing a homologous gene. The intron/exon structure of this gene was found to closely resemble that of the gene encoding a rat liver MBP (MBP A). The nucleotide sequence of the exons differed in several places from that of the human cDNA clone published by Ezekowitz, Day & Herman [(1988) *J. Exp. Med.* 167, 1034–1046]. The MBP molecule comprises a signal peptide, a cysteine-rich domain, a collagen-like domain, a 'neck' region and a carbohydrate-binding domain. Each domain is encoded by a separate exon. This genomic organization lends support to the hypothesis that the gene arose during evolution by a process of exon shuffling. Several consensus sequences that may be involved in controlling the expression of human serum MBP have been identified in the promoter region of the gene. The consensus sequences are consistent with the suggestion that this mammalian serum lectin is regulated as an acute-phase protein synthesized by the liver.

INTRODUCTION

Proteins capable of binding mannose-terminated glycoproteins have been isolated from the liver and serum of rats [1–3], rabbits [4,5] and humans [6–9]. At least two mannose-binding proteins (MBPs) are present in human serum [8]. The major human serum MBP (MBP1) is a homopolymer of approx. 700 kDa composed of subunits of 32 kDa and is secreted by the liver [7]. MBP1 has specificities for *N*-acetylglucosamine, *N*-acetylmannosamine, glucose and fucose as well as mannose [8]. The other MBP (MBP2) is a homopolymer of approx. 200 kDa composed of subunits of 28 kDa and is specific only for mannose and fucose [8]. Both lectins are Ca^{2+} -dependent. The function of these MBPs is unknown, but they have a high affinity for yeast mannan [1–9] and thus might act as a part of the immune system, binding and inactivating yeasts and other micro-organisms bearing mannose oligosaccharides. Support for this hypothesis comes from the finding that serum MBPs from rat, rabbit and human can activate complement through the classical pathway [10].

Cloning and sequencing of two rat liver MBPs has shown that they are divided into several domains: a signal peptide, a short cysteine-rich *N*-terminal domain, a collagen-like domain, a 'neck' region and a *C*-terminal carbohydrate-recognition domain [11]. The primary structure of a human liver MBP has recently been determined and found to be similar to that of the rat MBPs [12]. Northern-blot hybridization revealed marked

MBP mRNA accumulation in the liver of a trauma victim, suggesting that MBP may be an acute-phase reactant [12]. The sequences of carbohydrate-recognition domains of MBPs are highly homologous to those found in other carbohydrate-binding proteins, including membrane-bound lectins such as the rat asialoglycoprotein receptor [13]. By using the conserved amino acid residues found in the carbohydrate-recognition domains of these proteins as a guide, it has been possible to identify putative carbohydrate-recognition domains in proteins such as pulmonary-surfactant apoprotein [14] and the core protein of cartilage proteoglycan [15].

The gene structure of one rat liver MBP (MBP A) has been determined and found to share structural features with the genes for non-fibrillar collagens as well as with the genes for other carbohydrate-binding proteins [16]. In the present paper we give the structure of the gene encoding a human MBP, which we suggest is the gene encoding the major human serum MBP (MBP1). We show that the structure of this gene is very similar to that of the gene encoding rat liver MBP A. We have also identified consensus sequences in the promoter region of the human MBP gene that indicate that its expression is regulated as part of the acute-phase response.

MATERIALS AND METHODS

Protein sequence analysis

The major human serum MBP (MBP1) was isolated

Abbreviation used: MBP, mannose-binding protein.

† Present address: Department of Biochemistry, Columbia University, 630 West 168th Street, New York, NY 10032-3702, U.S.A.

§ Present address: Research Division 1, ICI Pharmaceuticals, Alderley Park, Cheshire SK10 4TG, U.K.

|| To whom correspondence should be addressed.

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession numbers X15954, X15955, X15956 and X15957.

from pooled serum collected from patients undergoing venesection for haemochromatosis by methods previously described [8]. Affinity chromatography with a monoclonal antibody against human serum MBP1 (Mab37) was used to purify MBP1 to homogeneity, as judged by SDS/polyacrylamide-gel electrophoresis. The *N*-terminal amino acid sequence of this serum MBP sample was determined on an Applied Biosystems gas-phase sequencer by Dr. K. Drickamer (Columbia University, New York, NY, U.S.A.).

Isolation of clones

A cosmid library of human genomic DNA was kindly provided by Dr. Dimitris Kioussis (National Institute for Medical Research, London, U.K.). This library was constructed in the cosmid vector cos 202 by using partially digested *Sau*3AI fragments of genomic DNA from the human acute-lymphocytic-leukaemia cell line HPB-ALL [17]. Some 500 000 colonies were screened for the presence of MBP-encoding inserts by the method of Grunstein & Hogness [18]. The probe used (probe 4) was a 40-base oligonucleotide with the sequence 5'-dGGGAATCA-CAGACGGCAGATGGGAGGTGGAAGCAAGGG-G-3' derived from the region of the cDNA clone [12] coding for the carbohydrate-binding domain of MBP (Fig. 3*d*, exon 4, bases 645–607, lower strand). The oligonucleotide was labelled with [γ -³²P]ATP to a specific radioactivity of 10⁹ c.p.m./ μ g by the use of T4 polynucleotide kinase [19]. Duplicate filters were prehybridized overnight at 55 °C in a solution composed of 6 $\times$ SSC, 5 $\times$ Denhardt's solution, 0.5% SDS and 100 μ g of denatured herring testis DNA/ml and then hybridized for 8 h at 55 °C in the same solution containing the labelled probe at 10⁶ c.p.m./ml. [1 $\times$ SSC is 0.15 M-NaCl/15 mM-sodium citrate buffer, pH 7.0, and 1 $\times$ Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin (Pentax fraction V).] The filters were washed at 65 °C for 30 min in 2 $\times$ SSC/0.1% SDS and autoradiographed. Colonies corresponding to positive signals were purified by two rounds of screening at low density.

Analysis of clones

Four further oligonucleotide probes were used, each having a sequence found in the cDNA clone [12]. Probe 1 was 31 bases long with the sequence 5'-dGGTG-TGAGAAACTCAGGGAAGGTTAATCTC-3' derived from the 5' untranslated region of the cDNA clone (Fig. 3*a*, exon 1, bases 57–27, lower strand). Probe 2 was 31 bases long with the sequence 5'-dGGCCCTTTG-GTCCTGGTGACCCAGAAGGCC-3' derived from the region of the cDNA clone encoding the collagen-like domain (Fig. 3*b*, exon 2, bases 580–550, lower strand). Probe 3 was 29 bases long with the sequence 5'-dCC-ATTCCTGGGGGTGGCCACAGAGGCTGG-3' derived from the region of the cDNA clone encoding the carbohydrate-binding domain (Fig. 3*d*, exon 4, bases 408–379, lower strand). Probe 5 was 28 bases long with the sequence 5'-dGGTCATGTCATTAGGGGAGCC-CTCTACC-3' derived from the 3' untranslated region of the cDNA clone (Fig. 3*d*, exon 4, bases 1065–1038, lower strand). The sequences of some probes differed slightly from the sequence in Fig. 3 since they were based on the reported sequence of the cDNA clone [12], which we believe contains errors (see below). Southern-blot analysis was used to identify the restriction fragments of

the cosmid insert that hybridized to the probes. Fragments containing sequence of interest were cloned into M13 mp18 and M13 mp19 and sequenced by the dideoxy method [20] with the use of Sequenase (United States Biochemical Corporation). The strategy for the nucleotide sequence analysis is shown in Fig. 2.

RESULTS AND DISCUSSION

Protein sequences of human liver and serum MBPs

Our aim was to isolate the gene encoding the major human serum MBP (MBP1) by using sequence data from the human liver MBP cDNA clone sequenced by Ezekowitz *et al.* [12]. However, since both human and rat liver contain at least two MBPs [9,11], we needed to show that the human liver cDNA sequenced by Ezekowitz *et al.* [12] encoded human serum MBP. The *N*-terminal amino acid sequence of serum MBP1 is shown in Fig. 1. At all 21 positions determined, this sequence matched perfectly the amino acid sequence predicted from the nucleotide sequence of the human liver MBP cDNA clone reported by Ezekowitz *et al.* [12]. This strongly suggests but does not prove absolutely that the cDNA clone encodes the major human serum MBP (MBP1).

Isolation of positive clones

Screening of 500 000 colonies with radiolabelled oligonucleotide probe 4 yielded one positive clone (cMBP1). This clone contained an insert of approx. 30 kb. Southern-blot analysis with oligonucleotide probes corresponding to sequences at the 5' and 3' ends of the cDNA clone (probes 1 and 4) showed that cMBP1 contained the whole of the transcribed region of the gene together with approx. 9 kb of 5' flanking sequence and 16 kb of 3' flanking sequence. Therefore no further screening was performed. A partial restriction map of the clone is shown in Fig. 2.

Intron/exon structure of the gene

The nucleotide sequence of the 5' flanking sequence, the exons and the intron/exon boundaries of the MBP gene contained within cMBP1 is shown in Fig. 3. Comparison of this nucleotide sequence with that of the cDNA clone [12] (Fig. 4) shows that the protein-coding region of the gene is interrupted by three introns of approx. 600 bp, 1350 bp and 800 bp in size. The sequences at all of the intron/exon boundaries contain the usual splice donor and acceptor consensus sequences [21]. The predicted amino acid sequence of the protein encoded by this gene is shown in Fig. 5.

The intron/exon structure and the sizes of the introns are almost identical with those of the rat MBP A gene [16]. Thus exon 1 encodes the signal peptide, a cysteine-rich domain and seven copies of the Gly-Xaa-Yaa motif typical of the triple-helix-forming region of collagen. At the site of the first intron the regular Gly-Xaa-Yaa pattern is interrupted by the sequence Gly-Gln-Gly. Exon 2 then encodes a further 12 Gly-Xaa-Yaa repeats. Both rat MBP A and rat MBP C have the same single Gly-Gln-Gly interruption in the Gly-Xaa-Yaa repeat structure. A regular Gly-Xaa-Yaa pattern encoded by two adjacent exons is typical of genes encoding other non-fibrillar collagens [22–25]. Exon 3 encodes a 'neck' region and exon 4 encodes a carbohydrate-recognition domain. This is also the case for the rat MBP A gene [16]. Moreover, the carbohydrate-binding domain of the rat

ND-Thr-Val-Thr-ND-Glu-ND-Ala-Gln-Lys-Thr-ND-Pro-Ala-ND-Ile-ND-ND-ND-ND-Pro-Gly-Ile-Asn-Gly-Phe-Pro-Gly-Lys-Asp

Fig. 1. N-Terminal amino acid sequence of the major human serum mannose-binding protein (MBP1)

Abbreviation: ND, not determined.

Fig. 2. Partial restriction map of cosmid genomic clone cMBP1 containing the gene encoding human MBP1

(a) Insert of cosmid clone cMBP1. (b) Region of cMBP1 sequenced. (c) Structure of MBP1 gene. (d) Sequencing strategy. Abbreviations: Ba, *Bam*HI; Bs, *Bst*EII; C, *Cla*I; E, *Eco*RI; P, *Pst*I; R, *Rsa*I; S, *Stu*I; X, *Xba*I; Xh, *Xho*I. Coding regions are indicated by black boxes. 5' and 3' untranslated regions are indicated by open boxes.

liver asialoglycoprotein receptor is also encoded by discrete exons separated from the exons encoding other functional domains [26]. This is consistent with the hypothesis that evolutionary shuffling has allowed carbohydrate-recognition domains to become associated with different effector molecules [13].

The sequence of the exons was found to be different from the published cDNA sequence [12] in several places, as shown in Fig. 4. The 3' untranslated region of the published cDNA clone [12] contains two stretches of 23 and 56 nucleotide residues that are absent from the genomic clone. In addition, nucleotide residues 975–1128, which are present in exon 4 (Fig. 3d, nucleotide residues 728–958) are absent from the 3' untranslated region of the cDNA clone. It is unlikely that this reflects the presence of an intron in the 3' untranslated region of the human MBP mRNA, since there are no consensus splice donor and acceptor sites at the appropriate positions in the genomic sequence (Fig. 3d). It is possible that these differences arose from reverse transcriptase errors in the construction of the cDNA clone or that they arose because the cDNA and genomic clones derive from two distinct genes. The remainder of the sequence differences are changes, deletions or insertions of single base-pairs that, in some cases, yield differences in the amino acid sequences predicted from the two nucleotide sequences.

The amino acid sequences predicted from our sequence, from the published cDNA sequence [12] and from the two rat MBPs [11] are shown in Fig. 5. As indicated, the signal sequence predicted from the genomic clone has one extra amino acid residue (residue 7) and four amino acid residue changes when compared with that predicted from the cDNA clone. The signal sequence predicted from the genomic clone retains the typical hydrophobic structure expected in a secreted eukaryotic

protein [27] and, in particular, has a serine residue at position 20, immediately preceding the *N*-terminal glutamic acid residue of the mature protein (Fig. 5). This conforms with the general observation that the amino acid residue immediately preceding the site of cleavage by the signal peptidase has a small side chain [28]. There are three amino acid residue differences in the first part of the collagen-like domain in exon 1, and in exon 3 only six of the 22 amino acid residues are the same as those published. Finally, there are 16 amino acid residue differences in the carbohydrate-binding domain (exon 4). The amino acid residue changes increase the overall homology of the human MBP to the rat MBPs [11], especially in the carbohydrate-recognition domain. Thus in 21 positions a residue is changed to that found in rat MBP C and in 16 positions a residue is changed to that found in rat MBP A. In particular, Thr-187 predicted by cMBP1 is one of the 18 invariant residues found in all mammalian carbohydrate-binding proteins [29] whereas Pro-187 predicted by the cDNA clone is not. These results give values of 61% for the homology of the human MBP to rat MBP C and of 52% for the homology of the human MBP to rat MBP A.

It is possible the sequence of the cDNA clone [12] differed from that of our genomic clone because the genomic clone derives from a second MBP gene. However, there are two lines of evidence that suggest that the differences arise because the sequence of the cDNA clone [12] was partially incorrect. Firstly, where there are stretches of amino acid residue differences (e.g. Fig. 5, residues 107–123) this reflects an alteration of the reading frame caused by a single-base deletion/insertion rather than extensive nucleotide differences (Fig. 4, nucleotides 389–440). Secondly, if the two clones were derived from different genes encoding closely related proteins we would

expect to find extensive nucleotide differences in the third-base position of codons, even when the predicted amino acid sequences were identical. This is seen when the rat and human nucleotide sequences are compared, but not when the human cDNA and genomic sequences are compared. For these reasons we suggest that the human cDNA [12] and genomic clones are derived from a single human MBP gene. In view of the identity of

their products' predicted *N*-terminal sequence with the *N*-terminal sequence of the major human serum MBP (MBP1) (Fig. 1), we suggest that they derive from the human serum MBP1 gene. However, we cannot exclude the possibility that they represent two different MBP genes, in which case our data do not allow us to identify which clone, if either, represents the human serum MBP1 gene. If the cDNA [12] and genomic clones are derived

(a) 5' flanking sequence, exon 1 and portion of intron 1

```

ctgcagtggagactgtcctttgttttcaaagggaaacttgaggcttagacctatggggct -784
aggctgctgaggtttcttaggggcaatagctggaagaaagctctgaAGAACAAATGaaa -724
ggttaatactgagaaatgggaggaggattcaaggcaagttttctaattgccagtggtttt -664
tgactcacAGAACTgggggaattcctgccagaaagtagagaggtatttagcactctgccca -604
gggccaaacgtagTAAGAAATTTCCAGagaaaaatgcttaccaggcaagcctgtctaaaac -544
accaaggggaagcaaaactccagtttaattctgggctgggttggtgactaaggttgagggtg -484
atctgaggttgagaccttcctcttttgatcaccagctttcagctcagggcctgccaatga -424
gtaaatgatagtttaacaggtcctggagggggaatcagctgccagatacaaagatgggatt -364
caggtggcagatggacccgaagaggacatggagagaaagaggaagctcctacagacacct -304
gggtttccactcattctcattccctaagctaacaggcataagccagctggcaatgcacGG -244
TCCCATTTGTTCTcactgccaccgaaagcatgtttatagtCTTCCAGCAGCAACGCCAGG -184
tgtctaggcacagatgaaccctccttaggatccccactgctcatcatagtgcctacctt -124
tgttaaagtactagtcacgcagtgtcacaaggaatgtttacttttCCAAATccccagcta -64
gaggccagggtgggtcatctatttcTATATAgcctgcacccagattgtaggacagaggg -4
catGCTCGGTAAATATGTGTTTCATTAAGTATTAACCTTCCTGAGTTTTCTCACACC 57
AAGGTGAGGACCATGTCCCTGTTTCCATCACTCCCTCTCCTTCTCCTGAGTATGGTGGCA 117
MetSerLeuPheProSerLeuProLeuLeuLeuLeuSerMetValAla
GCGTCTTACTCAGAACTGTGACCTGTGAGGATGCCCAAAGACCTGCCCTGCAGTGATT 177
AlaSerTyrSerGluThrValThrCysGluAspAlaGlnLysThrCysProAlaValIle
GCCTGTAGCTCTCCAGGCATCAACGGCTTCCAGGCAAAGATGGGCGTGATGGCACCAAG 237
AlaCysSerSerProGlyIleAsnGlyPheProGlyLysAspGlyArgAspGlyThrLys
GGAGAAAAGGGGAACAGgtacgtgttg 267
GlyGluLysGlyGluProG

```

(b) Exon 2 and flanking sequences

```

aacaacaactactagcaaaacaaatgcagtttaattttcactttgcaccctccctgcagca 60
acctccacgtggcaactttattttcttaagttattgctctcaggtgcacaccatacagtta 120
ttgagagcagtgctcagaaaggtcagtcctgggtcaaggtctcctttctcctgagaagga 180
ttggcatcaaactcttgaagagagagcaagaacatgagatattaagtcacattcctttgt 240
cttccaacagGCCAAGGGCTCAGAGGCTTACAGGGCCCCCTGGAAAGTTGGGGCTCCA 300
lyGlnGlyLeuArgGlyLeuGlnGlyProProGlyLysLeuGlyProPro
GGAAATCCAGGGCCTTCTGGGTACACAGGACCAAAGGGCCAAAAGGAGACCCTGGAAAA 360
GlyAsnProGlyProSerGlySerProGlyProLysGlyGlnLysGlyAspProGlyLys
AGTCCGGgtaaggacccc 378
SerProA

```

(c) Exon 3 and flanking sequences

tctagATGGTGATAGTAGCCTGGCTGCCTCAGAAAGAAAAGCTCTGCAACAGAAATGGC 60
 spGlyAspSerSerLeuAlaAlaSerGluArgLysAlaLeuGlnThrGluMetAl
 ACGTATCAAAAAGTgtaagctttttctcttactctccaggcagcttgaagtttggaataa 120
 aArgIleLysLysT
 atagatgcaacaaatatttgttgatgcatataattttctgtaccctgctaggcatttct 180
 catattcttacctcatgaaattctcacaacatttttggtagaaatggaggcaaagggaagt 240
 taaattacttgttcaaatgcacagagctaataaatggcaggggtggtttatagatggaag 300
 tcagtctgactcgagagaccctaatacctttaccgtctgatattgctcactgaaaatggga 360
 cttgagatttgggtggtcactggtattgagacctggccgtgggtctaacctgcctgggg 420
 caaatattttcaga 434

(d) Exon 4 and flanking sequences

accagtctgtctgttcacatttgggtgcatatttaacaaataaaaagacaaagcacca 60
 gttaaattggatttcagctgaaaaacatttttagcacaagtgtgtctgaaaaattatatg 120
 ggagatataactaaaatattcattgtttgtctgaaattcaaatttaactggacatactata 180
 ttttatccggcaactctactctagaagactttttcttgagaaataccttgagttgggctt 240
 aaggatgagtgagtttcacccactttttcacatttttagGGCTGACCTTCTCTCTGGGCAA 300
 rpLeuThrPheSerLeuGlyLy
 ACAAGTTGGGAACAAGTTCTTCCTGACCAATGGTGAAATAATGACCTTTGAAAAAGTGAA 360
 sGlnValGlyAsnLysPhePheLeuThrAsnGlyGluIleMetThrPheGluLysValLy
 GGCCTTGTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCAAGGAATGCTGCAGAGAATGG 420
 sAlaLeuCysValLysPheGlnAlaSerValAlaThrProArgAsnAlaAlaGluAsnGly
 AGCCATTGAGAATCTCATCAAGGAGGAAGCCTTCCTGGGCATCACTGATGAGAAGACAGA 480
 yAlaIleGlnAsnLeuIleLysGluGluAlaPheLeuGlyIleThrAspGluLysThrGly
 AGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCTACACAACTGGAACGAGGGTGA 540
 uGlyGlnPheValAspLeuThrGlyAsnArgLeuThrTyrThrAsnTrpAsnGluGlyGly
 ACCCAACAATGCTGGTTCTGATGAAGATTGTGTATTGCTACTGAAAAATGGCCAGTGGAA 600
 uProAsnAsnAlaGlySerAspGluAspCysValLeuLeuLeuLysAsnGlyGlnTrpAs
 TGACGTCCCCTGCTCCACCTCCCATCTGGCCGTCTGTGAGTTCCCTATCTGAAGGGTCAT 660
 nAspValProCysSerThrSerHisLeuAlaValCysGluPheProIleEnd
 ATCACTCAGGCCCTCCTTGTCTTTTACTGCAACCCACAGGCCACAGTATGCTTGAAAA 720
 GATAAATTATATCAATTTCTCATATCCAGCATTGTTTCTTTTGTGGGCAATCACTAAAA 780
 ATGATCACTAACAGCACCAACAAAGCAATAATAGTAGTAGTAGTAGTTAGCAGCAGCAGT 840
 AGTAGTCATGCTAATTATATAATATTTTAAATATATACTATGAGGCCCTATCTTTTGCAT 900
 CCTACAATTAATTATCTAGTTTAATTAATCTGTAATGCTTTTCGATAGTGTAACTTGCTG 960
 CAGTATGAAAAATAAGACGGATTATTTTTTCCATTACAAACAAACACCTGTGCTCTGTTGA 1020
 GCCTTCCTTTCTGTTTGGGTAGAGGGCTCCCCTAATGACATCACCACAGTTTAATACCAC 1080
 AGCTTTTTACCAAGTTTCAGGTATTAAGAAAATCTATTTTTGTAACTTTCTCTATGAACT 1140
 CTGTTTTCTTTCTAATGAGATATTAAACCATGTAaagaacataaataacaaatctcaagc 1200
 aaacagcttcacaaattctcacacacatacatatatactcacttttctaga 1254

Fig. 3. Partial nucleotide sequence of human MBP1 gene

Sequences from introns and the 5' flanking region are shown in lower case. The consensus sequences in the 5' flanking region are shown in upper case and are underlined (see the text for details). The four exons are shown in upper case. The sequences from which the five oligonucleotide probes were derived are marked with a line above them.

```

GCTCGGTAAATATGTGTTTCACTTAAGTGAATTAACCTTCCCTGAGTTTTCTCACACCAAGGTGAGGACCATGTCCCT
GCTCGGTAAATATGT-TTCATTAAGTGAATTAACCTTCCCTGAGTTTTCTCACACCAAGG-GAG-ACCATGTCC-T
154
GTTTCCATCACTCCCTCTCCTTCTCTGAGTATGGTGGCAGCGTCTTACTCAGAACTGTGACCTGTGAGGATGCCC
GTTTC-ATCACTCCCTCTC-TTCTCTGAGTATGGTGGCAGCGTCTTAC-CAGAACTGTGACCTGTGAGGATGCCC
231
AAAAGACCTGCCCTGCAGTGATTGCCCTGTAGCTCTCCAGGCATCAACGGCTTCCAGGCAAAGATGGGCGTGATGGC
AAAAGACCTGCCCTGCAGTGATTGCCCTGTAGCTCTCCAGGCATCAACGGCTTCCAGGCAAAGATGGGCGTGATGGC
1 2 306
ACCAAGGG-AGAAAAGGGGAACAGG-GCCAAGGGCTCAGAGGCTTACAGGGCCCCCTGGAAAGTTGGGGCTTCC
ACCAAGGGTAGAAAAGGGGAAC-AGGT--CAAGGGCTCAGAGGCTTACAGGGCCCCCTGGAAAGTTGGGGCTTCC
2 3 383
AGGAAATCCAGGGCTTCTGGGTCAACAGGACCAAAGGGCCAAAAGGAGACCTGGAAAAGTCCGGATGGTGATA
AGGAAATCCAGGGCTTCTGGGTCAACAGGACCAAAGGGCCAAAAGGAGACCTGGAAAAGTCCGGATGGTGATA
3 4 457
GTAGCC-T-GGCTGCCCTCAGAAAGAAAAGCTCTGCAACAGAAATGGCAGCTATCAAAAAGTGT-AAGCTGACCTTC
GTAGCCC-AGGCTGCCCTCAGAAAGAAAAGCTCTGCAACAGAAATGGCAGCTATCAAAA-GTG-C---CTGACCTTC
534
TCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCTGACCAATGGTGAAATAATGACCTTTGAAAAGTGAAGGCCTT
TCTCTGGGCAAACAAGTTGGGAACAAGTTCTTCTGACCAATGGTGAAATAATGACCTTTGAAAAGTGAAGGCCTT
610
GTGTGTCAAGTTCCAGGCCTCTGTGGCCACCCCCAGGAATG-CTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG
GTGTGTCAAGTTCCAG-CCTCTGTGGCCACCCCCAGGAATGGCTGCAGAGAATGGAGCCATTCAGAATCTCATCAAG
686
GAGGAAGCCTTCTCTGGGCAT-CACTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCT
GAGGAAGCCTTCTCTGGGCATGC-CTGATGAGAAGACAGAAGGGCAGTTTGTGGATCTGACAGGAAATAGACTGACCT
763
ACACAACTGGAACGAGGGTGAACCCAACAATGCTGGTCTGATGAAGATTGTGTATTGCTACTGAAAATGGCCAG
ACACAACTGGAACGAGGGTGAACCCAACAATGCTGGTCTGATGAACATTGTGTATTGCTACTGAAAATGGCCAG
828
TGAATGACGTCCCT-GCT-CCACCCTCCATCTGGCCGCTGTGAGTTCCCTATCTGAAGGGTCAT-----
TGAATGAC-TCCCTTGCTTCCACCCTCCATCTGCC-GTCTGTGAGTTCCCTATCTGAAGGGTC-TGTGAGTTCCC
886
-----ATCACTCA---GGCCCTCCTTGTCTTTTTACTGCA-ACCC-ACAGGCCACAGTATGCTTGAA
TATCTGAAGGGTCAT-A-TCACTCAGGCCCTCCTTGTCTTTTTACT-C-TACCACA---GCCAC-GTATGCTGAA
963
AAGATAATGATTATATCAATTTCTCATATCCAGCATTGTTCCTTTTGTGGCAATCACTAAAAATGATCACTAACA
A-GATAATGAT-----
1040
GCACCAACAAAGCAATAATAGTAGTAGTAGTTAGCAGCAGCAGTAGTAGTCATGCTAATTATATAATATTTT
-----
1117
ATATATACTATGAGGCCCTATCTTTTGCATCCTACAATTAATTATCTAGTTTAATTAATCTGTAATGCTTTTCGATAG
-----
1139
TGTTAACTTGC-----CAGTATGAAA-A
-----AATATATCATTCTCAGATCAGTACCTGCAGATGAAAATATCAGATCAGTACCTGCAG-ATGAA-GA
1214
TAAGACGGA-TTTATTTTCCATTTACAACAAACACCTGTGCTCTGTTGAGCCTTC-CTTCTGTTTGGGTAGAGGG
TAAGACGGCATTATTTTCCATTTACAACAAACACCTGTG-T--GTTGAGCCTT-ACTTCTGTTTGGGTAGAGGG
1290
CTCCCTTAATGACATCACCACAGTTTAATACCACAGCTTTTT-ACCAAGTTTCAGGTATTAAGAAAATCTATTTT
CTCCCTTAATGACATGACCACAGTTTAATACCACAGCTTTTTTACCAAGTTTCAGGTATTAAGAAAATCTATTTT-G
1342
TAACTTTCTCTATGAACCTCTGTTTTCTTTCTAATGAGATATTAACCATGTA
TAACTTTCTCTATGAACCTCTGTTTTCTTTCTAATGAGATATTAACCATGTAAAAAAAAAAAAAAAAAAAA

```

Fig. 4. Comparison of the nucleotide sequences of the exons of cMBP1 (upper) and the hepatic cDNA clone [12] (lower)

The cMBP1 sequence is numbered. The three intron/exon boundaries of cMBP1 are shown as vertical lines and are numbered. The translation initiation and termination codons are marked by horizontal lines. Gaps in the sequence are indicated by hyphens. See the text for details.

from the same gene, we suggest that the sequence of our genomic clone is the correct sequence because the predicted changes increase the overall homology of human MBP to rat MBP C and rat MBP A, as described above.

Upstream regulatory sequences

Examination of the 843 bp of sequence obtained upstream of the 5' end of the cDNA clone [12] showed the presence of several consensus sequences, which are indicated in Fig. 3. TATAA and CAAT

boxes [21] are found at -38 bp and -79 bp from the start of the cDNA clone respectively. There is a region similar to the *Drosophila* heat-shock promoter consensus sequence CTnGAAnnTTCnAG [30] at -592 bp (TAAGAAATTTCCAG). In *Drosophila* the presence of heat-shock consensus sequence in the 5' region of the heat-shock genes is necessary for heat-induced chromosome 'puffing' and the co-ordinate increase in transcription of the heat-shock gene products [30]. The heat-shock consensus element in the 5' region

Vol. 262

with another glucocorticoid-responsive element [35], and at -736 bp is a sequence (AGAACAAATG) similar to the glucocorticoid-responsive element AGAACAGATG [36]. Increased release of glucocorticoid hormones by the adrenals is an important part of the response to stress and results in elevated serum concentrations, principally of cortisol. Thus the glucocorticoid-responsive elements in the promoter region of the MBP gene may also serve to increase MBP expression during the acute-phase response.

The region from -204 bp to -184 bp has 90% homology with a part of the promoter region of the cloned serum amyloid A gene (*SAAg9*) [37], whose product is another acute-phase protein. The expression of the *SAAg9* gene is increased by interleukin 1 and tumour necrosis factor α [38]. The 5' region of the *SAAg9* gene, which contains the elements responsive to interleukin 1, also contains the MBP-like 5' sequence, although it is not known if this is the sequence that is responsive to interleukin 1 (P. Woo, personal communication). The presence of a highly homologous 20 bp sequence in the promoter regions of MBP and serum amyloid A protein suggests that MBP synthesis may also be induced by cytokines and is again consistent with MBP being an acute-phase protein. Regulatory sequences have not yet been identified in the rat MBP A gene [16].

Serum amyloid A protein and C-reactive protein are two major acute-phase proteins, synthesized by the liver, which increase in concentration up to 1000-fold or more in response to tissue injury or inflammation. The details of the regulation of expression of these genes is unclear, but their 5' regions contain heat-shock consensus elements [31] and sequences responsive to cytokines [38]. The report by Ezekowitz *et al.* [12] that hepatic MBP mRNA concentrations were low in a normal liver but were greatly increased in a human liver that had been exposed to acute stress first suggested that MBP might also be an acute-phase protein. The findings in the 5' region of the human MBP gene of a heat-shock consensus sequence, three glucocorticoid-responsive elements and a sequence with a high degree of homology to a similar sequence in the serum-amyloid-A-protein gene lends further support to the view that MBP is an acute-phase protein. Further work is required to establish which factors enhance MBP gene expression and which sequences mediate this effect.

These data also provide insight into the possible biological role of MBP. The signal peptide and collagen-like domain of MBP as well as the evidence that it is an acute-phase protein all indicate that MBP is a hepatic secretory protein. It seems unlikely that this MBP functions as an intracellular transport protein [39]. As an acute-phase protein like serum amyloid A protein or C-reactive protein, serum MBP may be a part of the host defences acting as a carbohydrate-specific immune system. Potential ligands include bacteria, viruses and yeasts that bear mannose-rich oligosaccharides. The binding of MBP to an organism may then result in its destruction by MBP-mediated activation of the complement cascade [10].

We thank Dr. A. Sano for the MBP monoclonal antibody, Dr. Dimitris Kioussis for the cosmid library, Dr. K. Drickamer for amino acid sequencing, Dr. M. Waterfield for oligonucleotide synthesis and Dr. J. Scott, Dr. A. Ezekowitz,

Dr. P. Woo and Dr. T. Lund for advice. This work was supported by the Wellcome Trust.

REFERENCES

1. Townsend, R. & Stahl, P. (1981) *Biochem. J.* **194**, 209-214
2. Mizuno, Y., Kozutsumi, Y., Kawasaki, Y. & Yamashina, I. (1981) *J. Biol. Chem.* **256**, 4247-4252
3. Maynard, Y. & Baenziger, J. U. (1982) *J. Biol. Chem.* **257**, 3788-3794
4. Kawasaki, T., Etoh, R. & Yamashina, I. (1978) *Biochem. Biophys. Res. Commun.* **81**, 1018-1024
5. Kozutsumi, Y., Kawasaki, T. & Yamashina, I. (1980) *Biochem. Biophys. Res. Commun.* **95**, 658-664
6. Kawasaki, N., Kawasaki, T. & Yamashina, I. (1983) *J. Biochem. (Tokyo)* **94**, 937-947
7. Summerfield, J. A. & Taylor, M. E. (1986) *Biochim. Biophys. Acta* **883**, 197-206
8. Taylor, M. E. & Summerfield, J. A. (1987) *Biochim. Biophys. Acta* **915**, 60-67
9. Wild, J., Robinson, D. & Winchester, B. (1983) *Biochem. J.* **210**, 167-174
10. Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T. & Yamashina, I. (1987) *J. Biol. Chem.* **262**, 7451-7454
11. Drickamer, K., Dordal, M. S. & Reynolds, L. (1986) *J. Biol. Chem.* **261**, 6878-6887
12. Ezekowitz, A. B., Day, L. E. & Herman, G. A. (1988) *J. Exp. Med.* **167**, 1034-1046
13. Drickamer, K. (1988) *J. Biol. Chem.* **263**, 9557-9560
14. Haagsman, H. P., Haegood, S., Sargent, T., Buckley, D., White, R. T., Drickamer, K. & Benson, B. J. (1987) *J. Biol. Chem.* **262**, 13877-13880
15. Halberg, D. F., Proulx, G., Doege, K., Yamada, Y. & Drickamer (1988) *J. Biol. Chem.* **263**, 9486-9490
16. Drickamer, K. & McCreary, V. (1987) *J. Biol. Chem.* **262**, 2582-2589
17. Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J. & Playfair, J. H. L. (1987) *EMBO J.* **6**, 355-361
18. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961-3965
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467
21. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349-383
22. Yamada, Y., Liau, G., Mudryj, M., Obici, S. & de Crombrughe, G. (1984) *Nature (London)* **310**, 333-337
23. Lozano, G., Ninomya, Y., Thompson, H. & Olsen, B. R. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4050-4054
24. Kurkinem, M., Bernard, M. P., Barlow, D. P. & Chow, L. T. (1985) *Nature (London)* **317**, 177-179
25. Sakurai, Y., Sullivan, M. & Yamada, Y. (1986) *J. Biol. Chem.* **261**, 6654-6657
26. Leung, J. O., Holland, E. C. & Drickamer, K. (1985) *J. Biol. Chem.* **260**, 12523-12527
27. von Heigne, G. (1982) *J. Mol. Biol.* **159**, 537-541
28. von Heigne, G. (1983) *Eur. J. Biochem.* **133**, 17-21
29. Drickamer, K. (1987) *Kidney Int.* **32**, S167-S180
30. Simon, J. A., Sutton, C. A., Lobell, R. B., Glaser, R. L. & Lis, J. T. (1985) *Cell* **40**, 805-817
31. Woo, P., Korenberg, J. R. & Whitehead, A. S. (1985) *J. Biol. Chem.* **260**, 13384-13388

32. Schutz, G. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 77–86
33. Martinez, E., Francoise, G. & Wahli, W. (1987) *EMBO J.* **6**, 3719–3727
34. Evans, R. M. (1988) *Science* **240**, 889–895
35. Payvar, F., De Franco, D., Firestone, G. L., Edgar, B., Wrange, O., Okret, S., Gustafsson, J.-A. & Yamamoto, K. R. (1983) *Cell* **35**, 381–392
36. Speck, N. A. & Baltimore, D. (1987) *Mol. Cell. Biol.* **7**, 1101–1110
37. Edbrook, M. R., Burt, D. W., Cheshire, J. K. & Woo, P. (1989) *Mol. Cell. Biol.* **9**, 1908–1916
38. Woo, P., Sipe, J., Dinarello, C. A. & Colten, H. R. (1987) *J. Biol. Chem.* **262**, 15790–15795
39. Mori, K., Kawasaki, T. & Yamashina, I. (1988) *Arch. Biochem. Biophys.* **264**, 647–656

Received 19 January 1989/3 April 1989; accepted 27 April 1989