THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 265, No. 25, Issue of September 5, pp. 15040-15048, 1990
Printed in U.S.A.

Molecular Structure, Chromosome Assignment, and Promoter Organization of the Human Matrix Gla Protein Gene*

(Received for publication, March 19, 1990)

Leonor Cancela‡, Chih-Lin Hsieh§, Uta Francke§, and Paul A. Price‡¶

From the ‡Department of Biology, University of California at San Diego, La Jolla, California 92093 and the §Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305-5428

Matrix Gla protein (MGP) is an 84-residue vitamin K-dependent protein initially isolated from bovine bone. MGP is also expressed at high levels in heart, kidney, and lung and is up-regulated by vitamin D in bone cells. To characterize the genomic sequences responsible for the regulated expression of this gene, we screened a human genomic library using a MGP cDNA probe and obtained two clones containing the MGP locus. The human MGP gene spans 3.9 kilobases of chromosomal DNA and consists of four exons separated by three large intervening sequences which account for more than 80% of the gene. Southern analysis of total human genomic DNA indicated the presence of a single copy of the MGP gene. Hybridization of the hMGP cDNA to a series of Chinese hamster × human hybrid clones assigned this gene to the short arm of the human chromosome 12 (12p). The N-terminal sequences of the known vitamin K-dependent vertebrate proteins reveal a transmembrane signal peptide, followed by a putative γ-carboxylation recognition site and a Gla-containing domain. Each of these regions correspond to a separate exon in MGP. MGP also contains a fourth exon of unknown function which codes for 11 residues and lies between the transmembrane signal peptide and the putative recognition site for the γ-carboxylase. This four-exon organization is essentially identical to that of bone Gla protein and is quite different from the two exon organization encoding this region in the other known vitamin K-dependent proteins. Analysis of the MGP gene promoter revealed, in addition to the typical TATA and CAT boxes, the presence of a number of putative regulatory sequences homologous to previously identified hormone and transcription factor responsive elements. In particular, two regions of the promoter were delineated containing possible binding sites for retinoic acid and vitamin D receptors.

Matrix Gla[1] protein (MGP) is one of two small vitamin K-dependent proteins which have been isolated from bone. Unlike bone Gla protein (BGP), which has proven to be specific to calcified tissues (1), MGP has recently been shown to be expressed in many different tissues including heart, lung and kidney. In each tissue, MGP synthesis is restricted to a discrete set of tissue-specific cells (2). Although MGP contains a high percentage of hydrophilic residues, it is exceptionally water-insoluble (solubility < 10 μg/ml). The MGP cDNA structure predicts an 84-residue mature protein and a 19-residue transmembrane signal peptide (3, 4), MGP being the only known vitamin K-dependent protein which lacks a propeptide. The two forms of MGP that have been so far isolated from bone contain 79 (3) and 83[2] residues and are missing 5 and 1 residues at the C terminus.

Vitamin D_3, through its hormonally active form 1,25-dihydroxyvitamin D_3, has been shown to produce a rapid, 15-fold increase in the level of MGP message in several bone cell lines (5). MGP is one of the few proteins to be significantly up-regulated by vitamin D_3, the others being BGP (6–9) and the calbindins (10, 11). As has also been shown for the calbindins (12, 13), the regulation of MGP by vitamin D appears to be tissue-specific.[3] A vitamin D-responsive element has recently been identified within the promoter of the rat (14) and human (8) BGP genes, and the structures of the calbindin genes have been determined (15–17). In order to compare the MGP gene with the genes for other vitamin D-regulated proteins, we have cloned the human MGP gene and characterized its molecular organization including 5′-promoter elements and 3′-flanking DNA. The promoter region of the hMGP gene contains several putative regulatory elements homologous to previously well characterized factors, including possible vitamin D- and retinoic acid-responsive elements. Comparison of the MGP gene with the genes for other vitamin K-dependent proteins has revealed that, although they may all share a common ancestor, only BGP seems to have followed the same exact pattern of exon evolution as MGP.

MATERIALS AND METHODS

Cloning of the Human MGP cDNA—Two rat MGP cDNAs, $rMGP_6$, and $rMGP_1$ (see Ref. 3) were used to screen a human lung λgt11 cDNA library (Clontech, Palo Alto, CA). Bacteriophage were plated and transferred to nitrocellulose filters (18). Two replica filters were lifted from each plate. Following hybridization to the ^{32}P-labeled rat cDNAs, positive phage plaques present in both filter replicas of the same plate were plaque-purified and DNA prepared from the phage stocks (18). The cDNA inserted into the phage was excised with *Eco*RI and subcloned into an appropriate plasmid vector for further characterization and sequencing.

Cloning of the Human MGP Gene—Both rat and human MGP

* This work was supported in part by United States Public Health Grant AR25921. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05572.

¶ To whom correspondence should be addressed.

[1] The abbreviations used are: Gla, γ-carboxyglutamic acid; bp, base pairs; kb, kilobase pairs; BGP, bone Gla protein, osteocalcin; h, human; MGP, matrix Gla protein; MRE, metal responsive element; HRE, hormone responsive element.

[2] J. E. Hale, M. K. Williamson, and P. A. Price, unpublished results.

[3] Results obtained with normal rat kidney cells show that the expression of the MGP gene in this system is not vitamin D-dependent (J. D. Fraser and P. A. Price, unpublished results).

cDNAs (see above) were used to screen a library derived from partially digested human leukocyte genomic DNA cloned into the bacteriophage EMBL-3 (Clontech). Approximately 7×10^5 phage plaques were screened by *in situ* hybridization (18). Positive clones were plaque-purified and DNA was prepared following established procedures (18). Further characterization of the genomic positive clones was accomplished by digestion with restriction endonucleases and Southern hybridization analysis (19). Selected genomic restriction fragments were subcloned into the *Bam*HI, *Hin*dIII, or *Xba*I site of a plasmid vector for further characterization and sequencing.

DNA Sequence Determination—Sequencing of the human MGP cDNA and gene was performed on DNA fragments subcloned into the Bluescript (Stratagene, La Jolla, CA) plasmid vector using a modified DNA polymerase (Sequenase, obtained from United States Biochemical Corp.) and the dideoxy chain termination method of Sanger (20). Sequence data were obtained from various subcloned restriction fragments as well as from 5′- and 3′-unidirectional deletions of cDNA and genomic DNA cloned into pBluescript and generated using the Erase-A-Base system (Promega, Madison, WI). The nearly full length cDNA, as well as the complete gene and its 5′- and 3′-flanking DNA, extending from −1063 to +4277, were sequenced from both strands at least twice. Genomic DNA sequences located further upstream from the start site of transcription (from −1064 to −3398) were sequenced from at least two different clones and approximately 40% was sequenced from both strands.

Computer Analysis—Identification of putative regulatory elements within the hMGP gene and flanking DNA was accomplished using the Staden ANALYSEQ program (21) and the Los Alamos SEQUENCE HOMOLOGY program (22). Identification of repetitive DNA elements was performed by comparing our sequence with those present in the GenBank data base using the Lipman-Pearson FASTA programs (23, 24). All programs were run on a Digital Equipment Corp. VAX 11780 computer system.

Isolation of Human Genomic DNA—Cells (human MG-63 osteosarcoma cells, American Type Culture Collection, CRL 1427), grown to confluency in the presence of newborn calf serum were trypsinized, washed twice in ice-cold phosphate-buffered saline, and pelleted. Genomic DNA was prepared following a previously described procedure (25). Briefly the cell pellet was resuspended in 1 ml of digestion buffer (100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 25 mM EDTA, pH 8.0, 0.5% sodium dodecyl sulfate) per 10^8 cells and 0.1 mg/ml proteinase K (Boehringer Mannheim). The samples were incubated with shaking overnight (18 h) at 50 °C in tightly capped tubes. Nucleic acids were extracted and cleaned as described previously (25), resuspended in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and kept at 4 °C until restriction enzyme digestions were performed.

Southern Analysis—Genomic DNA from human MG-63 osteosarcoma cells, prepared as described above, was digested with different restriction endonucleases, electrophoresed on a 0.8% agarose gel, and transferred to Nytran (0.45 μm, Schleicher & Schuell). Hybridization was performed at 42 °C under the conditions recommended by the manufacturer (26) using a random primed, ^{32}P-labeled human MGP cDNA probe at the concentration of 3×10^6 cpm/ml of hybridization solution (specific activity was approximately 10^8 cpm/μg of labeled DNA). The membrane was washed at 65 °C under high stringency conditions (26) and autoradiographed for 2–5 days at −70 °C with intensifying screens. For the chromosome localization, the *Pst*I-digested DNA fragments from each of the hybrid and parental control cell lines were separated on 0.8% agarose gels and Southern transferred to Hybond N membranes (Amersham Corp.). The filters were prehybridized, hybridized with the hMGP cDNA (^{32}P-labeled by random priming to a specific activity of 7×10^8 cpm/μg following the method of Feinberg and Vogelstein (27)), rinsed, and washed as described previously (28). All filters were exposed to Kodak X-Omat AR films with two intensifying screens at −70 °C for various lengths of time.

Labeled DNA Probes—Plasmid vectors containing MGP cDNA or genomic fragments were labeled with [^{32}P]dCTP (>3000 Ci/mmol, PB10205 from Amersham Corp.) using a Nick Translation System (Amersham Corp.). Gel-purified MGP cDNA inserts were labeled with [^{32}P]dCTP (>3000 Ci/mmol) using a Multiprime DNA labeling system (Amersham Corp.). An average of 10^6 to 10^7 cpm were used per ml of hybridization solution.

Ribonuclease Protection Assay—Mapping of the hMGP mRNA

```
                    (-19)                                (-10)
                    MetLysSerLeuIleLeuLeuAlaIleLeuAlaA
CTGAGACTGACCTGCAGGACGAAACCATGAAGAGCCTGATCCTTCTTGCCATCCTGGCCG      60
                (-1)(1)                              (10)
laLeuAlaValValThrLeuCysTyrGluSerHisGluSerMetGluSerTyrGluLeuA
CCTTAGCGGTAGTAACTTTGTGTTATGAATCACATGAAAGCATGGAATCTTATGAACTTA     120

            (20)                                 (30)
snProPheIleAsnArgArgAsnAlaAsnThrPheIleSerProGlnGlnArgTrpArgA
ATCCCTTCATTAACAGGAGAAATGCAAATACCTTCATATCCCCTCAGCAGAGATGGAGAG     180
            (40)                                 (50)
laLysValGlnGluArgIleArgGluArgSerLysProValHisGluLeuAsnArgGluA
CTAAAGTCCAAGAGAGGATCCGAGAACGCTCTAAGCCTGTCCACGAGCTCAATAGGGAAG     240
            (60)                                 (70)
laCysAspAspTyrArgLeuCysGluArgTyrAlaMetValTyrGlyTyrAsnAlaAlaT
CCTGTGATGACTACAGACTTTGCGAACGCTACGCCATGGTTTATGGATACAATGCTGCCT     300
            (80)        (84)
yrAsnArgTyrPheArgLysArgArgGlyThrLys***
ATAATCGCTACTTCAGGAAGCGCCGAGGGACCAAATGAGACTGAGGGAAGAAAAAAAATC     360

TCTTTTTTTCTGGAGGCTGGCACCTGATTTTGTATCCCCCTGTAGCAGCATTACTGAAAT     420
ACATAGGCTTATATACAATGCTTCTTTCCTGTATATTCTCTTGTCTGGCTGCACCCCTTT     480
TTCCCGCCCCCAGATTGATAAGTAATGAAAGTGCACTGCAGTGAGGGTCAAAGGAGAGTC     540
AACATATGTGATTGTTCCATAATAAACTTCTGGTGTGATACTTTC-poly(A)tail        585
```

FIG. 1. **Nucleotide sequence of the cDNA encoding human MGP.** The derived amino acid sequence is shown above the DNA sequence. Amino acid residues are numbered according to residue 1 of the mature protein. Numbering of nucleotides is shown on the *right margin*. The stop codon is marked by *asterisks* and the polyadenylation signal is *underlined twice*. Sequences present in our cDNA and not previously reported (4) are *underlined*. A single difference between our sequence data and the previously published sequence was noted and is marked in *bold* as well as the corresponding amino acid.

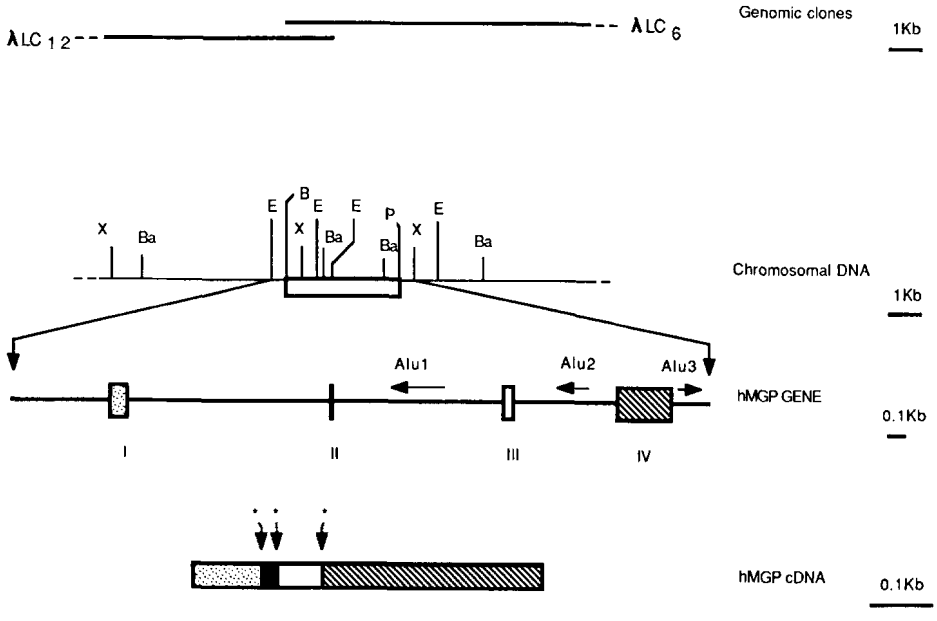

FIG. 2. **Map and exon/intron organization of the human MGP gene.** The *top two lines* represent the location of the overlapping phage clones λLC_{12} and λLC_6 containing the entire hMGP gene. The middle line shows a partial restriction endonuclease map of the MGP locus (*X*, *Xba*I; *Ba*, *Bam*HI; *E*, *Eco*RV; *B*, *Bgl*I; *P*, *Pst*I). The *box* indicates the location of the hMGP gene and it is expanded below to show its exon/intron organization and location of the *Alu*I repeats. Each exon encodes a different region of the protein. *Dotted box I* = 5′-untranslated region + pre-domain of the protein; *black box II* = α-helical domain; *white box III* = putative γ-carboxylase recognition site; *hatched box IV* = Gla-containing domain + 3′-untranslated region. The structure of the corresponding hMGP cDNA is shown at the *bottom*. The location of the splice junctions within the cDNA is indicated by the *vertical arrows* and the *asterisks*.

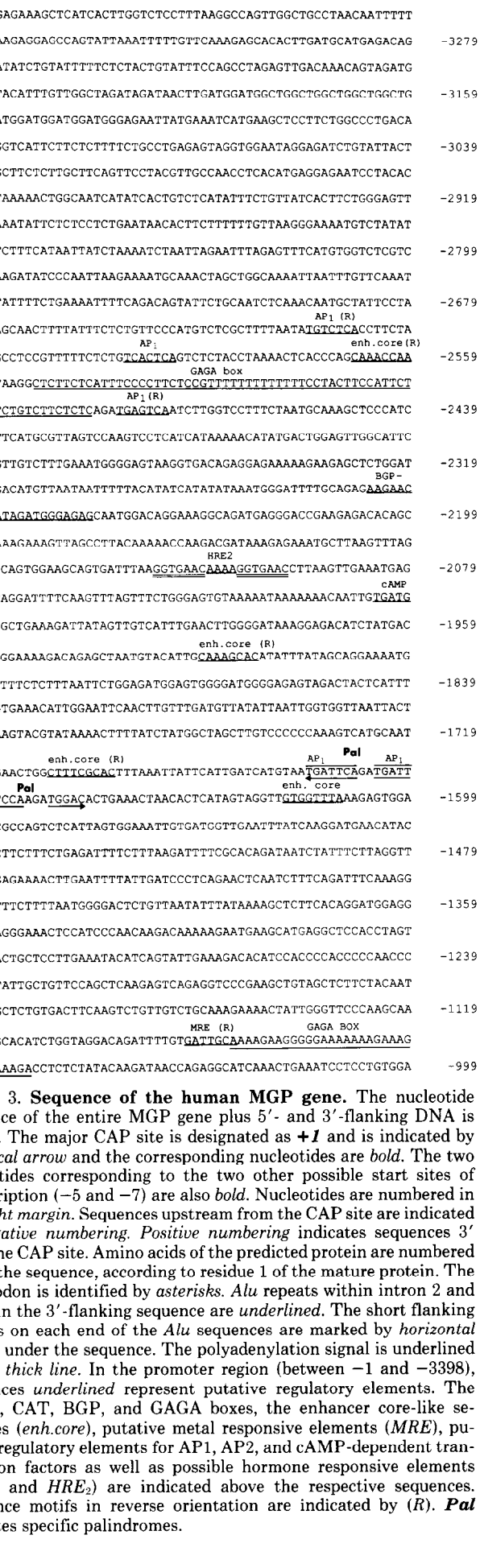

FIG. 3. **Sequence of the human MGP gene.** The nucleotide sequence of the entire MGP gene plus 5′- and 3′-flanking DNA is shown. The major CAP site is designated as ***+1*** and is indicated by a *vertical arrow* and the corresponding nucleotides are *bold*. The two nucleotides corresponding to the two other possible start sites of transcription (−5 and −7) are also *bold*. Nucleotides are numbered in the *right margin*. Sequences upstream from the CAP site are indicated by *negative numbering*. *Positive numbering* indicates sequences 3′ from the CAP site. Amino acids of the predicted protein are numbered above the sequence, according to residue 1 of the mature protein. The stop codon is identified by *asterisks*. *Alu* repeats within intron 2 and 3 and in the 3′-flanking sequence are *underlined*. The short flanking repeats on each end of the *Alu* sequences are marked by *horizontal arrows* under the sequence. The polyadenylation signal is underlined with a *thick line*. In the promoter region (between −1 and −3398), sequences *underlined* represent putative regulatory elements. The TATA, CAT, BGP, and GAGA boxes, the enhancer core-like sequences (*enh.core*), putative metal responsive elements (*MRE*), putative regulatory elements for AP1, AP2, and cAMP-dependent transcription factors as well as possible hormone responsive elements (HRE_1 and HRE_2) are indicated above the respective sequences. Sequence motifs in reverse orientation are indicated by (*R*). ***Pal*** indicates specific palindromes.

FIG. 3—*continued*

```
TyrArgLeuCysGluArgTyrAlaMetValTyrGlyTyrAsnAlaAlaTyrAsnArgTyr
TACAGACTTTGCGAACGCTACGCCATGGTTTATGGATACAATGCTGCCTATAATCGCTAC      +3682
      (80)        (84)
PheArgLysArgArgGlyThrLys***
TTCAGGAAGCGCCGAGGGACCAAATGAGACTGAGGGAAGAAAAAAAATCTCTTTTTTTCT

GGAGGCTGGCACCTGATTTTGTATCCCCCTGTAGCAGCATTACTGAAATACATAGGCTTA      +3802
TATACAATGCTTCTTTCCTGTATATTCTCTTGTCTGGCTGCACCCCTTTTTCCCGCCCCC
AGATTGATAAGTAATGAAAGTGCACTGCAGTGAGGGTCAAAGGAGAGTCAACATATGTGA      +3922
TTGTTCCATAATAAACTTCTGGTGTGATACTTTCATCTTGTAAATCTGCTTTCTTTTGGG
AAGATATTGAGATATTTAAATCATGGCCCACCTTACCCAAAATAGGAGATTCTGTTCATC      +4042
TCATATCTAGTATTAATTAGAAAAATAACTACATAAAAAGAAGGAAGCTAAGAAGGCACT
CACTCAGCCATAAATTCTCTAAACCCTCTCTACCTTGGAATCCGTGAATGGAATCTGGTA      +4162
TGTTTTTTGCAGGATTTTCCTATTGTAAATTGTGGCAAATACAGGGCTCCCTTCATTTGC
TTTTCATCTCTTATGCATCAAAGTCAAAAACATTTCTGAATCAAGATAATCTAGA           +4277
```

FIG. 3—*continued*

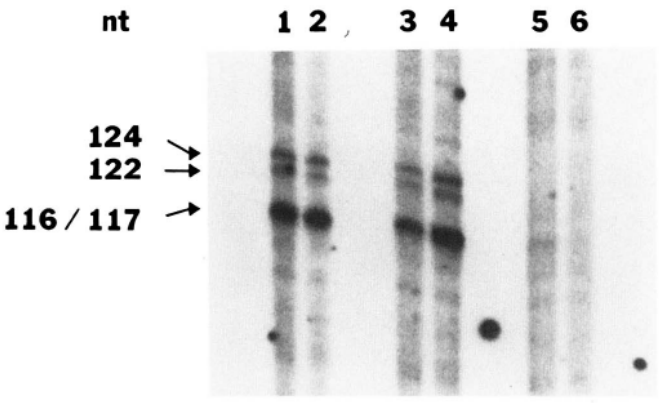

FIG. 4. **Mapping of the hMGP mRNA CAP site.** RNase protection assay was performed as described under "Materials and Methods." *Lanes 1* and *2*, 5 and 10 μg of cytoplasmic RNA from human fetal osteoblasts (FOB cells); *lanes 3, 4*, 5 and 10 μg of total RNA from FOB cells; *lanes 5, 6*, 5 and 10 μg of total RNA from MG 63 cells. The lengths of the protected fragments are shown on the *left* and were determined using ^{32}P-labeled fragments generated by *Msp*I digestion of pBR$_{322}$ as well as an unrelated sequencing ladder which were loaded on adjacent lanes and used as size markers. *nt*, nucleotides.

CAP site was performed essentially as described (29). Total and cytoplasmic RNA from FOB cells (human normal fetal osteoblasts) and total RNA from MG-63 cells (human adult osteosarcoma cells) were isolated by the guanidinium isothiocyanate/cesium chloride method (18, 30). A 5′-*Eco*RV/3′-*Xba*I genomic DNA fragment extending from −669 to +308 (according to the ATG start codon of the hMGP gene) was subcloned into pBluescript immediately adjacent to the T7 promoter. The construct was linearized by the restriction enzyme *Sal*I and transcribed *in vitro* using the T7 RNA polymerase (Stratagene). The resulting ^{32}P-labeled complementary RNA fragment was hybridized overnight to 5 or 10 μg of RNA, at 45 °C. RNA hybrids were RNase-treated for 40 min at 30 °C and then digested with proteinase K for 15 min at 37 °C, phenol/chloroform-extracted, and ethanol-precipitated. RNA pellets were resuspended in 5 μl of sequencing dye, heat-denatured (5 min at 80 °C), and loaded on a 6% acrylamide (38/2 acrylamide/bis-acrylamide, w/w), 7 M urea sequencing gel. The size of the protected fragments was determined from DNA markers loaded on adjacent lanes and consisting of ^{32}P-end-labeled *Msp*I restriction fragments of the plasmid pBR$_{322}$, and an unrelated sequencing ladder.

Somatic Cell Hybrids—Primary chromosomal assignment of the MGP gene in human was carried out with a panel of 11 hybrid clones which were derived from five independent fusion experiments between Chinese hamster and human cell lines (for summary see Ref. 31). Two Chinese hamster × human hybrids from series XXI containing different regions of human chromosome 12 were used for regional mapping.

RESULTS

Human MGP cDNA Structure—As an initial step toward the complete characterization of the human MGP gene, a human MGP cDNA clone was isolated from a human lung cDNA library by screening with ^{32}P-labeled rat MGP cDNA (3). A nearly full length clone was obtained and entirely sequenced in both directions. This 585-bp clone contains 26 bp of 5′-untranslated message and extends to the site of insertion of the poly(A) tail (Fig. 1).

Our sequence analysis revealed a cDNA structure nearly identical to the recently reported sequence for the hMGP cDNA (4). A single difference was noted at the site encoding amino acid 83 of the mature protein, which in our clone reads [ACC] and codes for *threonine* (see Fig. 1) and in the previously reported sequence (4) reads [GCC] and codes for *alanine*. The sequence of the hMGP gene reported here confirms the presence of a threonine and not alanine at position 83. Our clone also contains an additional 20 base pairs of sequence upstream from the 5′ end of the previously reported hMGP cDNA (Fig. 1, Ref. 4), and the insertion point of the poly(A) tail is located slightly further downstream from the AATAAA polyadenylation signal than previously reported (4).

Isolation of the Human MGP Gene—Screening of a human leukocyte genomic library resulted in the identification of two positive clones, λLC$_6$ and λLC$_{12}$ (see Fig. 2). Upon further characterization by Southern analysis, λLC$_6$ (spanning nearly 14 kb of chromosomal DNA) was found to contain a 3.9-kb *Xba*I fragment which hybridized to both the rat and human MGP cDNA clones. This fragment contains most of intron 1 and extends 340 bp 3′ past the polyadenylation signal. λLC$_{12}$ (spanning nearly 13 kb of chromosomal DNA) was found to overlap λLC$_6$ by approximately 1.5 kb and to extend more than 10 kb upstream from the human MGP gene. A 6.0-kb *Bam*HI fragment overlapping λLC$_6$ by approximately 1100 bp was subcloned into pBluescript and used to obtain the remaining sequence of the human MGP gene as well as the sequence of nearly 3.5 kb of 5′-flanking DNA (Figs. 2 and 3).

Organization of the Human MGP Gene—The human MGP gene spans 3937 bp of chromosomal DNA from the major start site of transcription (Figs. 3 and 4) to the polyadenylation signal and contains four exons separated by three large intervening sequences which account for nearly 85% of the total DNA of the MGP gene. The main CAP site of the hMGP mRNA was identified to a CC motif located 33 bp downstream

TABLE I

Exon and intron organization of the human MGP gene

The nucleotides surrounding the splice donor and acceptor sites for each of the three introns are indicated. Exon and intron lengths are given in base pairs.

Exon	Exon length	Splice donor	Intron no. and length	Splice acceptor
	bp		*bp*	
I	61[a] + 5′-NT[b]	TATG| GTGAGA	1 (1485)	C(T)$_{10}$ATTTCAG| AATC
II	33	CTTA| GTAAGT	2 (1165)	(T)$_4$CTGTTCTAAG| ATCC
III	76	AGAG| GTCAGT	3 (691)	CTTCCAC(T)$_3$CTAG| GATC
IV	139[a] + 3′-NT			

[a] Only protein-coding nucleotides are counted.
[b] NT, nontranslated region.

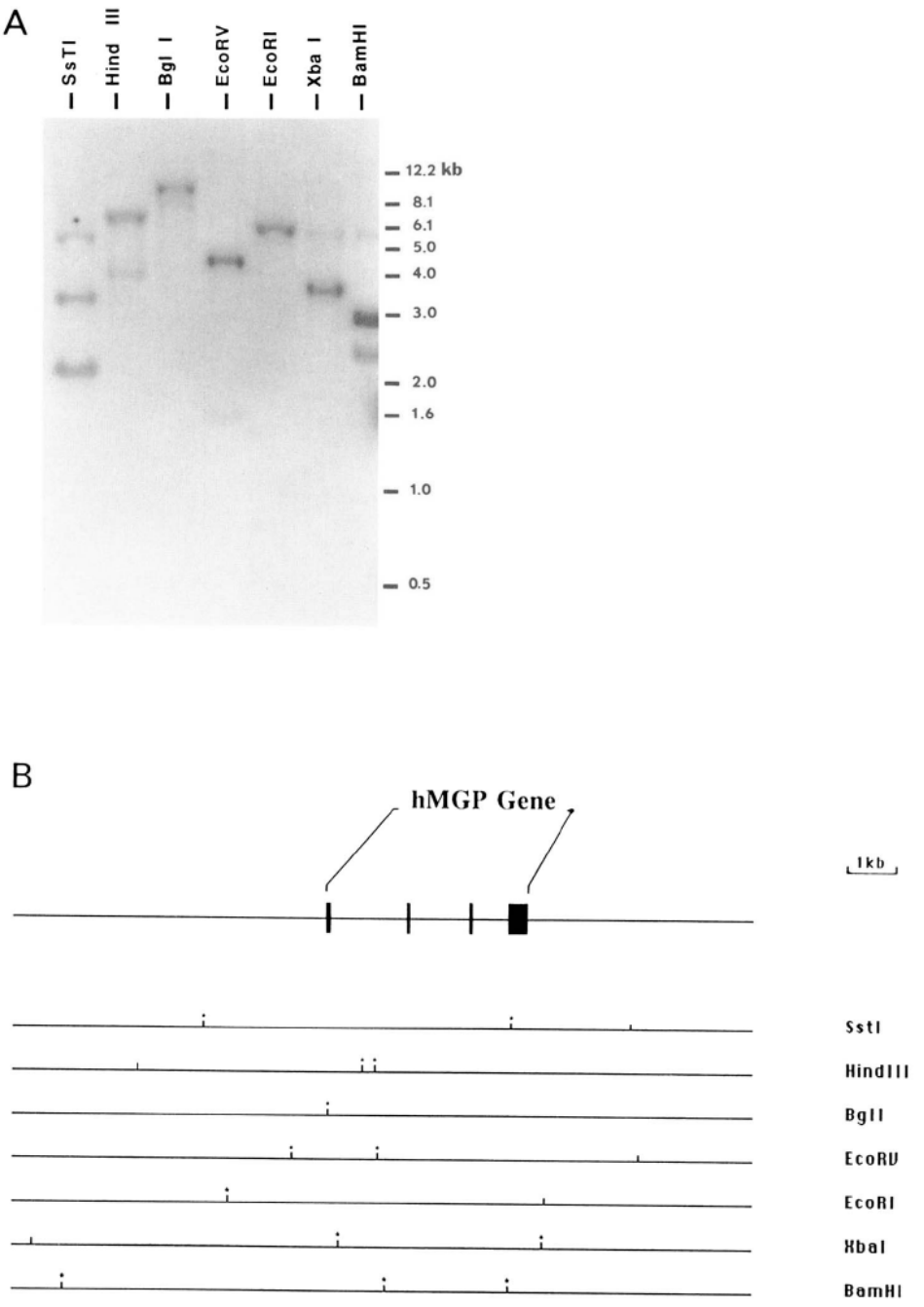

FIG. 5. **Southern analysis of the human MGP gene locus.** *A*, hybridization of the full length human MGP cDNA to digests of human genomic DNA. Following restriction enzyme digestion, DNA samples were electrophoresed on a 0.8% agarose gel, transferred to Nytran, and hybridized to the hMGP cDNA as described under "Materials and Methods." Size markers are shown on the *right margin*. *B*, localization of the restriction enzyme sites within and surrounding the human MGP gene as obtained from the mapping of the genomic clones. The positions of the sites marked with *asterisks* were confirmed by sequencing analysis.

from the TATA box and corresponding to protected fragments of 116/117 bp in length (Fig. 4). The first C residue was taken as the +1 nucleotide. Two additional fragments were found to be protected (122 and 124 bp in length) corresponding most likely to two weaker transcription start sites located at positions −5 and −7 (Figs. 3 and 4). No protected fragments were detected when using RNA from MG 63 cells which do not express MGP (unpublished results).

The nucleotide sequences flanking the exons of the hMGP gene are characteristic of eukaryotic genes. The three introns are bounded by consensus 5′-(GT) and 3′-(AG) splicing sequences following the rule of Breathnach and Chambon (32) (Table I). The existence of a unique polyadenylation signal is in agreement with the presence of a single size of message for MGP in human (results not shown) and rat tissues (2, 5).

Three *Alu*I repeats, a common feature of the transcribed portion of the human genome, are found within the MGP gene structure, one each in introns 2 and 3 and in the 3′-flanking DNA (see Figs. 2 and 3). The first two repeats share, respectively, 90 and 84% identity with the *Alu*I consensus sequence (33), while the last repeat is much less conserved. Also typical of *Alu*-repetitive DNA is the presence of short flanking repeats at each end of the *Alu* sequence (33). Such flanking sequences are found at the 5′ and 3′ ends of the *Alu* repeats in the hMGP gene (Fig. 3).

Identification of Putative Regulatory Elements within the MGP Gene Promoter—The genomic sequence flanking the 5′ end of the MGP gene is typical of RNA polymerase II transcribed genes. Immediately upstream from the transcription initiation site there is a consensus TATA sequence at position −33 bp (TTATAAAAA) (Fig. 3) (34). Two CAT boxes are located at −119 bp (CCAAT) and −527 bp (ATTGG = reverse motif) (35).

The human MGP gene promoter contains a number of sequence elements which are homologous to previously identified consensus sequences known to interact with well char-

acterized regulatory factors (Fig. 3). These include putative binding sites for the transcription factors AP1 (TGACTCA) (36) and AP2 (CCCAGGC) (37) as well as possible binding sites for cAMP-dependent transcription factors (TGACGTCA) (38) and a number of metal responsive elements ($TGC^{A}_{G}C^{T}_{C}C$) (39) (consensus sequences are in parentheses). Many of these putative regulatory sequences are flanked by enhancer-like core elements (consensus sequence = $GTGG^{TTT}_{AAA}G$, Ref. 40) which are known to exert a potent effect on gene expression (41).

Given the possibility that the 1,25-dihydroxyvitamin D_3 receptor complex is a trans-acting modulator of MGP gene expression, we have searched the MGP gene promoter sequence for the presence of a motif (GGTGA, CCACT) known to be part of the vitamin D receptor binding site in the human and rat BGP gene promoters (8, 14). Two regions of interest were located. One contains two copies of the reverse repeat CCACT in opposite orientation (−819 to −832) and the other contains two copies of the direct repeat GGTGA in tandem (−2095 to −2112). Interestingly, the second of these sequence motifs is nearly identical to the retinoic acid responsive element recently identified within the promoter of the retinoic acid receptor β-gene (42). The possibility that these sequences may indeed represent hormone binding sites is currently being investigated.

Other noteworthy motifs include: a perfect palindrome (CAGGCAAATGCCTG) overlapping an AP_2 consensus sequence and located immediately upstream from the reverse CAT box (−564 to −551). A palindrome (TTCCAAGATGGAC) located between −1654 and −1642 which is immediately preceded by two AP_1 sites in tandem. Both of these motifs share some homology to the consensus sequences for steroid hormone responsive elements previously identified (for a review, see Ref. 43).

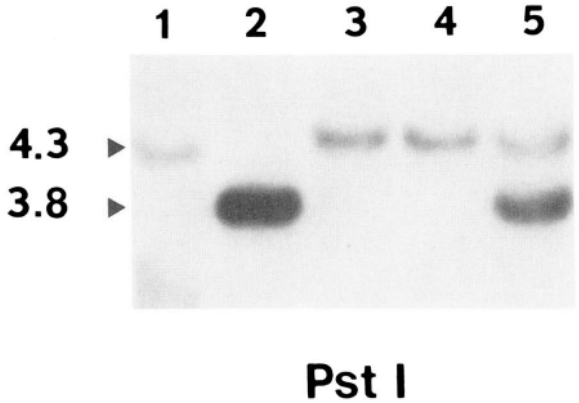

FIG. 6. **Hybridization of ^{32}P-labeled human cDNA probe to a Southern blot of *Pst*I-digested DNA from Chinese hamster × human hybrid cell lines and controls.** *Lane 1*, Chinese hamster cells V79/380-6; *lane 2*, human diploid lymphoblastoid cells; *lanes 3–5*, Chinese hamster × human hybrid cell lines. Only the hybrid in lane 5 is positive for the human fragment.

Two other types of sequence elements were identified. The MGP promoter contains three boxes showing a moderate homology (66–70% identity) with a sequence recently identified in both rat and human BGP promoters (6) and termed "osteocalcin or BGP box" (Fig. 3). Several [GA] motifs called GAGA boxes and first identified within the promoters of osteonectin (44, 45) and rat BGP (7) genes are also present in the MGP promoter (Fig. 3).

The MGP Gene Is a Single Copy Gene within the Human Genome—To further characterize the human MGP gene, samples of human genomic DNA were digested with several restriction endonucleases and the resulting fragments analyzed by Southern blot hybridization. For most of the enzymes, only those restriction fragments expected from the mapping of the human MGP locus were observed (Fig. 5). Since there is also full agreement between the exon sequences and the cDNA sequence for hMGP, these results strongly suggest the presence of a single copy of the MGP gene within the human genome. An additional restriction fragment was observed when the genomic DNA was digested with *Sst*I (only two fragments are predicted from the mapping of the hMGP gene; see Fig. 4, *A* and *B*), which is indicative of the existence of a restriction fragment length polymorphism for this enzyme. This possibility is currently being investigated.

The MGP Gene Is Located in Human Chromosome 12p—The MGP gene was mapped to human chromosome 12 by hybridizing the ^{32}P-labeled hMGP cDNA to *Pst*I-digested genomic DNA from Chinese hamster × human somatic cell hybrids and controls. A single *Pst*I fragment of approximately 3.8 kb was detected in human DNA (Fig. 6, *lane 2*) corresponding to the genomic fragment originated from *Pst*I sites located at +42 and +3887 bp in the hMGP gene, and a 4.3-kb fragment was observed in *Pst*I-digested Chinese hamster DNA (Fig. 6, *lane 1*). Both the 3.8-kb human and the 4.3-kb Chinese hamster signals were present in hybrid clones containing human chromosome 12 (Fig. 6, *lane 5*). The human band was absent in hybrids not containing human chromosome 12 (Fig. 6, *lanes 3* and *4*). The human MGP signal was in perfect concordance with the presence of human chromosome 12 in hybrid cell lines. All other human chromosomes were excluded by 2 to 6 discordant hybrids (Table II). We have further localized MGP to the short arm of chromosome 12 (12p) by studying two hybrids that contain only partially

TABLE II

Correlation of human MGP sequences with human chromosomes in rodent × human somatic cell hybrids

The numbers of hybrids that are concordant (+/+ or −/−) and discordant (+/− or −/+) with the human MGP sequence are given for each chromosome. Hybrids in which a particular chromosome was structurally rearranged or present in fewer than 10% of cells were excluded.

Hybridization/ chromosome	Human chromosomes																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
+/+	2	2	2	1	0	4	1	1	1	1	2	4	1	2	3	2	0	3	2	3	4	3	2
−/−	6	6	4	5	4	5	4	4	5	6	4	7	2	1	3	4	4	3	4	4	4	2	1
+/−	2	2	1	1	4	0	3	3	2	3	1	0	3	2	1	2	4	1	2	1	0	1	0
−/+	1	0	3	1	2	2	1	3	2	1	2	0	3	4	4	2	2	4	3	3	2	4	2
Discordant hybrids	3	2	4	2	6	2	4	6	4	4	3	0	6	6	5	4	6	5	5	4	2	5	2
Informative hybrids	11	10	10	8	10	11	9	11	10	11	9	11	9	9	11	10	10	11	11	11	10	10	5

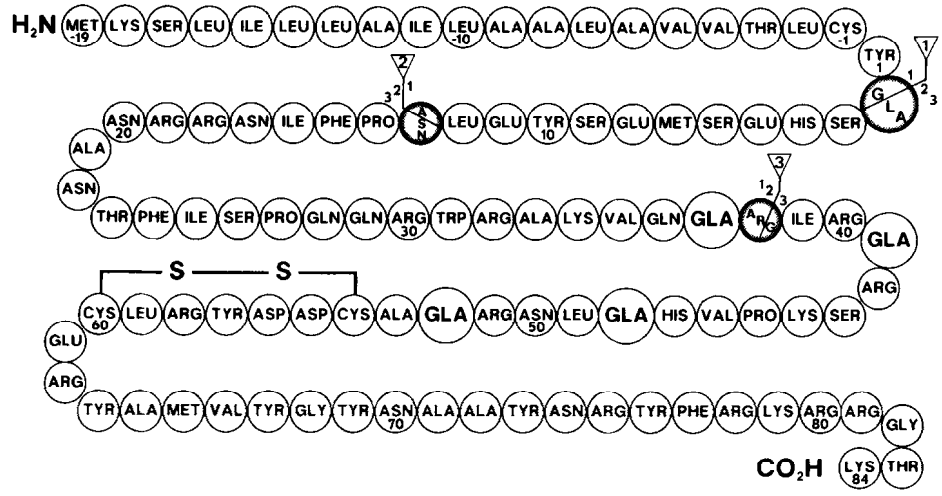

FIG. 7. **Localization of the intervening sequences within the structure of the human MGP.** The first amino acid of the mature protein is identified as residue +1. *Negative numbers* correspond to the pre-peptide region (−1 to −19). The five putative Gla residues are indicated by *larger circles* and *bold letters*. The position of the three introns within the protein sequence is indicated by the numbers enclosed within *open triangles*. Numbers adjacent to amino acid residues indicate codon positions. The location of the disulfide bond is noted.

TABLE III

Relationship between protein regions and genomic organization of the human MGP: comparison with gene structures of other vitamin K-dependent proteins

	Protein regions and corresponding exons			
	Transmembrane signal	α-Helical	γ-Carboxylase recognition site	Gla-containing domain
hMGP[a]	I {20}[b]	II {11}	III {26}	IV {46}
hBGP	I {21}	II {11}	III {24}	IV {42}
rBGP	I {21}	II {11}	III {24}	IV {43}
hFVII	I, I[c] {20, 22}		II {54}	
hFIX	I {29}		II {55}	
hFX	I {23}		II {54}	
hPC	I {24}		II {56}	
hPt	I {19}		II {54}	

[a] Abbreviations and references: hMGP, human MGP (Fig. 7); hBGP, human BGP (46); rBGP, rat BGP (6, 7, 9, 47); hFVII, human factor VII (48); hFIX, human factor IX (49); hFX, human factor X (50); hPC, human protein C (51); hPt, human prothrombin (52).

[b] Number of amino acid residues per exon; an amino acid residue was counted as member of a given exon if at least two nucleotides of the spliced codon were part of that exon.

[c] Due to alternative splicing, the leader sequence can be encoded by one or two exons in the factor VII gene.

overlapping regions of human chromosome 12. Hybrid I, containing the long arm of human chromosome 12 (12q), did not hybridize to the hMGP cDNA whereas hybrid II, spanning the entire short arm (12p) as well as a portion of the long arm (12_{q11}–12_{q21}) of human chromosome 12 gave a strong hybridization signal.

Relationship between Genomic and Protein Structures—Comparison between the organization of the hMGP gene and the corresponding protein structure shows that each exon corresponds to a separate functional region of the protein (Fig. 7, Table III). Exon I contains the 5′-nontranslated region as well as the sequence encoding the transmembrane signal peptide. Exon II encodes an α-helical domain defined by residues 2–12 of the mature protein. This region shares no homology with other proteins and is of unknown function. Exons III and IV encode, respectively, the putative γ-carboxylase recognition site and most of the Gla-containing domain. These two exons show some amino acid sequence identity with corresponding exons in the bone Gla protein gene and with a single exon spanning both protein domains in the other characterized vertebrate vitamin K-dependent proteins (3).

DISCUSSION

MGP as a Member of the Vitamin K-dependent Protein Family—The protein domain structures and genomic organization of the presently known vitamin K-dependent vertebrate proteins are compared in Table III. As can be seen, the five vitamin K-dependent proteins involved in blood coagulation have an identical two exon genomic organization of their N-terminal protein sequence (48–52), and the second exon in each case corresponds to the γ-carboxylase recognition site plus the Gla-containing domain. In contrast, MGP and BGP (46, 47) have a four exon organization of their N-terminal protein sequence, with the γ-carboxylase recognition site on exon three and the Gla-containing domain on exon four. The essentially identical genomic organization of MGP and BGP is consistent with the fact that the N-terminal regions of these two proteins have a somewhat higher level of sequence identity than either protein has to the vitamin K-dependent coagulation proteins, although the overall level of sequence identity is quite low (3). Exon two of MGP and BGP codes for eleven amino acid residues and has no counterpart in the other known vitamin K-dependent proteins (see Table III). The amino acid sequences encoded by exon two of BGP and MGP have no significant level of sequence identity (Refs. 46 and 47 and Fig. 7). Furthermore, while this sequence represents the N terminus of mature MGP, it is the N terminus of the propeptide in BGP. Therefore, it seems unlikely that exon two in BGP and MGP have analogous roles in biosynthesis or function.

Analysis of the MGP Promoter—The DNA sequence upstream from the hMGP gene contains a number of putative hormone regulatory elements (HRE) including potential vitamin D (HRE_1) and retinoic acid (HRE_2) responsive elements (see Fig. 3). The direct repeat (GGTGAACAAAAGGTGAAC) found within HRE_2 is also found in reverse orientation within the recently reported (42) retinoic acid responsive element of the retinoic acid receptor β-gene (GTTCACCGAAAGTTCACT). Although no data are available as yet concerning a possible regulation of MGP gene expression by retinoic acid, this hormone is a powerful morphogen and known to be involved in cell differentiation. MGP has been found to be expressed only in some lines of osteoblasts and its expression is thought to be associated with the development of a particular cell phenotype (5).

An interesting feature of the MGP promoter is that these putative hormone responsive elements are in each case directly flanked by sequences similar to known transcription factor binding sites. A similar promoter organization has been previously observed for other steroid hormone responsive genes such as the chicken lysozyme gene (53, 54) and the human metallothionein IIA gene (55). It has also been shown

that well characterized transcription factors are able to act synergistically with the steroid hormone receptor complex in order to induce activation of an adjacent promoter (56–58). By testing combinations of the DNA binding sites for steroid hormone receptors and for different transcription factors, Schüle *et al.* (57) were able to show that there is cooperativity in the binding of hormone receptors and other transcription factors to neighboring sites on DNA. Furthermore, these authors showed that the observed cooperativity is mediated by protein:protein interactions (58) and suggested that steroid hormone responsive units consist of one or several hormone responsive elements combined with other transcription factor binding sites. Since there is an analogous clustering of transcription factor binding sites near possible hormone responsive elements in the MGP promoter, it seems likely that these putative hormone receptor binding sites could be functionally significant and that the corresponding hormones will, in some cells, regulate MGP gene activity.

It is noteworthy that each putative regulatory cassette also contains a BGP box and a GAGA box. The function of these elements is unknown. BGP boxes have been described as a conserved sequence motif present in the promoter of rat and human BGP genes (6) and we have also located a possible BGP box within the human alkaline phosphatase genes.[4] The GAGA boxes were identified in the promoter of the osteonectin gene and found to be associated with a S1 nuclease binding site (45). These homopurine/homopyrimidine-rich sequences can form a "hairpin triplex" structure (61) and contribute to the generation of local regions of altered DNA conformation which are often found within regulatory regions of active genes and are thought to play an important role in the regulation of gene transcription (for review see Ref. 62).

Vitamin D is known to modulate the expression of MGP in bone cells. The vitamin D-dependent stimulation of MGP synthesis is consistent with a receptor-mediated primary gene response (63, 64), as evidenced by the rapid, 15-fold increase in MGP message after 1,25-dihydroxyvitamin D_3 treatment (5). The identification within the MGP gene promoter of at least one possible vitamin D receptor binding site which is located within a putative steroid hormone responsive unit is therefore quite exciting since very few proteins are known to be directly up-regulated by vitamin D. So far, only one vitamin D-responsive element has been conclusively identified within the promoter of the rat and human BGP genes (8, 14). The availability of the MGP promoter with its putative hormone binding sites will allow us to further investigate the regulation of the MGP gene expression and the involvement of the identified putative regulatory elements in both the basal and the tissue-specific, hormone-mediated transcription of this gene.

Chromosome Assignment—We have mapped the MGP gene to the short arm of human chromosome 12 by somatic cell hybrid analysis. Since only a single *PstI* fragment was present in human and in Chinese hamster DNA, we are convinced that MGP is a single copy gene in both species. This confirms the results obtained by Southern analysis of different digests of human genomic DNA (Fig. 5). Many other genes, A2M, C1R, C1S, CD4, CD9, ENO2, F8VWF, GAPD, GNB3, KRAS2, LDHB, MPE, PTHLH, PZP, and TPI1, have been mapped to the short arm of human chromosome 12 (65). The short arm of human chromosome 12 and mouse chromosome 6 appear to be conserved, since CD4, GAPD, TPI1, LDHB, PTHLH, and KRAS2 have also been mapped to mouse chromosome 6 (65). We were unable to map the MGP gene in mouse, because the human MGP cDNA probe did not cross-hybridize with mouse genomic DNA. However, it is very likely that MGP is also on mouse chromosome 6 since MGP seems to be a member of this conserved group of genes.

Acknowledgments—We wish to thank Dr. C. Delsert for his help in mapping the CAP site of the hMGP mRNA and Dr. R. Schüle for helpful discussions. We also thank S. Monaghan for technical assistance during the running of the sequencing gels, M. K. Williamson for assistance with computer searches, and C. Lewis for typing the manuscript. The isolation of the human MGP cDNA clone was achieved with the help of P. Hjorth.

REFERENCES

1. Price, P. A. (1985) in *Vitamins and Hormones* (Aurbach, G. D., and McCormick, D. B., eds) Vol. 42, pp. 65–108, Academic Press, New York
2. Fraser, J. D., and Price, P. A. (1988) *J. Biol. Chem.* **263,** 11033–11036
3. Price, P. A., Fraser, J. D., and Metz-Virca, G. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8335–8339
4. Kiefer, M. C., Bauer, D. M., Young, D., Hermsen, K. M., Masiarz, F. R., and Barr, P. J. (1988) *Nucleic Acids Res.* **16,** 5213
5. Fraser, J. D., Otawara, Y., and Price, P. A. (1988) *J. Biol. Chem.* **263,** 911–916
6. Lian, J., Stewart, C., Puchaz, E., Mackowiak, S., Shalhoub, V., Collart, D., Zambetti, G., and Stein, G. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1143–1147
7. Yoon, K., Rutledge, S. J. C., Buenaga, R. F., and Rodan, G. A. (1988) *Biochemistry* **27,** 8521–8526
8. Kerner, S. A., Scott, R. A., and Wesley Pike, J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 4455–4459
9. Demay, M. B., Roth, D. A., and Kronenberg, H. M. (1989) *J. Biol. Chem.* **264,** 2279–2282
10. Theofan, G., Nguyen, A. P., and Norman, A. W. (1986) *J. Biol. Chem.* **261,** 16943–16947
11. Dupret, J.-M., Brun, P., Perret, C., Lomri, N., Thomasset, M., and Cuisinier-Gleizes, P. (1987) *J. Biol. Chem.* **262,** 16553–16557
12. Clemens, T. L., McGlade, S. A., Garrett, K. P., Horiuchi, N., and Hendy, G. N. (1988) *J. Biol. Chem.* **263,** 13112–13116
13. Thomasset, M., Parkes, C. O., and Cuisinier-Gleizes, P. (1982) *Am. J. Physiol.* **243,** E483–E488
14. Demay, M., and Kronenberg, H. M. (1989) *J. Bone Mineral Res.* **4,** Suppl. 1, Abstr. 384
15. Perret, C., Lomri, N., Gouhier, N., Auffray, C., and Thomasset, M. (1988) *Eur. J. Biochem.* **172,** 43–51
16. Minghetti, P. P., Cancela, L., Fujisawa, Y., Theofan, G., and Norman, A. W. (1988) *Mol. Endocrinol.* **2,** 355–367
17. Wilson, P. W., Rogers, J., Harding, M., Pohl, V., Pattyn, G., and Lawson, D. E. M. (1988) *J. Mol. Biol.* **200,** 615–625
18. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual*, pp. 75–403, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
19. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
20. Sanger, R., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
21. Staden, R. (1986) *Nucleic Acids Res.* **14,** 217–231
22. Kanehisa, M. I. (1982) *Nucleic Acids Res.* **10,** 183–196
23. Lipman, D. J., and Pearson, W. R. (1985) *Science* **227,** 1435–1441
24. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 2444–2448
25. Srauss, W. M. (1987–1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) Unit 2.2, pp. 2.2.1–2.2.3, Greene Publishing Associates and Wiley-Interscience, Brooklyn, NY
26. *Transfer and Immobilization of Nucleic Acids to Schleicher & Schuell Solid Supports* (1987) pp. 6–9, Schleicher & Schuell, Keene, NH

[4] This sequence motif is located in both the intestinal and placental human alkaline phosphatase genes within the first 30 nucleotides following the 5′ splice junction of intron 7 (59, 60). It shares 75% sequence identity with the human BGP box motif (the sequence in the human alkaline phosphatase genes is: ATGACCCCCTTC-CTGCCCTGGCA; regions of identity are underlined).

27. Feinberg, A., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
28. Hsieh, C-L., Sturm, R., Herr, W., and Francke, U. (1990) *Genomics* **6,** 666–672
29. Gilman, M. (1987–1989) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) Unit 4.7, pp. 4.7.1–4.7.8, Greene Publishing Associates and Wiley-Interscience, Brooklyn, NY
30. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299
31. Yang-Feng, T. L., DeGennaro, L. J., and Francke, U. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 8679–8683
32. Breathnach, R., and Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383
33. Kariya, Y., Kato, K., Hayashizaki, Y., Himeno, S., Tarui, S., and Matsubara, K. (1987) *Gene* **53,** 1–10
34. Corden, J., Wasyluk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980) *Science* **209,** 1406–1414
35. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) *Nucleic Acid Res.* **8,** 127–142
36. Lee, W., Haslinger, A., Karin, M., and Tjian, R. (1987) *Nature* **325,** 368–372
37. Mitchell, P. J., Wang, C., and Tjian, R. (1987) *Cell* **50,** 847–861
38. Short, J. M., Wynshaw-Boris, A., Short, H. P., and Hanson, R. W. (1986) *J. Biol. Chem.* **261,** 9721–9726
39. Karin, M., Haslinger, A., Holtgreve, A., Richards, R. I., Krauter, P., Westphal, H. M., and Beato, M. (1984) *Nature* **308,** 513–519
40. Weiher, H., Konig, M., and Gruss, P. (1983) *Science* **219,** 626–631
41. Vanderbilt, J. N., Miesfeld, R., Maler, B. A., and Yamamoto, K. R. (1987) *Mol. Endocrinol.* **1,** 68–74
42. de The, H., del Mar Vivanco-Ruiz, M., Tiollais, P., Stunnenberg, H., and Dejean, A. (1990) *Nature* **343,** 177–180
43. Beato, M. (1989) *Cell* **56,** 335–344
44. Young, M. F., Findlay, D. M., Dominguez, P., Burbelo, P. D., McQuillan, C., Kopp, J. B., Gehron Robey, P. G., and Termine, J. D. (1989) *J. Biol. Chem.* **264,** 450–456
45. Nomura, S., Hashmi, S., McVey, J. H., Ham, J., Parker, M., and Hogan, B. L. M. (1989) *J. Biol. Chem.* **264,** 12201–12207
46. Celeste, A. J., Rosen, V., Buecker, J. L., Kriz, R., Wang, E. A., and Wozney, J. M. (1986) *EMBO J.* **5,** 1885–1890
47. Theofan, G., Haberstroh, L. M., and Price, P. A. (1989) *DNA (N. Y.)* **8,** 213–221
48. O'Hara, P. J., Grant, F. J., Haldeman, B. A., Gray, C. L., Insley, M. Y., Hagen, F. S. and Murray, M. J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 5158–5162
49. Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A., and Brownlee, G. G. (1984) *EMBO J.* **3,** 1053–1060
50. Leytus, S. P., Foster, D. C., Kurachi, K., and Davie, E. W. (1986) *Biochemistry* **25,** 5098–5102
51. Plutzky, J., Hoskins, J. A., Long, G. L., and Crabtree, G. R. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 546–550
52. Degen, S. J. F., MacGillivray, R. T. A., and Davie, E. W. (1983) *Biochemistry* **22,** 2087–2097
53. Renkawitz, R., Schütz, G., von der Ahe, D., and Beato M. (1984) *Cell* **37,** 503–510
54. Steiner, C., Muller, M., Baniahmad, A., and Renkawitz, R., (1987) *Nucleic Acids Res.* **15,** 4163–4178
55. Lee, W., and Haslinger, A. (1987) *Nature* **325,** 368–372
56. Cordingley, M. G., Riegel, A. T., and Hager, G. L. (1987) *Cell* **48,** 261–270
57. Schüle, R., Muller, M., Kaltschmidt, C., and Renkawitz, R. (1988) *Science* **242,** 1418–1420
58. Schüle, R., Muller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988) *Nature* **332,** 87–90
59. Henthorn, P. S., Raducha, M., Kadesch, T., Weiss, M. J., and Harris, H. (1988) *J. Biol. Chem.* **263,** 12011–12019
60. Knoll, B. J., Rothblum, K. N., and Longley, M. (1988) *J. Biol. Chem.* **263,** 12020–12027
61. Christophe, D., Cabrer, B., Bacolla, A., Targovnick, H., and Vassart, G. (1985) *Nucleic Acids Res.* **13,** 5127–5144
62. Wells, R. D. (1988) *J. Biol. Chem.* **263,** 1095–1098
63. Cancela, L., Theofan, G., and Norman, A. W. (1988) *New Comp. Biochem.* **18A,** 269–289
64. Haussler, M. R. (1986) *Annu. Rev. Nutr.* **6,** 527–562
65. Lalley, P. A., Davission, M. T., Graves, J. A. M., O'Brien, S. J., Womack, J. E., Roderick, T. H., Creau-Goldberg, N., Hillyard, A. L., Doolittle, D. P., and Rogers, J. A. (1989) *Cytogenet. Cell Genet.* **51,** 503–532