Structure and Chromosomal Location of the Human Gene Encoding Cartilage Matrix Protein*

(Received for publication, July 5, 1990)

Robert N. Jenkins‡, Sherri L. Osborne-Lawrence‡, Andrea K. Sinclair‡, Roger L. Eddy, Jr.§, Mary G. Byers§, Thomas B. Shows§, Allan D. Duby§§

From the ‡Harold C. Simmons Arthritis Research Center, §Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75235 and the †Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263

Cartilage matrix protein (CMP) is a major component of the extracellular matrix of nonarticular cartilage. The structure and chromosomal location of the human gene encoding CMP was determined by molecular cloning analysis. We used a partial chicken CMP cDNA probe to isolate three overlapping human genomic clones. From one of these clones, a probe containing 2 human CMP exons was isolated and used to map the gene to chromosome 1p35 and to screen a human retina cDNA library. Two overlapping cDNA clones were isolated. The predicted protein sequence of 496 amino acids includes a 22-residue signal peptide and a 474-residue mature protein of M, 51,344. The human CMP gene and polypeptide are strikingly similar to the chicken CMP gene and polypeptide. Human CMP is 79% identical to chicken CMP and contains two homologous domains separated by an epidermal growth factor-like domain. One potential N-glycosylation site is conserved between the two species. The human CMP gene spans 12 kilobase pairs with 8 exons and 7 introns which are similar in size to those of the chicken CMP gene. Both RNA splice junctions of intron G in the human and chicken CMP genes are nonconform- ing to the consensus splice sequences. This suggests that the CMP gene utilizes a new RNA splicing mechanism.

The extracellular matrix (ECM) is an insoluble network of protein and carbohydrate. The ECM of cartilage comprises 95% of its weight; the major components are Type II collagen and aggregates of the major proteoglycan of cartilage (1, 2). The biochemical and functional properties of many of the noncollagenous macromolecules of cartilage are incompletely understood (3). Cartilage matrix protein (CMP) is a molecule of unknown function found in the ECM of cartilage. It normally exists as a homotrimer of M, 148,000 with three subunits linked by disulfide bonds (4). The distribution of CMP has been studied by radioimmunoassay in bovine tissues where it was detected in tracheal, nasal septal, xiphisternal, auricular, and epiphysial cartilage, but not in articular cartilage or extracts of the intervertebral disc (5). The largest quantities of CMP were found in tracheal cartilage where it comprised up to 5% of the wet weight of aged tissue (6). Cartilage matrix protein has also been detected by immunocytochemistry in embryonic chick eyes (7). Antiserum raised in rabbits against bovine CMP detected large amounts of CMP in human trachea, but CMP was not detectable in articular cartilage from human knees (8).

The structure of chick CMP has been deduced by molecular cloning analysis (9, 10). It consists of two homologous repeated domains separated by a domain homologous to epidermal growth factor. The functional role of the epidermal growth factor-like domain has not been determined. The repeated domains, termed CMP1 and CMP2, are homologous to segments of von Willebrand factor, complement factors B and C2, the α chains of members of the integrin family, and Type VI collagen (10). Two of these three homologous domains in von Willebrand factor have the ability to bind collagen (11, 12). VLA-2, a member of the integrin family, is thought to be a collagen receptor. The α subunit of VLA-2 contains a domain homologous to CMP1 and CMP2 (13). The ability of CMP to bind collagen has not yet been demonstrated.

We report here the cloning from a retina cDNA library of cDNAs encoding human CMP, and the structure and chromosomal localization (1p35) of the human CMP gene. The deduced protein sequence of human CMP is 79% identical to the chicken sequence. There is striking similarity between the organization of the human and chicken CMP genes. Not only is there conservation of the structure of the exons, but intronic sizes are also conserved between the two species. Furthermore, one of the introns of the human gene displays the same novel RNA splice junctions that are found in the chicken CMP gene.

**EXPERIMENTAL PROCEDURES**

Isolation of Human CMP Genomic and cDNA Clones—A 770-base pair (bp) partial cDNA probe, pCMP1, encoding the entire CMP2 domain, the carboxyl-terminal domain, and approximately 50 bp of the 5'-untranslated region of the chick CMP gene was a generous gift of Dr. P. Goetinck (9). This probe was used to screen two human genomic libraries cloned into λ bacteriophage vectors (14, 15). Filters were hybridized at 40° C overnight in 50% formamide, 5 X SSC (16), 10 mM Tris, pH 7.6, 1 X Denhardt's (16), 0.1% SDS, and 10% dextran sulfate with heat-denatured, nick-translated probe (~2 x 10^6 dpm/150-mm nitrocellulose filter). Filters were washed nonstringently in 0.2 X SSC, 0.1% SDS at 45 °C and exposed overnight to film. Three overlapping clones, CM(C)1, CM(M)2, and CM(M)3, were isolated.
and mapped with restriction enzymes (Fig. 1A). DNA sequence analysis confirmed the identification of two human exons (exons 5 and 6). From one of the genomic clones, CM(C)1, a 2-kilobase pair (kb) PstI-PstI fragment, denoted HCMP5/6, containing exons 5 and 6 was isolated and used as a probe. The 5' PstI site of HCMP5/6 is derived from the left arm of the EMBL3 bacteriophage vector. A human retina cDNA library, kindly provided by Dr. Jeremy Nathans (Howard Hughes Medical Institute, Baltimore, MD), was screened with HCMP5/6 using the conditions described above, except the final wash was performed under stringent conditions (0.1 X SSC, 0.1% SDS, at 66°C). Two overlapping cDNAs, pHCMP1 and pHCMP2, were isolated (Fig. 1B).

**DNA Analysis and Sequencing**—Bacteriophage and plasmid DNA were isolated by standard procedures (16). Restriction enzyme digests were done per the manufacturer's instructions. A 7-kb XhoI-XhoI fragment of CM(M)2, a 6-kb KpnI-KpnI fragment of CM(C)1, and 3.1-, 3.5-, and 9-kb EcoRI-EcoRI fragments of CM(M)3 were subcloned into pBR322 or PUC18 plasmids for further analysis. The 5' KpnI site and 1.5 kb of CM(C)1 are derived from the EMBL3 bacteriophage vector. DNA fragments were subcloned into the M13-derived vectors MP18 and MP19 and sequenced by the chain-termination method (17). The coding regions of the genomic clones and the cDNA clones were sequenced at least once on both strands, except as noted in the figure legends.

**Computer Analysis of DNA and Protein Sequences**—The Beckman Microgenie (18) program was used for routine DNA sequence analysis and for generation of the deduced protein sequences. The University of Wisconsin Genetics Computer Group programs Wordsearch, FASTA, and TFASTA were used for searching data banks for homologous protein and DNA sequences (19).

**Hybrid Cell Lines**—A panel of hybrid cell lines derived from 19 unrelated human cell lines and 4 mouse cell lines was used to determine the chromosomal location of the CMP gene. Descriptions of the origin and construction of these cell lines are published elsewhere (20-22).

**Southern Blot Analysis**—High molecular weight genomic DNA was isolated from human, mouse, and hybrid cells by standard techniques (16). For each restriction endonuclease, 10 μg of DNA was digested to completion and fragments were separated by electrophoresis in 0.8% agarose gels. Nitrocellulose or Zetabind (AMP CUNO) filters containing the transferred DNA were hybridized under stringent conditions to the 32P-labeled human CMP probe HCMP5/6, as described above. Prior to autoradiography the filters were washed under stringent conditions, as described above.

**In Situ Hybridization**—These techniques are described in detail elsewhere (23). Briefly, phytohemagglutinin-stimulated, human male lymphocytes were synchronized in the cell cycle using bromodeoxyuridine. Radiolabeled HCMP5/6 probe was used for in situ hybridization prior to chromosome staining with Hoechst 33258/Giemsa.

**RESULTS**

**Isolation of Human Genomic and cDNA Clones**—A 770-bp cDNA fragment encoding the carboxyl-terminal portion of chick CMP (9) was used as a probe to screen two genomic libraries (14, 15) for the human CMP gene. Two overlapping clones, CM(M)2 and CM(M)3, were isolated from one genomic library (14) and a single clone, CM(C)1, was isolated from the other (15) (Fig. 1A). A 2.0-kb fragment of CM(C)1, HCMP5/6, which includes human exons 5 and 6 was used as a probe (Fig. 1A). Because it has been suggested that CMP binds collagen (9), we reasoned that the CMP gene might be expressed in tissues other than cartilage where Type II collagen is found. Accordingly, we screened a human, retina cDNA library because the retina and the vitreous, which is elaborated by the retina, are known to contain Type II collagen (24). Two overlapping cDNA clones, pHCMP1 and pHCMP2, that hybridized to HCMP5/6 were isolated. While our experiments were in progress, data showing expression of CMP by embryonic chick eyes were published (7).

**Analysis of the cDNA Clones**—Fig. 1B depicts the restriction enzyme map of the human CMP cDNA clones. The cDNA sequence and the deduced protein sequence are presented in Fig. 2. pHCMP1 is a 2.76-kb cDNA encoding 460 bp of coding region and 2.3 kb of 3'-untranslated region. A consensus polyadenylation signal (25), AATAAA, is located at the 3' end of this clone (position 3206-3211, Fig. 2). Although pHCM1 does not have a poly(A) tail, examination of the genomic sequence (see below) reveals a GT-rich segment downstream of the AATAAA motif, suggesting that it is likely to be a functional polyadenylation site.

pHCMP2 is a 1.37-kb cDNA encoding 1020 bp of coding region and 350 bp of 3' untranslated region. It overlaps the 5' end of pHCM1 by 800 bp. pHCMP2 ends in a 17-bp poly(A) tail and contains an acceptable polyadenylation signal (25), AATATA (position 1338-1343, Fig. 2). There are several nucleotide polymorphisms in the 3'-untranslated region of these two cDNA clones (Fig. 2). Together the two cDNA clones encode the carboxyl-terminal 340 amino acids of human CMP and 2.3 kb of 3' untranslated region. Comparison to the predicted protein sequence of the chick CMP precursor indicates that the human cDNA encodes part of CMP1, the EGF-like domain, CMP2, and the carboxyl-terminal domain. Because the chick CMP gene encodes a precursor of 493 amino acids including a 23-amino acid signal peptide, it is apparent that pHCM1 and pHCMP2 do not comprise the entire coding region.

**Identification of Exons 1 and 2 of the Human CMP Gene**—pHCMP1 and pHCMP2 were used as probes to find their corresponding exons in the genomic clones. Comparison of the genomic and cDNA sequences identified the 5 introns and 6 exons at the 3' end of the gene. Although our cDNA clones were not full length, we were able to identify the 5' portion of the coding region in the genomic clones by taking advantage of the high degree of organizational similarity between the human and chick CMP genes.

The location and boundaries of exons 1 and 2 and introns A and B were deduced as follows. We noted that the sizes of
Fig. 2. Nucleotide and deduced amino acid sequence of the human CMP cDNAs. Overlapping cDNA clones hCMP1 and hCMP2 were merged to give the depicted nucleotide sequence. Nucleotide numbering is shown to the right of the sequence. The one-letter amino acid code is used. The relationship of the cDNA sequence to the intron-exon organization of the CMP gene is indicated. The CTT codon (position 403-405), which differs from the genomic clones, is indicated by an asterisk. Polyadenylation signal sequences are underlined. The polyadenylation site of pHCMP2 is indicated by A.

Sequence polymorphisms: at position 805-807, (GAA) in pHCMP1 versus (GAG) in pHCM2 represents a silent polymorphism; position 1065 (G) in pHCMP1 versus (A) in pHCMP2; and beginning at position 1204, the number of (TG) repeats varied from 11 to 13.

Table I

<table>
<thead>
<tr>
<th>Species</th>
<th>Exon</th>
<th>Length</th>
<th>5' splice</th>
<th>Intron</th>
<th>Length</th>
<th>3' splice</th>
<th>Intron phase</th>
<th>Amino acid interrupted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>1d</td>
<td>94</td>
<td>bp</td>
<td>A</td>
<td>1067</td>
<td>bp</td>
<td>I</td>
<td>G (32)</td>
</tr>
<tr>
<td>Chicken</td>
<td>1</td>
<td>88</td>
<td>bp</td>
<td>A</td>
<td>2720</td>
<td>bp</td>
<td>I</td>
<td>G (32)</td>
</tr>
<tr>
<td>Human</td>
<td>2</td>
<td>237</td>
<td>bp</td>
<td>B</td>
<td>4840</td>
<td>bp</td>
<td>O</td>
<td>K/V (147/148)</td>
</tr>
<tr>
<td>Chicken</td>
<td>2</td>
<td>238</td>
<td>bp</td>
<td>C</td>
<td>2270</td>
<td>bp</td>
<td>O</td>
<td>K/V (145/146)</td>
</tr>
<tr>
<td>Human</td>
<td>3</td>
<td>223</td>
<td>bp</td>
<td>C</td>
<td>1017</td>
<td>bp</td>
<td>V</td>
<td>(222)</td>
</tr>
<tr>
<td>Chicken</td>
<td>3</td>
<td>223</td>
<td>bp</td>
<td>C</td>
<td>2270</td>
<td>bp</td>
<td>V</td>
<td>(220)</td>
</tr>
<tr>
<td>Human</td>
<td>4</td>
<td>196</td>
<td>bp</td>
<td>D</td>
<td>464</td>
<td>bp</td>
<td>V</td>
<td>(264)</td>
</tr>
<tr>
<td>Chicken</td>
<td>4</td>
<td>196</td>
<td>bp</td>
<td>D</td>
<td>464</td>
<td>bp</td>
<td>V</td>
<td>(264)</td>
</tr>
<tr>
<td>Human</td>
<td>5</td>
<td>417</td>
<td>bp</td>
<td>E</td>
<td>585</td>
<td>bp</td>
<td>I</td>
<td>G (403)</td>
</tr>
<tr>
<td>Chicken</td>
<td>5</td>
<td>418</td>
<td>bp</td>
<td>E</td>
<td>585</td>
<td>bp</td>
<td>I</td>
<td>G (400)</td>
</tr>
<tr>
<td>Human</td>
<td>6</td>
<td>153</td>
<td>bp</td>
<td>F</td>
<td>829</td>
<td>bp</td>
<td>E</td>
<td>(451)</td>
</tr>
<tr>
<td>Chicken</td>
<td>6</td>
<td>153</td>
<td>bp</td>
<td>F</td>
<td>829</td>
<td>bp</td>
<td>E</td>
<td>(451)</td>
</tr>
<tr>
<td>Human</td>
<td>7</td>
<td>81</td>
<td>bp</td>
<td>G</td>
<td>644</td>
<td>bp</td>
<td>I</td>
<td>L (481)</td>
</tr>
<tr>
<td>Chicken</td>
<td>7</td>
<td>81</td>
<td>bp</td>
<td>G</td>
<td>644</td>
<td>bp</td>
<td>I</td>
<td>L (481)</td>
</tr>
</tbody>
</table>

Consensus splice sequences:

<table>
<thead>
<tr>
<th>AGg_Bgagt</th>
<th>III.......nccagG</th>
</tr>
</thead>
</table>

* Exon sequences are in capital letters; intron sequences are in lower case.

* Lengths of the chicken CMP gene introns were determined by electron microscopy (10); human lengths were determined by DNA sequencing.

* Phase 0 introns lie between two codons; phase I introns lie between the first and second nucleotides; and phase II lie between the second and third nucleotides of a codon (40).

* The lengths of the 5'- and 3'-untranslated regions in exons 1 and 8 are not included.

* Shapiro and Senapathy (33).

Introns C-F were similar to those of the chicken gene (Table I). Reasoning that introns A and B at the 5' end of the human CMP gene might also be similar in size to the chicken introns, we sequenced the 5' end of the human genomic DNA in areas where the chicken CMP gene structure predicted the remaining human exons would be found. This strategy was successful in identifying DNA sequences (exons 1 and 2 whose deduced protein sequences are highly homologous to the chick CMP protein sequence. RNA consensus splice signals are present at both ends of the deduced boundaries of human introns A and B (Table I). The DNA and the predicted protein sequences of the human CMP introns and exons are presented in Fig. 3.

There are two potential translation initiation sites 27 bp apart (positions 490 and 517, Fig. 3). We suspect that the proximal ATG is the initiator codon because the 5'-proximal ATG triplet served as the initiator codon in >90% of the mRNAs examined by Kozak (26). Furthermore, the presence
Human Gene Encoding Cartilage Matrix Protein

FIG. 3. Nucleotide and deduced amino acid sequence of human genomic clones encoding CMP. Overlapping sequences from CM(C)1, CM(M)2, and CM(M)3 clones were merged to give the complete nucleotide sequence. Where the sequences from the different clones differ, the CM(M)2 sequence is depicted. The one-letter amino acid code is used. Numbering of the nucleotide sequence begins arbitrarily. The exons were sequenced on both strands with the following exceptions: Exon 3, 6 bp of exon 3 and the adjoining 5' end of intron C were sequenced on one strand twice; Exon 7, 20 bp of exon 7 and the adjoining 5' end of intron G were sequenced three times on one strand. The putative CAAT and TATA sites are underlined. The putative polyadenylation signals in the 3'-untranslated region are underlined. The downstream GT-rich sequences are double underlined. The polyadenylation sites of the cDNA clones are indicated by -1 for pHCM1 and -2 for pHCM2. The TTT codon (position 8280), which differs from the cDNA sequence, is indicated by an asterisk. Silent polymorphisms: in exon 5 (position 8469-8471), CM(M)2 = (ACT) versus CM(C)1 = (ACC); in exon 6 (position 9267-9269), CM(M)2 = (GAG) versus CM(C)1 = (GAA) (the same polymorphism is present in the cDNA clones at position 807, see Fig. 2 legend); in intron E at position 9042 CM(M)2 = (T) versus CM(C)1 = (G) (not shown); and in the 3'-untranslated region at 11,093 CM(M)2 = (T) versus CM(C)1 = (C). Additional differences between the genomic and cDNA sequence were noted at positions 11,093, 11,253, 11,788, 11,956, 12,118, 12,222, 12,241, 12,469, 12,490, 12,569, 12,585, 12,795, 12,798, 12,818, 12,906, 12,979, and 13,129.

of a purine nucleotide 3 bases upstream of the initiator codon is the most conspicuous conserved feature of initiation sequences (26), and the proximal ATG has the purine nucleotide adenyllic acid (A) in this position, whereas the distal ATG does not.

Deduced Amino Acid Sequence of Human CMP—The protein sequence deduced from the cDNA and genomic clones is presented in Fig. 4. The entire gene is predicted to encode a precursor protein of 496 residues. To determine the probable beginning of the mature polypeptide, we examined the sequence of the amino-terminal portion of the precursor protein for characteristics of a signal peptide and the likely, proteolytic cleavage site. The algorithm of von Heijne (27), applied by computer program (28), indicated that more than one acceptible signal peptidase processing site is present. The sites producing the highest scores were after residue 20 (alanine) and after residue 22 (cysteine). The former site would leave a cysteine (at position 22) in mature human CMP that is not present in mature chicken CMP. Therefore, we suspect that the cleavage site occurs after residue 22. This analysis suggests that the signal peptide consists of the first 22 amino acids, and mature human CMP begins with serine at position 23 and contains 474 residues with a calculated M, of 51,344.

There is a potential N-glycosylation site (NAT) (29) in CMPl at residues 76-78; such a site is found in a similar location of chicken CMP. There are two domains, CMPl and CMP2, of approximately 190 amino acids, encoded by exons 2 and 3, respectively. The protein sequences
of CMP1 and CMP2 are 42% identical. At each end of CMP1 and CMP2 there is a cysteine available for intradomain disulfide bonding. CMP1 and CMP2 are separated by a 41-amino acid, EGF-like region, encoded by exon 4. The carboxyl-terminal domain is encoded by exons 7 and 8 and contains 2 cysteine residues separated by a single amino acid.

Comparison to Chicken CMP and Other Proteins—The amino acid sequences of the mature polypeptides of human and chicken CMP are 79% identical (Fig. 5). The locations of the 14 Cys residues are completely conserved between species. Human and chicken CMP1 domains are 79% identical, while the CMP2 domains are 84% identical. The EGF-like region is 85% identical to the corresponding region of the chicken protein. The carboxyl-terminal domains of human and chicken CMP are 61% identical.

A computer search of the GenBank, EMBL, and NBRF data banks for sequences homologous to the CMP1 and CMP2 domains produced three tandem repeats of human von Willebrand factor, the LY subunits of human LFA-1, human Type VI collagen, and human complement factor B.

The human introns, with the exception of intron G, are 50—200% the size of the corresponding chicken introns. The RNA splice junctions are homologous to the consensus sequences (33), AGgtiagt, at the 5’ end of the introns and, ttctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctct, at the 3’ end of the introns with the exception of intron G. At the 5’ end of human intron G the dinucleotide “at” is found instead of the consensus dinucleotide “gt” and at the 3’ end “ac” is found instead of “ag.” Exceptions to the “gt-ag” rule are very rare; the identical substitutions are found in the RNA splice sites for intron G of the chicken CMP gene (Table I) (10). Interestingly, both species share 13 identical residues (ttctctctactctctctat) 7 bp upstream of 3’ splice site for intron G.

3’-Untranslated Region—The 3’-untranslated region for acceptable polyadenylation signals revealed seven possible sites (Fig. 3). In view of evidence that transcription termination in eukaryotes depends on “GT-rich” sequences about 15–40 bp downstream of the polyadenylation signal and within 30 bp of the actual site of polyadenylation (34), we examined the genomic sequence for GT-rich modules downstream of these seven sites. There is a GT-rich sequence downstream from the putative polyadenylation signal (AA-TATA, position 11,233–11,238) used by the cDNA clone downstream of 3’ splice site for intron G.

The 3’-untranslated region of the chicken CMP gene is approximately 1.5 kb long. No extended ranges of sequence
identity were found in the 3'-untranslated region between the chicken (400 bp are available for comparison) and human CMP genes.

Promoter Region—We examined the human DNA sequence 5' from the putative initiation codon for features similar to those noted in the same region of the chicken gene. This region is likely to be the promoter if the human gene, like the chicken gene, lacks introns within the 5'-untranslated sequence (10). Most eukaryotic promoters have a TATA box (35). The human CMP gene contains the sequence TTGATA (position 426-431), 65 bp upstream from the first of two potential initiation codons (see above). The chicken CMP gene has a similar sequence, TTAATA, 69 bp upstream from its initiation codon. Neither sequence conforms to the consensus sequence (10). Most eukaryotic promoters have a TATA box (positions 290-295, 331-336, 357-362, 411-416); the 5' pair is in the opposite orientation from the 3'.

Examining the human and chicken promoters for homologous sequences revealed the following shared motifs:

\[
\text{TATA}^A_{\text{TT}}^A_{\text{T}}^A_{\text{T}}\]

The GGCCAGGCT sequence (position 278-286) may represent a CAAT box as it is identical in 7 of 9 bp to the consensus sequence (36).

\[
\text{GG}^G_{\text{T}}\text{CAATCT}
\]

The CAAT-like sequence (10) in the chicken CMP gene (ATGCAATAT) is identical to the consensus sequence in 5 of 9 bp. Four characteristic binding sites (GGGCCGG) for the DNA binding protein Sp1 (37) are found in the putative human promoter region (positions 290-295, 331-336, 357-362, 411-416); the 5' pair is in the opposite orientation from the 3'.

In this paper we report the cloning of the gene encoding human CMP and its chromosomal location. Partial cDNA clones were isolated from a retina cDNA library. Overlapping λ clones were isolated from two human genomic libraries. The human gene encompasses approximately 12 kb and is composed of 8 exons and 7 introns, with a 2.3-kb 3'-untranslated region. The deduced protein sequence of human CMP contains 496 residues, has one potential N-glycosylation site, and is highly homologous to chicken CMP. We found one protein sequence polymorphism (position 291) where the genomic clones encode a phenylalanine residue, while the cDNA clone pHCM5 encodes a leucine. The physiological significance of this alteration remains to be determined. The isolation of cDNA clones encoding CMP from a retina library indicates that the human CMP polypeptide may be expressed in the retina and tissues other than nonarticular cartilage. In the embryonic chick eye, CMP was detected by immunocytochemistry in the cornea, sclera, choroid, lens capsule, lens epithelium, and lens fibers, but not in retina or pigment epithelium (7). If, in fact, CMP has a collagen binding function, it may be that the tissue distribution of CMP will

Chromosomal Location—Southern blot analysis of somatic cell hybrids and in situ chromosomal hybridization were used to map the human gene for CMP. The results of Southern blot analysis of genomic DNA from 45 mouse-human hybrid cell lines (Fig. 6) probed with HCM5/6 are shown in Table II. The discordancy scores localize the human CMP gene to chromosome 1. The hybrids containing translocations (Table II legend) localize the gene to the 1p12-pter region. One can also see from Fig. 6 that on HindIII-digested human DNA there is evidence of only one gene hybridizing to this probe (two bands are detected because HCM5/6 has an internal HindIII site). Further Southern blot analysis of human genomic DNA with the HCMP5/6 probe under nonstringent conditions where homologous human genes would be detected provided evidence for only one human CMP gene (data not shown).

In situ hybridization was used to define the subchromosomal location of the CMP gene. One hundred normal metaphase chromosomes were analyzed by hybridization with HCM5/6. Of 234 grains counted, 56 (24%) were associated with chromosome 1 (Fig. 7). 27% of the grains on chromosome 1 and 6.4% of the total grains were found at 1p35. These data indicate that the 1p35 region encodes the human CMP gene.

DISCUSSION

In this paper we report the cloning of the gene encoding human CMP and its chromosomal location. Partial cDNA clones were isolated from a retina cDNA library. Overlapping λ clones were isolated from two human genomic libraries. The human gene encompasses approximately 12 kb and is composed of 8 exons and 7 introns, with a 2.3-kb 3'-untranslated region. The deduced protein sequence of human CMP contains 496 residues, has one potential N-glycosylation site, and is highly homologous to chicken CMP. We found one protein sequence polymorphism (position 291) where the genomic clones encode a phenylalanine residue, while the cDNA clone pHCM5 encodes a leucine. The physiological significance of this alteration remains to be determined. The isolation of cDNA clones encoding CMP from a retina library indicates that the human CMP polypeptide may be expressed in the retina and tissues other than nonarticular cartilage. In the embryonic chick eye, CMP was detected by immunocytochemistry in the cornea, sclera, choroid, lens capsule, lens epithelium, and lens fibers, but not in retina or pigment epithelium (7). If, in fact, CMP has a collagen binding function, it may be that the tissue distribution of CMP will

![Fig. 6. Southern blot of somatic cell hybrids. A representative blot is shown. 10 μg of DNA prepared from the somatic cell hybrids, XER-7 (lane +) and XTR-3BSAgB (lane -), murine cells (RAG, lane M), and human cells (GM00131, lane H) were digested with HindIII, size-fractionated on a 0.8% gel, transferred to nitrocellulose, and hybridized to the HCM5/6 probe. The HCM5/6 probe hybridized to a 7.7-kb murine restriction fragment while it hybridized to 5.9- and 3.0-kb human restriction fragments. Somatic hybrid cells, XER, were scored as positive while somatic hybrid cells, XTR-3BSAgB, were scored as negative. Lane 1, hybrid positive for chromosome 5; lane 2, hybrid negative for chromosome 5; lane 3, mouse DNA; lane 4, human DNA.](http://www.jbc.org/)

Downloaded from http://www.jbc.org/ by guest on October 3, 2017
parallel that of a particular type of collagen, e.g. Type II collagen.

Application of algorithms designed to predict the signal peptidase processing site produced at least two possible sites. Beginning the mature human polypeptide with alanine at position 27 would match the alanine residue at the amino peptidase processing site produced at least two possible sites.


collagen. Although these are very common residues at these positions,

Interestingly, beginning the mature polypeptide at leucine at position 23 in the -3 position of the leader segment. Proline residues are rarely seen in positions -3, -2, and -1, respectively, of the signal peptide and proline at position 2 of the mature polypeptide.

This cleavage site would put a proline at position -2. Sequenc-

ing of human CMP polypeptides will resolve this issue and add to our understanding of signal peptidase processing sites.

the degree of discordant segregation for a marker and a chromosome. A 0% discordancy is the basis for chromosome assignment. HCM5/6 mapped to human chromosome 1.

The structure of the human CMP gene is quite similar to that of the chicken gene both in the exon/intron organization and the size of the introns (Table I). All of the human and mouse CMP domains contain the reverse sequence DGR (position 154, 38). Human and chicken CMP molecules contain the reverse sequence DGR (position 154, 38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.

include the tripeptide RGD (38). Human and chicken CMP do not contain this sequence, but both domains of both molecules contain the reverse sequence DGR (position 154, Fig. 4) in the same location. There is evidence that the DGR sequence may be involved in some mechanism of fibroblast adhesion to ECM glycoproteins. Yamada et al. (39) showed that the peptide SDGR inhibited spreading of fibroblasts on fibronectin, vitronectin, laminin, and collagen substrates. Whether the DGR sequence in the CMP domains is involved in its interaction with other ECM components or with cell surface ECM receptors remains to be determined.

The table below summarizes the analysis of 100 normal human metaphase cells from phytohemagglutinin-stimulated peripheral blood lymphocytes that were hybridized to the HCM5/6 probe. Each dot represents one labeled site observed in the corresponding band. 27% (15/56) of the labeled sites on chromosome 1 were located at 1p35; this cluster represented 6.4% (15/234) of all labeled sites.
untranslated region revealed no extended ranges of homology. Interestingly, other ECM proteins, such as cartilage link protein, show large regions of homology in the 3′-untranslated regions of homologous human and chicken genes (44). This suggests that regulatory regions in the 3′-untranslated region of the CMP gene differ between humans and chickens.

Neither pHCMP1 or pHCMP2 used the polyadenylation signal ATTAAA found in chicken CMP cDNA and human \( \alpha_1(I) \) collagen cDNA (45). However, this sequence is present in the 3′-untranslated region and has a downstream GT-rich segment. Thus it is possible that additional cDNA clones would reveal this to be a functional polyadenylation signal in the human.

The poly(TG) direct repeat found in the 3′-untranslated region is apparently one of the estimated 106 copies of poly(TG) sequences distributed throughout the human genome (46, 47). Some of these have been shown to play a role in the regulation of gene expression through formation of Z-DNA structures (48). The potential role of these DNA sequences in the mRNA stability and translational efficiency of the human CMP gene can now be addressed with the cDNA clones.

Mouse-human hybrid cell lines and in situ hybridization were used to map the human CMP gene to 1p35. Other loci in the region of chromosome 1 near the CMP gene include GALE (UDP-galactose-4-epimerase), FUCAl (fucosidase, \( \alpha-I \) collagen cDNA (45). However, this sequence is present in the 3′-untranslated region and has a downstream GT-rich segment. Thus it is possible that additional cDNA clones would reveal this to be a functional polyadenylation signal in the human.

The poly(TG) direct repeat found in the 3′-untranslated region is apparently one of the estimated 106 copies of poly(TG) sequences distributed throughout the human genome (46, 47). Some of these sequences have been shown to play a role in the regulation of gene expression through formation of Z-DNA structures (48). The potential role of these DNA sequences in the mRNA stability and translational efficiency of the human CMP gene can now be addressed with the cDNA clones.

Mouse-human hybrid cell lines and in situ hybridization were used to map the human CMP gene to 1p35. Other loci in the region of chromosome 1 near the CMP gene include GALE (UDP-galactose-4-epimerase), FUCAl (fucosidase, \( \alpha-I \) collagen cDNA (45). However, this sequence is present in the 3′-untranslated region and has a downstream GT-rich segment. Thus it is possible that additional cDNA clones would reveal this to be a functional polyadenylation signal in the human.

The poly(TG) direct repeat found in the 3′-untranslated region is apparently one of the estimated 106 copies of poly(TG) sequences distributed throughout the human genome (46, 47). Some of these sequences have been shown to play a role in the regulation of gene expression through formation of Z-DNA structures (48). The potential role of these DNA sequences in the mRNA stability and translational efficiency of the human CMP gene can now be addressed with the cDNA clones.
Structure and chromosomal location of the human gene encoding cartilage matrix protein.

R N Jenkins, S L Osborne-Lawrence, A K Sinclair, R L Eddy, Jr, M G Byers, T B Shows and A D Duby


Access the most updated version of this article at http://www.jbc.org/content/265/32/19624

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/32/19624.full.html#ref-list-1