

Isolation, Characterization, and Chromosomal Assignment
of the Human Colligin-2 Gene (CBP2)

ヒト colligin-2 遺伝子 (CBP2) の単離、構造解析、
及び染色体マッピングに関する研究

池 川 志 郎

Isolation, Characterization, and Chromosomal Assignment
of the Human Colligin-2 Gene (CBP2)

ヒト colligin-2 遺伝子 (CBP2) の単離、構造解析、
及び染色体マッピングに関する研究

池 川 志 郎

ABSTRACT

Colligin, a collagen-binding glycoprotein localized to endoplasmic reticulum, belongs to the serpin (serine protease) superfamily. Colligin has been identified and cloned in a number of species including mouse, rat, chicken, and man. We have identified a novel member of the colligin gene family, human colligin-2 (CBP2: collagen binding protein 2). The full-length cDNA of CBP2 encodes a 418-amino-acid peptide that is highly homologous (97% identity) to the previously reported human colligin (CBP1: collagen binding protein 1). The predicted peptide sequence of CBP2 contains the cardinal motifs of the colligin gene family. Its expression is ubiquitous among normal human tissues except for brain and peripheral leukocytes. We have also isolated a genomic clone containing the entire coding sequence and the 5'- and 3'- non-coding sequences, and determined the exon-intron organization and sequence of the 5'-flanking region of the gene. The promoter sequence of CBP2 shows significant homology to that of its mouse counterpart; it contains a heat-shock element and a possible retinoic acid-responsive element. We have mapped this novel gene to chromosome 11q13.5 by Fluorescence *in situ* hybridization (FISH). Our data will facilitate investigations into the regulation, evolution and function of colligins.

INTRODUCTION

Isolation and characterization of novel genes associated with disease contribute information of vast significance to modern biology and medicine; cloning of a previously unknown gene can provide dramatic new insights into the molecular mechanisms involved in a disease process and lead to better diagnosis and treatment. Although the human genome is estimated to contain 50,000 to 100,000 genes (Fields et al. 1994), only a small fraction of this number have been characterized so far. At present, an international effort to sequence the entire genome (the Human Genome Project) is progressing very rapidly, and many novel genes have been identified by participants in this program.

As a part of the Human Genome Project, toward the identification of biologically interesting genes and especially genes associated with diseases, we have been determining nucleotide sequences of cDNA clones randomly selected from a human fetal-lung cDNA library (Sudo et al. 1994). In the course of this work, we identified a cDNA clone highly homologous to, but distinct from, the human colligin reported previously by Clarke and Sanwal (1992). Because the degree of homology between the two genes is high, we designated new gene human colligin-2 (CBP2: collagen-binding protein 2), and now refer to the gene described previously as CBP1 (collagen binding protein 1).

Colligins are glycoproteins that bind specifically to collagen I, collagen IV, and gelatin (Kurkinen et al. 1984; Cates et al. 1987). Their amino acid structure includes an N-terminal hydrophobic signal sequence and two putative N-linked

oligosaccharide attachment sites (Clarke and Sandwal 1992). C-terminal RDEL sequences present in these proteins act as retention signals in the endoplasmic reticulum (ER) (Pelham 1990). Colligins also contain a reactive site characteristic of proteins belonging to the serpin (serine protease inhibitor) family (Hunt and Dayhoff 1980; Carrell and Boswell 1986; Clarke et al. 1991), and possess a P-F-X-F sequence downstream of this reactive site, a sequence that is important for binding serine-protease complexes to membrane receptors (Joslin et al. 1991).

Colligin was originally described as a 47-kDa glycoprotein isolated from murine parietal endoderm cells (Kurkinen et al. 1984), but similar proteins were later identified in cell lines established from rat myoblast (termed gp46; Cates et al. 1984), murine embryonal carcinoma (termed J6; Wang and Gudas 1990), and chick embryo fibroblast (termed HSP46; Hirayoshi et al. 1991). Clarke and Sanwal (1992) subsequently isolated and characterized a full-length cDNA clone encoding a human homologue of these colligins (CBP1) from skin fibroblasts. The high degree of homology in amino acid sequence and conservation of characteristic motifs among these colligins allow them to be classified as a definite family of proteins.

To characterize the novel form of human colligin, CBP2, we isolated a full-length cDNA clone and a genomic clone which contains the entire coding sequence of the cDNA. We have determined the genomic structure, including the promoter region, and its chromosomal localization to 11q.

MATERIALS AND METHODS

Cloning and Sequencing of cDNA

Nearly 2500 clones were randomly selected from a human fetal-lung cDNA library, and the 5' portion of each clone was sequenced as previously described (Sudo et al. 1994). Briefly, Poly(A) RNA isolated from pooled lung tissues of 20- to 25-week-old human fetuses (Clontech) was used for construction of a cDNA library with a unizAP-cDNA synthesis kit (Stratagene) according to the manufacturer's instructions. An aliquot of the cDNA library was *in vivo* excised with the Exissist helper phage and transfected into *E. coli*, XL1-Blue. Plasmid DNAs were isolated using an automatic plasmid isolation system (Kurabo Model PI-100; Kurabo Co.). Sequencing reactions were performed with the Taq Dye Primer cycle sequencing kit (Applied Biosystems) and were analyzed on an ABI 373A automated DNA sequencer (Applied Biosystems). Comparisons of these partial DNA sequences with known genes in public data bases by the FASTA program (Pearson and Lipman 1988) revealed that one 2.1-kb clone (termed L13932) was highly homologous to CBP1 (Clarke and Sanwal 1992). We determined the nucleotide sequence of this cDNA on both strands by the dideoxy chain-termination method (Sanger et al. 1977), using universal or synthetic primers.

Northern-Blot Analysis

A [³²P]-randomly labeled, 1.5-kb *Eco*RI fragment of L13932 which contained the entire coding region of CBP2 was used as a probe in a human Multiple-Tissue Northern blot system (MTN) (Clontech). The blot was prehybridized in

Prehybridization/hybridization Solution 2 (5x SSPE; 10x Denhardt's solution; 50% formamide; 2% SDS; 100 mg/ml freshly denatured, sheared salmon sperm DNA) at 42°C for 6 hours. Hybridization was performed in prehybridization/hybridization solution 2 containing the radiolabeled probe at 42°C for 18 hours. The blot was then washed in Wash Solution 3 (0.1x SSC, 0.1% SDS) at 50°C for 40 min. The membrane was autoradiographed at -80°C for 7 hours with intensifying screens. A 2-kb human β -actin cDNA included in MTN was used as a control for sample loading.

Reverse-transcription-Polymerase Chain Reaction (RT-PCR)

RT-PCR was performed as described previously (Horii et al. 1993). In brief, 30-mg aliquots of total RNAs were used to produce cDNAs, which were amplified by PCR with primers designated 5C2 (5'-ATGCGCTCCCTCCTGCTTCT-3') and 3X (5'-GCTCGAGAGGCTCCACGTG-3') (Fig. 1) for 30 cycles as follows: denaturation at 94°C for 30 sec., annealing at 58°C for 30 sec., and extension at 72°C for 1 min. This primer set was designed to amplify CBP2 cDNA from nucleotides 88 to 925 (Fig. 1). The sequence of primer 5C2 was present in CBP2 cDNA, but not in CBP1 cDNA. Primers ACT1 (5'-CCAACTGGGACGACATGG-3') and ACT2 (5'-AGGATCTTCATGAGGTAGTC-3') were used simultaneously to measure the expression of the human β -actin gene as a control for this method. The PCR products were electrophoresed in a 2% agarose gel.

Genomic Library Screening

A human genomic library cloned in the cosmid vector pWEX15 was screened by a standard procedure (Sambrook et al. 1989). A 1.5-kb *Eco*RI fragment of the cDNA clone (L13932) which contained the entire coding region of CBP2 was used as a probe. Prehybridization and hybridization was performed in hybridization solution (7% PEG, 10% SDS; 200 mg/ml sheared human placenta DNA) at 65°C. Washing was performed in washing buffer (0.1x SSC, 0.1% SDS) at 65°C for 40 min. The membrane filters were autoradiographed at -80°C for 8 hours with intensifying screens.

Characterization of the Cosmid Clone

After digestion with *Eco*RI and/or *Xho*I, the cosmid DNA was resolved by electrophoresis through a 0.8% agarose gel, then transferred to a nitrocellulose membrane. The blot was hybridized to several synthetic oligonucleotide probes corresponding to the coding sequence and to the 5'- and 3'- untranslated regions of CBP1 or CBP2, according to standard procedures (Sambrook et al. 1989).

Fluorescence in situ Hybridization (FISH)

We performed FISH for chromosomal assignment according to the method of Takahashi et al. (1990). Briefly, a cosmid DNA obtained by screening the genomic library was used as a probe. The probe DNA (0.6 mg) was labeled with nick-translation kit (Boehringer) using biotin-16-dUTP (Boehringer) and was hybridized with metaphase chromosomes from peripheral blood lymphocytes of a healthy individual. Signal detection was achieved with avidin-fluorescein isothiocyanate (FITC). To

localize the signal, chromosomes were simultaneously banded with propidium iodide (Sigma).

Glutathione-S Transferase (GST)-Fusion Protein

The entire coding sequence of CBP2 was amplified by PCR with primers 5A (5'-TTTGGATCCATGCGCTCCCCTCCTGCTT-3') which contained a *Bam*HI site, and 3E (5'-TTTGAATTCCTATAACTCGTCTCGCATCT-3') which contained an *Eco*RI site. PCR conditions were as follows: denaturation at 94°C for 30 sec., annealing at 58°C for 30 sec., and extension at 72°C for 1 min., in 35 cycles. The PCR product was digested with *Bam*HI and *Eco*RI, and cloned into a *Bam*HI-*Eco*RI site of pGEX-2T expression vector (Pharmacia). The recombinant CBP2 was introduced into competent *E. coli*, DH5a by transformation. Expression and purification of the fusion protein were as described by Smith and Corcoran (1994). In brief, transformants were selected on LB/ampicillin plates; positive transformant colonies were picked and cultured in LB/ampicillin medium at 37°C for 12-15 hours. The bacteria culture was diluted 1:10 with LB/ampicillin medium and grown 3 hours at 37°C. Expression of fusion protein was induced by adding IPTG to 0.1 mM. Cells were collected and lysed by sonication on ice. The fusion protein was purified with glutathion-agarose beads (Pharmacia). The purified protein was analyzed on 10% SDS-PAGE with Coomassie blue staining.

RESULTS

Isolation of CBP2 cDNA Clone

The 2.1-kb plasmid clone (L13932) isolated from the human fetal lung library was considered to contain an almost full-length cDNA, because 1) the size of the clone corresponded to that of its mRNA as shown by Northern experiment (Fig. 3); 2) the predicted initiating methionine codon of this clone resided within a context favorable for translation initiation (Kozak 1986); and 3) this codon was in a position analogous to the initiation site of other colligins, and was followed by a characteristic signal sequence (Fig. 1, 2).

Clone L13932 contained a poly(A) tail, preceded by a polyadenylation signal, at the 3' end. This cDNA encoded an open reading frame of 1254 bp and contained 87 bp of the 5'-untranslated region and the entire 3'-untranslated region of 727 bp (Fig. 1).

Primary Structure of CBP2

L13932 encoded a polypeptide of 418 amino acids (Fig. 1). A putative hydrophobic signal sequence of 18 amino acids was present at the N-terminal, followed by the mature protein with a C-terminal RDEL sequence. The deduced polypeptide had two putative N-linked oligosaccharide attachment sites, one at amino acid residues 120-122 and the other at 125-127. It also contained a reactive site characteristic of the serpin family at amino acids 362-391, with a P-F-X-F motif downstream.

Comparison of the Nucleotide Sequence of CBP2 with CBP1

The entire nucleotide sequences of the two human colligin genes exhibited 92% identity. In particular, the regions encoding mature protein were highly conserved (97% homology in amino acid sequence), whereas the signal peptides shared only 82% identity (Fig. 2). The 5'-untranslated regions were even less homologous (67% identity). In the 3'-untranslated region of CBP2, the first 131 bp after the translation stop codon were 100% identical to those of CBP1; a 60-bp unique region followed, after which the sequences were highly homologous (94% overall identity over the rest of the 3' end), although the homology decreased toward the terminal.

Expression of CBP2 in Various Tissues

Northern-blot analysis showed a transcript about 2 kb long (Fig. 3) under high-stringency condition. The strongest hybridization signals were found in tissue samples from placenta, lung, skin, and ovary. Tissues from heart, kidney, and colon showed high levels of expression, but no expression was detectable in brain or in peripheral leukocytes.

RT-PCR (Fig. 4) yielded PCR products of the expected size (about 800 bp). The observed pattern of expression was substantially identical to that achieved in the Northern-blot experiment. However, this technique detected a trace level of expression in fetal and adult brain tissues.

Isolation and Characterization of a Genomic Clone of CBP2

A genomic clone (c3932) was isolated from a human cosmid library. Southern hybridization analysis using oligonucleotide

probes specific to CBP2 sequence showed that this clone encompassed the entire coding sequence, the entire 3'-untranslated region, and at least a part of the 5'-untranslated region. In contrast, oligonucleotide probes specific to CBP1 sequence revealed no hybridization signals under the same conditions.

The restriction map of the clone was analyzed (Fig. 4). The CBP2 gene, approximately 11 kb in length, consisted of 5 exons. Four EcoRI fragments of 5, 5, 4 and 2 kb covered the entire gene.

Exon-intron Structure of CBP2

The genomic DNA sequence of CBP2 was compared with the cDNA sequence, and this comparison enabled us to determine exon-intron boundaries as well as investigate the 5'- and 3'-flanking regions of the gene (Table 1). Exons I-V were composed of 53, 656, 99, 234, and 1005 bp respectively. Sizes of introns 1-4 were approximately 4.0, 2.0, 0.1, and 2.5 kb respectively. Sequences at exon-intron junctions were consistent with the consensus sequences reported for splice junctions (Breathnach and Chambon 1981).

Nucleotide Sequence of the 5'-flanking Region of CBP2

The nucleotide sequence was determined for about one kilobase of the 5'-flanking region of CBP2 (Fig. 5). The promoter sequence showed significant homology (74% identity) to that of the murine colligin gene, HSP47 (Hosokawa et al. 1993) by the FASTA program (Pearson and Lipman 1988). A TATA box was positioned at -45, but no CAT box was found. There was a

potential heat shock-response element, which contained alternative arrays of blocks XGAAX and XTTCX (Sorger 1991), spanning positions -91 to -78. No Sp1 or AP1 binding sites were present, but a purine-rich stretch extended for 35 nucleotides between nt.-767 and -802. A GATA binding site consensus sequence (WGATAR; Evans et al. 1988) was present at -914 (5'-CTATCA-3', complementary to TGATAG on the other strand). No complete consensus sequence of reinoic acid-responsive element (RARE) (GGTCAXntGACC; Beato 1989) was found, but direct repeats of TGACC or TGACC-like sequence (Vasios et al. 1989; Wang 1992; Liu et al. 1994) existed at -1012 to -983, -708 to -682, and -657 to -630.

Chromosomal Localization of CBP2

More than 100 metaphase cells were examined after *in situ* hybridization. Specific hybridization signals were identified at band 11q13.5 in almost all cells (Fig. 6); no significant background was observed at any other chromosomal sites.

GST-Fusion Protein

The entire coding sequence of L13932 was cloned into pGEX-2T vector in the correct reading frame, which was confirmed by restriction digestion analysis and sequencing of the cloning site. The transformant expressed a GST-fusion protein of the expected size (about 75 kDa) in the soluble fraction (Fig. 7).

DISCUSSION

We have isolated and characterized a novel human colligin gene, CBP2. Amino acid identities of the CBP2 product with its counterparts in rat, mouse and chicken are 94%, 93%, and 77%, respectively (Fig. 2). This evolutionary conservation underscores the likely functional importance of the colligins as collagen-binding proteins. In addition to the high degree of sequence homology, the predicted gene product contains features in common with CBP1 (Fig. 1, 2): 1) an N-terminal hydrophobic signal sequence, 2) two putative N-linked oligosaccharide attachment sites, 3) a reactive-site region characteristic of the serpin family, in particular the P-F-X-F sequence, and 4) a C-terminal RDEL sequence. Therefore, this novel gene is clearly a member of the family of colligin genes.

Although colligin has been found in various kinds of cell lines, tissue distribution of colligin *in vivo* is still unclear. The expression pattern of CBP1 has not been reported. Although we performed simultaneous RT-PCR experiments using specific primers for CBP1, we were unable to detect the expression of CBP1. The distribution of gp46 (rat colligin) in newborn and adult rats assayed by immunoblotting (Nandan et al. 1990) implied that muscle, kidney, heart, and skin in neonatal rat contain substantial amounts of colligin, but that none is present in liver or brain. In those experiments, colligin was completely absent in the adult rat. In contrast, our results indicate a pattern of ubiquitous expression in humans; we have demonstrated strong expression of CBP2 in various adult tissues, including the liver. The discrepancies in expression patterns

between rat and human tissues may result from the method of detection, (i.e., protein vs mRNA level) and/or reflect species differences. Another possible explanation is that developmental regulation may differ according to the form of colligin being measured. As we have shown that at least two forms of colligin exist in humans, molecular heterogeneity of this protein may exist in other species also. The significance of molecular heterogeneity of colligin in the human is not clear; a major question is whether the two colligins differ in some functional aspect, such as regulation of expression in various tissues and/or developmental stages, binding properties, or substrate specificities as serpins.

In genomic structure, the CBP2 gene is very similar to the murine colligin gene except in the 5'-untranslated region, where alternatively-spliced variants have been reported in the mouse (Wang 1992; Hosokawa et al. 1993). The promoter sequence of CBP2 is also very similar to the mouse counterpart (J6, Wang 1992; HSP47, Hosokawa et al. 1993) and contains several regulatory elements including a purine-rich stretch, a possible RAREs, a GATA binding site consensus sequence, and a heat-shock element. The promoter sequence of CBP1 has not been reported to date, nor has its genomic structure. We have tried genomic PCR using primers for specific sequences in the 3'-untranslated region of CBP1, but we were unable to obtain any PCR-products.

The function of colligin has not been clarified at present. Possible functions include acting as a molecular 'chaperon' (Ellis and Van der Veijs 1991; Nakai et al. 1992) for the synthesis or assembly of types I and IV collagens, and/or protection of α -chains of collagen from denaturation. Its

serpin-like structure supports the possibility that it could be involved in protection of the nascent procollagen chain from attack by proteinase in the ER (Klausner and Sitia 1990). On the other hand, the presence of a heat-shock element indicates that colligin is a heat-shock protein (HSP); HSPs are reported to play an important role as molecular chaperones in the process of folding or assembling newly synthesized or malformed proteins (Ellis and Van der Veis 1991; Gething and Sambrook 1992). Colligin specifically binds to collagens and the expression of colligin always correlates with expression of collagen genes (Nagata and Yamada 1986; Saga et al. 1987). Messenger RNA levels of colligin and type I collagen also change concurrently during differentiation and transformation (Clarke et al. 1993), and respond in a similar fashion to treatment with various chemicals (Nandan et al. 1990, Wang and Gudas 1990) and growth factors (Clarke et al. 1993). This body of evidence indicates that colligin is a collagen-specific molecular chaperon.

The possible RARE in the promoter region of CBP2 is of particular interest. Retinoic acid responsiveness of a similar sequence in murine colligin has been demonstrated by means of transfection experiments (Wang 1992). In laboratory animals, regulation of colligin by retinoic acid is developmental: Gp46, rat colligin, is absent in undifferentiated F9 embryonal carcinoma cells, but the protein is produced when cellular differentiation is induced by retinoic acid (Nandan et al. 1990). Colligin was induced in parallel with collagen IV in these rats (Wang and Gudas 1990). The role of colligin in the process of retinoic acid-induced differentiation remains to be determined.

REFERENCES

Beato M: Gene regulation by steroid hormones. *Cell* 56: 335-344 (1989)

Breathnach R, Chambon P: Organization and expression of eukaryotic split genes coding for proteins. *Annu Rev Biochem* 50: 349-383 (1981)

Carrell RW, Boswell DR: Serpins, the superfamily of plasma serine proteinase inhibitors. In *Proteinase Inhibitors*. Barrett AJ, Salvesen G (eds). Elsevier, Amsterdam. 403-420 (1986)

Cates GA, Brickenden AM, Sanwal BD: Positive involvement of a cell surface glycoprotein in the differentiation of skeletal myoblasts. *J Biol Chem* 259: 2646-2650 (1984)

Cates GA, Nandan D, Brickenden AM, Sanwal BD: Differentiation defective mutants of skeletal myoblasts altered in a gelatin binding glycoprotein. *Biochem Cell Biol* 65: 767-775 (1987)

Clarke EP, Cates GA, Ball EH, Sanwal BD: A collagen-binding protein in the endoplasmic reticulum of myoblasts exhibits relationship with serine proteinase inhibitors. *J Biol Chem* 266: 17230-17235 (1991)

Clarke EP, Sanwal BD: Cloning of a human collagen-binding protein, and its homology with rat gp46, chick HSP47 and mouse J6 proteins. *Biochem Biophys Acta* 1129: 246-248 (1992)

Clarke EP, Jain N, Brickenden AM, Lorimer IA, Sanwal BD: Parallel regulation of collagen I and colligin, a collagen-binding protein and a member of the serine proteinase inhibitor family. *J Cell Biol* 121: 193-199 (1993)

Ellis RJ, Van der Veis SM: Molecular chaperons. *Annu Rev Biochem* 60: 321-347 (1991)

Evans T, Reitman M, Felsenfeld G: An erythrocyte-specific DNA binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proc Natl Acad Sci USA* 85: 2444-2448. (1988)

Fields C, Adams MD, White O, Venter JC: How many genes in the human genome? *Nature Genet*: 345-346 (1994)

Gething MJ, Sambrook J: Protein folding in the cell. *Nature* 355: 33-45 (1992)

Hirayoshi K, Kudo H, Takechi A, Nakai A, Iwamatsu K, Yamada KM, Nagata K: HSP47, a tissue-specific transformation-sensitive, collagen-binding heat shock protein of chicken embryonal fibroblasts. *Mol Cell Biol* 11: 4036-4044 (1991)

Horii A, Nakatsuru S, Ichii S, Nagase H, Nakamura Y: Multiple forms of the APC transcripts and their tissue-specific expression. *Hum Molec Genet* 2: 283-287 (1993)

Hosokawa N, Takechi H, Yokota S, Hirayoshi K, Nagata K: Structure of the gene encoding the mouse 47-kDa heat shock protein (HSP47). *Gene* 126: 187-193 (1993).

Hunt LT, Dayhoff MO: A surprising new protein superfamily containing ovalbumin, antithrombin-3, and alpha 1-proteinase inhibitor. *Biochem Biophys Res Commun* 95: 864-871 (1980)

Joslin G, Fallon RJ, Bullock J, Adams SP, Perlmutter DH: The SEC receptor recognizes a pentapeptide neodomain of α 1-antitrypsin-protease complexes. *J Biol Chem* 266: 11282-11288 (1991)

Klausner RD, Sitia R: Protein degradation in the endoplasmic reticulum. *Cell* 62: 611-614 (1990)

Kozak M: Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44: 283-292 (1986)

Kurkinen M, Taylor A, Garrels JI, Hogan BLM: Cell surface-associated proteins which bind native type IV collagen or gelatin. *J Biol Chem* 259: 5915-5922 (1984)

Liu Y, Chen H, Dong J-M, Chiu J-F: Cis-acting elements in 5'-flanking region of rat α -fetoprotein mediating retinoic acid responsiveness. *Biochem Biophys Res Commun* 205: 700-705 (1994)

Nagata K, Yamada KM: Phosphorylation and transformation-sensitivity of a major collagen-binding protein of fibroblasts. *J Biol Chem* 261: 7531-7736 (1986)

Nakai A, Satoh K, Hirayoshi K, Nagata K: Involvement of the stress protein HSP47 in collagen processing in the endoplasmic reticulum. *J Cell Biol* 117: 903-914 (1992)

Nandan D, Cates GA, Ball EH, Sanwal BD: Partial characterization of a collagen-binding, differentiation-related glycoprotein from skeletal myoblasts. *Arch Biochem Biophys* 278: 291-296 (1990)

Nandan D, Clarke EP, Ball EH, Sanwal BD: Ethyl-3,4-dihydroxybenzoate inhibits myoblasts differentiation: evidence for an essential role of collagen. *J Cell Biol* 110: 1673-1679 (1990)

Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85: 2444-2448 (1988)

Pelham HRE: The retention signal for soluble proteins of endoplasmic reticulum. *Trends Biochem Sci* 15: 483-486 (1990)

Saga S, Nagata K, Chen W-T, Yamada KM: PH-dependent function, purification, and intracellular location of a major collagen-binding glycoprotein. *J Cell Biol* 105: 517-527 (1987)

Sambrook J, Fritsch EF, Maniatis T (eds): Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY (1989)

Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain termination inhibitors. Proc Natl Acad Sci USA 74: 5463-5467 (1977)

Smith DB, Corcoran LM: Expression and purification of glutathione-S-transferase fusion proteins. In Current Protocols in Molecular Biology, eds. Ausubel FM et al. John Wiley and Sons, NY (1994)

Sorger PK: Heat shock factor and heat shock response. Cell 65: 363-366 (1991)

Sudo K, Chinen K, Nakamura Y: 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. Genomics 24: 276-278 (1994)

Takahashi E, Hori T, O'Connell P, Leppert M, White R: R-banding and nonisotopic in situ hybridization: precise localization of the human type I collagen gene (COL2A1). Hum Genet 86:14-16 (1990)

Vasios GW, Gold JD, Retkovich M, Chambon P, Dudas LJ: A retinoic acid responsive element is present in the 5'flanking region of the laminin B1 gene. Proc Natl Acad Sci USA 86: 9099-9103 (1989)

Wang SA, Gudas LJ: A retinoic acid-inducible mRNA from F9 teratocarcinoma cells encodes a novel proteinase inhibitor homologue. *J Biol Chem* 265: 15818-15822 (1990)

Wang S-Y: Structure of the gene and its retinoic acid-regulatory region for murine J6 serpin. *J Biol Chem* 267: 15362-15366 (1992)

Figure 1. Nucleotide and deduced amino acid sequences of human colligin-2 (CBP2). Amino acids of CBP1 that differ from CBP2 are indicated beneath the aligned sequences. The initiating methionine (ATG) and the polyadenylation signal (AATAAA) of this gene are underlined; the termination codon (TAG) is indicated by an asterisk. Characteristic amino acid sequences are also underlined and identified with numbers as follows: 1, N-terminal hydrophobic signal sequence; 2, putative N-linked oligosaccharide attachment sites (N-X-S/T motif); 3, P-F-X-F motif; 4, RDEL sequence. Sequences of the primers used for RT-PCR are overlined. Numbering in the side columns refers to the nucleotide sequence.

1	GTCGCGGCTCGAGAGCGAGAGTACAGTCCCGCGCTAGCCCAGCCGACCCAGGCCACC	60
61	GTGGTGACGCAAACTTCTCTGGCC ^{5C2} <u>ATGCGCTCCCTCTGCTTCTCAGCGCCTTCTGC</u>	120
	<u>M R S L L L L L S A F C</u>	
121	CTCCTGGAGGCGGCCCTGGCGCGGAGGTGAAGAACTCGCAGCCGAGCAGCTCCTGGC	180
	<u>L L E A A L A A E V K K P A A A A A P G</u>	
181	ACTGCGGAGAAGTTGAGCCCCAAGCGGCCACGCTTGCCGAGCGCAGCGCGGCTGGCC	240
	<u>T A E K L S P K A A T L A E R S A G L A</u>	
241	TTCAGCTTGTACAGGCCATGGCCAAGGACAGGCACTGGAGAATCTGCTGTCAACC	300
	<u>F S L Y Q A M A K D Q A V E N I L V S P</u>	
301	GTGGTGTGGCCTCGTCGCTGGGGCTCGTGTGCTGGGCGGCAAGGCGACCCAGCGCTCG	360
	<u>V V V A S S L G L V S L G G K A T T A S</u>	
361	CAGGCCAAGGCAGTGTGAGCGCCGAGCAGCTGCGCGACGAGGAGGTGCACCGCGGCTG	420
	<u>Q A K A V L S A E Q L R D E E V H A G L</u>	
421	GGCGAGCTGCTGCGCTCACTCAGCACTCCACGGCGCGCAACGTGACCTGGAAGCTGGGC	480
	<u>G E L L R S L S N S T A R N V T W K L G</u>	
481	AGCCGACTGTACGACCCAGCTCAGTGAGCTTCGCTGATGACTTCGTGCGCAGCAGCAAG	540
	<u>S R L Y G P S S V S F A D D F V R S S K</u>	
541	CAGCACTACAAGTGGGAGCACTCCAAGATCAACTCCGCGACAAGCGCAGGCGGCTGCAG	600
	<u>Q H Y N C E H S K I N F R D K R R P L Q</u>	
601	TCCATCAACGAGTGGGCGCGCAGACCCGACGGCAAGCTGCCCGAGGTACCAAGGAC	660
	<u>S I N E W A A Q T T D G K L P E V T K D</u>	

661 GTGGAGCGCACGGACGGCGCCCTGTTAGTCAACGCCATGTTCTTCAAGCCACACTGGGAT 720
 V E R T D G A L L V N A M F F K P H W D
 721 GAGAAATCCACCACAAGATGGTGGACAACCGTGGCTTCATGGTGACTCGGTCCCTATACC 780
 E K F H H K M V D N R G F M V T R S Y T
 781 GTGGGTGTCTGATGATGACACCGGACAGGCCTTCAACTACTACGACGACGAGAAGGAA 840
 V G V M M M H R T G L Y N Y Y D D E K E
 841 AAGCTGCAAAATCGTGGAGATGCCCCCTGGCCCCACAAGCTCTCCAGCCTCATCCTCATG 900
 K L Q I V E M P L A H K L S S L I I L M
 901 ^{3X}CCCCATCACGTGGAGCCTCTCGAGCGCCTTGAAAAGCTGCTAACCAAGAGCAGCTGAAG 960
 P H H V E P L E R L E K L L T K E Q L K
 961 ATCTGGATGGGGAAGATGCAGAAGAAGGCTGTGTGCCATCTCCTTGCCCAAGGTGTGGTG 1020
 I W M G K M Q K K A V A I S L P K G V V
 1021 GAGGTGACCCATGACCTGCAGAAACACCTGGCTGGGCTGGGCTGACTGAGGCCATTGAC 1080
 E V T H D L Q K H L A G L G L T E A I D
 1081 AAGAACAAGCCGCACTTGTCAAGCATGTCAGGCAAGAAGGACCTGTACCTGCCCAGCGTG 1140
 K N K A D L S R M S G K K D L Y L A S V
 1141 TTCCACGCCACCGCCTTGAGTTGGACACAGATGGCAACCCCTTTGACCAGGACATCTAC 1200
 F H A T A F E L D T D G N P F D Q D I Y
 1201 GGGCGCGAGGAGCTGCGCAGCCCCAAGCTGTTCTACGCCGACCAACCCCTTCATCTTCTTA 1260
 G R E E L R S P K L F Y A D H P E I F L
 1261 GTGCGGACACCCAAAGCGGCTCCCTGCTATTCTATTTGGGCGCTGGTCCGCGCTAAGGT 1320
 V R D T Q S G S L L F I G R L V R P K G
 1321 GACAAGATGCGAGACGAGTTATAGGGCCTCAGGTGCACACAGGATGGCAGGAGGCATCC 1380
 D K M R D E L *
 4
 1381 AAAGGCTCCTGAGACACATGGGTGCTATTGGGGTTGGGGGGGAGGTGAGGTACAGCCTT 1440
 1441 GGATACTCCATGGGGTGGGGTGGAAAAGCAGACCGGGGTTCCCGTGTGCCTGAGCGGAC 1500
 1501 CTTCCAGCTAGAAATCACTCCAATTGGACATGGGCCCCAGATACCATGATGCTGAGCCC 1560
 1561 GGAAACTCCACATCCTGTGGACCTGGGCCATAGTCATTCTGCCTGCCCTGAAAGTCCCA 1620
 1621 GATCAAGCCTGCCTCAATCAGTATTATATTATAGCCAGGTACCTTCTCACCTGTGAGA 1680
 1681 CCAAATTGAGCTAGGGGGTGCAGCCAGCCCTCTTCTGACACTAAACACCTCAGCTGCCT 1740
 1741 CCCCAGCTCTATCCCAACCTCTCCCAACTATAAACTAGGTGCTGCAGCCCTTGGGACCA 1800
 1801 GGCACCCCCAAGATGACCTGGCCCGCAGTGAGCGGATTGAGAAGGAGCTCCCGAGAGGGG 1860
 1861 CTTCTGGGCAGACTCTGGTCAAGAAGCATCGTGCTGCGCGTTGTGGGGATGAACTTTTTG 1920
 1921 TTTTGTTTCTCTCTTTTTTTAGTTCTTCAAAGATAGGAGGGAAGGGGGAACATGAGCCTT 1980
 1981 TGTGTCTATCAATCCAAGAACTTATTTGTACATTTTTTTTTTCAATAAAACTTTTCCAAT 2040
 2041 GACATTT

241 280
 CBP2 GLYNYYDDEKEKLQIVEMPLAHKLSSLIILMPHHRVEPLER
 CBP1 L
 GP46 L
 J6 M I A S
 HSP47 A V M F I N

281 320
 CBP2 LEKLLTKEQLKIWMGKMCKKAVAIISLPKGVVEVTHDLQKH
 CBP1 T
 GP46 A
 J6 T A S K R S V L S
 HSP47 V N R

321 360
 CBP2 LAGLGLTEAIDKNKADLSRMSGKKDLYLASVFHATAFELD
 CBP1 W
 GP46 W
 J6 W
 HSP47 D T K I S N A L W

361 400
 CBP2 TDGNPPFDQDIYGREELRSPKLFYADHPFIELVRDTQSGSL
 CBP1 N
 GP46 E N
 J6 E N
 HSP47 Y A M N M I K S K T N I

401
 CBP2 LFIGRLVRPKGDKMRDEL*
 CBP1 L
 GP46
 J6
 HSP47

Figure 3. Northern blotting. Top: Expression of human colligin-2 mRNA in various human tissues. Lanes: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas; 9, spleen; 10, thymus; 11, prostate; 12, testis; 13, ovary; 14, small intestine; 15, colon; 16, peripheral blood leukocytes. Molecular sizes are indicated in kb at left. Bottom: Expression of human β -actin gene as a control.

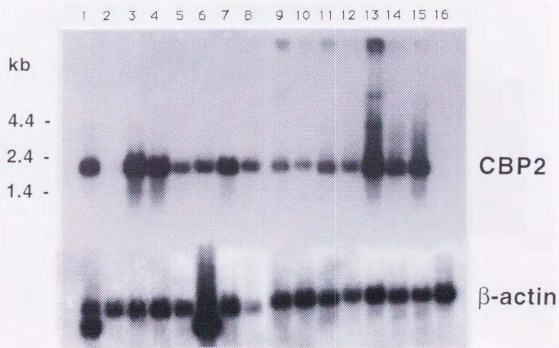


Figure 4. Reverse-transcriptase polymerase chain reaction of human colligin-2 in various human tissues. Lanes: 1, prostate; 2, ovary; 3, uterus; 4, brain; 5, pancreas; 6, spleen; 7, kidney; 8, stomach; 9, fetal lung; 10, heart; 11, small intestine; 12, breast; 13, skeletal muscle; 14, liver; 15, testis; 16, colon; 17, fetal brain; 18, lymphoblastoid cell; 19, skin fibroblast; 20, negative control; 21, L13932 (positive control).

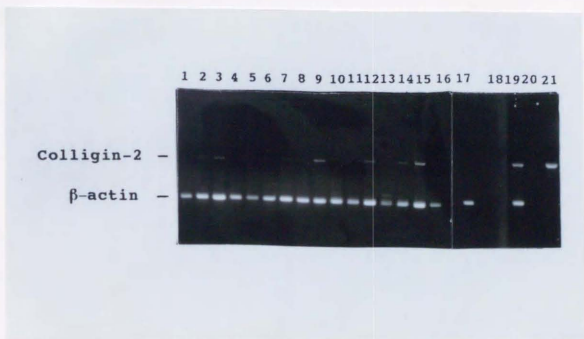


Figure 5. Restriction mapping of the human colligin-2 gene. The gene consists of 5 exons (I-V) separated by 4 introns (1-4). Open boxes indicated the 5'- and 3'-untranslated regions, and closed boxes the coding regions. Restriction sites: E, *EcoRI*; X, *XhoI*. The scale in base pairs is given on the right, under the map.

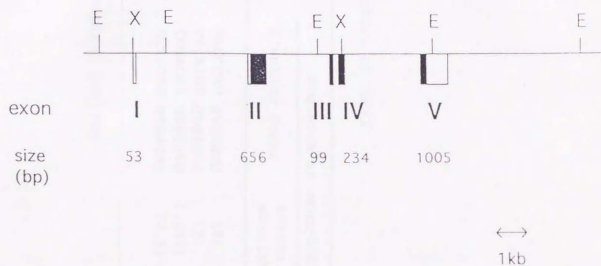


Table 1. Exon-Intron Structure of CBP2

Exon		Intron location	Sequence at exon-intron junction		
No.	size (bp)		5'splice donor	intron size(kb)	3'splice acceptor
1	53	53/54	GACCCAG gtgaggg	(4)	ctcacag GCCCACC
2	656	709/710	TTCAAGC gtgagtc	(2)	actacag CACACTG
3	99	808/809	CGGACAG gtaggtg	(.109)	ctcccag GCCTCTA
4	234	1042/1043	CCTGCAG gtaaggg	(2.5)	cccacag AAACACC
5	1005				

Exon sequences are in capital letters.

Figure 6. Nucleotide sequence of the 1-kb 5'-flanking region of the human colligin-2 gene. Sequence motifs are underlined as follows: 1, TATA box; 2, heat shock-response elements; 3, purine-rich stretch; 4, GATA-binding site consensus sequence; 5, possible retinoic acid-responsive elements.

```

-1042 CCTACTAACAGGTACATACCCACAGATGGACATCGCACAGGGCAAGGAC
                                     5
-992 TTTGTTTCAGCTCTCAGCTGTATCCTCAGCACCTAGAACAGTAATGAATAC
-942 CCTAGCTTAACTTGGAGGTCAAGGAGCTATCAGTTTGCAGGGGTGGGTA
                                     4
-892 GGAATTGACAGTGAGACCTGAGGCTGTGGGAGGGGACCCAAAGAGGGAG
-842 GGGATGCAATAGGGAGGGGGCCAGGGGTGACAAGGATTGAGGAAGGGAGA
                                     3
-792 GAGGGGGGAAAAAAGCAAGGGATGCCTTAGAACCACATTTACAGCCAA
-742 GGGAACAGAGGCCAGAAAGGGAAGTAACCTGCTTAGGGTCACACAGCA
                                     5
-692 CCTTGCTCAGTGGAGAGCCAGGTTTTTCCTTCTGTGCACTCTCTCAAGCC
                                     5
-642 CAGCCAGAGCCACCTGAAGTTCCCCAGGCATCTCTGCCTCTATTACTCCAG
-592 GACTTGAACTTTCCGGGTGCCCGGCAGGTACCGGGTCTGGTCTGCTCCCT
-542 CTCCCTCTGGCCATCGCTGAGGTTGAGGTTTTTGTATGTACAAGTATGGA
-492 GAAGGGCACTCCCTTCAGAAGCCTGAACGTCTCCCTGAAAGGAGGGGGT
-442 GCACAGGACTCARTTGTTCAGCTTGAAAAATGGGGGAGAGCGGGGAGAAG
-392 GGGAGATGGCTCTGCTTGGGCAGAAGCCTGCGGGGAAAGGGGCGCTGAA
-342 AGGACGTGCGATTTCGAGTGGGCTAGCTTATGCAGAGACCTTGGGGTGG
-292 GAGGAAGCTCGCACTCTGAAGGACACGCTGATCCCCGTGGGGACTCCCCG
-242 CGCCCCGCAGCCCGGGCCGCCGAGGAGGCAGTAGGACCCAGGGGCCGGG
-192 AGCGCGCCGCAGAGGAGGGGCGGGGGCCGGGAGGTTTTGAGGAGGT
-142 CTTTGGCTTTTTTTGGCGGAGCTGGGGCGCCCTCCGGAAGCGTTTCCAAC
-92 TTTCCAGAAAGTTTCTCGGGACGGGCAGGAGGGGTGGGGACTGCCATATA
    2                                     1
-42 TAGATCCCGGAGCAGGGGAGCGGCTAAGAGTAGAATCGT

```

Figure 7. Chromosomal localization of human colligin-2 by fluorescence in situ hybridization (FISH). Specific hybridization signals were identified at chromosome 11q13.5 (arrow).

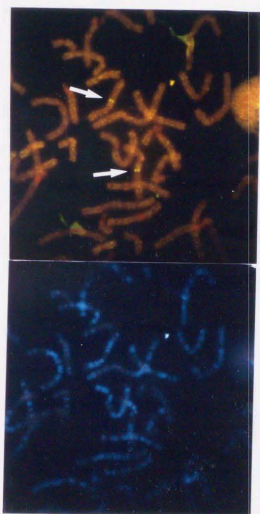


Figure 8. Expression of glutathione-S transferase (GST)-fusion protein assayed by 10% SDS-PAGE. Lane 1, GST-CBP2 (human colligin-2) fusion protein; lane 2, GST (vector only); lane 3, molecular weight markers (sizes are denoted at right in kDa).

