J. Biochem. 122, 983-990 (1997)

# Isolation and Characterization of the Plasma Hyaluronan-Binding Protein (PHBP) Gene (*HABP2*)[1]

**Jun-ichi Sumiya,* Shuichi Asakawa,† Takashi Tobe,*,[2] Ken Hashimoto,* Ken-ichi Saguchi,* Nam-Ho Choi-Miura,* Yoshiko Shimizu,‡ Shinsei Minoshima,† Nobuyoshi Shimizu,† and Motowo Tomita***

**Department of Physiological Chemistry, School of Pharmaceutical Sciences, Showa University, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142; †Department of Molecular Biology, Keio University School of Medicine, Shinjuku-ku, Tokyo 160; and ‡Department of Medical Genetics, School of Health Sciences, Kyorin University, Hachioji, Tokyo 192*

Received for publication, June 24, 1997

**PHBP is a novel human plasma hyaluronan-binding protein that shows significant homology in amino acid sequence to hepatocyte growth factor activator. Two overlapping clones that encode the human plasma hyaluronan-binding protein (PHBP) gene (*HABP2*) were isolated and characterized. The PHBP gene spans 35 kb and is composed of 13 exons from 37 to 1,394 bp in size with consensus splice sites. The gene's regulatory sequences contain putative promoter elements, but no typical TATA box. Some exons of this gene showed significant similarities to those of coagulation factor XII, tissue-type plasminogen activator, and urokinase genes in nucleotide length and in intron phasing. We also report the chromosome mapping of this gene by fluorescence *in situ* hybridization (FISH) using a genomic DNA fragment as a probe. The PHBP gene (*HABP2*) was located on chromosome 10q25-q26.**

**Key words: chromosome 10q25-q26, gene structure, hyaluronan-binding protein, hepatocyte growth factor activator, serine protease.**

Previously, we have reported that a novel plasma hyaluronan-binding protein was purified from human plasma by affinity chromatography with hyaluronan-conjugated Sepharose (*1*). Therefore, we named this protein plasma hyaluronan-binding protein (PHBP). PHBP was found to possess protease activity, which was inhibited by PMSF, indicating that it could be classified as a serine protease. PHBP showed a single 70 kDa band on SDS-PAGE under non-reducing conditions, and 50 and 25 kDa bands under reducing conditions. Thus, PHBP appeared to be a heterodimeric protein composed of 50 and 25 kDa subunits which were linked by disulfide bonds.

The cloning of the cDNA encoding PHBP revealed that this protein was a translation product of an HGFA-like mRNA that had been registered in DDBJ by Kitamura *et al.* (accession number, D49742). The amino acid sequence predicted from the nucleotide sequence of cloned PHBP cDNA showed significant homology to that of hepatocyte growth factor activator (HGFA) (*2*). The predicted structure of PHBP showed three epidermal growth factor (EGF) domains, a kringle domain, and a serine protease domain. In contrast, HGFA has a fibronectin type II domain, an EGF domain, a fibronectin type I domain, an EGF domain, a kringle domain, and a serine protease domain. PHBP has structural regions homologous to those found in HGFA, although it lacks the fibronectin type I and type II domains found in HGFA. Many families of proteins exist which bear structural and functional homologies but share only small degrees of amino acid sequence identity. For example, PHBP shares only 36% sequence identity with HGFA, but this amount of homology may translate to 66% homology in terms of tertiary structure.

The intron/exon organization for the human tissue-type plasminogen activator (tPA) gene has been described (*3, 4*). Separate exons encode distinct structural domains. It has been suggested that the exons encoding protein domains may be transferred within the gene, through an exon shuffling mechanism (*5-8*). In this model, protein domains encoded on separate exons can be exchanged between genes to confer a new structure and possibly a new function upon the recipient molecule. In this paper, we describe the PHBP gene as one such example at the nucleotide level. The PHBP gene is similar in organization to many serine protease genes.

## MATERIALS AND METHODS

*Reagents*—Restriction endonucleases, T4 phage DNA ligase, recombinant Taq DNA polymerase, Moloney murine leukemia virus reverse transcriptase, and pUC119 vector were purchased from Takara Shuzo (Otsu). Nitrocellulose filters (BA85) were purchased from Schleicher

[1] This study was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Sciences, Sports and Culture of Japan and by a Project Research Grant from Kyorin University.

[2] To whom correspondence should be addressed. Tel: +81-3-3784-8215, Fax: +81-3-3784-8216, E-mail: ttobe@pharm.showa-u.ac.jp

Abbreviations: ITI, inter-$\alpha$-trypsin inhibitor; ITI-H1 and H3, inter-$\alpha$-trypsin inhibitor heavy chain 1 and heavy chain 3; IHRP, inter-$\alpha$-trypsin inhibitor family heavy chain-related protein; PHBP, plasma hyaluronan-binding protein; FXII, coagulation factor XII; tPA, tissue-type plasminogen activator; UK, urokinase.

and Schull. [$\alpha$-$^{32}$P]dCTP, [$\gamma$-$^{32}$P]ATP, and nylon filters were obtained from Amersham Japan. A random-primer labeling kit was obtained from Du Pont-New England Nuclear.

*Isolation of PHBP Gene by the BAC System*—The genomic clones encoding the PHBP gene were isolated from the bacterial artificial chromosome library (*9*) by a PCR screening method with the primer set corresponding to the 3′ untranslated region of PHBP cDNA.

*Characterization of the Positive Clones*—Positive clones were analyzed by Southern blotting. The DNAs of genomic clones and human peripheral blood cells were completely digested with a set of restriction enzymes, including *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, or *Sal*I. The DNA fragments were separated by 0.7% agarose gel electrophoresis. After depurinating, denaturing, and neutralizing the gel, the DNA fragments were transferred onto a nylon membrane (Hybond N, Amersham Japan). The membrane was probed with $^{32}$P-labeled cDNA which contained the entire reading frame of PHBP.

*DNA Sequence Analysis*—DNA inserts were characterized by partial restriction enzyme mapping and then subcloned into pUC119 vector. DNA sequencing was performed by using the dideoxy-chain-termination method with a Shimadzu model DSQ1000 DNA sequencer (*10*). The sequences were aligned and decoded with the Genetyx analysis program (SDC, Japan).

*Primer Extension*—A synthetic oligonucleotide (P1) corresponding to nucleotides 82 to 121 of the PHBP cDNA (*1*) (5′-GGAGATCAGACATCCTGGCAAACATCTTTGC-AGTTTAAGG-3′) near the 5′-end of the PHBP cDNA was labeled with [$\gamma$-$^{32}$P]ATP using polynucleotide kinase. The primer ($2\times10^6$ cpm) was annealed to 10 $\mu$g of poly(A)-rich RNA from HepG2 cells for 16 h at 50°C in 15 $\mu$l of hybridization buffer containing 400 mM NaCl, 40 mM Pipes pH 6.4, 1 mM EDTA, and 80% formamide. The annealed RNA/primer mixture was precipitated and resuspended in 20 $\mu$l of buffer containing: 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol, actinomycin D (50 $\mu$g/ml), and 0.5 mM deoxyribonucleoside triphosphate, and extended with 100 units of Moloney murine leukemia virus reverse transcriptase at 42°C for 60 min. The RNA strand was then digested with 500 ng of DNase-free RNase A in the presence of 16 $\mu$M EDTA for 45 min at 37°C. Ammonium acetate was added to give a final concentration of 2 M, and the sample was extracted with phenol-chloroform-isoamyl alcohol (25 : 24 : 1). After precipitation with two volumes of ethanol and lyophilization, the sample was dissolved in 1 mM EDTA, 10 mM Tris-HCl, pH 8.0. The primer extension product was analyzed by acrylamide gel electrophoresis in parallel with DNA sequence reactions of the 5′ genomic subclone. These DNA sequencing reactions were prepared using the same 40-bp oligonucleotide primer.

*Fluorescence In Situ Hybridization (FISH) Analysis*—Metaphase chromosome preparations were made from normal peripheral lymphocytes according to standard methodology. Chromosomes were prepared by R-banding (*11*), and FISH was carried out as previously described with several modifications (*12*). The probe used to identify the PHBP gene was the BAC clone 234-C3 DNA containing human PHBP gene. The genomic DNA was labeled with digoxigenin-11-dUTP (BRL) by nick translation. The reaction mixture (50 $\mu$l) contained 1 $\mu$g of the BAC DNA; dCTP, dATP, dGTP, and digoxigenin-11dUTP (0.2 mM each); DNA polymerase I (2 units, BRL); and DNase I (2 ng). Probe DNA (300 ng) was mixed with 10 $\mu$l of hybridization solution and used for one slide. Immunological detection of hybridization was carried out using FITC-conjugated sheep anti-digoxigenin antibody (Boehringer Mannheim, Germany) and FITC-conjugated donkey anti-sheep immunoglobulin antibody (Silenus, Australia). Chromosomes were counter-stained with propidium iodide (PrI, 20 ng/ml). The fluorescence image of probe hybridization was analyzed by a confocal laser scanning microscope, MRC-600 (Bio-Rad, Richmond, CA). To confirm the chromosome assignment, chromosome 10-specific satellite DNA (Oncor) was used as a probe.

## RESULTS

A primer set was designed based on the nucleotide sequence of the 3′ untranslated region of human PHBP cDNA and was used to screen a BAC library. Two independent BAC clones, 56-B9 and 234-C3, were isolated by several rounds of PCR using this primer set. BAC DNA was isolated from large-scale lysates of each of the two BAC clones and analyzed by restriction enzyme digestion. Analysis of the fragments by agarose gel electrophoresis showed that the inserts contained DNA sequences which overlapped, but were not identical (data not shown). The restriction fragments were transferred onto nylon membranes and hybridized with $^{32}$P-labeled PHBP cDNA. The clone 234-C3 contained two hybridizing bands in addition to those found

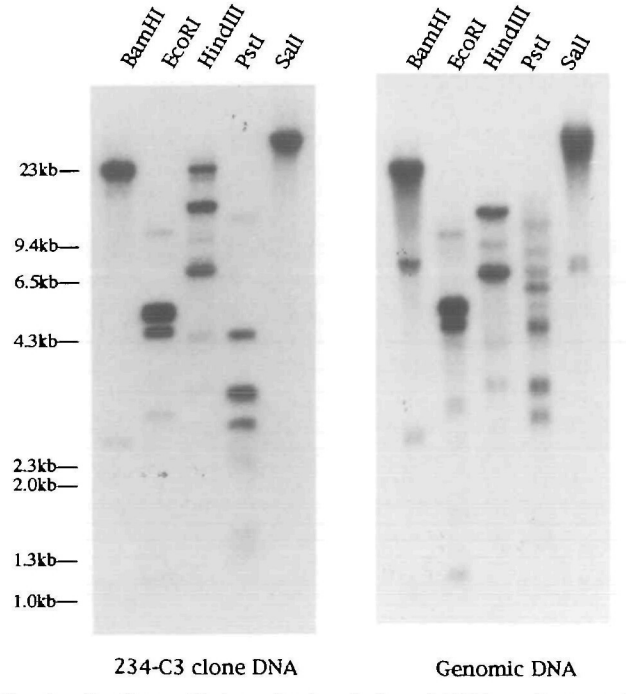


Fig. 1. **Southern blot analysis of cloned DNA and genomic DNA.** The DNAs from the BAC clone 234-C3 and human peripheral blood cells were digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, and *Sal*I, and separated on a 0.7% agarose gel. The DNA fragments were transferred onto nylon filters and hybridized with PHBP cDNA containing the entire coding region.

in clone 56-B9 (data not shown). Figure 1 shows the results of Southern blot analysis of clone 234-C3 and genomic DNA. Although there were some polymorphic differences between 234-C3 clone and genomic DNAs digested with *Bam*HI, *Hin*dIII, *Pst*I, or *Sal*I, all of the hybridizing bands in the *Hin*dIII-digested genomic DNA were found in clone 234-C3, indicating that this clone covered the entire PHBP gene.

The DNA fragments which hybridized to the human PHBP cDNA were subcloned into the pUC119 plasmid vector. DNA sequencing of these inserts was performed by the chain-termination method (*10*). The intron-exon boundaries of the PHBP gene were identified by comparison with the previously reported cDNA sequence (*1*). The putative end of the last exon, exon 13, was defined as the last base of the human hepatocyte growth factor activator (HGFA)-like mRNA registered in DDBJ by Kitamura *et al.* The intron/exon junction sequences for all exons were obtained on both strands and are given in Fig. 2. The positions of the exons were determined by a combination of restriction enzyme analysis, PCR, and nucleotide sequencing. Generally, the introns follow the GT-AG rule, with the exceptions of introns 7, 8, and 10 (*13, 14*). In addition, the intron/exon junctions of the PHBP gene were found to be in agreement with the consensus sequences for the splice junctions of RNA polymerase II-transcribed genes (*14*). The PHBP gene, approximately 35 kb in length, was organized into 13 exons interrupted by 12 introns (Fig. 3). As shown in Fig. 3, comparison of the intron/exon organization with the protein structure revealed that the protein-encoding region of PHBP began at the beginning of exon 1 and ended at the beginning of exon 13. Exons ranged in size from 37 (exon 2) to 1,394 bp (exon 13). Exon 13 was the largest exon (1,394 bp) and corresponded to the 3′-untrans-

| EXON (bp) | 5' Donor Sequence | Intron(kb) | EXON | 3' Acceptor Sequence |
|---|---|---|---|---|
| | 166 | | | 167 |
| 1 (166) | AAGACAGCCTGTGGG gtgaagtgtt<br>K T A C G | I (12) | 2 | tgtttttcag TTCTCCCTGATGTCT<br>F S L M S |
| | 203 | | | 204 |
| 2 (37) | AAAGCCTGGACCCAG gtaagtgtgc<br>S L D P D | II (6.5) | 3 | atccctgcag ACTGGACCCCTGACC<br>W T P D Q |
| | 320 | | | 321 |
| 3 (117) | CTGAGGACCAAGCTG gtaggtacca<br>E D Q A D | III (1.5) | 4 | tgtgtcttag ATCCATGCCAGCCCA<br>P C Q P N |
| | 428 | | | 429 |
| 4 (108) | ATAAGTGTCAGAAAG gtgagtccgt<br>K C Q K V | IV (1.3) | 5 | tctcctacag TGCAAAATACGTGCA<br>Q N T C K |
| | 545 | | | 546 |
| 5 (117) | CCAGCTGCTCCCAAG gtaagtggtg<br>S C S Q V | V (0.8) | 6 | gtgttcacag TGGTTCCTGTATGCA<br>V P V C R |
| | 665 | | | 666 |
| 6 (120) | AATTCTGTGAAATAG gtatgggtct<br>F C E I G | VI (0.5) | 7 | tctctcggag GTTCTGATGACTGCT<br>S D D C Y |
| | 837 | | | 838 |
| 7 (172) | ACACAATTTCTGCAG gcatgcaagc<br>H N F C R | VII (1.9) | 8 | tgtattttag AAACCCAGATGCGGA<br>N P D A D |
| | 935 | | | 936 |
| 8 (98) | CCTGCTCAGCCCAGG gaaaggcttg<br>C S A Q D | VIII(1.2) | 9 | tgtcccacag ACGTTGCCTACCCAG<br>V A Y P E |
| | 1191 | | | 1192 |
| 9 (256) | TGCCCACTGCACCGA gtaggnscgc<br>A H C T D | IX (1.1) | 10 | tcbcttgcag CATAAAAACCAGACA<br>I K T R H |
| | 1334 | | | 1335 |
| 10 (143) | CCCACAATGATATTG gcaagktcct<br>H N D I A | X (0.9) | 11 | ctccacctag CATTGCTCAAGTTAA<br>L L K L K |
| | 1469 | | | 1470 |
| 11 (135) | GTGTTACAGAAACAG gtgagtcggc<br>V T E T G | XI (1.5) | 12 | ttccccacag GAAAAGGGTCGCGCC<br>K G S R Q |
| | 1615 | | | 1616 |
| 12 (146) | CAAGACACCTGCCAG gtcagagact<br>Q D T C Q | XII (2.3) | 13 | tttcccttag GGTGACTCTGGAGGC<br>G D S G G |

3009

13 ----------GACCTTTGGCCAAAAATAAACTTTGAAAAG AAACAATGAGNTTGTCTTTCCCC
(1394)

Fig. 2. **Exon/intron structure of human PHBP gene (*HABP2*).** Intron (lowercase letters) and exon (capital letters) junction boundaries of 13 exons and 12 introns are displayed. Approximate intron sizes were determined by restriction enzyme mapping, PCR, or direct sequencing. The polyadenylation signal is underlined, and the arrowhead indicates the site of poly(A) addition.

lated end of the message. The average length of the remaining exons was 135 bp, similar to the average length of eukaryotic exons of 140 bp (*15*). Introns varied in length from 500 to 12,000 nucleotides, with an average size of approximately 2.6 kb (Fig. 2). The size of the introns was determined by restriction enzyme digestion or by PCR. The predicted size of the mRNA is 2,350 bp or 3,010 bp, depending on the utilization of alternative polyadenylation signals in exon 13. This anticipated size is in close agreement with the size of the mRNA as determined by Northern blot analysis [3,000±100 or 2,300±100 bp, including a poly(A) tail] (*1*). The nucleotide sequence of the protein-encoding region and the cDNA (*1*) were identical except for two nucleotides at positions 281 and 1481 in the cDNA sequence, in which the third base changes of these triplet codons were silent.

To determine the transcription initiation site of the PHBP gene, primer extension experiments were performed using RNA from HepG2 cells and an oligonucleotide primer complementary to the PHBP mRNA. For the primer, we used a 40-base-long oligonucleotide complementary to the mRNA sequence that contained 15 untranslated nucleotides followed by 25 nucleotides in the translated region. The position of the transcription initiation site was determined by analyzing the primer extension product in parallel with DNA sequence analysis of a 2.5-kb genomic DNA fragment that includes the transcription initiation site. The pattern of extension is shown in Fig. 4. As shown in Fig. 4, a major primer extension product was identified. Alignment of the longest PHBP primer extension product with the sequence of the PHBP genomic DNA demonstrated that the human PHBP mRNA was initiated at sequence TCCTGAA, which is 97 nucleotides upstream from the translation initiation codon. Although this observation was not consistent with the fact that gene transcription tends to initiate at a purine nucleotide (*16*), our data implicate this thymidine residue as the transcription initiation site and suggest that it is the 5′ end of exon 1. This experiment indicated that the isolated genomic DNA fragment contained the promoter region of the PHBP gene.

To analyze the promoter region of the PHBP gene, we sequenced 450 nucleotides of the 2.5 kb fragment which were 5′ to the transcription initiation site and contained no consensus acceptor splice sites. Analysis of the 5′-flanking sequence of the PHBP gene revealed that this gene did not have a typical TATA box (Fig. 5). By inspection, several regulatory element motifs were identified in this region of the PHBP gene (Fig. 5). An Sp1 element was located at nucleotide −250, and three potential TCF-1 elements were found at positions −46, −100, and −277 bp upstream from the putative cap site. Consensus sequence motifs for the AP-1 and AP-2 transcription factors were found at nucleotides −117 and −108.

Comparison of the PHBP gene organization with that of multiple members of the serine protease family, human coagulation factor XII (FXII), tissue-type plasminogen activator (tPA), and urokinase (UK), revealed several striking similarities (Fig. 6) (*3, 4, 17, 18*). Ten exons, from exons 4 through 13 of the PHBP gene, were very similar to exons 5 through 14 of the genes encoding factor XII, tPA, and UK in terms of intron phasing, in encoding structural

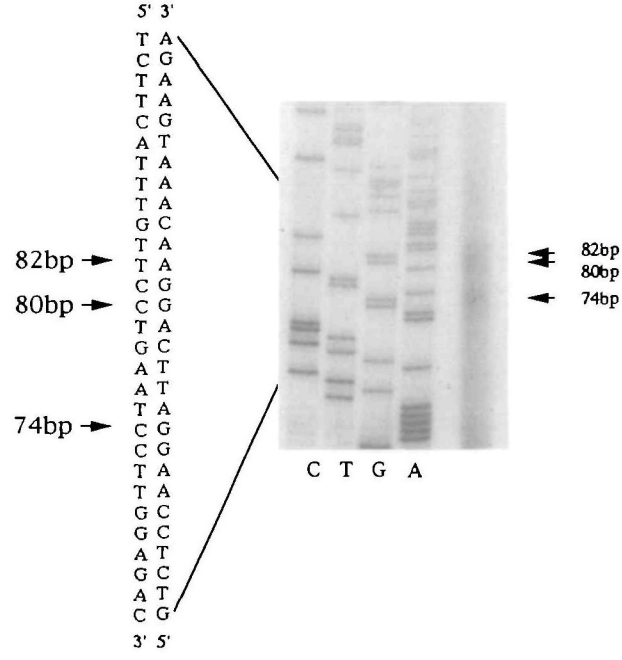


Fig. 4. **Mapping the 5′ end of the human PHBP mRNA.** A 40 bp end-labeled oligonucleotide primer (nucleotides 82-121 of the PHBP cDNA) was hybridized with HepG2 RNA. Hybrids were used as substrates for reverse transcriptase. Extension products were analyzed by electrophoresis in denaturing polyacrylamide gels along-side a sequencing ladder of PHBP genomic DNA prepared using the same primer. A section of the nucleotide sequence of both strands is presented, and the proposed transcription initiation site is marked with an arrow.

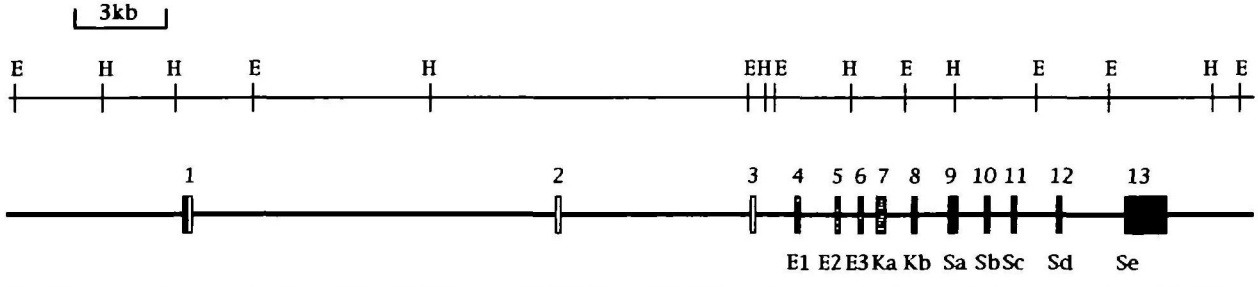


Fig. 3. **The genomic organization of the PHBP gene (*HABP2*).** The boxes representing exons are numbered. Filled-in boxes represent untranslated regions, whereas open, striped, or stippled boxes represent translated regions. The recognition sites for *Eco*RI and *Hin*dIII are shown under the complete gene structure. E1, E2, and E3 represent the exons encoding the EGF domain. Ka and Kb represent the exons encoding the kringle domain. Sa, Sb, Sc, Sd, and Se represent the exons encoding the serine protease domain.

```
agggcccaagaatcaatgtttgtcaatagcagttttgcatcctttaatttatgtctcatgaatataaagtgtgatgttattaggagccca   -362
                                                                  TATA-like
tccatgccttaattctgtttgagatttgatgaaccagaagtctattaacattctgtacaaatcagcnctggcagaggcctgtggcacagg   -272
                                                                                      TCF-1
cttcacctgatcttttatcctgaggctgggatgctagtcacgctgtcaggcttcctgctcggaacacccagcagctgtcatgtgcaccgt   -182
                     Sp1
ggagccaggggctatgtcagctactttctttggagagtttctccccagaagagaaaccacttcatgactcagcccagccccaaagttgc    -92
                                                                  AP-1     AP-2     TCF-1
catttgtttaatcaagatagttagtttgcaaaacattaaacattaaacagtatttgatgcctagttaataatttgcatcctcttcatttg     -2
                                                TCF-1
t TCCTGAATCCTTGGAGACTGACATTTTTCCCCCCTAAAGGCATAGACAACAAAAGAAATTTTATTGAGAGGAAAACACAAGTCCTTAA    +87
ACTGCAAAGATGTTTGCCAGGATGTCTGATCTCCATGTTCTGCTGTTAATGGCTCTGGTGG                               +148
         Met
```

Fig. 5. **DNA sequence of the 5′ flanking region of the PHBP gene (*HABP2*).** The transcribed sequence is in boldface type, and the translation initiation codon is shown by outlined letters. The proposed mRNA cap site is indicated by an arrowhead. The numbering uses the transcription start site upstream of exon 1 as the reference point.

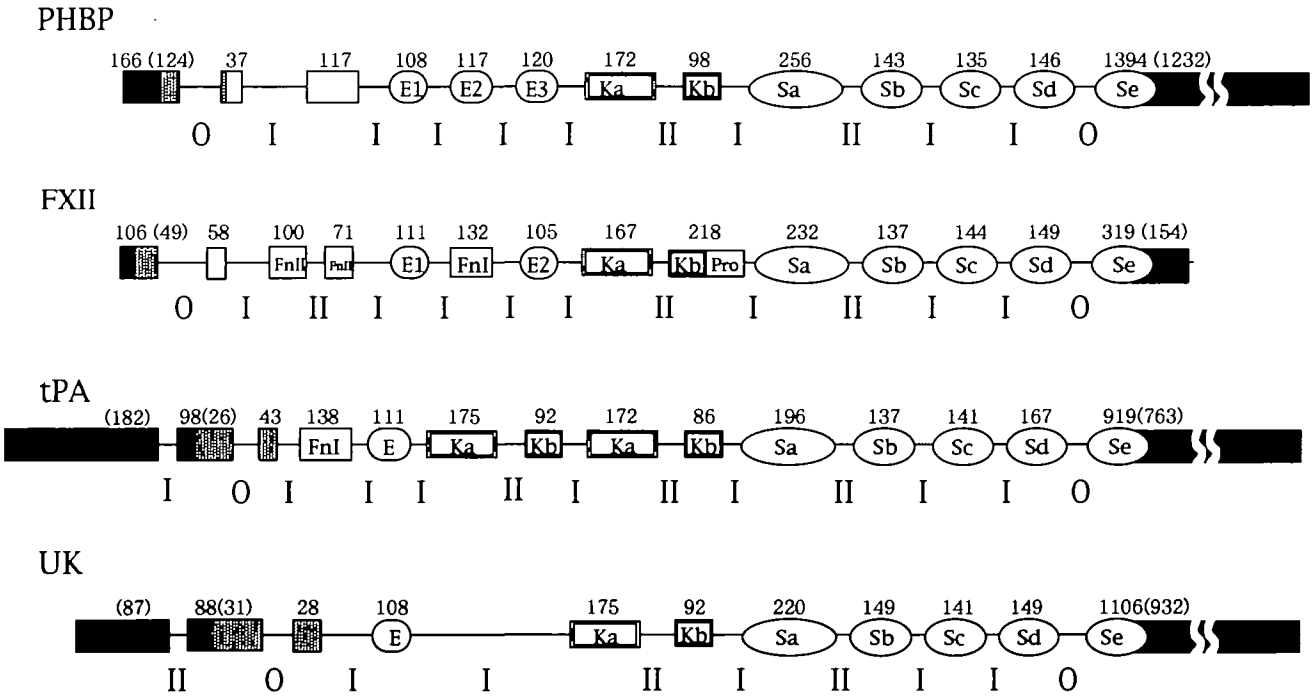


Fig. 6. **Structural comparison of human PHBP gene (*HABP2*) with those of coagulation factor XII, tissue-type plasminogen activator, and urokinase.** The structures of human PHBP, coagulation factor XII (FXII), tissue-type plasminogen activator (tPA), and urokinase (UK) genes are schematically shown. E1, E2, and E3 represent the exons encoding an EGF domain. Ka and Kb represent the exons encoding a kringle domain. Sa, Sb, Sc, Sd, and Se represent the exons encoding a serine protease domain. FnI and FnII represent the exons encoding a fibronectin type I and type II domain, respectively. Pro represents the proline-rich region. The numbers above the exons represent the nucleotide length, and the bracketed numbers represent untranslated nucleotide length of the exon. The Roman numbers indicate the intron phasing. The filled-in boxes represent the untranslated regions, and shaded boxes indicate the region encoding a signal peptide.

domains, and in nucleotide length (Fig. 6).

We performed FISH analysis using the genomic DNA of the PHBP gene to map this gene's chromosomal location. Among more than 100 chromosome spreads examined, 21 well-R-banded chromosome spreads showed clear fluorescence signals on chromosome 10q25-q26 (Fig. 7A). No significant hybridization signals were found on other chromosomes. To confirm the mapping, we used chromosome 10-specific satellite DNA probe (Oncor) and found signals for both satellite DNA and genomic BAC DNA of the PHBP on the same chromosome (Fig. 7B). The most likely localization is within the negative band q26, based on further analysis of elongated chromosomes and computer magnification (Fig. 7C).

## DISCUSSION

We have previously reported the isolation and sequencing

of a cDNA clone encoding the human plasma hyaluronan-binding protein (PHBP) (*1*). This cDNA was identical in sequence to the hepatocyte growth factor activator-like mRNA (accession number, D49742), and the predicted amino acid sequence of PHBP shows significant homology to hepatocyte growth factor activator (*2*). In this study, we have characterized the human PHBP gene (*HABP2*).

Of two overlapping genomic BAC clones, one (named 234-C3) covered the entire PHBP gene. The human PHBP gene spans 35 kb and contains 13 exons, which range in size from 37 to 1,394 bp. The translation initiation codon resides in exon 1, and the stop codon resides in exon 13. The average size of the introns (ranging from 500 to 12 kb) is 2.6 kb and is much larger than that normally encountered in vertebrates (*15*). This is particularly true of introns 1 and 2, which are 12 and 6.5 kb in length, respectively. This makes the genomic PHBP a large gene, with a low coding/non-coding ratio. The coding capacity of this gene is calculated to be about 9%.

The primer extension experiment demonstrated that the transcription of the PHBP gene is initiated from the thymidine residue at nucleotide 1 in Fig. 4. As shown in Fig. 4, the primer extension product was detected as a broad signal, suggesting that the gene of PHBP contains multiple transcription initiation sites around this thymidine residue. Transcription usually starts at several points in TATA-less promoters (*19, 20*). Analysis of the 5′ flanking region of the PHBP gene suggested that the PHBP gene is a TATA-less gene. Although a sequence containing TATA is found at −388 bp in the 5′ flanking region of the PHBP gene, it is not thought to be a TATA-box, because it is too far from the putative transcription initiation site (Fig. 5). Further, the sequence of AATAAT is found at −25 bp, which is a reasonable position for a TATA box. However, it is improbable that the sequence is a TATA box, based on the data of Breathnach and Chambon (*13*).

The promoters of genes that are regulated without the usual TATA or CCAAT sequences have been divided into two classes (*21*). The first has generally been associated with constitutively active housekeeping genes that have G+C-rich promoters with CpG islands, Sp1 binding sites, and often multiple start points (*19, 20*). The second group includes promoters that also lack a TATA element and are not G+C-rich. Smale and Baltimore have suggested that the second class of promoters is differentially or developmentally regulated. They describe a 17-bp motif, the initiator (Inr), that includes the transcription initiation site and is sufficient for basal transcription of the lymphocyte-specific terminal deoxyribonucleotidyl transferase gene (*21*). Another transcription initiator element (HIP1) has been identified in the G+C-rich promoter of the dihydrofolate reductase gene (*22*). Although the 5′ flanking region of the PHBP gene contains neither the consensus sequence for initiation nor the HIP1 element, we consider that the PHBP gene can be classified into the second class of TATA-less promoters.

The transcription initiation sites in TATA-less promoters appear to be determined by three factors. The first is local sequence elements or cap sequences which may in some cases be an initiator (*21, 23, 24*) and may require a bound protein factor (*25-28*). The second is the inducing effect of a factor such as Sp1 at a defined distance, and the

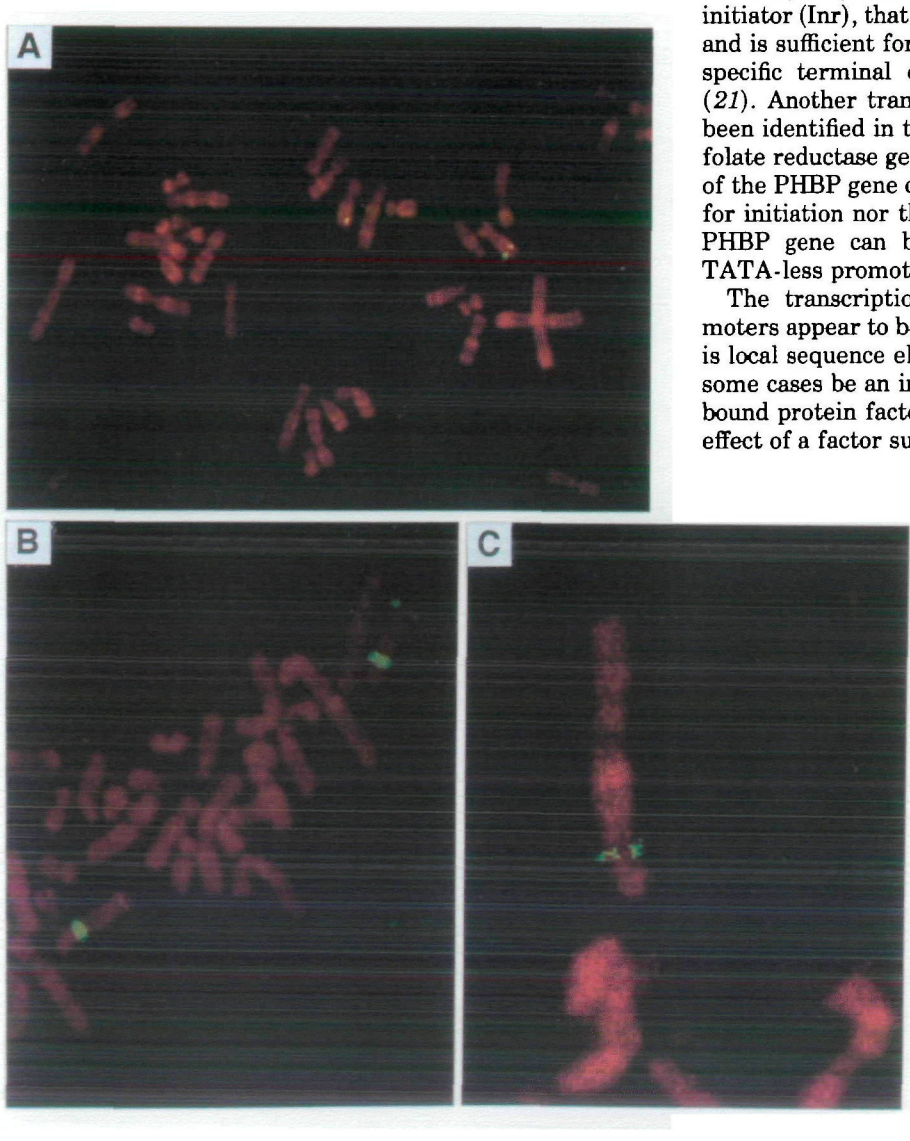


Fig. 7. **Mapping of the PHBP gene by FISH using the genomic BAC DNA clone 234-C3.** Chromosomes of the cultured lymphocytes were R-banded and processed for FISH analysis. The probe was labeled with digoxigenin-11-dUTP and subjected to hybridization. Hybridization signals (yellow to green color) were immunologically detected. The R-banded chromosomes were counter-stained with propidium iodide (red color). The fluorescence image was analyzed by confocal laser scanning microscopy. A, hybridization signals of the PHBP gene on 10q25-q26; B, hybridization signals of the PHBP gene and chromosome 10-specific satellite DNA probe (large green signals on centromere); C, computer magnification of hybridization image of the PHBP gene on elongated chromosome 10.

third is one or more activating factors which bind to the regulatory elements of the promoter and whose effect extends uniformly to the whole start site region. Regarding these determinants, it is possible that the sequence upstream of that analyzed here acts as an initiator. Further analyses by sequencing, mutation, footprinting, and *in vitro* transcription with purified factors will be needed to test this hypothesis.

The organization of the PHBP gene was compared to that of other serine proteases in order to determine the evolutionary history of the gene (Fig. 6). Although the exon number of the PHBP gene is different from that of other serine protease family members, a structural comparison of the human PHBP gene with the coagulation factor XII (FXII), tissue-type plasminogen activator (tPA), and urokinase (UK) genes demonstrates that these genes have many significant similarities (Fig. 6). The serine protease domain of these genes contains five exons which are flanked by introns inserted into the same triplet codon reading frame in all four genes. In many serine proteases, including chymotrypsin, elastase, trypsin, kallikrein, thrombin, tPA, and UK, the coding region outside the protease domain is split by at least one intron in phase I of the reading frame and the various extra domains within the region are encoded by separate exons or groups of exons, also separated by introns in phase I (*8*). These genes are thought to be a serine protease family derived from the same ancestor gene. Serine proteases can be divided into five different families of genes based on their intron/exon gene organization (*8*). Although there are differences, PHBP is more closely related to FXII, tPA, and UK since each of these genes contain introns 3′ to codons for both the active site histidine and aspartate and contain an intron 5′ to the codon for the active site serine. However, the placement of all four introns in almost identical locations, the number of exons encoding a kringle domain, and the presence of an EGF domain in the UK and PHBP genes suggests that these two genes may be more closely related to one another than to the other serine proteases of this family.

It was revealed by FISH analysis that the gene of PHBP is located on human chromosome 10. Interestingly, although the gene for PHBP showed significant similarities to those for FXII, tPA, and UK, the genes of FXII, tPA, and UK, are located on chromosomes 5, 8, and 10, respectively (*29-35*). This also suggests that the PHBP gene may be more closely related to UK. The gene of PHBP seems to have diverged from the genes encoding FXII, tPA, and UK. Given the genes' structural similarities and locations, the PHBP gene probably diverged from that of UK. Although the amino acid sequence of PHBP shows similarity to that of HGFA, the gene encoding HGFA may have diverged from that encoding FXII, as these genes have both fibronectin type I and II domains in addition to EGF domains.

## REFERENCES


1. Choi-Miura, N.-H., Tobe, T., Sumiya, J., Nakano, Y., Sano, Y., Mazda, T., and Tomita, M. (1996) Purification and characterization of a novel hyaluronan-binding protein (PHBP) from human plasma: It has three EGF, a kringle and a serine protease domain, similar to hepatocyte growth factor activator. *J. Biochem.* **119**, 1157-1165
2. Miyazawa, K., Shimomura, T., Kitamura, A., Kondo, J., Morimoto, Y., and Kitamura, N. (1993) Molecular cloning and sequence analysis of the cDNA for a human serine protease responsible for activation of hepatocyte growth factor. *J. Biol. Chem.* **268**, 10024-10028
3. Ny, T., Elgh, F., and Lund, B. (1984) The structure of the human tissue-type plasminogen activator gene: Correlation of intron and exon structures to functional and structural domains. *Proc. Natl. Acad. Sci. USA* **81**, 5355-5359
4. Degen, S.J.F., Rajiput, B., and Reich, E. (1986) The human tissue plasminogen activator gene. *J. Biol. Chem.* **261**, 6972-6985
5. Gilbert, W. (1978) Why genes in pieces? *Nature* **271**, 501
6. Blake, C.C.F. (1978) Do gene-in-pieces imply proteins-in-pieces? *Nature* **273**, 267
7. Blake, C. (1983) Exons—present from the beginning? *Nature* **306**, 535-537
8. Rogers, J. (1985) Exon shuffling and intron insertion in serine protease genes. *Nature* **315**, 458-459
9. Asakawa, S., Abe, I., Kudoh, Y., Kishi, N., Wang, Y., Kubota, R., Kudoh, J., Kawasaki, K., Minoshima, S., and Shimizu, N. (1997) Human BAC library: Construction and rapid screening. *Gene* **191**, 69-79
10. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
11. Takahashi, E., Hori, T.O., Connell, P., Leppert, M., and White, R. (1990) R-banding and nonisotopic *in situ* hybridization. *Hum. Genet.* **86**, 14-16
12. Minoshima, S., Wang, Y., Ichida, K., Nishino, T., and Shimizu, N. (1995) Mapping of the gene for human xanthine dehydrogenase (oxidase) (XDH) to band p23 of chromosome 2. *Cytogenet. Cell Genet.* **68**, 52-53
13. Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349-383
14. Mount, S.M. (1982) A catalogue of splice junction sequence. *Nucleic Acids Res.* **10**, 459-472
15. Naora, H. and Deacon, N.J. (1982) Relationship between total size of exons and introns in protein-coding genes of higher eukaryotes. *Proc. Natl. Acad. Sci. USA* **79**, 6196-6200
16. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563-578
17. Cool, D.E. and MacGillivrary, R.T.A. (1987) Characterization of the human coagulation factor XII gene. *J. Biol. Chem.* **262**, 13662-13673
18. Riccio, A., Grimaldi, G., Verde, P., Sebastio, G., Boast S., and Blasi, F. (1985) The human urokinase-plasminogen activator gene and its promoter. *Nucleic Acids Res.* **13**, 2759-2771
19. Dynan, W.S. (1986) Promoter for housekeeping gene. *Trends Genet.* **6**, 196-197
20. Sehgal, A., Patil, N., and Chao, M. (1988) A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell. Biol.* **8**, 3160-3167
21. Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. *Cell* **57**, 103-113
22. Means, A.L. and Farnham, P.J. (1990) Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. *Mol. Cell. Biol.* **10**, 653-661
23. Roy, A.L., Meisterernst, M., Pognonec, P., and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. *Nature* **354**, 245-248
24. O'Shea-Greenfield, A. and Smale, S.T. (1992) Role of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *J. Biol. Chem.* **267**, 1391-1401
25. Hariharan, N. and Perry, R.P. (1990) Functional dissection of a mouse ribosomal protein promoter: Significance of the polypyrimidine initiator and an element in the TATA-box region. *Proc. Natl. Acad. Sci. USA* **87**, 1526-1530
26. Seto, E., Shi, Y., and Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* **354**, 241-245

27. Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II: Initiator-directed formation of transcription-competent complexes. *FASEB J.* **6**, 3300-3309
28. Irwin, D.M. (1986) Ph.D. thesis, University of British Columbia, Vancouver
29. Tripputi, P., Blasi, F., Ny, T., Emanuel, B.S., Letosfsky, J., and Croce, C.M. (1986) Tissue-type plasminogen activator gene is on chromosome 8. *Cytogenet. Cell Genet.* **42**, 24-28
30. Verheijen, J.H., Visse, R., Wijnen, J.T., Chang, G.T., Kluft, C., and Meera, P. (1986) Assignment of the human tissue-type plasminogen activator gene (PLAT) to chromosome 8. *Hum. Genet.* **72**, 153-156
31. Benham, F.J., Spurr, N., Povey, S., Brinton, B.T., Goodfellow, P.N., Solomon, E., and Harris, T.J. (1984) Assignment of tissue-type plasminogen activator to chromosome 8 in man and identification of a common restriction length polymorphism within the gene. *Mol. Biol. Med.* **2**, 251-259
32. Rajput, B., Degen, S.F., Reich, E., Waller, E.K., Axwlrod, J., Eddy, R.L., and Shows, T.B. (1985) Chromosomal locations of human tissue plasminogen activator and urokinase genes. *Science* **230**, 672-674
33. Stein, P.M., Stass, S.A., and Kagan, J. (1993) The human urokinase-plasminogen activator gene (PLAU) is located on chromosome 10q24 centromeric to the HOX11 gene. *Genomics* **16**, 301-302
34. Tripputi, P., Blasi, F., Verde, P., Cannizzaro, L.A., Emanuel, B.S., and Croce, C.M. (1985) Human urokinase gene is located on the long arm of chromosome 10. *Proc. Natl. Acad. Sci. USA* **82**, 4448-4452
35. Citarella, F., Tripodi, M., Fantoni, A., Bernardi, F., Romeo, G., and Rocchi, M. (1988) Assignment of human coagulation factor XII (FXII) to chromosome 5 by cDNA hybridization to DNA from somatic cell hybrids. *Hum. Genet.* **80**, 397-398