Characterization of the Gene Encoding a Folate-binding Protein Expressed in Human Placenta

IDENTIFICATION OF PROMOTER ACTIVITY IN A G-RICH SP1 SITE LINKED WITH THE TANDEMLY REPEATED GGAAG MOTIF FOR THE ets ENCODED GA-BINDING PROTEIN*

(Received for publication, May 12, 1993, and in revised form, October 13, 1993)

Easwara Sadasiviant, Marisol M. Cedeno, and Sheldon P. Rothenberg§
From the Division of Hematology/Oncology, Department of Medicine, State University of New York Health Science Center and Veterans Affairs Medical Center, Brooklyn, New York 11209

The gene encoding a folate-binding protein (FBP) expressed in human placenta has been cloned by screening a genomic library with the KB cell cDNA. This gene, contained in a 10-kilobase EcoRI fragment of this genomic clone, has 5 exons, 4 introns, the AATAA polyadenylation signal in the 3'-untranslated region, and a 5'-flanking sequence which contains the promoter elements, all of which span approximately 5 kilobases. Transcription initiation was mapped by RNase protection to a site 73 base pairs downstream from a G-rich sequence linked to a tandemly repeated GGAAG sequence which is a motif that the ets oncogene encoded GA-binding protein (GABP) transcription factor binds. Gel-shift and supershift mobility assays indicate that the G-rich sequence and the ets motif bind specifically to SP1 and GABP, respectively. These cis regulatory elements in tandem drive expression of the chloramphenicol acetyltransferase reporter gene in transiently transfected mouse 3T3 cells. The location of these elements upstream of transcription initiation in this gene, which lacks an appropriately located TATA box promoter, indicates that this SP1-GA binding region most probably regulates expression of this placental FBP. The gene encoding this placental FBP has been assigned the FBP/PL-1 gene because it is a member of a multigene family that includes a gene encoding a FBP expressed in both KB cells and placenta and its unprocessed pseudogene.

Folate-binding proteins (FBPs) have now been identified and characterized in a variety of mammalian tissues and cultured cell lines (see Ref. 1 for a recent review). The FBP(s) can be distinguished by structure, orientation on the plasma membrane or in the cell, and by immunoreactivity, even though their ligand binding properties are quite similar. Structural features distinguish a hydrophobic membrane-associated FBP (m-FBP) from a hydrophilic soluble form of this protein (2-5). This m-FBP is anchored to the external surface of the plasma membrane of cultured human KB (6) and CaCo-2 cells (7) by a glycosylphosphatidylinositol (GPI) tail which can be released as a soluble form from the cell membrane by digestion with phosphatidylinositol-specific phospholipase C.

Cellular orientation can also distinguish the m-FBP on the plasma membrane from a nonhydrophobic FBP which has been identified in the cytoplasm of human leukemia cells (8) and normal granulocytes (9). Finally, these FBP(s) can also be distinguished by monospecific antiserums raised to these different forms of the FBP (10).

Corresponding to these properties that distinguish these functional proteins, genetic differences have also been identified. Although the complementary DNA (cDNA) encoding the FBP in KB cells (11, 12) and CaCo-2 cells (7) is identical, a different cDNA has been cloned that encodes a FBP isolated from human placenta (13) and recently Page et al. (14) have cloned and characterized a gene with a similar coding sequence as the cDNA encoding this placental FBP. In addition, we have isolated and characterized an unprocessed pseudogene that appears to have arisen by duplication of the gene encoding the FBP expressed in KB cells (15).

It is now apparent that the FBPs identified in human cells and tissues are encoded by a family of related genes which are expressed uniquely or selectively in different tissues. In order to understand more fully the biological significance of a group of functionally similar FBP(s) encoded by a family of genes in which expression of the protein may be tissue specific, it is necessary to characterize the genomic structure and regulatory elements of each gene. In a previous report (15), we described the structural organization of the gene encoding the FBP expressed in KB cells. In this report we characterize the structure and organization of a gene encoding a FBP expressed in human placenta. This gene lacks an upstream proximal promoter region with TATA or CAAT elements but contains instead an SP1 binding sequence linked to tandemly repeated GGAAG motifs that bind an ets oncogene-derived GA-binding protein (GABP) (16) that may regulate transcription.

MATERIALS AND METHODS

Restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase, and the Rlenov fragment of Escherichia coli DNA polymerase were purchased from New England Biolabs (Beverly, MA). The Sequenase kit and random primer labeling kit were products of U. S. Biochemical Corp. The GeneAmp DNA amplification kit was purchased from Perkin Elmer. Library efficiency and subcloning efficiency DH5a competent cells were obtained from Life Technologies Inc. [α-32P]dCTP (specific activity 6000 Ci/mmole), [γ-32P]ATP (specific activity 6000 Ci/mmole), and [α-32P]ATP (specific activity 1000 Ci/mmole) were purchased from DuPont-New England Nuclear.
Screening of Genomic Library—The ACharon 4A human genomic library was obtained from ATCC, Rockville, MD. This library was constructed by Maniatis et al. (17) from human fetal liver genomic DNA partially digested with Alul/HaeIII and, after EcoRI linker ligation and EcoRI digestion, was cloned into the ACharon 4A vector. The library was screened with a cDNA that was synthesized from a cDNA library prepared from KB cell poly(A)+ RNA (11). The size of the linking clone was estimated by electrophoresis on a 1.5% agarose gel. Each clone was then hybridized with the cloned probe and subsequent plaque purification were carried out using standard procedures (18). For some initial screening of the library, the high stringency wash (0.1 x SSC, 1% SDS at 65 °C) was omitted. The DNA isolation was carried out using both the plate lysis and the liquid culture methods (18).

Plasmid Cloning—The selected clones were separately digested with EcoRI and BamHI and the fragments (ranging from 3 to 15 kb) were isolated following separation by agarose gel electrophoresis and then ligated respectively, into EcoRI-digested and phosphatase-treated, and BamHI-digested and phosphatase-treated pUC18. To subclone the product of the polymerase chain reaction (see below), the amplified DNA fragment was blunt-ended using the Klenow fragment of DNA polymerase I, phosphorylated with T4 polynucleotide kinase, and then isolated by agarose gel electrophoresis and the DNA was purified from the gel using GeneClean kit from BTO 101, Inc. (San Diego, CA). The DNA fragments were then ligated to Smal-digested and phosphatase-treated pUC18. Plasmid DNA was prepared by alkaline-SDS lysis followed by precipitation with polyethylene glycol 8000 as described by Sambrook et al. (18).

DNA Sequencing—DNA sequencing was performed directly on isolated recombinant clones using the Sequenase chain termination method of Sanger et al. (19) provided with the Sequenase kit from U. S. Biochemical Corp. The universal and reverse primers were used to obtain the sequence of each strand beginning at the linker site and these sequences were then extended using 18-24-mer Oligonucleotides synthesized to the unambiguous sequence determined in the preceding run and from the published sequence of the cDNA for the placental FBP (13). Both strands of introns 3 and 4, and the 5′-region of the gene were also sequenced. DNA sequences were analyzed using PC gene software supplied by IntelliGenetics, Mountainview, CA.

Polymerase Chain Reaction (PCR) and Reverse Transcription—PCR was determined to determine the size of introns in the cloned fragments of the gene and to amplify the 5′-region of the gene from normal human genomic DNA. PCR was also used to amplify fragments of putative upstream regulatory elements to insert into the pCAT vector constructs to analyze for promoter activity.

The size of the introns was determined using 18-25-mer sense and antisense amplimers to the exons flanking each intron (see Fig. 1). The PCR reaction mixture contained 200 ng of each dNTP, 50 pmol of each amplimer (sense, 5′-(-246)→3′(-222)); antisense, 5′(+42)→3′(-18)), 10 μl of 10 x Taq buffer (supplied by Perkin Elmer), 2 ng of DNA from the cloned gene fragments, in a total volume of 100 μl. The mixture was boiled for 2 min and then rapidly cooled in ice. Taq polymerase (0.5 μl) was added to the reaction mixture and this was overlaid with 50 μl of mineral oil. PCR was carried out (Perkin Elmer Thermal Cycler) with a denaturation cycle at 94 °C for 1 min, annealing cycle at 50 °C for 1 min, and extension at 72 °C for 4 min. This was repeated for 20 cycles with the last cycle having an additional 7-min extension time. For 18-mer amplimers (sense, 5′(-73)→3′(+91)); antisense, 5′(+244)→3′(-226)), annealing time was 2 min at 42 °C with no additional extension for the last cycle.

The 5′-region of the gene was amplified from genomic DNA using the same PCR conditions described in the preceding paragraph with the same 5′(-246)→3′(-222) sense amplimer but with a 5′(-58)→3′(-35) antisense amplimer upstream from the first intron. The PCR fragments were generated were processed and cloned as described in the section on plasmid cloning (see above).

Reverse transcriptase coupled PCR was carried out using a sense amplimer, 5′(-145)→3′(-125) and an antisense amplimer, 5′(-27)→3′(-49). First strand synthesis was carried out in a 20-μl reaction volume containing 80 pmol of antisense primer, 1 x Taq buffer, 1 m dNTPs, 40 units of RNAsin, and 100 units of Moloney murine leukemia virus reverse transcriptase. The reaction temperature was 37 °C for 30 min and then denatured at 95 °C for 5 min and cooled on ice. The reaction volume was adjusted to 100 μl with 1 x Taq buffer after adding 80 pmol of sense primer, and 2 units of Taq polymerase. PCR amplification cycles were 1 min at 94 °C and 3 min at 65 °C for 30 cycles. This was confirmed by Southern blot using a 24-mer internal probe, 5′(-67)→3′(-91).

Mapping of the Transcription Start Sites by RNase Protection.—The antisense RNA fragments for this assay encompassed the putative TATA box and a G-rich sequence selected by computer analysis as a GC box, 5′(-434)→3′(-181)), and a 191-nucleotide sequence extending from 5′(-216)→3′(-25)). These fragments were prepared by PCR using 24-mer sense and antisense primers to the termini of these regions and then subcloned into the transcription vector, pGEM Z, which contains the Sphl and T7 RNA polymerase promoters. Following digestion with Sphl to linearize the pGEM Z vector, the T7 transcript, BamHI for the plasmid containing the 191-bp fragment, the antisense and sense strands were labeled with [32P]UTP by in vitro transcription using the SP6 and T7 polymerases, respectively, with the kit from Promega (Madison, WI). The RFA kit from Ambion (Austin, TX) was used for the RNA protection assay with 8 μg of plasmid DNA and, for the negative control, 80 μg of URNA. The actin RNA antisense transcript as provided by the manufacturer was used with liver RNA as the positive control. The protected fragment(s) was separated by electrophoresis on a 6% urea-polyacrylamide gel for 90 min, the gel was then dried and exposed to Kodak X-AR5 film.

Identification of Transcriptional Regulatory Regions of the Gene—Three regions of the gene extending upstream from the ATG start codon of the first coding exon were analyzed for promoter activity (see legend of Fig. 4). These regions were prepared either by PCR amplification using the genomic clone as the template and 24-mer sense and antisense primers, or by digestion of the genomic clone with a restriction enzyme and the elution of the specific fragment following agarose gel electrophoresis.

Each fragment was blunted using DNA polymerase (Klenow fragment), purified by gel electrophoresis, and subcloned into the XmaI linker site upstream of the chloramphenicol acetyltransferase (CAT) reporter gene in the pCAT Basic plasmid (pCAT Basic, Promega). An aliquot (3 μl) of the ligation mixture was used to transform DH5α competent E. coli (Life Technologies Inc.). Several transformed colonies were selected from each plate by hybridization to the corresponding primer. Plasmid DNA was prepared and the orientation of each insert was established by DNA sequencing.

For transient expression of CAT activity, the transcription protocol described by Sambrook et al. (18) was used to co-transfect NIH 3T3 fibroblasts with the pCAT constructs and, as an internal control to monitor transfection efficiency, the plasmid containing the β-galactosidase gene was also cotransfected. The CAT reporter gene in the pCAT Basic plasmid (pCAT Basic, Promega). An aliquot (3 μl) of the ligation mixture was used to transform DH5α competent E. coli (Life Technologies Inc.). Several transformed colonies were selected from each plate by hybridization to the corresponding primer. Plasmid DNA was prepared and the orientation of each insert was established by DNA sequencing.

Transcriptional Regulatory Regions of the Gene—Three regions of the gene extending upstream from the ATG start codon of the first coding exon were analyzed for promoter activity (see legend of Fig. 4). These regions were prepared either by PCR amplification using the genomic clone as the template and 24-mer sense and anti-
100 mM β-mercaptoethanol, and 1.33 mg/ml O-nitrophenyl-β-D-galacto-
pyranoside). The samples were mixed and incubated at 37 °C for 30 min. The reactions were terminated by the addition of 500 μl of 1 mM Na2CO3 and the absorbance at 420 nm was determined. The protein concentration of the cytosol fractions was assayed using the BCA pro-
tein assay reagent supplied by Pierce Chemical Co. and the β-galacto-
sidase activity was determined by the absorbance of the released β-ni-
trate (5). The radioactivity (cpm/μg protein) comprising the sum of the mono- 
and diacylated chloramphenicol (CAT activity) was normalized to 0.1 A unit of 
β-galactosidase activity for that sample and the ratio of the CAT activity of the construct to the CAT activity of the pCAT Basic vector control was computed to be the fold stimulation.

Analysis of DNA-protein Binding by Gel-shift Mobility Assay—A sense oligonucleotide sequence, 5′-3′(-233)GAAGAGGTTGTTGTCG-
GAAGGGAAAGAGAAGGAAAGGAAATACG-3′(-186) and the complementry antisense sequence were prepared for the gel-shift mobi-
ity assays. This sequence contains the G-rich region (5′-
GGGTGGG-3′), which is a computer-selected putative SP1 binding 
motif, linked to the tandemly duplicated GAGA sequence, the consen-
sus motif for binding a number of related ets oncogene encoded nuclear 
proteins (16). A second sense of sense and antisense oligomers were pre-
pared to the sequence, 5′(-220) - 3′(-186) so that this probe will 
contain only the tandemly repeated GAGA GAAG sequences. These oligomers were gel 
and the antisense strand was end-labeled with [γ-32P]ATP and 
buffer (25 mM Tris-HCl, pH 8.0, 3 mM MgCl₂, 0.5 mM EDTA, 0.5 mM dithio-
tritol, 10% glycerol). The oligomers were annealed to the bound probes and the 
probe was incubated at room temperature for 30 min in binding 
buffer (25 mM Tris-HCl, pH 8.0, 1.5 mM dithio-
tritol, 25 μg/ml bovine serum albumin, 10% glycerol) for 
30 min at 4 °C. The GABP binding reaction was carried out in the same buffer with 
the addition of 10 ng of the GABP, subunit with or without 10 ng of the 
GABP, antisense primer extending from 5′(-85) to 3′(-190), and which mutants four GA dinucleotides to CT dinucleotides in the motif which is the putative GABP binding sequence. This double 
stranded oligonucleotide was used for concentration studies in the gel-shift 

RESULTS

Identification and Isolation of the Gene Encoding a Placental 
FBP—A total of 29 clones in the λChiron 4A genomic library hy-
bridized to the KB cell FBP cDNA. The clones were then 
distinguished by separate digestions of the DNA in order to 
determine the size of the insert in each clone (EcoRI) and to 
identify independent isolates on the basis of a difference in the 
fragments generated (BamHI). Southern blotting with the 
KB cell FBP cDNA as the probe was used to identify the homolo-
gene fragments. The EcoRI digest of 4 clones generated 
seven fragments of which only one, a 10-kb fragment, hybrid-
ized to the KB cell FBP cDNA. This 10-kb fragment was sub-
cluded into pUC18 for sequencing.

Nucleotide Sequence Analysis of this Gene—Fig. 1 shows the nucleotide sequence of the gene with coding exons 
aligned with placental FBP cDNA (13), and the deduced amino 
acid sequence. The gene has five exons (129, 174, 189, 136, 
and 483 bp(s) in length) interrupted by four introns (~1200, ~2200, 
117, and 158 bp(s) in length) and spans at least 5 kb. The exons 
of this gene have greater than 95% homology with the placental 
FBP cDNA as reported by Ratnam et al. (13) including a Kozak 
consensus sequence flanking the putative translation initiation 
site (22). There are, however, a number of important differ-
ences. First, the 3′-terminal nucleotide (~25, shaded) of exon 1 
is a G instead of a C contained in the placental FBP cDNA and 
this provides an AG/gc sequence which may be an intron donor 
splice site. Second, exon 4 also differs from this placental FBP 
cDNA, containing C instead of A at nucleotide 357 (shaded). 
Third, 144 nucleotides of the 5′-untranslated region of the gene 
beginning at nucleotide ~118 is completely divergent from the 
corresponding region of placental FBP cDNA (shaded region). 
The 3′-untranslated region of this gene contains the polyade-
nylation signal sequence which begins 164 nucleotides 
following the stop codon and 16 nucleotides upstream from the 
begning of the poly(A) sequence reported for the cDNA (13). 

Except for this divergence in the nucleotide sequence of this 
5′-region of the gene from the corresponding untranslated 
region of the placental FBP cDNA, the evidence that this gene encodes the placental FBP is compelling. Our first consider-
ation to explain the observed divergence was that the discrep-
ancy may be the result of a cloning artifact occurring in the 
preparation of either the cDNA or the genomic libraries. To 
resolve this question we used a number of strategies with PCR 
and reverse transcriptase-PCR to amplify this region of 
genomic DNA. For the first strategy, we prepared an antisense 
primer common to this placental FBP gene and the placental 
FBP cDNA sequences, and a sense primer complementary to 
the unique 5′-divergent region of the gene. For the second strat-
egy, we used the same antisense primer but with a sense primer 
unique to the 5′-region of the cDNA that was divergent from 
this gene. A 213-bp fragment was amplified using the first 
strategy with the sense primer to the unique 5′-extremity of the 
gene which was subcloned into pCRII and the sequence ob-
gained was identical to this divergent region of the cloned gene 
given in Fig. 1. With the second strategy, no identifiable frag-
ment was obtained using the sense primer complementary to 
the unique sequence of this region of the placental FBP cDNA.

For the third strategy, we used reverse transcriptase-PCR to 
generate the first strand cDNA from placental RNA with an 
antisense primer extending from 5′(-27) to 3′(-49) and contin-

2 The GA-binding protein subunits, GABPα and GABPβ, and the polyclonal antisera to each protein were kindly provided by Dr. Steven McKnight, Tuliarik, Inc., South San Francisco, CA.
FIG. 1. The nucleotide sequence of the FBP/PL-1 gene aligned with the placental FBP cDNA and the deduced amino acids of the coding regions. The shaded areas indicate the divergence of the genomic and cDNA nucleotide sequences. Line 1 represents the sequence of the genomic clone; line 2 is the cDNA sequence published by Ratnam et al. (13); and line 3 is the deduced amino acid sequence in single letter abbreviations. The numbers above the nucleotide sequence begin at the first base (A) of the methionine codon for the start of the signal peptide and are positive and negative in the 3' and 5' directions, respectively. The numbers in the right margin are the sum of amino acids from the initial methionine to the end of the corresponding line. A sequence containing a putative TATA box and a G-rich sequence (GR) containing a computer selected cap sequence. The arrowhead over base -153 is the approximate transcription start site determined by RNase protection. The AATAAA polyadenylation sequence is indicated by the box enclosure (15'(927) - 3' (933)). The terminal 3' AATAAA sequence is followed by the poly(A) sequence in the placental FBP cDNA (12).
ued the amplification with the addition of the sense primer extending from 5'(-148) → 3'(-125) in the divergent region of the gene. This generated a 121-nucleotide fragment (Fig. 2) that was predicted from the location of the amplimers indicating that this “divergent” region is in the transcribed RNA. No transcriptase was omitted. Southern blotting analysis of the product was obtained in a control reaction in which reverse transcriptase-PCR of placental RNA was also carried out using the product was obtained in a control reaction in which reverse transcriptase was omitted. Southern blotting analysis of the same site obtained with the RNase protection assay using the 316-nucleotide probe. This result was confirmed by replica analyses.

Identification of Putative Transcriptional Promoter Elements —Selected domains in the 5‘-region of the gene upstream from the ATG start site were analyzed for promoter activity by subcloning each fragment into a pCAT Basic plasmid containing the chloramphenicol acetyltransferase reporter gene for transient expression in NIH 3T3 mouse fibroblasts. The results obtained with the region 5'(-341) → 3'(-118) containing the putative GC box and the tandem repeated ets motifs are shown in Fig. 4. There was a 6-fold stimulation of acetylation of chloramphenicol by the 3T3 cells transfected with this construct (lanes 3 and 4) over the basal activity obtained with the pCAT Basic vector control (lanes 1 and 2) or the pCAT Basic antisense construct (lanes 5 and 6). No stimulation of CAT activity was observed with the putative TATA box alone, 5'(-434) → 3'(-299), or with the putative TATA box linked to the 5‘-flanking 2.2-kb fragment, (5'(-2.2 kb) → 3'(-469)).

Gel-shift Mobility Assay to Identify DNA-binding Proteins—A mobility gel-shift was obtained (Fig. 5A) with the 32P-labeled oligomer, 5‘(-233) → 3‘(-186), containing the G-rich-GGAAG tandem duplicated motif and purified SP1 protein (lane 2) and this shift was partially competed by the unlabeled G-rich-GGAAG oligomer (lane 3) and completely competed out by the SP1 specific oligomer (lane 4) indicating that this G-rich sequence contains a GC box even though it lacks the GGGCGG hexanucleotide that is the classical SP1 binding motif (23). When the same radiolabeled probe was incubated with the nuclear extract (Fig. 5B), four distinct retarded bands were observed (indicated by the arrowhead, open and closed triangles and open arrowhead in lanes 4, 6, and 8). Two of these retarded bands (arrowhead and open triangle) were competed out by the SP1 specific oligonucleotide (lanes 5, 7, and 8) indicating that this G-rich motif binds the SP1 protein contained in the nuclear extracts. A supershift assay using an antiserum to the human SP1 transcription factor that cross-reacts with mouse SP1 (Fig. 5C, lane 4) provides additional evidence that the component of the nuclear extract binding the probe is the SP1 protein. The mobility shift assay was also used to establish whether the tandem linked duplicated GGAAG ets motif in the 5‘-flanking region bind specifically to the GA-binding protein (GABP) (both the GABPα and GABPβ subunits) and whether a nuclear extract prepared from 3T3 cell nuclei contains similar DNA-binding proteins. For this experiment a 35-bp fragment was

### Figure 2

**Reverse transcriptase-PCR to characterize the 5′-region of the gene divergent from the cDNA.** An antisense primer (5’(-27) → 3’(-49)) was used to generate the first strand copy of the mRNA and amplification was continued with a sense primer (5’(-148) → 3’(-125)) in the divergent sequence. The left lane shows the HaeIII digest of the dX174 DNA marker. The right lane shows the reverse transcriptase-PCR of placental RNA. The 121-nucleotide reverse transcriptase-PCR product is indicated by the asterisk and corresponds to the 118-nucleotide dX174 HaeIII DNA marker. The size of the product is that predicted from the location of the primers.
prepared which included only the GGAAG sequences (5′(-220) → 3′(-186)). The autoradiograph in Fig. 6A shows two retarded bands (arrows 1 and 2, lane 2) following incubation with the GABPα subunit and these shifted bands were competed by the unlabeled oligonucleotide containing only the GGAAG sequences (lane 3).

When the probe was incubated with both the GABPα and GABPβ subunits, the retardation of the probe differs. A more slowly migrating band (Fig. 6A, band 3, lane 4) appears with the persistence of band 2, but band 1 that was observed in lane 2 is now missing. Competition with the unlabeled GGAAG containing oligonucleotide markedly reduces the intensity of bands 2 and 3 (lane 5). A similar retardation of two complexes was observed by Thompson et al. (24) in the gel-shift assay using recombinant GABPα and a DNA substrate derived from the enhancer of an immediate early gene of herpes simplex virus, and the GABPβ1 subunit reacted with the GABP, on both complexes.

The same probe (Fig. 6A) was incubated with the 3T3 nuclear extract and shows the mobility shift of a fragment (lanes 7 and 9, arrow 4) which is competed out by the GGAAG containing oligonucleotide (lanes 8 and 10). Fig. 6B is a longer exposure of the gel and shows a doublet (lanes 7 and 9, arrow 5) that is seen as lighter bands in Fig. 6A and which is also competed out (lanes 8 and 10) by the GGAAG containing oligomer. We do not know whether this doublet is due to some additional factor in this crude nuclear extract which is interacting with the putative primary GABP transcription factor or whether it is an artifact (although it was observed in replicate assays) but it is evident by the competition studies that the GGAAG repeat sequences compete out both components of this doublet.

A mobility shift of the longer probe containing both the G-rich and GGAAG motifs (5′(-233) → 3′(-186)) with the nuclear extract is seen in Fig. 6C and shows that just the lower band (arrow) is competed out by the unlabeled oligomer containing only the tandem linked GGAAG pentamers. The upper band (arrowhead), that is not competed out with this oligomer, is the analogous band in Fig. 5B which is the mobility shift due to the SP1 protein in the nuclear extract binding to the G-rich sequence. With this longer probe we did not identify the doublet shown in Fig. 6B (band 5) that appears to be a GGAAG motif for the GABPβ by the competition studies. Fig. 6D shows the supershift of the 35-bp probe (5′(-220) → 3′(-186)) with the nuclear extract when the specific antisera for the GABPα (lane 4) and the GABPβ (lane 5) were added to the reactions. A control normal rabbit serum (lane 3) retarded a smaller component of the probe (arrowhead) and this shift is distinctly different than observed with the antisera. The lighter intensity of the shifted probe with the nuclear extract alone (lane 2) is likely to be due to some technical artifact in the reaction mixture or the
Human Folate-binding Protein Gene

**FIG. 5. Gel-shift mobility and supershift assays to identify DNA binding by SP1 and 3T3 cell nuclear extract.**

A: lane 1, $^{32}$P-labeled probe, 5'(−233) → 3'(−186) which includes the G-rich sequence linked to the tandemly repeated GGAAG motif; lane 2, the probe incubated with 1 footprinting unit of purified SP1 protein (Promega); lanes 3 and 4, contain, respectively, 25 ng of the unlabeled G-rich-GGAAG containing fragment and 25 ng of SP1 binding oligonucleotide (Promega).

B: lanes 1 and 2, the $^{32}$P-labeled probe as in A; lane 3, the $^{32}$P-labeled probe plus nuclear extract (1 μg of protein); lane 4, the same $^{32}$P-labeled probe plus 2 μg of poly(dI-dC) and 3.5 μg of nuclear extract protein; lane 5, the same as lane 4 plus 25 ng of SP1 binding oligonucleotide; lane 6, $^{32}$P-labeled probe, 2 μg of poly(dI-dC), 7 μg of nuclear extract protein; lane 7, same as lane 6 plus 25 ng of SP1 binding oligonucleotide; lane 8, same as lane 6 but containing 14 μg of nuclear extract protein; lane 9, same as lane 8 plus 25 ng of the SP1 binding oligonucleotide.

C: lane 1, same $^{32}$P-labeled probe as in A and B; lane 2 contains this probe plus 14 μg of nuclear extract protein; lane 3, same as lane 2 plus 1.5 μl of normal rabbit serum; lane 4, same as lane 2 plus 1.5 μl of polyclonal rabbit anti-human SP1 antiserum.

Electrophoresis because the shift of the same probe with the control non-immune serum (lane 2) is more intense.

Fig. 6E shows the mobility gel-shift of the shorter $^{32}$P-labeled GGAAG oligonucleotide (i.e. lacking the G-rich sequence obtained with the nuclear extract (lane 2)) and this is again competed out by the unlabeled GGAAG oligomer (lane 3). This mobility shift, however, was not competed out by the oligonucleotide in which four GA dinucleotides of the GGAAG pentamers were mutated to CT dinucleotides (lane 4).

It is of interest that the purified GABP$_{GABPD1}$ complex and the 3T3 nuclear extract differ in the retardation of the GGAAG containing oligonucleotide. This could be the result of an intrinsic property of the ets-encoded GABP-related proteins expressed in 3T3 cells which differs from the purified GABP/GABP$_{GABPD1}$ subunits, or it could be secondary to some alteration of the GA-binding protein(s) during the preparation of the nuclear extract. The competition observed with competing unlabeled GGAAG containing oligonucleotide, however, and the failure of the mutant oligomer to compete, establishes the specificity of the interaction of a component of the nuclear extract with this ets binding motif.

The gel-shift mobility assay (Fig. 7) was also used to estab-
Fig. 6. Gel-shift mobility assay and supershift assays to identify DNA binding by GABP/GABP in the 3T3 cell nuclear extract. 
For this assay a 35-bp 32P-labeled probe, 5'(-220) → 3'(-186), was prepared. A: lane 1, probe alone; lane 2, probe plus 10 ng of GABP; lane 3, same as lane 2 plus the addition of 50 ng of unlabeled oligonucleotide sequence; lane 4, same as lane 2 plus the addition of 10 ng of GABP subunit; lane 5, same as lane 4 plus the addition of 50 ng of the unlabeled oligonucleotide sequence. For lanes 6-10, the same 32P-labeled oligonucleotide probe was incubated with the 3T3 cell nuclear extract. Lane 6, 32P-labeled probe plus 7 μg of nuclear extract protein; lane 7, same as lane 6 plus 2 μg of poly(dl-dC), and 3.5 μg of nuclear extract protein; lane 8, same as lane 7 plus 50 ng of unlabeled oligonucleotide; lane 9, same as lane 6 plus 2 μg of poly(dl-dC); lane 10, same as lane 9 plus 50 ng of unlabeled oligonucleotide. B, same as A but a longer exposure of the gel. C, mobility shift of the longer probe, 5'(-233) → 3'(-186), with the 3T3 cell nuclear extract. Lane 1, the 32P-labeled probe and 3.5 μg of nuclear extract protein; lane 2 same as lane 1 plus 25 ng of the GGAAG containing oligonucleotide; lane 3 same probe plus 7 μg of nuclear extract protein; lane 4, same as lane 3 plus 25 ng of the GGAAG containing oligomer. D, supershift of the shorter probe (5'(-220) → 3'(-186)) incubated with the nuclear extract and the antisera to the GABP protein. Lane 1, 32P-labeled probe alone; lane 2, probe plus 0.8 μg of nuclear extract protein; lane 3, same as lane 2 plus 1.5 μl of normal rabbit serum; lanes 4 and 5, same as lane 2 plus 1.5 μl of antiserum to GABP, and GABP, respectively. E, gel-shift mobility assay of the same probe as in B with the competing GGAAG containing oligomer and the oligomer containing the GA dinucleotides mutated to the CT dinucleotides. Lane 1, probe alone; lane 2, probe plus 14 μg of nuclear extract protein; lane 3, same as lane 2 plus 6.5 ng of GGAAG containing oligomer; lane 4, same as lane 2 plus 6.5 ng of the GA → CT mutated oligomer.

lish the validity of the computer selection of the putative TATA box as the binding site for the TATA protein (TFIID) even though this sequence was not located in the usual site 25-30-bp upstream from the transcription start site (23). This TATA box containing the sequence 5'(-434) → 3'(-299) (lane 1) has been shifted to the application site of the gel electrophoresis (lane 2) following incubation with the TFIID protein and this shift has been blocked by the competing unlabeled TATA sequence (lanes 3-5).

Organization of this Gene Encoding this Placental FBP—
This placental FBP gene has been assigned the FBP/PL-1 gene because it encodes one of the FBPs first identified in human
cubated with 32P-labeled TATA sequence by WID (TATA-binding protein).

identical to the organization of the gene encoding the FBP in a FBP expressed in KJ3 cells as well as its related pseudogene (15). Exon 2 is intron of the FBP/KB gene. Moreover, the introns are inter-

KJ3 cell FBP (11, 12) and corresponding cDNA (11, 12) is shown in Fig. 9. Those sequences deduced from exon 2 and 4 have the highest identity (77%) compared to those sequences deduced from exon 2 (62%) and exon 5 (73%). There are a number of clustered amino acid sequences having a positive charge that are common to the KB cell FBP (11, 12), the placenta FBP (13), the human milk FBP (26), and the bovine milk FBP (27). These include (from Fig. 1) -K35-H36-H37-K38-, K91-R92-H93-, and -R119-K120-E121, R122; such charged clusters may be the site of ligand binding as suggested by Svendsen et al. (27).

**DISCUSSION**

There are two FBP(s) expressed in human placenta. One is identical to the FBP in KB cells and the cDNA encoding this FBP has been cloned from a library prepared from KB cell (11) and placental RNA (12). Ratnam et al. (13) purified a second FBP from human placenta and cloned the cDNA encoding this protein which proved to be distinct from the cDNA encoding the GPI-FBP. Page et al. (14) have recently reported the cloning of a similar cDNA and the corresponding gene which, however, lacked details of the structural and functional organization of the 5'-flanking promoter region. The proximal upstream region of this FBP/PL-1 gene is of considerable interest for a number of reasons. First, an SP1 binding region that is contained within a 13-nucleotide G-rich sequence, approximately 78 bp upstream from transcription initiation, is not a classical GC box in which the hexamer, GGCGCG, is the SP1 binding motif (23). However, the gel-shift mobility assay using purified SP1 protein and competing SP1 binding oligonucleotide, and the supershift assay with anti-SP1 antisera has established that this sequence, with T substituting for the C in the hexamer, is a motif which binds the SP1 transcription factor. Second, this G-rich sequence is fused to a tandemly duplicated GGAAG pentamer which is the binding motif for a number of related ets oncogene-encoded nuclear-binding proteins that are believed to be transactive transcription factors (16). The gel-shift mobility assay using purified GABPα and GABPβ and the specificity established by competition with the unlabeled oligomer containing this tandem duplicated GGAAG pentamer is evidence that this motif is the sequence to which the ets-encoded proteins bind. A similar gel-shift observed with the nuclear extract from 3T3 cells that was competed out with the unlabeled oligonucleotide containing these GGAAG sequences but not with an oligomer in which the GA dinucleotides were mutated to CT oligonucleotides, provides additional evidence that these pentameric motifs are the binding site(s) for the ets related proteins.

A similar organization of the SP1 binding sequences with the downstream tandemly duplicated GGAAG ets motifs has recently been identified by Carter et al. (28) as a basal transcription promoter element in the nuclear-encoded cytochrome oxidase subunit IV (COIV) gene. More recently, Virbasius et al. (29) have purified and sequenced from HeLa cells nuclear respiratory factor 2 which is involved in the transcriptional regulation of the proximal promoter of this COIV gene. Nuclear respiratory factor 2 is a multimeric protein comprised of 5 subunits with amino acid homology to GABPα, GABPβ1, and GABPβ2 subunits (24). Virbasius et al. (29) have also proposed that nuclear respiratory factor 2 is the human homolog of the mouse GABP and, therefore, is involved in regulating the expression of cellular genes. The similar organization of the SP1-GGAAG tandem repeats in the proximal promoter region of the FBP/PL-1 gene and the promoter activity of this region in driving expression of the CAT reporter gene suggests that this domain may also regulate expression of this placental FBP.

These G-rich and tandemly linked GGAAG cis elements provide a mechanism for tissue specific expression of the FBP(s). The GABPα and GABPβ1 subunits of the GA-binding protein complex form a heteromeric α2β2 tetramer which binds to the GGAAG motif for most efficient induction of transcription (30). Whereas the GABPα binds to the GGAAG sequence in the absence of the GABPβ1 (or GABPβ2) subunit, GABPβ1 (and GABPβ2) does not bind to the DNA directly but rather, the α2β2...
pressed in human placenta and thus, only tissues and cells which express both subunits of this FBPPL1 gene is the location of a TATA box between -360 and -380, well upstream from the G-rich SP1 binding sequence at -213. Since this was not observed, we concluded that this TATA sequence is not a regulatory element for this gene. However, we could not confirm this transcription start site by repeated RNase protection assays and we, therefore, concluded that this TATA box promoter and/or tissue specificity. The evolutionary modification of the ancestral FBP gene to a TATA-less "housekeeping" gene (31) for low level constitutive expression of the placental FBP could be an appropriate adaptation to the development of the placenta in mammalian reproduction since this would ensure a mechanism to transfer folate from maternal to fetal circulation. Ragoussis et al. (32) have located the family of FBP genes in a tandem arrangement within a 140-kb region on chromosome 11 (q13.3-q13.5) and this organization could be more permissive for such a regulatory evolutional adaptation following duplication of the ancestral FBP gene so that different forms of the mammalian FBPs may have specialized functions in folate metabolism.

A puzzling finding initially was the divergence of the nucleotide sequence of the 5'-region upstream of the first exon in the FBP/PL-1 gene from the 5'-untranslated region of the placental FBP cDNA as reported by Ratnam et al. (13). A number of subsequent studies, however, have clearly established that the 5'-untranslated sequence in the cDNA differed from this corresponding region of the gene as a consequence of a cloning artifact. First, PCR using a sense amplimer to a divergent sequence of the FBP/PL-1 gene and an antisense amplimer to a sequence common to the gene and the cDNA, amplified a fragment from the genomic DNA precisely the size predicted from the distance between the two amplimers (213 bp) and it has the same nucleotide sequence as this divergent region of the gene. Second, the RNase protection assay to locate the transcription start site used a probe extending from 5'-117 to 3'-25 which includes both the sequences in the gene in common and divergent from the cDNA cloned by Ratnam et al. (13) and we located the cap site within the divergent region of the gene at base -153. If there was another species of mRNA encoded by the sequence common to the gene and the cDNA, amplified a fragment from genomic DNA precisely the size predicted from the distance between the two amplimers (1,213 bp) and it has the same nucleotide sequence as this divergent region of the gene. Second, the RNase protection assay to locate the transcription start site used a probe extending from 5'-117 to 3'-25 which includes both the sequences in the gene in common and divergent from the cDNA cloned by Ratnam et al. (13) and we located the cap site within the divergent region of the gene at base -153. If there was another species of mRNA encoded by the sequence common to the gene and the cDNA, amplified a fragment from genomic DNA precisely the size predicted from the distance between the two amplimers (1,213 bp) and it has the same nucleotide sequence as this divergent region of the gene.

The primary sequence of the 5'-region upstream of the first exon in the FBP/PL-1 gene shows some similarities to the primary sequence of the FBP gene and an antisense amplimer to a sequence common to the gene and the cDNA, amplified a fragment from genomic DNA precisely the size predicted from the distance between the two amplimers (1,213 bp) and it has the same nucleotide sequence as this divergent region of the gene.

Another interesting organization of the 5'-flanking region of this FBP/PL1 gene is the location of a TATA box between -360 and -380, well upstream from the G-rich SP1 binding sequence. In our initial studies of this gene,1 primer extension using a primer to the region 5'-190 to 3'-213 and RNA from placenta identified a transcription initiation site in a putative cap consensus sequence 33 bp downstream from this TATA box. However, we could not confirm this transcription start site by repeated RNase protection assays and we, therefore, concluded that this TATA sequence is not a regulatory element for this FBP/PL-1 gene. Since the multigene family encoding the human FBP(s) has arisen by duplication of an ancestral gene (see below) (15), evolutionary divergence that follows such duplication could modify regulatory elements that provide high expression (i.e. TATA box promoter) and/or tissue specificity. The evolutionary modification of the ancestral FBP gene to a TATA-less "housekeeping" gene (31) for low level constitutive expression of the placental FBP could be an appropriate adaptation to the development of the placenta in mammalian reproduction since this would ensure a mechanism to transfer folate from maternal to fetal circulation. Ragoussis et al. (32) have located the family of FBP genes in a tandem arrangement within a 140-kb region on chromosome 11 (q13.3-q13.5) and this organization could be more permissive for such a regulatory evolutional adaptation following duplication of the ancestral FBP gene so that different forms of the mammalian FBPs may have specialized functions in folate metabolism.

A puzzling finding initially was the divergence of the nucleotide sequence of the 5'-region upstream of the first exon in the FBP/PL-1 gene from the 5'-untranslated region of the placental FBP cDNA as reported by Ratnam et al. (13). A number of subsequent studies, however, have clearly established that the 5'-untranslated sequence in the cDNA differed from this corresponding region of the gene as a consequence of a cloning artifact. First, PCR using a sense amplimer to a divergent sequence of the FBP/PL-1 gene and an antisense amplimer to a sequence common to the gene and the cDNA, amplified a fragment from genomic DNA precisely the size predicted from the distance between the two amplimers (213 bp) and it has the same nucleotide sequence as this divergent region of the gene. Second, the RNase protection assay to locate the transcription start site used a probe extending from 5'-117 to 3'-25 which includes both the sequences in the gene in common and divergent from the cDNA cloned by Ratnam et al. (13) and we located the cap site within the divergent region of the gene at base -153. If there was another species of mRNA encoded by the sequence common to the gene and the cDNA, amplified a fragment from genomic DNA precisely the size predicted from the distance between the two amplimers (1,213 bp) and it has the same nucleotide sequence as this divergent region of the gene.

The primary sequence of the 5'-region upstream of the first exon in the FBP/PL-1 gene shows some similarities to the primary sequence of the FBP gene and an antisense amplimer to a sequence common to the gene and the cDNA, amplified a fragment from genomic DNA precisely the size predicted from the distance between the two amplimers (1,213 bp) and it has the same nucleotide sequence as this divergent region of the gene.
molecule. Moreover, there are 9 cysteine residues (positions 83, 90, 99, 103, 123, 130, 146, and 160) that encompass this cluster and which can form stabilizing disulfide bridges that are essential for ligand binding function (33).

The GPI-FBP isolated from KB cells is one of a recently described class of proteins that is anchored to the external surface of the cell membrane by a glycosylphosphatidylinositol tail (34). This GPI tail is attached to the cell membrane by a transmembrane stretch and would, therefore, not be identified by this method. The high degree of homology between exons 3 and 4 and the amino acid sequence encoded by the FBPPL-1 gene; i.e. histidine (228), valine (229), and asparagine (230) replace serine, glycine, and alanine, respectively, in the KB cell FBP and these amino acids have not been reported to be the carboxyl terminus of GPI-tail proteins (35). Since there is little overall amino acid sequence homology in this carboxyl-terminal region of the KB cell FBP cDNA and this FBPPL-1 gene, it may very well be anchored to the cell membrane by a transmembrane stretch of hydrophobic amino acids having the α-helical configuration. In fact, Antony and co-workers (36) have previously shown that a FBP lacking a GPI tail would not incorporate precursor substrates into this structure and would, therefore, not be identified by this method.

The high degree of homology between exons 3 and 4 and the type of intron/exon junctions between all the defined introns of this FBP/PL-1 gene and the FBP/KB genes indicate that these genes are derived by duplication of an ancestral gene and they have remained on chromosome 11 (q13.3-q13.5) during evolution (32). The nucleotide divergence between the FBP/PL-1 and FBP/KB genes suggests that the duplication occurred about 350 million years ago, based on the estimate of 10⁶ years for each 0.17% divergence (38).

Finally, the fact that two different molecular forms of functionally similar proteins are found in human placenta (13) suggests that tissue-specific factors are likely to be involved in the expression of the genes encoding these FBPs. Placenta contains tissue components derived from both maternal and fetal origin and it will be of interest to establish the source of the GABPα₁ (or GABPα₂) subunit of the GABP heteromeric complex that may be necessary for constitutive expression of the FBP/PL-1 gene.

Acknowledgments—We express our appreciation to Annette Regec and Bijoyesh Mookerjee for preparing the oligonucleotides; to Anne Erickson for the computer generated graphics; and to Mary Ann Sacchoccio for secretarial assistance.

REFERENCES

1. Antony, A. C. (1992) Blood 79, 2897–2830
2. Antony, A. C., Uley, C., VanHorne, R. C., and Kolhouse, J. F. (1981) J. Biol. Chem. 256, 9684–9692
3. Antony, A. C., Uley, C. S., Marcell, P. D., and Kolhouse, J. F. (1982) J. Biol. Chem. 257, 10081–10089
4. Elwood, P., Kane, M. A., Portillo, R. M., and Kolhouse, J. F. (1986) J. Biol. Chem. 261, 15416–15425
5. Lubins, C. A., Sadasivan, E., da Costa, M., and Rothenberg, S. P. (1986) Arch. Biochem. Biophys. 250, 94–105
6. Luhrs, C. A., and Slonimsky, B. L. (1989) J. Biol. Chem. 264, 21446–21449
7. Lacey, S. W., Sanders, J. M., Rothenberg, K. G., Anderson, R. G. W., and Kamien, B. A. (1989) J. Clin. Invest. 84, 715–720
8. Rothenberg, S. P., and da Costa, M. (1971) J. Clin. Invest. 50, 719–726
9. Rothenberg, S. P., and da Costa, M. (1978) Clin. Hematol. 5, 569–567
10. Sadasivan, E., Rothenberg, S. P., da Costa, M., and Brink, L. (1987) Biochem. Biophys. Acta 926, 36–47
11. Sadasivan, E., and Rothenberg, S. P. (1989) J. Biol. Chem. 264, 5806–5811; Correction (1990) J. Biol. Chem. 265, 1821
12. Elwood, P. C. (1989) J. Biol. Chem. 264, 14983–14991
13. Ratnam, M., Marquardt, H., Duhring, T., and Freisheim, J. H. (1989) Biochemistry 28, 8549–8554
14. Page, S. T., Owen, W. C., Price, K., and Elwood, P. C. (1993) J. Mol. Biol. 229, 1175–1182
15. Dignam, J. D., Lebowitz, R. M., and Reeder, R. G. (1983) Nucleic Acids Res. 11, 1475–1489
16. Kozak, M. (1989) J. Cell Biol. 108, 229–241
17. Dynan, W. S., and Tjian, R. (1985) Nature 316, 774–778
18. Thompson, C. C., Brown, T. A., and McKnight, S. L. (1991) Science 253, 762–768
19. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Sailer, S., and Sharp, P. A. (1988) Annu. Rev. Biochem. 55, 1119–1150
20. Svensson, I., Hansen, S. I., Holm, J., and Lyngbye, J. (1988) Carlsberg Res. Commun. 47, 371–376
21. Svensson, I., Hansen, S. I., Holm, J., and Lyngbye, J. (1984) Carlsberg Res. Commun. 49, 125–131
22. Carter, B. S., Bhut, N. K., Basu, A., and Avadhani, N. G. (1992) J. Biol. Chem. 267, 23418–23426
23. Vierbosi, J. V., Vierbosi, C. A., and Scarpulla, R. C. (1993) Genes & Dev. 7, 380–392
24. Brown, T. A., and McKnight, S. L. (1992) Genes & Dev. 6, 2502–2512
25. Wasylyk, B. (1988) Biochem. Biophys. Acta 961, 17–35
26. Ragona, J., Senger, G., Trowsdale, J., and Campbell, I. G. (1992) Genomics 14, 423–430
27. Lubins, C. A. (1991) Blood 77, 1177–1180
28. Ferguson, M. A. J., and Williams, A. F. (1988) Annu. Rev. Biochem. 57, 295–320
29. Udrescu, S., Micanovic, R., and Kodukula, K. (1991) Cell Biol. Int. Rep. 15, 735–739
30. Antony, A. C., Verma, R. S., Unane, A. R., and Luhrs, C. A. (1989) J. Biol. Chem. 264, 1911–1914
31. Verma, R. S., Guillapalli, S., and Antony, A. (1992) J. Biol. Chem. 267, 4119–4126
32. Sabina, R. L., Morinaki, T., Clarke, P., Eddy, R., Shows, T. B., Morton, C. C., and Holmes, E. W. (1990) J. Biol. Chem. 265, 9423–9433