The Divergent 5′ Termini of the α Human Folate Receptor (hFR) mRNAs Originate from Two Tissue-Specific Promoters and Alternative Splicing: Characterization of the α hFR Gene Structure‡

Patrick C. Elwood,*,§ Kirsten Nachmanoff,§ Yutaka Saikawa,|| Stephanie T. Page,⊥ Paulo Pacheco,▽ Susan Roberts,§ and Koong-Nah Chung§

Section of Experimental Hematology, Medicine Branch, National Cancer Institute, Bethesda, Maryland 20892, Kanazawa University, Kanazawa, Japan, University of Washington Medical School, Seattle, Washington 98105, and Brown University Medical School, Box G 8045, Providence, Rhode Island 02912

Received August 16, 1996; Revised Manuscript Received November 22, 1996⊗

ABSTRACT: The human KB cell or α folate receptor (α hFR) is a membrane glycoprotein of 42 kDa that participates in the internalization of folates and antifolates. Seven independent α hFR cDNA isoforms have been reported that contain unique 5′ termini but share a common open reading frame (ORF). To investigate the molecular basis of these heterogeneous 5′ sequences, we determined the sequence of the α hFR gene from two clones isolated from a human lymphocyte λ DASH genomic library. The gene is composed of seven exons that span 6.8 kb. The ORF is encoded by exons 4 through 7 while the reported 5′ termini of the cDNA isoforms (including two novel cDNAs designated KB2 and KB4) are encoded by exons 1 through 4. Using RNase protection assays, we demonstrate that transcripts corresponding to the KB1 and KB4 cDNAs originate from promoters upstream from exon 1 and exon 4, designated P1 and P4, respectively, and that these mRNA isoforms are the most abundant transcripts expressed in KB cells and selected normal tissues (including kidney, lung, and cerebellum). We observed a heterogeneous start site within exon 1 from the P1 promoter while transcripts from the P4 promoter originate from a single site. In addition, we detected tissue specificity for the P1 and P4 promoter utilization. Transcripts originating from the P1 promoter are the most abundant transcripts expressed by human cerebellum and kidney. In contrast, transcripts from the P4 promoter are the most abundant transcripts expressed by human KB cells and lung. Total RNA from KB cells also protects a 66 bp fragment of an exon 3 riboprobe that is consistent with an alternatively spliced transcript. To examine the functional activity of the predicted P1 and P4 promoters, α hFR promoter-CAT chimeric plasmids were constructed using sequences flanking exon 1 and exon 4. We observed a 7.5- and 10-fold increase in CAT activity in HeLa cells transiently transfected with the P1 and P4 promoter constructs, respectively. These data demonstrate that a single gene encodes the divergent 5′ termini of the α hFR cDNAs and that the α hFR transcripts are transcribed from two promoters that are activated in a tissue-specific manner.

The human α folate receptor (α hFR)[1] is a member of the hFR gene family that includes the α, *β*, and *γ* isoforms (Elwood, 1989; Sadasivan & Rothenberg, 1989; Lacey et al., 1989; Coney et al., 1991; Campbell et al., 1991; Ratnam et al., 1989; Shen et al., 1994). These proteins are single-chained glycoproteins that range in size from 32 to 42 kDa and that contain a single high-affinity binding site for folic acid and other folates. The primary structures of the α, *β*, and *γ* hFR have been deduced from cDNAs isolated from human KB cell (Elwood, 1989; Sadasivan & Rothenberg, 1989), placental (Ratnam et al., 1989), and CML cell (Shen et al., 1994) cDNA libraries, respectively. The α hFR mediates the cellular transport of folates (Antony et al., 1985; Kamen & Capdevila, 1986), an essential vitamin, and antifolates (Deutsch et al., 1989; Saikawa et al., 1993; Kane et al., 1986) in selected cell lines. The level of expression of the α hFR may be an important determinant in acquired methotrexate resistance in KB cells (Saikawa et al., 1993; Hsueh & Dolnick, 1994) and antifolate cytotoxicity (Chung et al., 1993). For these reasons, further characterization of the functional properties and the factors regulating the expression of the hFRs are of physiologic and pharmacologic relevance.

The hFR proteins (Antony, 1992; Kane & Waxman, 1989; Weitman et al., 1992a,b; Mantovani et al., 1994; Franklin et al., 1994) and transcripts (Elwood, 1989; Weitman et al., 1992b; Page et al., 1993; Ross et al., 1994; Hsueh & Dolnick, 1993) are widely expressed in mammalian tissues and physiologic fluids. The factors involved in the regulation of hFRs are unknown, although several lines of evidence suggest that their expression is modulated by available folate. First, the serum levels of folic acid binding proteins are increased in pregnancy and in disease states associated with

‡ The following sequences have been submitted to GenBank/EMBL Data Bank: KB2 cDNA nucleotide, accession number U78794; α hFR gene, accession number U20391; and KB4 cDNA nucleotide, accession number U78793.

* Address correspondence to this author at MB/DCS/NCI, Building 10 Room 12N226, 9000 Rockville Pike, Bethesda, MD 20892. Phone: (301) 496-4150. FAX: (301) 402-0172. E-mail: pce@helix.nih.gov.
§ NCI.
|| Kanazawa University.
⊥ University of Washington Medical School.
▽ Brown University Medical School.

⊗ Abstract published in *Advance ACS Abstracts,* January 15, 1997.

[1] Abbreviations: hFR, human folate receptor; ORF, open reading frame; UTR, untranslated region; CAT, chloramphenicol acetyltransferase.

S0006-2960(96)02070-3

folate deficiency (Kane et al., 1988). Second, α hFR protein and transcript levels are inversely proportional to the intracellular folate concentration in KB cells (Kane et al., 1988; Ross et al., 1994), a cell line that only expresses the α hFR isoform. Recent studies suggest that the regulation of α hFR expression may include both transcriptional and post-transcriptional mechanisms (Hseuh & Dolnick, 1994; Luhrs et al., 1992). The structure and sequence of the 3′ terminus of the α hFR gene have been determined from PCR-amplified DNA fragments that span the open reading frame (ORF) and 3′ untranslated region (UTR) of the cDNA (Sadasivan et al., 1992). In other studies, we have recently identified and characterized a promoter within the genomic sequence immediately upstream from the translational start site of the α hFR cDNA in KB cells (Saikawa et al., 1995). However, the 5′ terminal and flanking sequences and other potential regulatory elements of the α hFR gene have not been reported. Seven independent α hFR cDNA isoforms have been isolated from libraries from human colon carcinoma [CaCo-2 (Lacey et al., 1989) and HT29 (Campbell et al., 1991) cells], ovarian carcinoma [IGROV1 (Coney et al., 1991) and SKOV3 (Campbell et al., 1991) cells], and nasopharyngeal epidermoid carcinoma (KB) (Elwood, 1989; Sadasivan & Rothenberg, 1989) cells. Although each of these cDNAs shares an identical ORF and 3′ UTR, their 5′ termini are heterogeneous in length and sequence. The molecular basis of the observed sequence divergence is unknown but may result from alternative splicing of transcripts from a single gene, from transcription from more than one promoter, or from polymorphic or other homologous α hFR genes.

To identify and further characterize the transcriptional elements involved in regulation of the α hFR gene, and to investigate the origin of the reported α hFR cDNA isoforms, we now report the complete sequence and organization of the α hFR gene. We discovered that the reported 5′ termini of the α hFR cDNA isoforms are homologous to sequence contained in exons 1, 2, and 3 of the α hFR gene and that the common ORF and 3′ UTR are contained in exons 4 through 7. Finally, evidence is presented suggesting that the heterogeneous 5′ termini of the α hFR cDNA isoforms originate from transcription from two independent promoters that are activated in a tissue-specific manner and from alternative splicing of 5′ exons.

EXPERIMENTAL PROCEDURES

General Reagents and Enzymes. Human MCF-7 mammary carcinoma and CaCo-2 colon carcinoma cells were purchased from ATCC (Rockville, MD), and human OvcarIII ovarian carcinoma cell line was from Dr. Ira Pastan (LMB/DCBD/NCI, Bethesda, MD). HeLa cells and the CMV-Luc vector (Liu et al., 1991) containing the luciferase gene downstream from a CMV promoter were gifts from Dr. Maria Zajac-Kaye (Bethesda, MD). Tissue culture cells were propagated as described (Elwood et al., 1986). Plasmids (pGEM4Z, pCAT basic, and pCAT control) were purchased from Promega (Madison, WI). Competent JM109 *Escherichia coli* were purchased (Promega) or prepared as described (Elwood, 1989). Radionucleotides including [γ-^{32}P]ATP (3000 Ci/mmol), 5′-[α-^{35}S]dATP, 5′-[α-^{32}P]dGTP, and dCTP (800 Ci/mmol), and SP6 grade 5′-[α-^{32}P]CTP (3000 Ci/mmol) were obtained from Amersham Corp. The [^{14}C]-chloramphenicol (57 mCi/mmol) was purchased from DuPont-New England Nuclear. *Spe*I and *Dra*III were obtained from Stratagene. Except as noted, all other enzymes were from Promega (Madison, WI). Reagents for radiolabeling DNA and for *in vitro* transcription were purchased from Promega. Oligonucleotides were synthesized on a model 392 Applied Biosystems Inc. DNA synthesizer. All other reagents were purchased from Sigma or Fisher.

Molecular Cloning. Recombinants [$2 \times (10^6)$] from a human lymphocyte λ DASH genomic library (Stratagene, La Jolla, CA) were screened with a radiolabeled cDNA-specific 5′ *Eco*RI-*Hin*cII restriction fragment of KB1 (c32 clone) α hFR cDNA as described (Elwood, 1989). Hybridization positive clones ($n = 3$) were plaque-purified, and recombinant DNA was purified from liquid or plate lysates using Lambdasorb (Promega).

DNA Sequencing. DNA sequence was determined from both DNA strands by the primer-directed dideoxynucleotide chain termination method (Sanger et al., 1977) using modified T7 DNA polymerase (United States Biochem. Corp.). The complete sequence of the α hFR gene contained in genomic clones S18 and S28 was determined from plasmids containing subcloned restriction fragments and from nested sets of deletion mutants of these constructs. For construction of 5′ terminal deletions, 5′ *Eco*RI fragments containing putative exons 1, 2, and 3 and the *Bam*HI fragments containing putative exons 4 and 5 of S18 and S28 were subcloned into pGEM4Z. Constructs (10 μg) were acid−phenol extracted, digested with *Sph*I to generate a 3′ overhang that is resistant to exonuclease III, and then digested with either *Sal*I (5′ *Eco*RI constructs), or with *Xba*I (*Bam*HI constructs). Nested deletions were generated from *Sal*I or *Xba*I termini of the linearized plasmid (10 μg) using exonuclease III and the reagents supplied in an Erase-a-Base kit (Promega).

Southern Analysis. Human genomic DNA (30 μg each) and cloned DNA (1 μg) were digested with restriction enzymes as specified, and Southern blots were prepared as described (Elwood, 1989). To facilitate the construction of restriction maps and to orient hybridization-positive fragments, blots containing cloned genomic DNA were probed sequentially with radiolabeled 5′ *Eco*RI-*Hin*cII or 3′ *Eco*RI-*Pst*I restriction fragments of KB1 or KB2 cDNA.

RNase Protection Assays. To determine the structure and relative abundance of α hFR transcripts and to map the 5′ termini of potential 5′ and/or alternatively spliced exons, we performed RNase protection assays as described (Page et al., 1993). The 5′ *Eco*RI-*Hin*cII KB1 and KB2 cDNA restriction fragments and genomic restriction fragments spanning the sequences upstream from exons 1 through 4 were subcloned into pGEM4Z. Genomic fragments included the following: (1) a 410 bp *Bam*HI-*Apa*I fragment containing exon 1; (2) a 443 bp *Hin*fI-*Sac*I fragment extending from 45 bp upstream of exon 2 through the 5′ 101 bp of exon 3; (3) a 365 bp *Sac*I-*Bam*HI fragment containing the 3′ sequence of exon 3; and (4) a 565 bp *Spe*I-*Ava*I fragment containing exon 4. The constructs were linearized, and antisense riboprobes were transcribed using SP6 RNA polymerase. The size and integrity of each riboprobe were verified on a 6% sequencing gel prior to each assay. Total RNA from cultured cells (KB, MCF-7, CaCo-2, and OvcarIII cells) was isolated as described (Elwood, 1989). Normal human tissue RNA was purchased from Clontech. The integrity and quantity of RNA was verified by Northern analysis.

Primer Extension. Oligonucleotides (EP1, 5′-$_{+17}$GTC-ATCCGCTGAGCCATGTCTGTCC$_{-8}$-3′, and EP3, 5′-

$_{+82}$ATGCAATCCTTGTCTGAGCCTCCCCTACT$_{+54}$-3′) were end-labeled. An aliquot (100 000 cpm) was annealed (42 °C for 3 h) with 30 μg of heat-denatured total RNA or wheat germ tRNA, and extended with AMV reverse transcriptase at 42 °C for 1.5 h as described (Ausubel et al., 1991). The RNase protection assays and primer extension reactions were resolved on a 6% sequencing gel in 0.5× TBE and 8 M urea at 60 W. End-labeled *Hae*III-digested φX174 DNA and sequencing reactions served as size markers.

Chimeric Promoter−Reporter Gene Constructs. Restriction fragments containing the 5′ flanking sequences of the α hFR gene were subcloned into the promoter-less pCAT basic vector. To analyze the sequences upstream from exons 1 through 3, the 5′ *Eco*RI restriction fragment of the α hFR gene was subcloned into pGEM4Z and named hKB1-3E. A 2.3 bp *Pst*I-*Eco*RI fragment from hKB1-3E construct (the 3′ *Pst*I site is from the MCS of pGEM4Z) that contains 1009 bp upstream from exon 1 was subcloned into the *Pst*I cloning site of pCAT basic and is named hKB1-3PE. The hKB1-3PE construct was digested with *Dra*III *(*contained in intron A) and *Acc*I (contained in the MCS of KB1-PE), blunt-ended, and re-ligated to form the hKB1PD construct. The 3.2 kb *Hin*dIII restriction fragment containing exon 4 and approximately 1300 bp 5′ from exon 4 was subcloned into pCAT basic and is named hKB4H. The hKB4H was digested with *Ava*I (contained in exon 4) and *Acc*I (contained in the MCS of pCAT basic), blunt-ended, and re-ligated to generate the hKB4HA construct. The hKB4HA-DSP construct was derived from hKB4HA by digestion with *Dra*III and *Spe*I followed by re-ligation of blunt ends. The orientation and sequence of each construct were verified by sequence analysis.

Transfection. HeLa cells were plated at a density of 1 (10^6) cells per 100 mm tissue culture dish in minimal essential media without added folic acid (DMEM) (Gibco) containing L-glutamine, Earle's salts, and supplemented with 10% fetal calf serum. After the cells reached 70% confluency, supercoiled CAT constructs (20 μg) were transfected by the calcium phosphate method (Cullen et al., 1987). CMV-Luc (0.5 μg) was co-transfected with each CAT construct to control for transfection efficiency. The medium was replaced 18 h after transfection. After incubation for 48 h, the cells were rinsed twice with ice-cold PBS and scraped into 1 mL of PBS. The cells were pelleted at 865*g* for 2 min at 4 °C and resuspended in 100 μL of 100 mM potassium phosphate, pH 7.8, containing 1 mM DTT. The cells were lysed by three cycles of freeze−thaw and centrifuged at 11000*g* for 5 min at 4 °C to remove particulate cellular debris, and the supernatant was assayed for CAT (Yu et al., 1993) and luciferase activity. To measure CAT activity, the cytosolic extract (50 μg of protein in 50 μL of 100 mM potassium phosphate, pH 7.8, containing 1 mM DTT), 0.1 μCi of [^{14}C]chloramphenicol, and 40 μg of fresh acetyl CoA (Pharmacia-LKB) was incubated at 37 °C for 3.5 h. After extraction into ethyl acetate, the reactions were resolved by TLC on silica gel 1B sheets (T. J. Baker, Inc.). After drying, the gel was autoradiographed, and the extent of acetylation of chloramphenicol was determined using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). To determine luciferase activity, 50 μg of the cytosolic extract was added to 250 μL of fresh reaction buffer (glycyl glycine, pH 7.8, containing 5 mM ATP, 25 mM $MgSO_4$, and 10% glycerol). Fresh D-luciferin−potassium salt (1 mM stock in H_2O) purchased from Analytic Luminescence Laboratory (San Diego, CA) was added, and fluorescence was determined using a monolight 2010 luminometer (Analytic Luminescence Lab). All experiments were performed in duplicate on at least three separate occasions.

Protein Assays. Protein assays were performed according to the method described by Lowry (1951) or with the Micro BCA protein assay kit (Pierce, Rockford, IL).

RESULTS

Southern Analysis and Cloning of the KB Cell (α) hFR Gene. We have previously reported that Southern analysis of human genomic DNA using a radiolabeled full-length α human folate receptor (hFR)[1] cDNA yields a complex hybridization pattern (Elwood, 1989). This observation results from cross-hybridization of the probe with other hFR genes (Page et al., 1993) and is consistent with the reported homology between the α hFR cDNA and the β (Ratnam et al., 1989; Page et al., 1993) and γ (Shen et al., 1994) hFR isoforms. To determine the sizes of restriction fragments containing the α hFR gene, we hybridized a Southern blot containing human KB cell genomic DNA with radiolabeled 5′ *Eco*RI-*Hin*cII α hFR cDNA restriction fragments (200 and 386 bp, respectively) from the KB2 cDNA or from clone c32 (Elwood, 1989), hereafter designated KB1 cDNA. These probes specifically hybridize with the α hFR cDNA under the conditions described (data not shown). The 5′ terminal 203 and 17 nucleotides of the KB1 and KB2 cDNA probes, respectively, are unique, whereas their 3′ 183 nucleotides, corresponding to residues −8 to 175 of the KB1 cDNA, are common to all α hFR cDNA isoforms. As shown in Figure 1A, the 5′ KB2 (lanes 1−3) and KB1 (lanes 1−3) probes hybridize with 9.0 kb *Eco*RI, 4.9 kb *Bam*HI, and 3.2 kb *Hin*dIII restriction fragments. The longer KB1 probe also hybridizes with 6.2 kb and 5.5 kb *Eco*RI, 1.3 kb *Bam*HI, and 8.0 kb *Hin*dIII restriction fragments (Figure 1A, lanes 1−3, respectively). These results demonstrate that the 5′ end of the ORF is contained in 9.0 kb *Eco*RI, 4.9 kb *Bam*HI, and 3.2 kb *Hin*dIII restriction fragments, and that the unique 5′ UTR of KB1 cDNA is contained within a 1.3 kb *Bam*HI restriction fragment and larger *Eco*RI and *Hin*dIII fragments. Since all α hFR cDNAs share the ORF sequence contained in the KB1 and KB2 probes, these results suggest that the gene is single copy and encodes each of the cDNAs. This hybridization pattern was used to identify genomic clones containing the α hFR gene.

A total of 2 × (10^6) PFUs contained in a normal human lymphocyte genomic λ DASH library (Stratagene) were screened using the 5′ *Eco*RI-*Hin*cII KB1 probe. Three hybridization-positive clones, designated S18, S27, and S28, were plaque-purified for further characterization. Southern analyses of the DNA from these clones following digestion with *Bam*HI and *Hin*dIII endonucleases were identical. The Southern blot of S18 and the ethidium bromide-stained gel containing DNA from S18 and S28 prior to Southern transfer are shown in Figure 1B and 1C, respectively. The 5′ KB2 and KB1 probes (Figure 1B, lanes 1−3 and 4−6, respectively) hybridize with a 4.9 kb *Bam*HI fragment and a 3.2 kb *Hin*dIII fragment. The KB1 probe also hybridizes with a 1.3 kb *Bam*HI fragment (Figure 1B, lanes 4 and 6). The sizes of these fragments are identical to the corresponding homologous restriction fragments contained in genomic DNA (Figure 1A) and to those contained in S28 (see Figure 2C) suggesting that both clones contained the α hFR gene. As

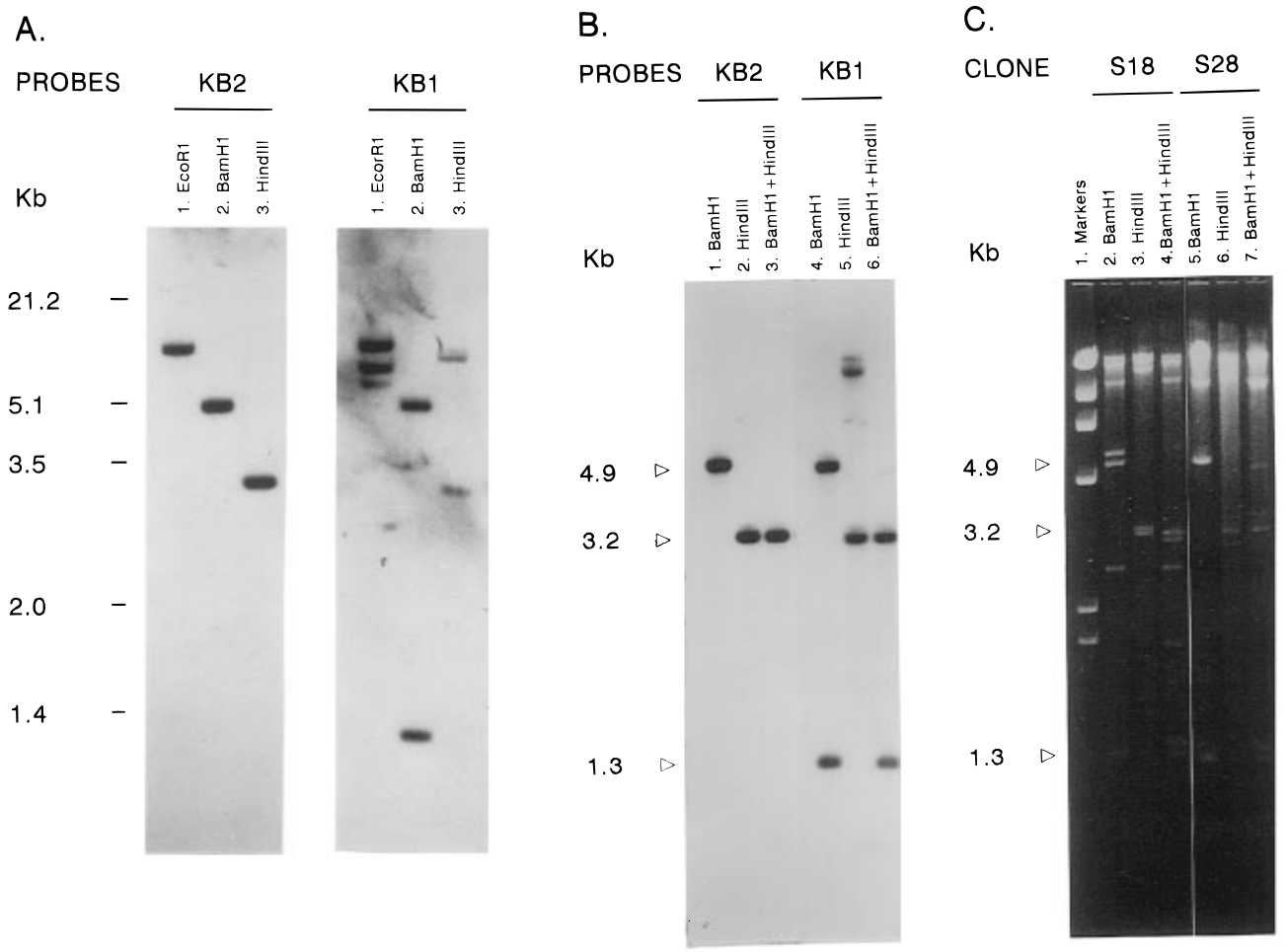

FIGURE 1: Southern analysis of the α hFR gene. (A) KB genomic DNA: Southern blots were prepared from genomic DNA isolated from KB cells that was digested with *Eco*RI (lane 1), *Bam*HI (lane 2), and *Hin*dIII (lane 3) endonucleases. The blots were hybridized with radiolabeled 5′ *Eco*RI-*Hinc*II restriction fragments of the KB2 or KB1 cDNA as indicated. The size markers were λ DNA digested with *Hin*dIII. (B) Genomic clones. Southern blots were prepared using DNA from clone S18 that was digested with *Bam*HI (lanes 1 and 4), *Hin*dIII (lanes 2 and 5), or both (lanes 3 and 6) endonucleases. The blots were hybridized with radiolabeled 5′ *Eco*RI-*Hinc*II restriction fragments of the KB2 (lanes 1−3) or KB1 (lanes 4−6) cDNA as indicated. (C) Restriction digests of genomic clones. Aliquots of S18 and S28 DNA were digested with *Bam*HI (lanes 2 and 5), *Hin*dIII (lanes 3 and 6), or both (lanes 4 and 7) endonucleases. The samples were resolved by electrophoresis, and the restriction fragments were stained with ethidium bromide. λ DNA digested with *Hin*dIII (lane 1) or with *Hin*dIII and *Eco*RI (not shown) served as size markers.

shown in Figure 1B, lane 5, the KB1 probe hybridizes with a 15 kb *Hin*dIII fragment. Although the 15 kb *Hin*dIII fragment is larger than the corresponding 8.0 kb *Hin*dIII fragment contained in genomic DNA (Figure 1A), the sequence homologous to the KB1 probe is contained within the 1.3 kb *Bam*HI fragment (Figure 1B, lane 6) present in both the genomic clones (Figure 1B, lane 4) and genomic DNA (Figure 1A).

The restriction and hybridization maps of clones S18 and S28 are shown in Figure 2A. The restriction map of S27 is identical to that of S28 indicating that these clones contain overlapping DNA inserts. Although S18 and S28 clones contain hybridization-positive restriction fragments of identical size, each clone contains unique restriction sites (including *Hin*dIII, *Bam*HI, and *Eco*RI sites) upstream from the 5′ 1.3 kb *Bam*HI fragment homologous with the KB1 probe. These restriction site polymorphisms likely explain the differences in the size of the *Hin*dIII restriction fragment contained in genomic DNA (Figure 1A, lane 6) compared to clone S18 (Figure 1B, lane 5). The authenticity of the genomic clones was verified by determining the sequence of the hybridization-positive restriction fragments. Taken together, these results demonstrate that S18 and S28 are alleles that contain a copy of the α hFR gene.

Organization and Sequence of the α hFR Gene. The sequence of both DNA strands of the α hFR gene was determined by primer-directed dideoxynucleotide sequencing according to the strategy shown in Figure 2B. The complete sequence including 951 bp of 5′ flanking sequence is available from GenBank. Based on comparison to the reported α hFR cDNA sequences, the α hFR gene is composed of 7 exons flanked by consensus splice site sequences and spans 6738 bp (Figure 2C). Figure 3 contains the sequences from the α hFR gene that encode the exons and that are upstream from exons 1, 2, 3, and 4. As illustrated in Figure 2C and shown in Figure 3, the divergent 5′ terminal sequences of the α hFR cDNAs are collinearly arranged in tandem within the 1.3 kb *Bam*HI fragment. Exon 1 contains the 5′ terminal residues of the KB1 (Elwood, 1989), IGROV1 4/6 and IGROV1 31 (Coney et al., 1991), and CaCo-2 (Lacey et al., 1989) cDNAs. The sequences of exons 2 and 3 encode the 5′ termini of KB2 and SKMOv 18 (Campbell et al., 1991) cDNAs, respectively. Immediately downstream from a potential splice acceptor site (double underline, Figure 3), the terminal 66 bp of exon 3 are also homologous to sequences within the divergent 5′ UTR of CaCo-2 and IGROV1 31 and 4/6 cDNAs. The 5′ terminus of KB4 cDNA is encoded by the first 37 bp of

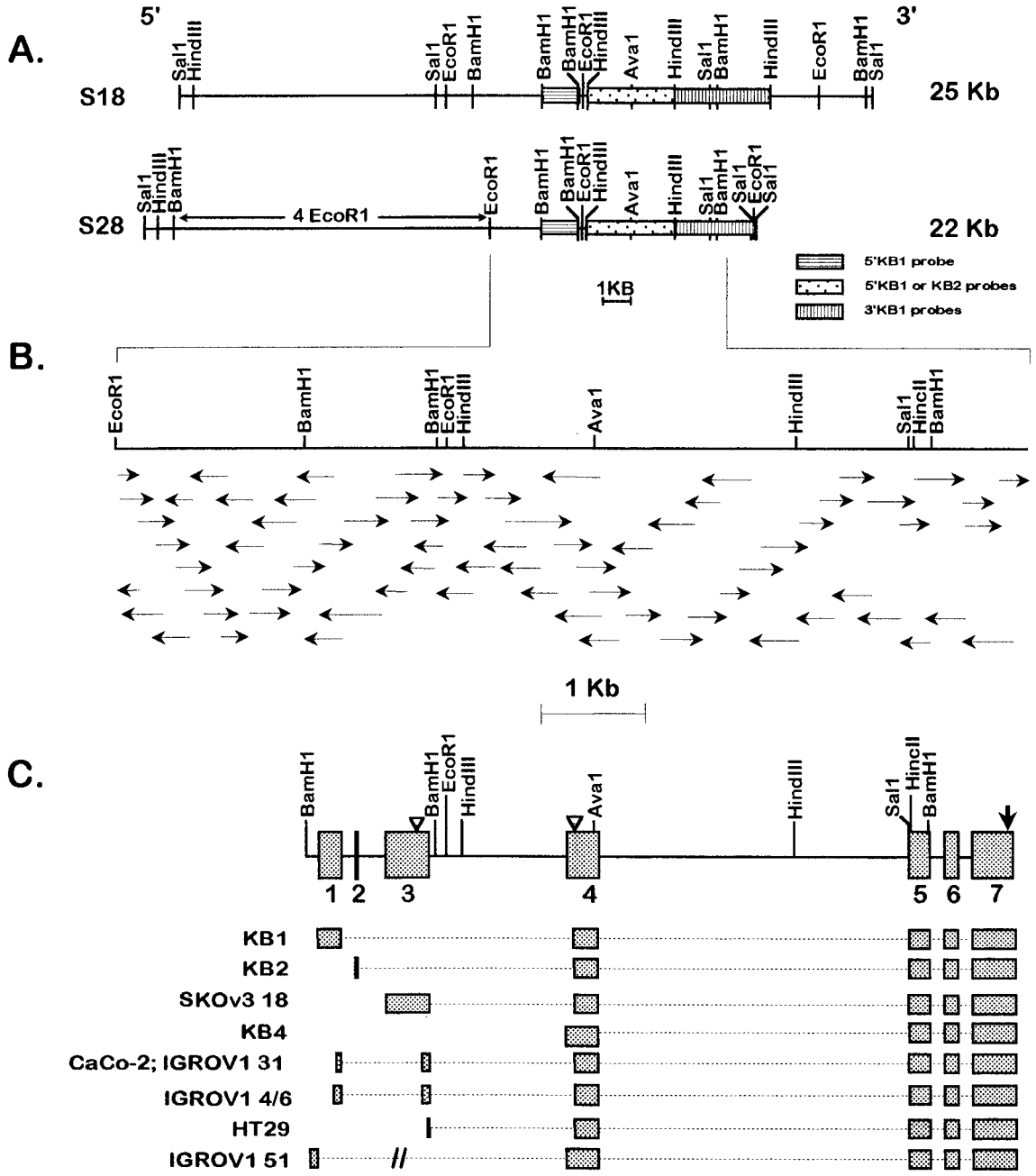

FIGURE 2: (A) Partial restriction and hybridization maps of genomic clones. Recombinant genomic DNA from clones S18 and S28 was digested with *Eco*RI, *Bam*HI, *Hin*dIII, *Hin*cII, *Ava*I, and *Sal*I and analyzed by Southern analysis using 3′ *Pst*I-*Eco*RI and 5′ *Eco*RI-*Hin*cII restriction fragments of the KB1 and KB2 cDNA probes. The *Bam*HI or *Hin*dIII restriction fragments that hybridize with the cDNA probes are indicated by filled rectangles as indicated in the legend. The map of S28 contains four as yet unmapped 5′ *Eco*RI sites as indicated by "<4 *Eco*RI>". (B) Sequencing strategy of the α hFR gene. The arrows indicate the direction and extent of sequence of clone S28 obtained with each primer. (C) Organization of the α hFR gene. The exons are represented by the rectangles (labeled 1 through 7), and the introns are indicated by the thin line. Potential consensus splice acceptor sites within exon 3 and exon 4 are depicted by the solid arrowhead. The arrow above exon 7 marks the 5′ junction of the unique 3′ terminal sequence of the HT29 cDNA. The names of the reported α hFR cDNAs are shown at the bottom left. The sequences from each cDNA that are homologous to the sequences contained in exons 1 through 7 of the α hFR gene are represented by the rectangles below each exon and are connected by a hashed line. The location of the genomic sequence encoding the 5′ terminus of the IGROV1 51 cDNA relative to exon 4 is unknown as indicated by the "//".

exon 4 and is immediately upstream from a consensus splice site (double underline, Figure 3). The 5′ terminal sequence of exon 4 is identical to residues with the coordinates of −54 to −9 of the IGROV1 51 cDNA (Coney et al., 1991); however, the 5′ terminal 97 nucleotides of IGROV1 51 cDNA are not contained in the 5′ *Eco*RI fragments of S18 or S28. The location of this sequence is unknown as indicated by the double slash ("//") in Figure 2C. The sequences common to all α hFR cDNAs including the 8 bp upstream from the translational start site (Figure 3, bold **ATG** with overline at residue 2617), the ORF, and the 3′ UTR are encoded by the sequence of exon 4 downstream from this splice site and exons 5 through 7. The 3′ terminus of exon 7 also contains the 3′ terminal 21 bp of HT29 cDNA. As shown in Figure 3, minor sequence differences are present in exons 2 and 3 relative to the sequence contained in the corresponding cDNA sequences. The nature of these minor sequence differences was not investigated further; however, the genomic sequence was verified on both DNA strands of clones S18 and S28.

Mapping the 5′ Transcript Boundary. We have determined the length of the 5′ termini and the transcriptional

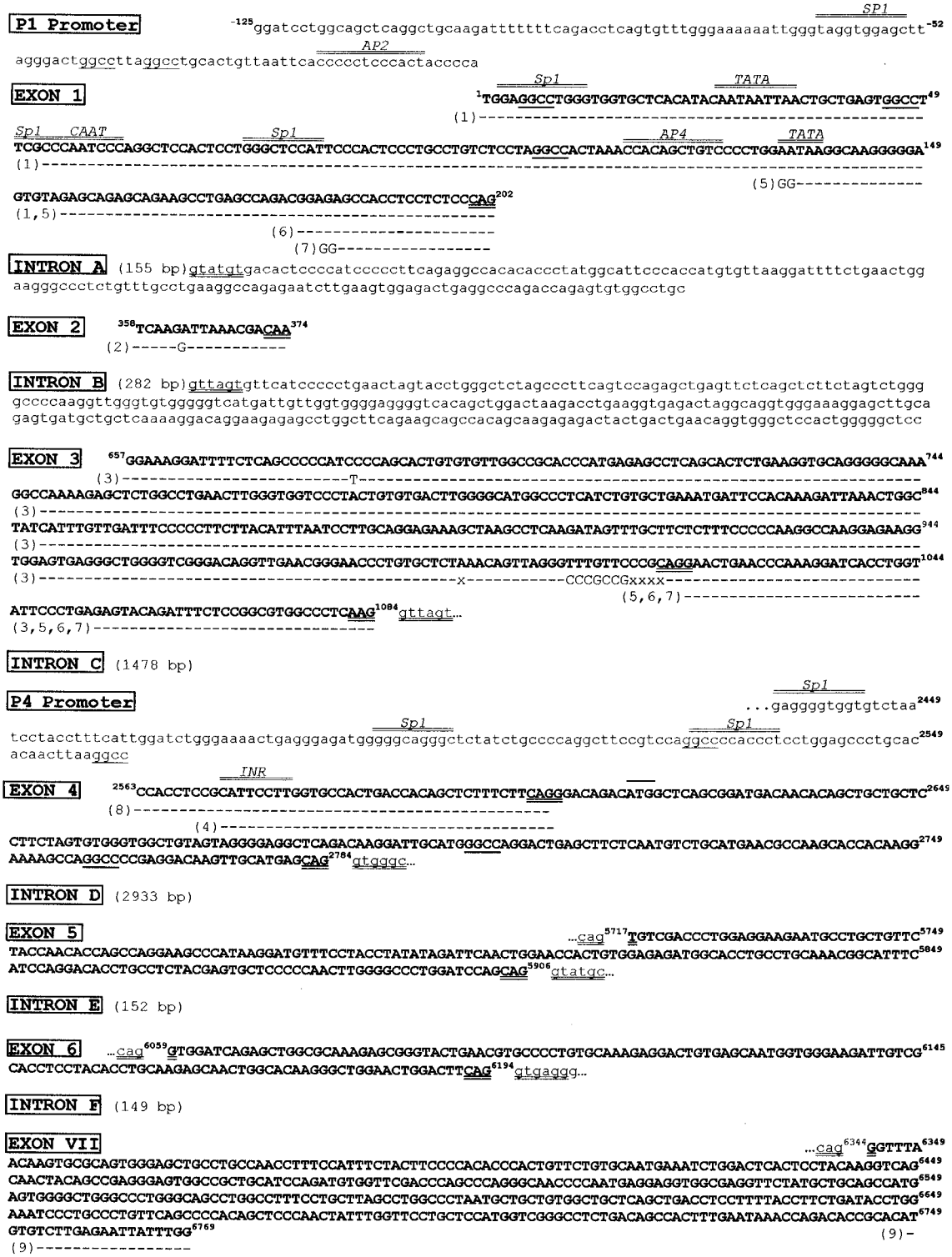

FIGURE 3: Sequence of the α hFR gene. The complete nucleotide sequence including 951 bp of 5′ flanking sequence of the α hFR gene is available from GenBank (accession number U20391). A partial sequence including the exons, splice junctions, and the sequences upstream from exons 1 through 4 is shown in Figure 3. The nucleotides in exons and in introns are in capital and small letters, respectively. The size of each intron is contained in parenthesis. The superscripts indicate the coordinates of the adjacent nucleotides relative to the first residue of exon 1. Consensus splice acceptor and donor sequences are marked by a double underline, and CpG doublets are marked with a single underline. The locations of potential DNA binding sites upstream from exons 1 and 4 are identified by name and by a double overline. The sequences of the heterogeneous 5′ termini of the α hFR cDNAs are shown beneath the homologous genomic sequence contained in exons 1, 2, 3, and 4. The translational start site in exon 4 is indicated by the single overline. The cDNA sequence shared by all α hFR cDNAs including the ORF and 3′ UTR is not shown. The "−" and "×" symbols represent homologous and deleted residues, respectively, while nucleotide substitutions are indicated by the respective nucleotide letter. The α hFR cDNA sequences are identified by the numbers shown to the left of their respective sequences as follows: 1 = KB1 (Elwood, 1989); 2 = KB2; 3 = cSKMOv 18 (Campbell et al., 1991); 4 = KB4; 5 = IGROV1 31 (Coney et al., 1991); 6 = CaCo-2 (Lacey et al., 1989); 7 = IGROV1 4/6 (Coney et al., 1991); 8 = IGROV1 51 (Coney et al., 1991); and 9 = HT29 (Campbell et al., 1991).

start sites of the α hFR gene by primer extension and RNase protection assays. The results of primer extension assays using two antisense primers, EP1 and EP3, complementary to sequences within the common α hFR ORF that are 65

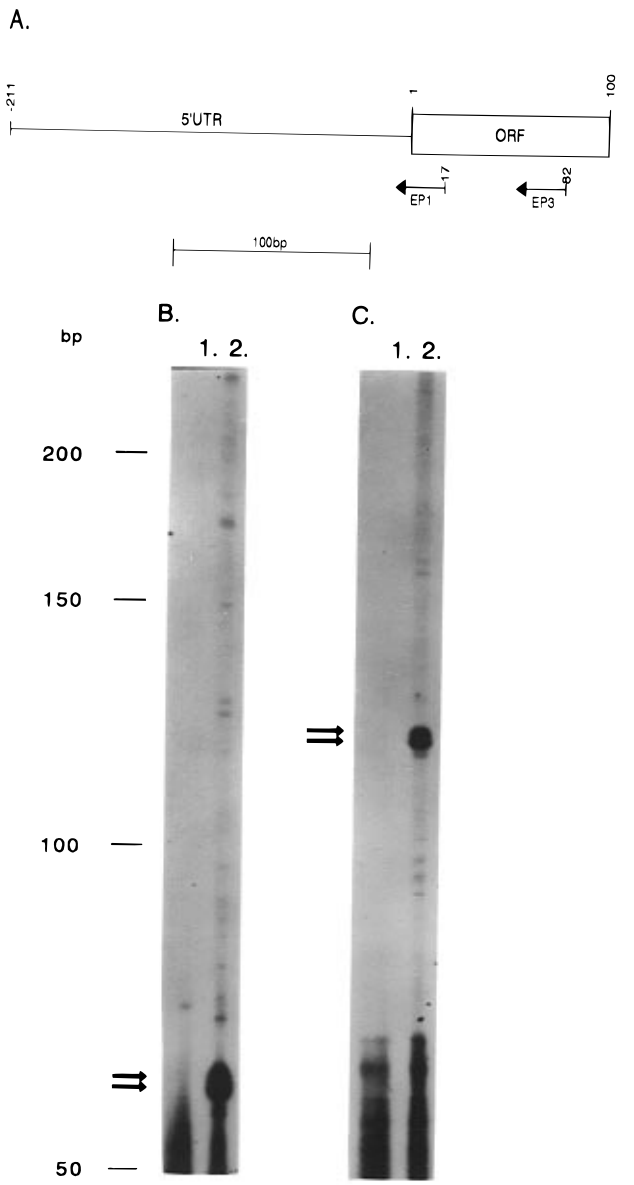

FIGURE 4: Primer extension assays. The origin of the EP1 and EP3 antisense primers relative to the translational start site of the common ORF of the KB1 α hFR cDNA is shown in A. End-labeled primers, EP1 (B) and EP3 (C), were hybridized with 30 μg each of wheat germ tRNA (lane 1) and KB cell total RNA (lane 2) and extended with AMV reverse transcriptase as described. The major extended products, a doublet, is indicated by arrows. Size markers included a sequencing reaction and end-labeled *Hae*III-digested øX174 DNA.

residues apart are contained in Figure 4. Following hybridization with total KB cell RNA, extension from the EP1 and EP3 primers yields major products (a doublet, indicated by arrows) of approximately 62 and 127 bp in length, respectively, as shown in Figure 4B and 4C, lane 2. In contrast, we did not observe extended products following hybridization of these primers with control wheat germ tRNA (Figure 4B and 4C, lane 1). The difference in size (∼65 bp) of these major products extended from EP1 and EP3 primers corresponds closely to their different origins (see Figure 4A) and confirms the specificity of the products. The length of the extended products suggest that the major transcriptional start site of the α hFR gene in KB cells is ∼45 bp upstream from the translational start site.

To map the transcriptional start site(s) of exons 1, 2, 3, and 4 and to determine the relative abundance of each transcript, we performed RNase protection assays using riboprobes corresponding to the 5′ terminal *Eco*RI-*Hin*cII fragment of KB1 and KB2 cDNAs (Figure 5A) and to genomic sequence containing exons 1 through 4 (Figure 5B). Wheat germ tRNA and RNA from human MCF-7 cells, a non-expressing cell line, served as negative controls. A major ∼180 bp fragment (Figure 5A, lanes 2, 4, 7 and 9) of each cDNA riboprobe is specifically protected by KB cell total RNA. Larger, less abundant protected fragments, ranging in length from 180 to <400 bp, of KB1 riboprobe are observed after longer exposure (lane 4). Since the common 3′ terminus of each cDNA riboprobe is 183 bp downstream from the alternate splice acceptor site that is contained in exon 4 (see Figure 3), the observation that the most abundant protected fragment of each riboprobe is ∼180 bp together with the results of primer extension assays suggests that a transcript other than KB1 and KB2 is the major isoform expressed by KB cells.

The results of RNase protection assays using riboprobes from the genomic constructs are shown in Figure 5B. KB cell RNA protects multiple fragments of the Ex 1 riboprobe that range in size from ∼50 to 205 bp (Figure 5B, lane 2). The length of longest protected fragment is consistent with the length (203 bp) of the 5′ termini of KB1 cDNA. The shorter protected fragments suggest the presence of other transcriptional start sites within exon 1. The 443 bp Ex 2−3 riboprobe, which contains exon 2, intron B, and the 5′ 101 bp of exon 3, is not protected by wheat germ tRNA or by total RNA from these tissue culture cells (lanes 5−8). Several fragments of the 365 bp Ex 3 riboprobe are protected by each RNA including wheat germ tRNA consistent with nonspecific hybridization, or self-hybridization (lanes 9−12). "Snap back" transcription resulting in self-hybridization is commonly observed after linearization of a construct with a restriction enzyme that leaves a 3′ overhang (e.g., *Sac*I). However, a specific fragment of ∼66 bp in length (see arrow) is protected by KB cell RNA (lane 10) and, to a much lesser extent, by CaCo-2 RNA (lane 11). The length of this protected fragment is very similar to the length of the sequence at the 3′ terminus of exon 3 that is homologous to the CaCo-2 and IGROV1 31 and 4/6 cDNAs (see Figure 3). KB cell RNA protects a major 190 bp fragment and a minor ∼155 fragment (see arrow) of the 565 bp Ex 4 riboprobe (lane 14). Since the *Ava*I cloning site is 156 bp downstream from the alternate splice site within exon 4 (see Figure 5B), these results demonstrate that the 5′ terminus of exon 4 is ∼35 bp upstream from the alternate splice site and that the 190 bp fragment corresponds to the major α hFR transcript expressed by KB cells. The less abundant 155 bp fragment represents transcripts originating from other 5′ exons such as exon 1.

Transfection Analysis of hFR Gene Promoter Elements. To determine if the regions flanking the transcriptional start sites of exon 1 and exon 4 contain promoter activity, restriction fragments spanning exon 1 through exon 4 were subcloned into pCAT basic (Figure 6A) for reporter gene analysis (Figures 6B and 6C). Compared to the negative control, pCAT basic (Figure 6C, lane 11), the hKB1PD (Figure 6C, lanes 3 and 4), hKB4HA (Figure 6C, lanes 7 and 8), and hKB4HA-DSP (Figure 6C, lanes 9 and 10) constructs contain promoter activity that is 7.5-, 6.8-, and 10-fold above background (Figure 6B). These two promoter regions upstream from exons 1 and 4 are designated P1 and P4, respectively. Interestingly, longer constructs (e.g., hKB1-

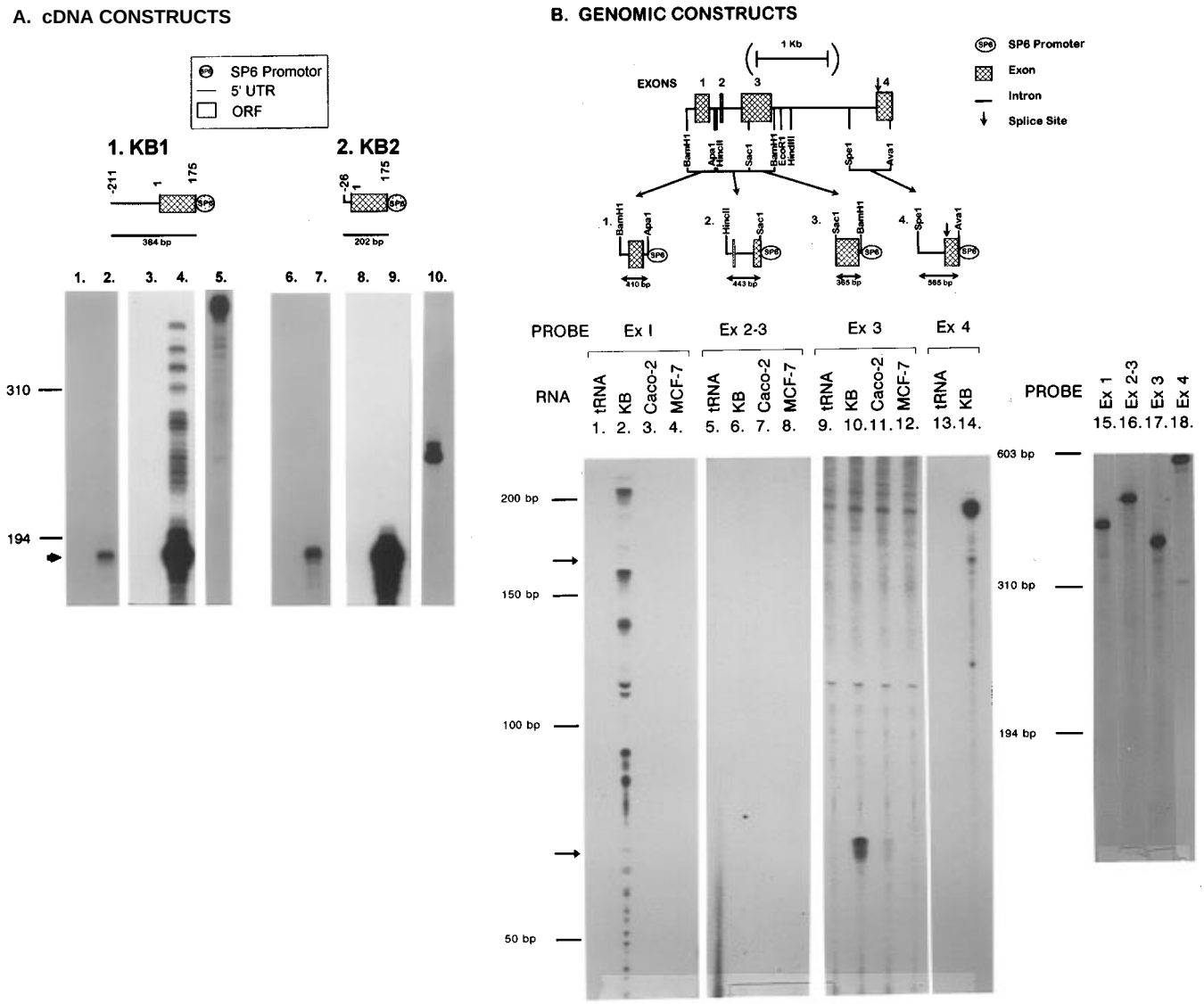

FIGURE 5: Determination of the transcriptional start sites by RNase protection assays. The constructs containing cDNA and genomic restriction fragments that were used to synthesize riboprobes are shown in A and B, respectively. The symbols are defined in the legends to the right of each model. The arrow marks the potential splice site within exon 4. The cDNA riboprobes (A) were hybridized with 30 μg each of wheat germ tRNA (lanes 1, 3, 6, and 8), and KB cell total RNA (lanes 2, 4, 7, and 9). Undigested KB1 and KB2 riboprobes are contained in lanes 5 and 10, respectively. The autoradiograph was exposed for 1 day (lane 1, 2, 6, and 7) or 10 days (lane 3, 4, 8, and 9). Genomic riboprobes (B) were hybridized with 30 μg of each RNA as indicated above lanes 1 through 14. Lanes 15−18 contain undigested genomic riboprobes. The arrows point to a 66 bp Ex 3 fragment (lanes 10 and 11) and the 156 bp Ex 4 fragment (lane 14) representing alternatively spliced transcripts. Size markers included a sequencing reaction and end-labeled *Hae*III-digested ϕX174 DNA.

3PE and hKB4H) that contain the identical P1 and P4 promoter sequences plus 1−1.5 kb of 3′ flanking sequence, do not exhibit detectable promoter activity (Figure 6C, lanes 1 and 2 and lanes 5 and 6, respectively). These results suggest that sequences downstream from exons 1 and 4 may contain elements that suppress transcription from the P1 and P4 promoters.

The P1 and P4 promoter sequences were analyzed using IBI Pustell Sequence Analysis or Hitachi HIBIO DNASIS software to identify potential *cis*-acting elements and to determine their GC content. Consensus DNA binding sequences (Locker & Buzard, 1990) for nuclear proteins contained within or immediately upstream from exons 1 and 4 are shown in Figure 3. Several potential promoter elements including five *Sp1* sites, two TATA boxes, and one CAAT box are present upstream from or within exon 1 of the P1 promoter region. An array of three *Sp1* sites is upstream of exon 4. We have shown that the three *Sp1* sites contained in the P4 promoter form complexes with *Sp1* and, together with an INR sequence flanking the start of transcription, are required for optimum promoter activity in KB and HeLa cells (Saikawa, 1995). The sequences upstream from and including exons 1 and 4 are relatively rich in GC residues (55% and 55%, respectively) and contain 5 CpG doublets (Figure 3, double underline).

Tissue Differential Promoter Utilization. Since the 5′ flanking regions of exons 1 and 4 contain promoter activity, we next performed RNase protection assays (Figure 7) using RNA from selected human tissues (e.g., kidney, lung, and cerebellum) to examine the relative activity of P1 and P4. These tissues are known to express α hFR transcripts (Elwood, 1989; Antony, 1992; Kane et al., 1989; Weitman et al., 1992a,b; Ross et al., 1994). The riboprobes were synthesized from the constructs illustrated in Figure 5A and 5B. For the 5′ *Eco*RI-*Ava*I KB1 cDNA and Ex 4 genomic probes, the putative splice site within exon 4 is 156 bp 5′ of the *Ava*I site; therefore, protected fragments of >156 bp are homologous to KB1 and KB4 transcripts, respectively, and

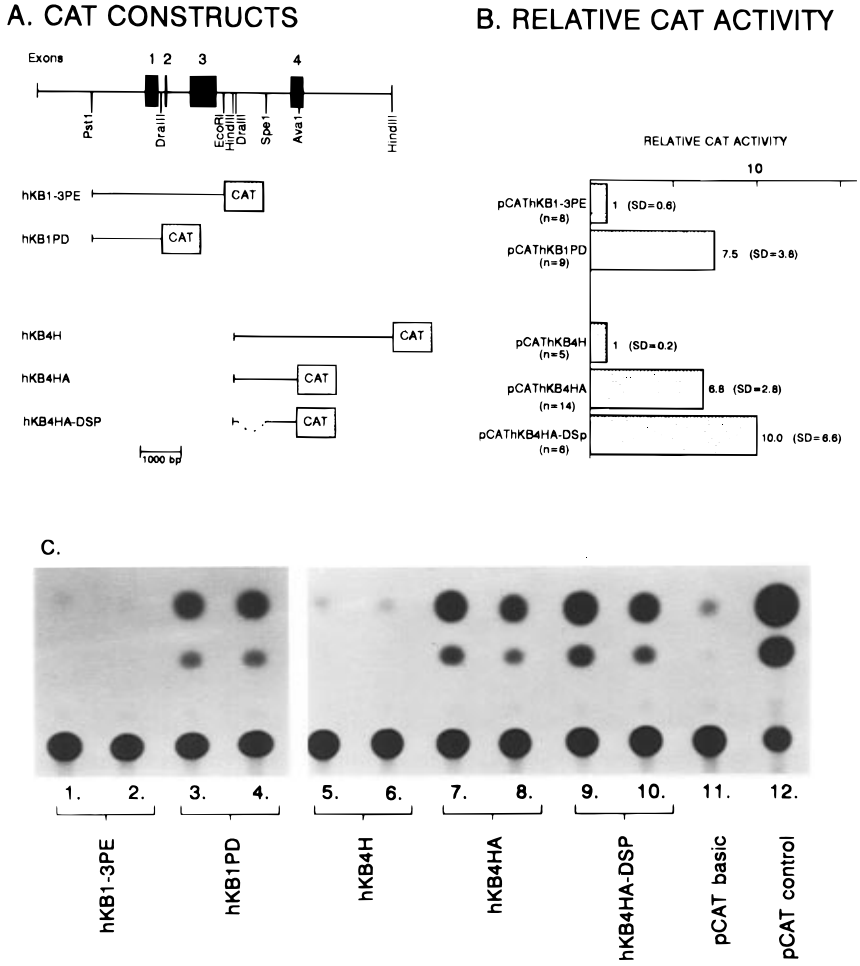

FIGURE 6: Functional activity of P1 and P4 promoter regions of the α hFR gene. Restriction fragments spanning potential 5′ exons including exons 1 through 4 were subcloned into pCAT basic as shown in A. The chimeric α hFR-CAT reporter gene constructs (20 μg) were co-transfected with CMV-luc plasmid (0.1 μg) into HeLa cells. CAT and luciferase activity were determined as described. The relative CAT activity (normalized to luciferase activity) of each construct is shown in B. Each experiment was repeated on three separate occasions in duplicate. A representative TLC plate containing each construct in duplicate is shown in C.

protected fragments of 156 bp represent alternatively spliced transcripts. Since the *Hinc*II restriction site is 27 bp downstream from the *Ava*I site in the α hFR cDNA, the putative splice site is 183 bp upstream from the *Hinc*II site of the 5′ *Eco*RI-*Hinc*II constructs. For each riboprobe, the protected fragment representing alternatively spliced transcripts are indicated by arrow in Figure 7. RNA from KB cells, lung tissue, and, to a lesser extent, kidney tissue protects two major fragments of the genomic Ex 4 probe (Figure 7A). The larger fragment (190 bp) corresponds to the transcriptional start site of P4, while the shorter fragment (155 bp) corresponds to protected fragments of alternatively spliced transcripts. The relative abundance of these protected fragments is different among these tissues. While lung and KB cell α hFR transcripts originate primarily from the P4 promoter region, the predominant transcripts in kidney and cerebellum RNA originate from a different promoter.

Since the sequence contained in the 5′ KB1 cDNA riboprobe (see Figure 3) is identical to exon 1 and to the sequence downstream from the splice within exon 4 (see Figure 3), and since this riboprobe spans the potential alternative splice site, we chose the KB1 riboprobe to analyze transcripts originating from P1 promoter region (Figure 7B). On the basis of the relative abundance of bands >156 bp (or >183 bp in the lane that contains human cerebellar RNA) compared to the signal intensity of the 156 bp protected fragment representing the alternatively spliced transcripts, human cerebellar and kidney transcripts originate primarily from the P1 promoter region. The variation in length of the protected fragments is similar to the results of RNase protection assays using the Ex 1 probe (see Figure 5B, lane 2) and suggests that transcription initiates from multiple sites within exon 1. Interestingly, the sites of transcription initiation within the P1 region varies in these tissues, which suggests that tissues may regulate transcription from the P1 region differently or that the P1 region contains more than one promoter element. In contrast, lung tissue contains less abundant transcripts originating from the P1 region (lane 4). The pattern of promoter usage by KB cells (Figure 7A and 7B, lane 2) resembles that observed in human lung tissue.

DISCUSSION

In the current study, we sought (1) to determine the organization of the human α folate receptor gene; (2) to determine the molecular basis for the heterogeneity in length and sequence of the 5′ termini of the α hFR cDNAs; and (3) to identify potential *cis*-regulatory elements within the 5′ flanking sequence of the gene involved in the transcriptional regulation of the human α folate receptor gene.

Structure of the α hFR Gene. Seven human α hFR cDNAs have been reported (Elwood, 1989; Sadasivan & Rothenberg, 1989; Lacey et al., 1989; Campbell et al., 1991; Coney et al., 1991) that are characterized by unique 5′

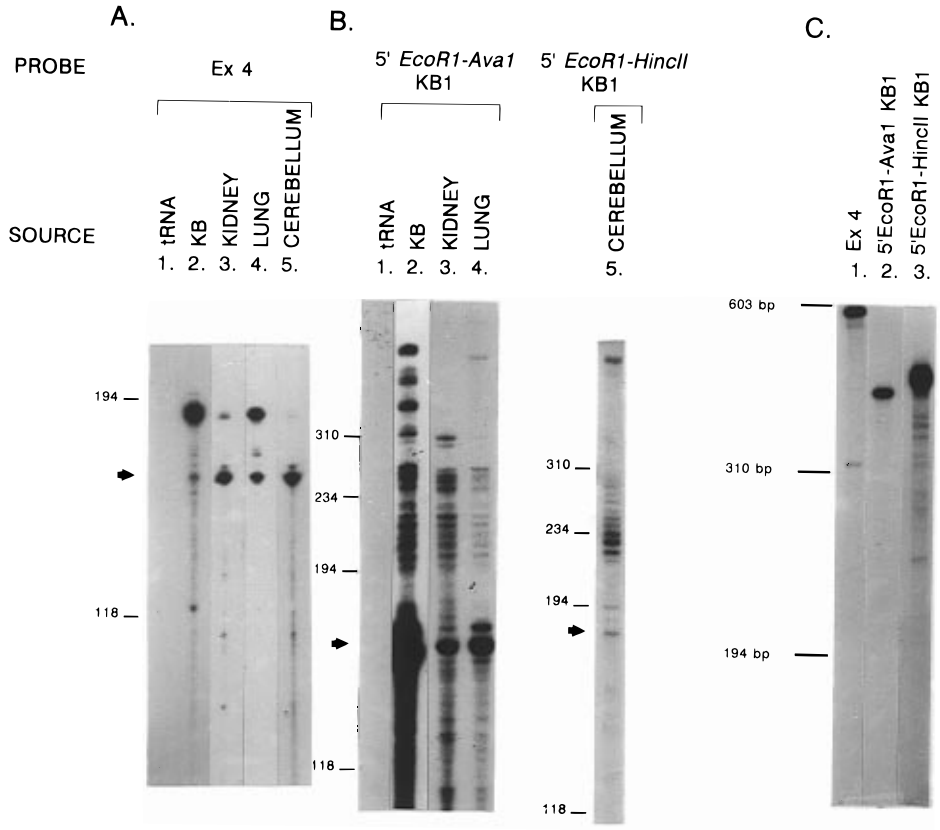

FIGURE 7: Tissue-specific expression of α hFR transcripts. Total RNA (30 μg) from the sources indicated above each lane was hybridized with the genomic Ex 4 riboprobe (see A) or with the 5′ *Eco*RI-*Ava*I or 5′ *Eco*RI-*Hin*cII KB1 riboprobes (B) and analyzed by RNase protection assays. Panel C contains an autoradiograph of undigested riboprobes. For each probe, the arrowhead indicates the location of the potential splice site within exon 4. Size markers included a sequencing reaction and end-labeled *Hae*III-digested ϕX174 DNA.

termini. We considered the possibilities that these cDNAs were encoded by homologous genes or that they originated from a single gene via alternative splicing or from multiple promoters. To determine the molecular basis of this heterogeneity, we determined the structure of the α hFR gene by Southern analysis and sequence analysis of two independent human genomic clones (S18 and S28) that contain the α hFR gene. Several observations from these experiments support the notion that the α hFR cDNAs are encoded by one gene. First, the divergent 5′ terminal sequences of the reported cDNAs, except for IGROV 51 cDNA, are contained in putative exons 1, 2, 3, and 4 of the α hFR gene (see Figures 2C and 3). Second; exon 4 contains a consensus splice acceptor site at the precise nucleotide where the sequences of the cDNAs diverge. Third, each of the putative 5′ exons are flanked by consensus splice site sequences (Figure 3). Fourth; the 3′ 66 nucleotides of exon 3 are homologous to sequence within the 5′ UTR of CaCo-2 and IGROV1 31 and 4/6 cDNAs and are flanked by a consensus splice acceptor site. Finally, the sizes of genomic *Bam*HI and *Hin*dIII restriction fragments containing the common ORF and 3′ UTR are virtually identical to those observed by Southern analysis of normal and KB cell genomic DNA (Figure 1) and genomic DNA of other human cell lines (Knight et al., 1989).

The origin of the 5′ terminus (99 bp) of IGROV1 51 (Coney et al., 1991) and its location relative to α hFR gene are unclear from these studies. Although the 46 nucleotides upstream from the alternative splice within exon 4 (Figure 3) are homologous to residues −54 to −9 of IGROV1 51 cDNA, the genomic sequence does not contain a consensus splice acceptor site surrounding the point of sequence divergence at −54. Furthermore, the 5′ terminal 99 nucleotides of IGROV1 51 cDNA are not present within approximately 5000 bp upstream from exon 1 of the α hFR gene (P. C. Elwood, unpublished data). It is possible that the IGROV1 51 cDNA is encoded by a polymorphic gene, represents an aberrantly spliced transcript, is a fusion transcript resulting from a genomic rearrangement in IGROV1 cells, or represents a ligation artifact.

The organization and sequence of the 3′ terminal exons of the α hFR gene are very similar to those of the β hFR gene (Page et al., 1993; Sadasivan et al., 1994), and the murine FBP1 (Brigle et al., 1992) and FBP2 (Brigle et al., 1994) genes. In each of these genes, the open reading frame is encoded by four 3′ exons that are homologous to putative exons 4 through 7 of the α hFR gene. Although the exons encoding the amino- and carboxyl-termini of the folate receptors (e.g., exons 4 and 7 of the α hFR gene) are more heterogeneous in terms of size and sequence, the exon junctions occur at identical sites between conserved residues in each of the FRs when their cDNA or amino acid sequences are aligned to maximize homology (Chung et al., 1994). Furthermore, the sizes of exons of the β hFR and murine FBP genes that are homologous to exons 5 and 6 of the α hFR gene are identical except for exon 3 of the murine FBP2 gene that is three nucleotides shorter than exon 6 of the α hFR gene. These conserved molecular features are consistent with the extensive homology observed at the protein and cDNA level and the fact that the folate receptors share biochemical properties (Antony, 1992). The conserved features also support the hypothesis (Page et al., 1993) that this gene family arose by gene reduplication and suggest that each of the exons encodes a peptide that has important

functional properties or that contributes to the overall function or stability of the receptor.

In contrast, the organization of the 5′ termini of the α hFR gene is different from that of the β hFR gene as expected given the differences observed in the relative abundances of their transcripts in normal and malignant tissues (Page et al., 1993; Ross et al., 1994). The β hFR gene (Page et al., 1993) contains a single exon that encodes the 5′ UTR of the β hFR cDNA while the heterogeneous 5′ termini of the α hFR cDNAs are encoded by at least four exons. These differences in the 5′ flanking sequences and the promoter regions (as discussed below) of the α and β hFR genes likely have important implications for the independent expression of these genes in tissue and cell lines as determined by Northern analysis (Page et al., 1993) and quantitative RT-PCR (Ross et al., 1994).

Expression and Regulation of the α hFR. The hFRs (Antony, 1992; Kane & Waxman, 1989; Weitman et al., 1992a,b) and their transcripts (Page et al., 1993; Ross et al., 1994) are widely expressed in human tissues and tissue culture cells. In the present study, we determined the relative abundance and structure of each of the reported α hFR transcripts in selected normal tissues and tissue culture cells, and identified potential regulatory regions upstream from the gene by means of RNase protection assays. We found that the expression of the α hFR gene transcript is variable in these samples (see Figures 5 and 7). These results are qualitatively similar to recent studies of α hFR expression employing Western (Weitman et al., 1992a,b) and Northern analysis (Page et al., 1993) or quantitative RT-PCR (Ross et al., 1994) and suggest that expression of the α hFR is regulated and not constitutive. We also observed that the most abundant α hFR transcripts originate from exon 1 and from exon 4, corresponding to KB1 and KB4 cDNAs, respectively, whereas transcripts corresponding to the other α hFR cDNAs are expressed at low levels (e.g., CaCo-2 and IGROV1 4/6 and 31) or are not detectable under these conditions. Furthermore, although the KB1 and KB4 transcripts are frequently co-expressed, their expression appears to be tissue specific (see Figure 7). For example, mRNAs transcribed from within exon 1 are the most abundant isoform expressed by human cerebellar and kidney tissue. In contrast, the mRNA transcribed from exon 4 is the most abundant isoform expressed by human KB cells and normal lung tissue. Thus, the KB1 and KB4 transcripts appear to be the most commonly expressed cDNA isoforms in the samples studied and are expressed in a tissue-specific manner.

The results of RNase protection assays suggested that the α hFR gene contained promoters upstream from exons 1 and 4. To test this hypothesis, we measured the functional promoter activity of sequences flanking these exons using a standard reporter gene assay. We discovered that sequences flanking exon 1 and exon 4, designated herein as P1 and P4 promoters, activated CAT transcription while the sequences flanking exons 2 and 3 did not exhibit detectable promoter activity in KB or HeLa cells. The differences in relative transcript abundance of the KB1 and KB4 isoforms among selected tissues and tissue culture cells and the presence of two independent promoters flanking exons 1 and 4 strongly support the notion that the α hFR mRNAs are transcribed from two independent promoters that are activated in tissue- or cell-specific fashion. The overall structure of the α hFR gene is similar to other genes such as the retinoic acid receptor α (Leroy et al., 1991) and β (Zelent et al., 1991) genes, α amylase gene (Hagnenbuchle et al., 1981), myosin alkalai light chain gene (Leff & Rosenfeld, 1986), IGF-II gene (Frunzio et al., 1986), IGF-I gene (Fu et al., 1991), and adenomatous polyposis coli gene (Thliveris et al., 1994) in which transcripts containing heterogeneous 5′ termini are generated via alternative splicing of 5′ exons or originate from independent tissue-specific promoters.

Transcription of eukaryotic genes may be regulated in a tissue-specific manner via activation of tissue-specific promoters or through other tissue-specific elements such as enhancers or suppressors (Andrin & Spencer, 1994). In order to understand the overall regulation of the α hFR gene, we are interested in further defining the elements of the α hFR basal promoters. We have recently shown that sequence of the P4 promoter region isolated from KB cells, a cell line that expresses high levels of the KB4 transcript (Figure 5A) and hFR protein (Antony, 1992), is identical to the sequence of the normal α hFR gene and that the basal P4 promoter activity is contained within sequence (∼200 bp) immediately upstream from the transcriptional start site (Saikawa et al., 1995). The sequence of the P4 promoter contains three clustered GC boxes that bind *Sp1* or *Sp1*-related nuclear proteins contained in nuclear extracts prepared from HeLa and KB cells and an INR region at the transcriptional start site. Each of these elements is required for optimum promoter activity. The minimal P1 promoter has not been characterized; however, the results reported herein indicate that P1 promoter is structurally and functionally distinct from the P4 promoter. In contrast to the single transcriptional initiation site of the P4 promoter, transcription from the P1 promoter initiates from multiple sites within exon 1 (Figures 5 and 7). The sequence upstream from the P1 promoter transcription start sites contains several GC boxes and potential *Sp1* binding sites, two TATA boxes, a CAAT box, and other potential *cis* elements as shown in Figure 3. The stuttering of transcription initiation, a feature of TATA-less promoters (Lewin, 1990), from the P1 promoter suggests that these TATA boxes are not functional or, alternatively, that the P1 promoter region contains more than one promoter element. In contrast to the P1 and P4 promoters of the α hFR gene, the β hFR gene contains a single promoter (Sadasivan et al., 1994) that contains an array of three GA binding protein motifs immediately upstream from the transcriptional start site. Similar to the α hFR promoters, the β hFR promoter contains an *Sp1* binding site and lacks canonical eukaryotic promoter elements. We are currently determining the components of the minimal P1 promoter and investigating the regulatory elements of the α and β hFR genes that result in the tissue-specific expression of their transcripts.

These results demonstrate that the molecular basis for the heterogeneity of the 5′ termini of the α hFR cDNAs includes transcription from at least two promoters and alternative splicing of 5′ exons. Although the sequence divergence within the 5′ UTRs of the α hFR cDNAs does not alter the protein structure, these differences may modulate the net expression of the α hFR at a post-transcriptional level. The functional relevance of 5′ UTRs in human RNA in general (Kozak, 1992) and of different 5′ UTRs such as observed in the α hFR isoforms is unclear. It is possible that the differences in 5′ UTRs may be important in post-transcriptional regulation such as transcript processing and stability or translational efficiency. Long UTRs such as observed in

the KB1 and SKOv3 18 α hFR cDNAs are relatively uncommon in eukaryotic transcripts (Lowe et al., 1987) and are more likely to form secondary structures. These stable structures usually negatively affect translational efficiency (Frunzio et al., 1986). In preliminary experiments, we have observed that the KB4 transcript is more efficiently translated than the longer KB1 transcript *in vitro* and *in vivo*. Thus, the regulated expression of the α hFR protein may also involve post-transcriptional mechanisms.

In conclusion, the α hFR gene is a complex transcriptional unit that generates diverse transcripts via simple or complex alternative splicing involving three 5′ exons downstream from at least two independent promoters that are activated in a tissue specific manner. These studies form the basis for further analysis of the P1 and P4 promoter regions and their flanking sequences to characterize the minimal promoters and other regulatory elements of the α hFR gene. The elucidation of the functional elements of each promoter responsible for the observed tissue-specific expression of the α hFR gene and its transcripts may ultimately shed light on the *cis*- and *trans*-acting elements regulating tissue-specific gene expression and the factors regulating alternative splice choices in human tissues.

REFERENCES

Andrin, C., & Spencer, C. (1994) *Biochem. Cell Biol. 72*, 377−380.

Antony, A. (1992) *Blood 79*, 2807−2820.

Antony, A. C., Kane, M. A., Portillo, R. M., Elwood, P. C., & Kolhouse, J. F. (1985) *J. Biol. Chem. 260*, 14911−14917.

Brigle, K. E., Westin, E. H., Houghton, M. T., & Goldman, I. D. (1992) *J. Biol. Chem. 267*, 22351−22355.

Brigle, K. E., Seither, R. L., Westin, E. H., & Goldman, I. D. (1994) *J. Biol. Chem. 269*, 4267−4272.

Buist, M. R., Kenemans, P., Hollander, W., Vermorken, J. B., Molthoff, C. J. M., Burger, C. W., Helmerhorst, T. J. M., Baak, J. P. A., & Roos, J. C. (1993) *Cancer Res. 53*, 5413−5418.

Campbell, I. G., Jones, T. A., Foulkes, W. D., & Trowsdale, J. (1991) *Cancer Res. 51*, 5329−5338.

Chung, K. N., Saikawa, Y., Paik, T. H., Dixon, K. H., Mulligan, T., Cowan, K., & Elwood, P. C. (1993) *J. Clin. Invest. 91*, 1289−1294.

Chung, K. N., Paik, T. H., Roberts, S., Kim, C. H., Kirassova, M., Weinstein, J. N., Trepel, J. B., & Elwood, P. C. (1994) *Arch. Biochem. Biophys. 315*, 407−414.

Coney, L. R., Tomassetti, A., Carayannopoulos, L., Frasca, V., Kamen, B. A., & Zurawski, V. R. (1991) *Cancer Res. 51*, 6125−6132.

Cullen, B. R. (1987) *Methods Enzymol. 152*, 684−704.

Deutsch, J. C., Elwood, P. C., Portillo, R. M., Macey, M. G., & Kolhouse, J. F. (1989) *Arch. Biochem. Biophys. 274*, 327−337.

Elwood, P. C. (1989) *J. Biol. Chem. 264*, 14893−14901.

Elwood, P. C., Kane, M. A., Portillo, R. M., & Kolhouse, J. F. (1986) *J. Biol. Chem. 261*, 15416−15423.

Franklin, W. A., Waintrub, M., Edwards, D., Christensen, K., Prendegrast, P., Woods, J., Bunn, P. A., & Kolhouse, J. F. (1994) *Int. J. Cancer 8*, 89−95.

Frunzio, R., Chiariotti, L., Brown, A. L., Graham, D. E., Rechler, M. M., & Bruni, C. B. (1986) *J. Biol. Chem. 261*, 17138−17149.

Fu, L., Ye, R., Browder, L. W., & Johnston, R. N. (1991) *Science 251*, 807−810.

Hagnenbuchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauen, P. K., & Young, R. A. (1981) *Nature 289*, 643−646.

Hsueh, C.-T., & Dolnick, B. J. (1993) *Biochem. Pharmacol. 45*, 2537−2545.

Hsueh, C. T., & Dolnick, B. J. (1994) *Biochem. Pharmacol. 47*, 1019−1027.

Kamen, B. A., & Capdevila, A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5983−5987.

Kane, M. A., Elwood, P. C., Portillo, R. M., Najfeld, V., Finley, A., Waxman, S., & Kolhouse, J. F. (1988) *J. Clin. Investig. 81*, 1398−1406.

Kane, M. A., Portillo, R. M., Elwood, P. C., Antony, A. C., & Kolhouse, J. F. (1986) *J. Biol. Chem. 261*, 44−49.

Kane, M. A., & Waxman, S. (1989) *Lab. Invest. 60*, 737−746.

Kingston, R. E. (1991) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K., Eds.) pp 4.8.1−4.8.2, Greene Publishing Associates and Wiley-Interscience, New York.

Knight, C. B., Elwood, P. C., & Chabner, B. A. (1989) *Adv. Enzyme Regul. 29*, 3−12.

Kozak, M. (1992) *Annu. Rev. Cell Biol. 8*, 197−225.

Lacey, S. W., Sanders, J. M., Rothberg, K. G., Anderson, R. G. W., & Kamen, B. A. (1989) *J. Clin. Invest. 84*, 715−720.

Lee, R. J., & Low, P. S. (1994) *J. Biol. Chem. 269*, 3198−3204.

Leff, S. E., & Rosenfeld, M. G. (1986) *Annu. Rev. Biochem. 55*, 1091−1117.

Leroy, P., Krust, A. Zelent, A., Mendelsohn, C., Garnier, J.-M., Kastner, P., Dierich, A., & Chambon, P. (1991) *EMBO J. 10*, 59−69.

Lewin, B. (1990) *Genes IV*, 4th ed., pp 545−550, Cell Press, Cambridge, MA.

Liu, J. M., Fujii, H., Green, S. W., Komatsu, N., Young, N. F., & Shimada, T. (1991) *Virology 182*, 361−364.

Locker, J., & Buzard, G. (1990) *DNA Sequence 1*, 3−11.

Lowe, W. L., Roberts, C. T., Lasky, S. R., & LeRoith, D. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8946−8950.

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265−275.

Luhrs, C., Raskin, C., Durbin, R., & Rothenberg, S. P. (1992) *Blood 80*, 242a.

Mantovani, L. T., Miotti, S., Menard, S., Canevari, S., Raspagliesi, F., Bottini, C., Bottero, F., & Colnaghi, M. I. (1994) *Eur. J. Cancer 30A*, 363−369.

Page, S. T., Owen, W. C., Price, K., & Elwood, P. C. (1993) *J. Mol. Biol. 229*, 1175−1183.

Ratnam, M., Marquardt, H. Duhring, J. L., & Freisheim, J. H. (1989) *Biochem. 28*, 8249−8254.

Ross, J. F., Chaudhuri, P. K., & Ratnam, M. (1994) *Cancer 73*, 2433−2443.

Sadasivan, E., & Rothenberg, S. P. (1989) *J. Biol. Chem. 264*, 5806−5811.

Sadasivan, E., Cedeno, M., & Rothenberg, S. P. (1992) *Biochim. Biophys. Acta 1131*, 91−94.

Sadasivan, E., Cedeno, M. M., & Rothenberg, S. P. (1994) *J. Biol. Chem. 269*, 4725−4735.

Saikawa, Y., Knight, C. B., Saikawa, T., Page, S. T., Chabner, B. A., & Elwood, P. C. (1993) *J. Biol. Chem. 268*, 5293−5301.

Saikawa, Y., Price, K., Hance, K. W., Chen, T. Y., & Elwood, P. C. (1995) *Biochemistry 34*, 9951−9961.

Sanger, F., Nicklen, S., & Coulson, R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463−5467.

Shen, F., Ross, J. F., Wang, X., & Ratnam, M. (1994) *Biochemistry 33*, 1209−1215.

Thliveris, A., Samowitz, W., Matsunami, N., Groden, J., & White, R. (1994) *Cancer Res. 54*, 2991−2995.

Weitman, S. D., Weinberg, A. G., Coney, L. R., Zurawski, V. R., Jennings, D. S., & Kamen B. A. (1992a) *Cancer Res. 52*, 6708−6711.

Weitman, S. D., Lark, R. H., Coney, L. R., Fort, D. W., Frasca, V., Zurawski, V. R., & Kamen, B. A. (1992b) *Cancer Res. 52*, 3396−3401.

Yu, B. W., Ichinose, I. Bonham, M. A., & Zajac-Kaye, M. (1993) *J. Biol. Chem. 268*, 19586−19592.

Zelent, A., Mendelsohn, C., Kastner, P., Krust, A., Garnier, J.-M., Ruffenach, G., Leroy, P., & Chambon, P. (1991) *EMBO J. 10*, 71−81.

BI962070H