Structural and Functional Analysis of the Human KB Cell Folate Receptor Gene P4 Promoter: Cooperation of Three Clustered Sp1-Binding Sites with Initiator Region for Basal Promoter Activity

Yutaka Saikawa, Kirsten Price, Kenneth W. Hance, Trina Y. Chen, and Patrick C. Elwood*

Medicine Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received February 2, 1995; Revised Manuscript Received June 1, 1995[⊗]

ABSTRACT: The human folate receptors (hFRs) are important in the cellular accumulation of folates and antifolates. We described the structure of the human KB cell FR (hFR-KB) gene and identified two discrete promoter regions (P1 and P4) upstream from exons 1 and 4, respectively (Elwood et al., 1993). To further understand the molecular basis of hFR expression, we have now analyzed the basal transcription of the P4 promoter localized upstream of a major transcription start site. The sequence upstream from exon 4 contains several potential transcriptional factor-binding sites and a consensus initiator region sequence at the transcription start site but does not contain canonical TATA or CAAT boxes. While deletion of a 5′ flanking sequence from nt −1023 to nt −605 of P4 promoter region decreases the luciferase reporter gene expression in KB cells to 54−70% of control construct, the removal of the sequence between nt −292 and nt −46 markedly decreases the activity to 3%. DNase I footprints and competitive mobility shift and supershift mobility assays indicate that Sp1 or Sp1-related nuclear protein(s) bind to three clustered GC-rich regions within the sequence between nt −292 and nt −46 of the hFR-KB P4 promoter. Both *in vitro* and *in vivo* analyses of the expression of promoter constructs containing site-specific mutation(s) of these three Sp1-binding sites and initiator sequence demonstrate that each of three Sp1 sites and the initiator sequence are required for optimum promoter activity and that they interact cooperatively in this P4 promoter of the hFR-KB gene.

The human folate receptors (hFR)[1] are a family of membrane-bound glycoproteins that share biochemical (Kane & Waxman, 1989; Henderson, 1990; Antony, 1992) and molecular (Elwood, 1989; Sadasivan & Rothenberg, 1989; Ratnam et al., 1989; Lacey et al., 1989; Campbell et al., 1991; Coney et al., 1991; Page et al., 1993; Shen et al., 1994) properties. These proteins mediate the transport of folates/antifolates [reviewed in Antony (1992), Henderson, 1990] and are important pharmacologic determinants of antifolate cytotoxicity (Saikawa et al., 1993; Chung et al., 1993). The hFRs are widely distributed in normal fetal (Page et al., 1993) and adult tissues as well as malignant tissues (Page et al., 1993; Elwood et al., 1993; Weitman et al., 1992a,b). The molecular structures and organization of the hFR gene family in a tandem array on chromosome 11q13 (Ragoussis et al., 1992) are complex. The primary structures of three hFR gene family members [KB cell (hFR-KB or hFR-α), placental (hFR-P or hFR-β), and FR-γ] have been deduced from the nucleotide sequences of cDNAs isolated from placenta (Elwood, 1989; Ratnam et al., 1989; Page et al., 1993), human tissue culture cells (Elwood, 1989; Sadasivan & Rothenberg, 1989; Lacey et al., 1989; Campbell et al., 1991; Coney et al., 1991), and human hematopoietic tissue (Shen et al., 1994). The hFRs are homologous glycoproteins that are immunologically cross-reactive (Kane & Waxman, 1989; Henderson, 1990; Antony, 1992) and evolutionarily conserved (Page et al., 1993). Despite these similarities, the hFR-KB and the hFR-P exhibit unique ligand-binding stereospecificities and are encoded by independent genes that are expressed in a restricted, independent, tissue-specific-manner (Page et al., 1993; Elwood et al., 1993). Recently, the organization and sequence of the hFR-P gene and the three hFR pseudogenes have been reported (Page et al., 1993; Sadasivan et al., 1994).

The organization of the gene(s) encoding the hFR-KB transcripts is even more complex on the basis of reported differences in the sequences of eight KB cell cDNA isoforms (Elwood, 1989; Sadasivan & Rothenberg, 1989; Lacey et al., 1989; Campbell et al., 1991; Coney et al., 1991; Elwood et al., 1993). We recently reported the structure of two polymorphic genomic clones containing the full hFR-KB gene that were isolated from a human lymphocyte genomic library (Elwood et al., 1993). On the basis of comparison of the nucleotide sequence of 5′ untranslated regions (UTR) of KB cell cDNA isoforms to the hFR-KB gene sequences, we observed that the 5′ UTR heterogeneity of reported hFR-KB cDNAs results from alternative splicing involving exons 1−3, while another novel, abundant cDNA isoform (Elwood et al., 1993) is encoded by a sequence upstream from the splice site within exon 4 utilized in 5′ alternative splicing. In addition, at least two discrete promoters are present in the hFR-KB gene adjacent to exons 1 and 4 (Elwood et al., 1993). Interestingly, hFR-KB transcripts in KB cells and lung tissue appear to originate primarily from the promoter located upstream of exon 4 (designated P4), whereas the transcripts in normal kidney and cerebellum tissues are driven

* Address correspondence to this author at: Medicine Branch, Building 10, 12N226, National Institutes of Health, Bethesda, MD 20892. Tel/Fax: 301-496-6840/301-402-0172.

[⊗] Abstract published in *Advance ACS Abstracts,* July 15, 1995.

[1] Abbreviations: hFR, human folate receptor; hFR-KB, human KB cell folate receptor; Sp1, transcription factor Sp1; INR, initiator region; Δ, mutation; PCR, polymerase chain reaction; bp, base pair; nt, nucleotide.

Table 1: Oligonucleotide Sequences

oligonucleotides	sequence[a]	coordinates[b]
Sp1	GATCGATCGGGGCGGGGCGAT	(consensus)
Sp1.1	TCAAGCCAGGGAGGGGTGGTGTCTAATCCTACCT	−192/−159
Sp1.2	ACTGAGGGAGATGGGGGCAGGGCTCTATCTGCCCCA	−140/−105
CACC/Sp1.3	GGTTCCGTCCAGGCCCCACCCTCCTGGAGCCCTG	−104/−71
ΔSp1.1	AGCCAGGGAGAAATGGTGTCTAATCCTACC	−189/−160
ΔSp1.2	GGAGATGGGGAAAGAAACTCTATCTGCCCCA	−134/−105
ΔCACC/Sp1.3	TCCGTCCAGGCAACAACCTCCTGGAGCCCT	−101/−72
ΔINR	CCACCTCCGCGTAGCTTGGTGCCACTGACC	−54/−15

[a] The sense strand sequences of the oligonucleotides are shown in a 5′ to 3′ orientation. The sequences of potential transcription factor-binding sites and INR region are underlined. Mutated residues are typed in bold. [b] The numbers are relative to the translation start site at +1.

primarily by the promoter located upstream of exon 1 (designated P1) (Elwood et al., 1993). These observations suggest that the differential expression of the hFR-KB gene is at least in part regulated by means of tissue differential promoter utilization.

In this paper, we now describe the functional elements of the hFR-KB P4 promoter. We chose to focus on the P4 promoter since the most abundant transcript expressed in KB cells and contained in RNAs isolated from several human tissues originates from the P4 promoter (Elwood et al., 1993). The nucleotide sequence of the hFR-KB gene cloned from KB cells is identical to the normal hFR-KB gene sequence (Elwood et al., 1993). The upstream sequence of exon 4 of the gene does not contain recognizable TATA or CAAT boxes but does contain several GC boxes and potential binding sites for transcription factors. We show that the 5′ flanking sequence (approximately 200 bp) upstream from the translation start site contains the basal P4 promoter of the hFR-KB gene and that each of three clustered GC boxes within this region are critical elements for transcriptional activation through a Sp1 or Sp1-related factor(s) binding to these sites. Finally, we also demonstrate that the sequence surrounding the major transcription start site in exon 4 is homologous to the initiator region and cooperates with upstream GC box elements for activity of the hFR-KB gene P4 promoter.

EXPERIMENTAL PROCEDURES

General Reagents and Enzymes. Wild type human nasopharyngeal epidermoid carcinoma (KB) cells, passage number 369, were obtained from American Type Culture Collection (Rockville, MD). HeLa cells and pCMV-β-gal vector containing the β-galactosidase gene downstream from a CMV promoter were gifts from Dr. Maria Zajac-Kaye (Bethesda, MD). Plasmids (pGEM4Z, pGL-2basic, and pGL-2control) were purchased from Promega (Madison, WI). All radionucleotides were obtained from Amersham Corp. All restriction enzymes, RNasin, RQ1 DNase, calf intestinal alkaline phosphatase, DNA polymerase (Klenow fragment), T4 polynucleotide kinase, T4 DNA ligase, and AMV reverse transcriptase were from Promega. Protease inhibitors were from Sigma. DNA-alternating copolymers [poly(dI-dC)-poly(dI-dC)] and Sure Track Footprinting kit were from Pharmacia LKB Biotechnology Inc. The HeLa cell nuclear extract and recombinant transcription factor Sp1, purified from HeLa cells (Kadonada et al., 1987), were purchased from Promega. Rabbit antisera directed against residues 520−538 of human Sp1 (Jackson & Tijan, 1988) and against human TFIID (Roeder, 1991) were obtained from Santa Cruz Biotechnology (Santa Cruz, CA). Oligonucleotides were synthesized on a model 392 Applied Biosystems Inc. DNA synthesizer. PCR amplifications were performed in a thermal cycler (Perkin-Elmer Cetus Instruments) with reagents supplied in a Geneamp PCR kit (Perkin-Elmer Cetus Instruments). All other reagents were of reagent grade or higher and purchased from Sigma, Fisher, or T. J. Baker (Phillipsburg, NJ).

Construction of hFR Promoter-Luciferase Plasmids. The hFR-KB gene P4 promoter deletion mutants used for transfection experiments are shown in Figure 2. The 5′ deletion constructs with termini at −1023, −840, and −605 (numbers relative to the translation start site at +1) were generated by PCR amplification from the hFR genomic clone (Elwood et al., 1993) using sense primers with coordinates −1023/−997, −840/−824, and −605/−583, respectively, and an antisense primer with coordinates +119/+145. Respective PCR products were sucloned into the *Sac*1 and *Nhe*1 multicloning sites of pGL-2basic by standard techniques. Each PCR cycle ($n = 16$) consisted of a denaturation step (93 °C for 2 min), an annealing step (65 °C for 1 min), and an extension step (72 °C for 2 min). The final extension step was 5 min to insure complete extension of PCR products. The internal deletion constructs were created by recombinant PCR methods (Higuchi, 1990) from the −1023/+145 (KB4-ScA, see Figure 2) parent construct. The 39−42-mer sense and antisense inside primers were designed to delete the restriction fragments *Pst*1/*Hae*III, *Hae*III/*Spe*1, *Spe*1/*Ava*II, and *Ava*II/*Bsm*1 with coordinates −840/−605, −605/−420, −420/−292, and −292/−46, respectively. The sense and antisense outside primers were synthesized with coordinates −1023/−997 and +119/+145, respectively. The primary PCR cycle ($n = 16$) was identical to that described above. The secondary PCR cycle ($n = 10$) consisted of a denaturation step (93 °C for 2 min), an annealing step (50 °C for 1 min), and an extension step (72 °C for 2 min) for the first two cycles followed by a denaturation step (93 °C for 2 min), an annealing step (65 °C for 1 min), and an extension step (72 °C for 2 min). The final extension step was 10 min to insure complete extension of PCR products. The site-specific promoter mutation constructs designated as ΔSp1.1, ΔSp1.2, ΔCacc/ΔSp1.3, and ΔInr and the combination of these mutations (shown in Figure 6) were also created by recombinant PCR (Higuchi, 1990). The procedures were identical as described above except for the primers (see Table 1). For each site-specific mutation, a set of complementary primers was synthesized. The sequence of each primer is contained in Table 1, and the mutated residues are in bold type. The PCR products from secondary amplification were gel-purified and subcloned into the *Sac*1/*Nhe*1 site of pGL-2basic. The sequences of each construct including the deletions, specific mutations, and subcloning

site of each construct were verified by sequencing (Sanger et al., 1977) prior to transfection.

Transient Transfection Assay. KB and HeLa cells were cultured under low-folate conditions as described (Saikawa et al., 1993; Elwood et al., 1993). Cells at a density of 5×10^5/60 mm of tissue culture dish were plated 24 h prior to transfection in 5 mL of minimal essential media without added folic acid (DMEM) (Gibco) containing L-glutamine, Earle's salts, and 10% fetal calf serum (Biofluid Inc., Rockville, MD). The cells were cotransfected with 15 μg of hFR promoter-luciferase construct and 3 μg of pCMV-β-gal, which served as the internal control, by the calcium phosphate method (Sambrook et al., 1989). After 48-h incubation with DNA, the transfected cells were scraped into the reporter lysis buffer provided in the luciferase assay system (Promega) and centrifuged at 14000*g* for 5 min at 4 °C. The supernatant was assayed for luciferase and β-galactosidase activity and total protein concentration. Luciferase and β-galactosidase activity were determined by a chemiluminescent assay using the reagents from the luciferase assay system (Promega) and a Galacto-Light kit (Tropix Inc., Bedford, MA), respectively, and a monolight 2010 luminometer (Analytic Luminescence Laboratory) following the protocol provided by each vender. The luciferase activity was corrected for protein measured by the Bradford (Bio-Rad) assay and normalized to β-galactosidase activity to correct for differences in transfection efficiency. All experiments were performed in duplicate in at least three separate experiments.

Preparation of Nuclear Extract. Nuclear extracts from KB cells cultured in DMEM (see above) were prepared by standard methodology (Ausubel et al., 1991). All buffers and procedures were at 4 °C. KB cell pellets were resuspended in hypotonic buffer [10 mM Hepes, pH 7.9, containing 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM phenylmethanesulfonyl fluoride (PMSF), and 0.5 mM dithiothreitol (DTT)] and placed in ice water for 15 min. Then, the nuclei were isolated using a glass Dounce homogenizer, pelleted by centrifugation at 3300*g* for 15 min, and resuspended in $^1/_2$ volume of low-salt buffer (20 mM Hepes, pH 7.9, containing 25% glycerol, 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, and 0.5 mM DTT). To prepare nuclear extracts, $^1/_3$ volume of high-salt buffer (20 mM Hepes, pH 7.9, containing 25% glycerol, 1.5 mM $MgCl_2$, 1.2 M KCl, 0.2 mM EDTA, 0.2 mM PMSF, and 0.5 mM DTT) was added and the nuclei were incubated with continuous gentle mixing for 30 min. The nuclear extract was separated from nuclei by centrifugation at 25000*g* for 30 min and dialyzed for 5 h against 50 volumes of 20 mM Hepes, pH 7.9, containing 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, and 0.5 mM DTT. The dialyzed extract was stored in 50−100 μL aliquots at −80 °C. Protein concentrations were determined by the Bradford assay (Bio-Rad). Typical protein concentrations ranged from 2 to 5 μg/μL.

DNase I Footprinting Assay. Each restriction fragment [*Pst*1/*Hae*III (−840/−605), *Hae*III/*Spe*1 (−605/−420), *Spe*1/*Ava*II (−420/−292), or *Ava*II/*Bsm*1 (−292/−46)] was gel-purified, blunt-ended, and subcloned into the *Sma*1 cloning site of pGEM4Z. To synthesize probes, each fragment was excised from pGEM4Z by restriction at flanking *Sac*1 and *Bam*H1 sites. After gel purification of restriction fragments, the coding strand was radiolabeled with [α-^{32}P]dATP using the Klenow fragment of DNA polymerase I. KB cell nuclear extract (25−100 μg), or an equal amount of bovine serum albumin as the control, and radiolabeled probes (10 000 cpm) were incubated for 30 min at room temperature in 50 μL of binding mix [10 mM Tris-HCl, pH 7.5, containing 50 mM NaCl, 0.5 mM DTT, 10% glycerol, 0.05% Nonidet P-40, and 5 μg of poly(dI-dC)-poly(dI-dC)]. The digestion with DNase I was performed for 1 min at room temperature and terminated by adding 140 μL of stop solution (192 mM sodium acetate, 32 mM EDTA, 0.14% sodium dodecyl sulfate, and 64 μg/mL yeast RNA). DNA samples were isolated, resuspended in formamide loading buffer, and resolved on an 8% polyacrylamide−urea sequencing gel. A Maxam and Gilbert G + A sequencing reaction (Sambrook et al., 1989) of the target probe, performed using reagents provided in the Sure Track Footprinting kit (Pharmacia), was run in parallel to determine the footprinting sequences.

Mobility Shift Assay. To radiolabel probes corresponding to either strand, ^{32}P-end-labeled primers were added to the PCR reaction as described (Hooft van Huijsduijnen, 1992) and the probes were gel-purified. Complementary oligonucleotides containing potential binding sites for the transcription factor Sp1 and their site-specific mutations (Δ) (see Table 1) were synthesized. Equimolar quantities of complementary oligonucleotides were denatured in TE buffer (10 mM Tris, pH 8.0, and 1 mM EDTA) by heating at 100 °C for 2 min and annealed by cooling to room temperature. Two micrograms of nuclear extract or 20 ng of purified human Sp1 was incubated with radiolabeled probes (10 000 cpm) at room temperature for 30 min in 20 μL of binding mix [10 mM Tris-HCl, pH 7.5, containing 50 mM NaCl, 0.5 mM DTT, 10% glycerol, 0.05% Nonidet P-40, and 5 μg of poly-(dI-dC)-poly(dI-dC)] with or without cold competitors. In order to immunologically identify the protein component of the protein−DNA complexes, nuclear extracts were first incubated with probes for 15 min at room temperature followed by the addition of an aliquot (0.5 and 2 μL) of anti-human Sp1 or anti-human TFIID antibodies. The reaction was then incubated for 1 h at 4 °C and resolved on prerun (100 V at 4 °C for 1 h) 5% nondenaturing polyacrylamide gel (37.5:1, acrylamide:bisacrylamide) in 0.5× TBE at 20 mA for 1.5 h at 4 °C. Gels were dried and autoradiographed at −70 °C.

In Vitro Transcription Assay and Primer Extension Analysis. In vitro transcription reactions and primer extension analysis of the transcripts were carried out essentially as described (Kadonaga, 1990). The transcription reaction (25 μL) contained 20 mM Hepes, pH 7.9, 50 mM KCl, 6 mM $MgCl_2$, 0.5 mM DTT, 0.05 mM EDTA, 5% (v/v) glycerol, 0.4 mM each of four ribonucleoside triphosphates, 40 units of RNasin (Promega), 500 ng of supercoiled plasmid DNA, and 72 μg of HeLa cell nuclear extract (Promega). The reaction mixtures were incubated at 30 °C for 1 h followed by digestion of DNA templates with 1 unit of RQ1 DNase (Promega) for 10 min at 37 °C and the reactions stopped by addition of 175 μL of stop mix [0.3 M Tris-HCl, pH 7.4, containing 0.3 M sodium acetate, 0.5% (w/v) sodium dodecyl sulfate, 2 mM EDTA, and 5 μg/mL carrier transfer RNA (Sigma)]. Isolated RNA products were dissolved in 10 μL of reverse transcription buffer (50 mM Tris-HCl, pH 8.3, containing 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, and 0.5 mM spermidine), and an aliquot (100 000 cpm, 0.03 pmol) of end-labeled antisense primer with coordinates +75/+102 was added and annealed at 55 °C for 1 h. For primer extension, 10 μL of reverse transcriptase

FIGURE 1: Nucleotide sequence of the human KB cell folate receptor gene promoter P4 and 5′ flanking region. The genomic sequence (−1023−+145) of the hFR-KB gene includes the upstream intron and exon 4 sequences. The intronic and exonic sequences are designated with capital and lowercase letters, respectively. Residue numbers are relative to the atg translation start site (underlined) at +1 in exon 4. The asterisks denote the transcription start site at nt −45 and nt −44. The sequences of potential sites for known transcription factors and regulatory elements are indicated by brackets with the following abbrevations: Ap2, transcription factor Ap2; Sp1, transcription factor Sp1; HLH, helix−loop−helix transcription factors (or E box motif); ERE, estrogen/steroid receptor-binding element; CACC, CACC box motif; INR, initiator region. The palindromic motifs are shown by the opposing arrows.

mix (AMV-RT buffer, containing 25 mM each of four deoxyribonucleoside triphosphates, 40 units of RNasin, and 8 units of AMV-RT) was added and the samples were incubated at 42 °C for 1.5 h. Extended products were resolved through a 6% denaturing polyacrylamide−urea sequencing gel. The gels were dried and autoradiographed. Quantitation of signals was performed by PhosphorImager (Molecular Dynamics, Inc.) analysis. All experiments were carried out three independent times.

RESULTS

Human KB Cell Folate Receptor (hFR-KB) Gene Promoter (P4) and Its 5′ Flanking Sequence. To determine if human KB cells express high levels of hFR due to mutations, we have now amplified and sequenced the hFR-KB gene from KB cell genomic DNA. The nucleotide sequence is identical to that of the gene isolated from normal DNA (Elwood et al., 1993). Figure 1 contains the KB cell nucleotide sequence from nt −1023 to nt +145 relative to the translational start site. The position of the major transcriptional start site for the hFR-KB gene P4 promoter is indicated with the asterisks at nt −45 and nt −44. Although the P4 promoter does not contain a TATA box, transcription initiation from this promoter is precise in contrast to the multiple transcriptional start sites observed from the P1 promoter upstream from exon 1 of the hFR-KB gene (Elwood et al., 1993). The nucleotide sequence upstream from the P4 promoter region of the hFR-KB gene is rich in potential transcriptional regulatory elements. As shown in Figure 1, this region contains five potential binding sites for transcription factor Ap2 (Faisst & Meyer, 1992), five sites for Sp1 (Faisst & Meyer, 1992; Kadonaga et al., 1986), several E box motifs (indicated as HLH), which are known to bind a family of helix−loop−helix transcription factors important in tissue-specific gene expression (Blackwell & Weintraub, 1990), two sites (one-half of a consensus sequence) for the estrogen/steroid receptor-binding element (Faisst & Meyer, 1992), and a CACC motif (deBoer et al., 1988; Schule et al., 1988; Yu et al., 1991). Five palindromic motifs are found within this region as indicated by facing arrows. Furthermore, the nucleotide sequence surrounding the major transcription start site shows striking homology to the consensus sequences of reported initiator elements (Roeder, 1991).

Identification of Functional Regions for Basal Activity of the hFR-KB Gene P4 Promoter. To localize the sequences containing the basal P4 promoter of the hFR-KB gene, we constructed a series of 5′ terminal and internal deletion mutants as shown in Figure 2. The −1023/+145 hFR-luciferase construct (KB4-ScA) exhibited 50-fold higher promoter activity compared to the promoterless control plasmid. The deletion of 5′ flanking regions from nt −1023 to nt −605 (KB4-PA and KB4-HA) which contain five Ap2 sites and two Sp1 sites decreased promoter activity modestly (54.2 ± 11.1% and 69.6 ± 7.6%, respectively, relative to that of KB4-ScA construct). The internal deletions between nt −840 and nt −605 (KB4-ScA-PH235 Del), nt −605 and nt −420 (KB4-ScA-HS185 Del), and nt −420 and nt −292 (KB4-ScA-SA128 Del), which also contain several potential binding sites for transcription factors, did not affect the promoter activity (86.2 ± 7.9%, 103.0 ± 16.2%, and 91.3 ± 12.9% of control, respectively). However, deletion of the region between nt −292 and nt −46 (KB4-ScA-AB246 Del) significantly decreased promoter activity (3.1 ± 0.5% of control). This region includes two potential E motifs (HLH), two potential Sp1 sites, and a CACC motif. The transcription start site (nt −45 and nt −44) and the INR region are intact in the KB4-ScA-AB246 Del construct. These experiments suggest that the region between nt −1023 and nt −840 may contain an enhancer element(s) and indicate that the region located between nt −292 and nt −46 contains the necessary elements for supporting basal P4 promoter activity.

KB Cell Nuclear Extract Protects the Sequences for Transcription Factor Sp1 and a CACC/Sp1.3 Motif in the Basal P4 Promoter. DNase I footprinting assays (Figure 3) were utilized to identify the specific sequences within restriction fragments (see Figure 2) from the P4 promoter region that interact with potential trans-acting factors. KB cell nuclear extract (50 μg) did not protect any regions when the distal restriction fragments including *Pst*I/*Hae*III (−840/−605) and *Hae*III/*Spe*I (−605/−420) were used as probes (data not shown). However, four regions were protected

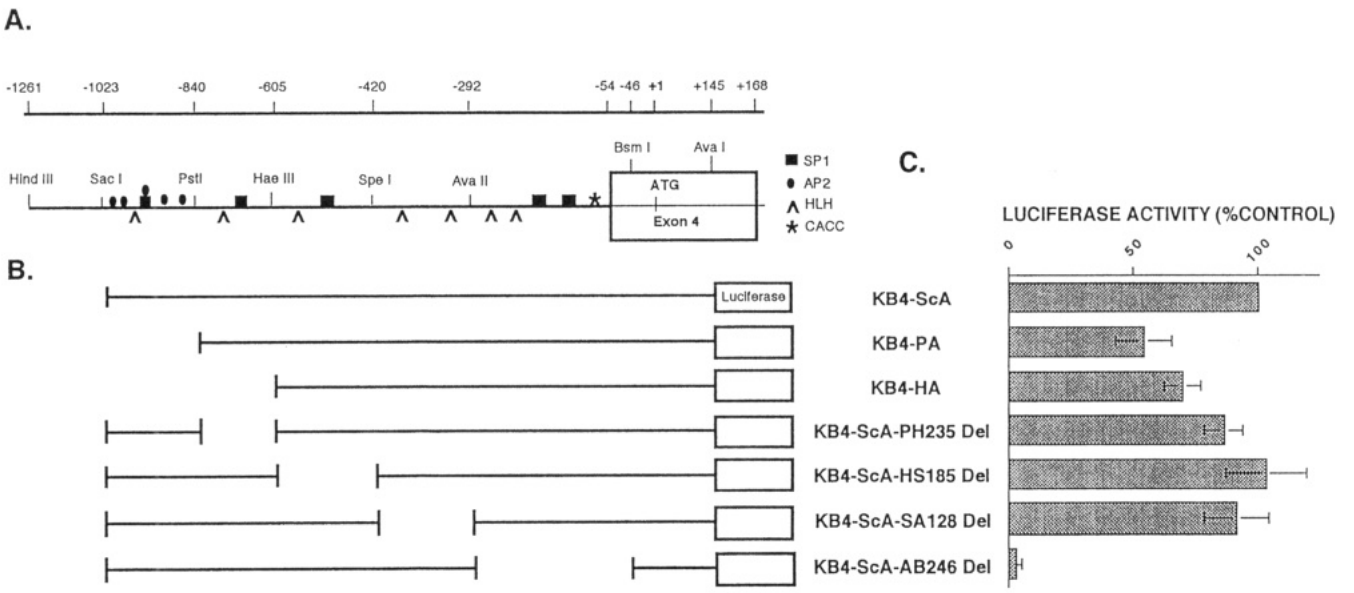

FIGURE 2: Promoter activity of hFR-KB gene P4 promoter deletion constructs. (A) Diagram of the hFR-KB gene promoter P4. The map coordinates are relative to the translation start site at +1 as indicated on the top bar. The relative locations of restriction enzyme sites and the potential sites for transcription factors and regulatory elements (Figure 1) are illustrated by the symbols that are defined in the legend. ERE sites and palindromic motifs are not shown. (B) 5′ Flanking deletion and internal deletion constructs schematically represented by brackets adjacent to the promoterless luciferase reporter gene (Luciferase) depicted by open boxes. The 3′ end of each promoter fragment is the *Ava*I restriction site at nt +145. (C) Analysis of the deletion hFR-KB P4 promoter luciferase constructs. Luciferase and β-galactosidase activities were measured by luminometry in KB cell extracts following cotransfection with hFR promoter-luciferase construct (15 μg) and pCMV-β gal plasmid (3 μg). Transfection efficiency was controlled for the measurement of β-galactosidase activities, and the luciferase activity is expressed as the percent control of activity in the KB4-ScA construct. The bar graph shows the mean of at least three independent transfections, and the bracket represents one standard deviation from the mean.

within the more proximal sequences between nt −420 and nt −46. The *Spe*I/*Ava*II (−420/−292) restriction fragment contained a 29-bp sequence between nt −401 and nt −373 (CCCAGATATACACTTGATTATTGGGTATA) (Figure 3A, lane 2). The protected region contains an E box motif (underlined) and is AT rich. Although the specificity of the KB cell nuclear extract binding to this region was confirmed by competitive mobility shift assays (data not shown), this region does not appear to be relevant to basal P4 promoter activity since the KB4-ScA-SA128 Del construct excluding this sequence exhibited promoter activity equivalent to the control (KB4-ScA) construct (Figure 2). The *Ava*II/*Bsm*I (−292/−46) restriction fragment contained three protected regions (Figure 3B, lane 6): the 23-bp protected region between nt −185 and nt −163 (AGG GAGGGGTGGTGTCTAATCCT) containing a potential transcription factor Sp1-binding site (underlined), the 27-bp protected region between nt −135 and nt −109 (GGGAGATGGGGGCAGGGCTCTATCTGC) containing a second potential Sp1-binding site (underlined), and the 23-bp protected region between nt −98 and nt −76 (GTCCAG GCCCCACCCTCCTGGAG) containing a CACC motif (underlined) that corresponds to a third Sp1 site on the antisense strand. Since the nucleotide sequences of these Sp1-binding sites are different from canonical Sp1-binding site sequence and from each other, we have designated these sites as Sp1.1, Sp1.2, and CACC/Sp1.3, respectively. These data, taken together with the observation that the deletion of the region between nt −292 and nt −46 from the P4 promoter construct resulted in loss of promoter activity (Figure 2), suggest that the transcriptional factor Sp1 and the CACC/Sp1.3 motif account for the transcriptional activation of the hFR-KB gene P4 promoter.

KB Cell Nuclear Factors Bound to the Basal Elements of the P4 Promoter Include Transcription Factor Sp1 and/or Sp1-Related Proteins. To further characterize the potential

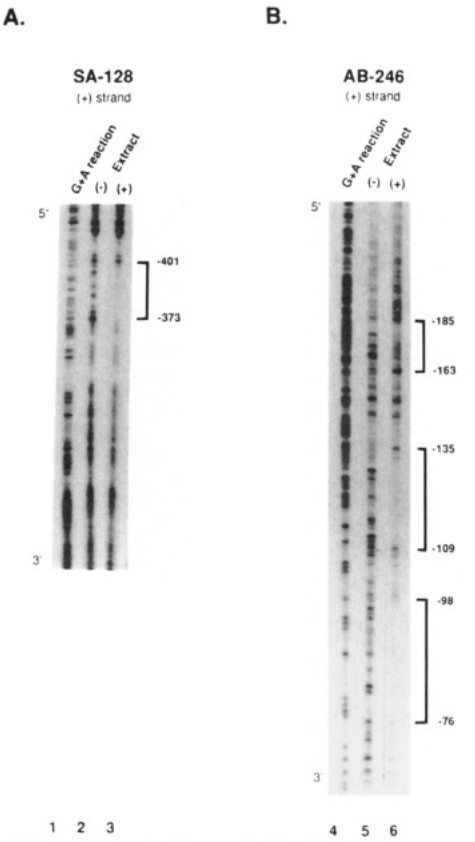

FIGURE 3: DNase I footprinting analysis of the P4 promoter regions. Radiolabeled promoter fragments with coordinates −420/−292 (SA-128, panel A) and −292/−46 (AB-246, panel B) were digested with DNase I in the presence of 50 μg of KB cell nuclear extract (lanes 3 and 6, respectively) or 50 μg of bovine serum albumin as control (lanes 2 and 5). The protected sites are bracketed; their positions relative to the translation start site are shown. The locations of the nucleotide sequences in the protected regions were determined from the Maxam−Gilbert (G + A reaction) sequence of the target probes (lanes 1 and 4) run in parallel.

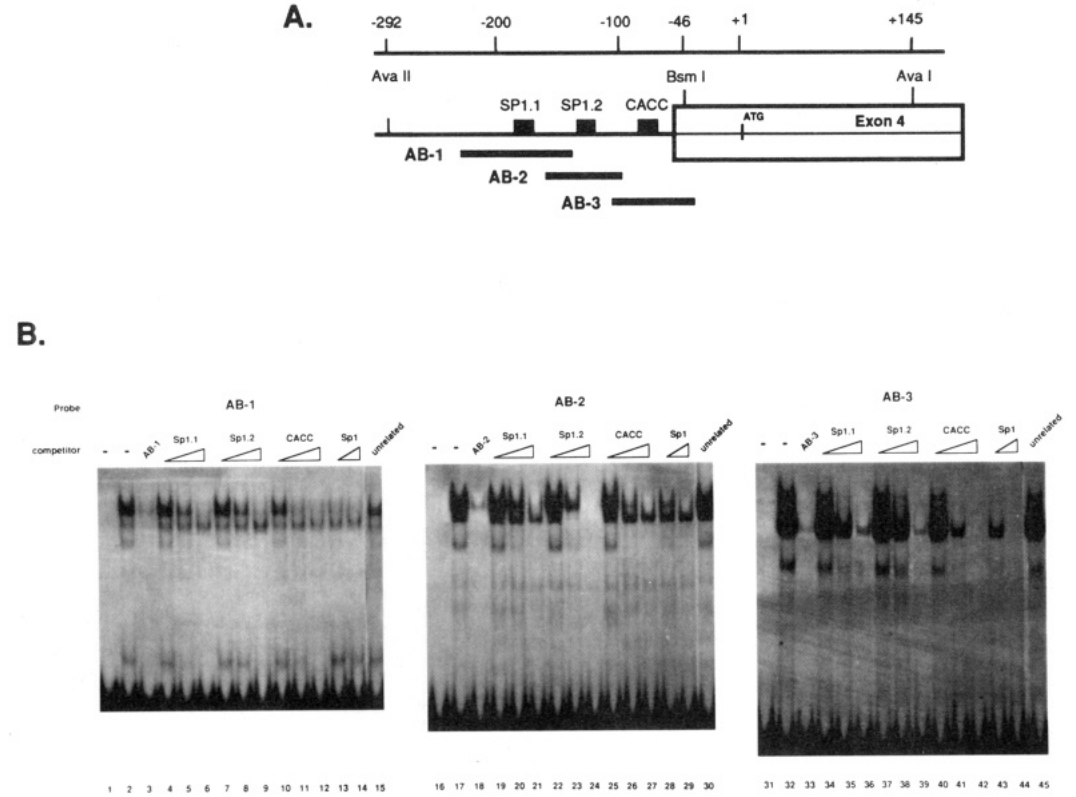

FIGURE 4: Competitive mobility shift analysis of the hFR-KB P4 gene promoter. (A) The structure of the hFR-KB gene from nt −292 to nt +145 is schematically shown. The protected sites identified by the DNase I footprinting are represented by solid boxes. The probes used in the mobility shift assay are illustrated by the solid bar and designated as AB-1, AB-2, and AB-3, corresponding to nt −231−nt −136, nt −158−nt −100, and nt −107−nt −41, respectively. (B) Each of the radiolabeled fragments (10 000 cpm) were incubated with 2 μg of KB cell nuclear extract in the absence (lanes 2, 17, and 32) or presence of competitors. Lanes 1, 16, and 31 contain the control reactions in which the probe was incubated in the absence of nuclear extract. Competitors used were as follows: 100-fold molar excess of unlabeled AB-1, AB-2, and AB-3 DNA fragments (lanes 3, 18, and 33, respectively); increasing molar excess (10-, 100-, and 1000-fold) of oligonucleotides containing Sp1.1 (lanes 4−6, 19−21, and 34−36), Sp1.2 (lanes 7−9, 22−24, and 37−39), and CACC/Sp1.3 motif (lanes 10−12, 25−27, and 40−42) sequences (see Table 1); 100- and 1000-fold molar excess of the oligonucleotide containing the Sp1 consensus sequence (lanes 13−14, 28−29, and 43−44); and 1000-fold molar excess of oligonucleotides (30 bp) containing unrelated sequences (lanes 15, 30, and 45).

protein-binding sites contained in the P4 promoter region, a competitive mobility shift assay (Figure 4) was performed using PCR-generated, overlapping radiolabeled probes, designated as AB-1, AB-2, and AB-3 (corresponding to coordinates −231/−136, −158/−100, and −107/−41, respectively), that spanned the binding sites identified by DNase I footprinting assays (Figure 3). To confirm protein-binding specificity, oligonucleotides containing the hFR-KB gene sequence of Sp1.1, Sp1.2, and CACC/Sp1.3 motifs (see Table 1) were synthesized for use as competitors. Radiolabeled probe was incubated with KB cell nuclear extract in the absence or presence (10−1000-fold molar excess as indicated) of competitor oligonucleotides. As shown in Figure 4B, KB cell extracts formed a major and two or more minor protein−DNA complexes with each of the AB-1, AB-2, and AB-3 probes (lanes 2, 17, and 32, respectively) compared to probe only (lanes 1, 16, and 31, respectively). In each case, the protein−DNA binding was specific since the formation of these complexes was inhibited by 100-fold molar excess of unlabeled probes (lanes 3, 18, and 33, respectively) but not by 1000-fold molar excess of the unrelated oligonucleotides (lanes 15, 30, and 45). These protein-radiolabeled probe complexes were also specifically inhibited (lanes 4−14, 19−29, and 34−44) in a concentration-dependent manner by each of the Sp1.1-, Sp1.2-, CACC/Sp1.3- (10-, 100-, and 1000-fold molar excess), and Sp1- (100- and 1000-fold molar excess) oligonucleotides. Interestingly, the oligonucleotide containing the CACC/Sp1.3 motif sequence was found to inhibit the formation of protein−DNA complexes with the AB-1 and AB-2 probes (lanes 10−12 and 25−27, respectively). Likewise, the Sp1.1-, Sp1.2-, and Sp1-oligonucleotides competitively inhibited protein−AB-3 complex formation (lanes 34−39 and 43−44). The inhibition by the CACC/Sp1.3-oligonucleotide was virtually equivalent to that by the Sp1-oligonucleotide. Thus, the KB cell nuclear extract contains a factor(s) that shares the binding site with Sp1 and that can bind to a CACC/Sp1.3 motif.

The specificity of the binding site within each oligonucleotide for the KB cell nuclear proteins was further examined by competitive mobility shift assays using shorter oligonucleotides with and without site-specific mutations within the core Sp1-binding sites of the Sp1.1, Sp1.2, and CACC/Sp1.3 sequences (see Table 1). As shown in Figure 5A, each oligonucleotide probe formed a major and one or more minor complexes with KB cell nuclear extract (lanes 2, 9, and 16) compared to the probe only (lanes 1, 8, and 15, respectively). The complexes with Sp1.1-oligonucleotide probe (lane 2) were inhibited by 100-fold molar excess of unlabeled Sp1.1-oligonucleotides (lane 4) but not by 1000-fold molar excess of ΔSp1.1-oligonucleotides (lane 7). Likewise, neither ΔSp1.2- nor ΔCACC/Sp1.3-oligonucleotides inhibited the DNA−protein complex formation (lanes 12−14 and lanes 19−21, respectively). Furthermore, radiolabeled mutated oligonucleotide probes did not form the same specific protein−DNA complexes as compared to the complexes

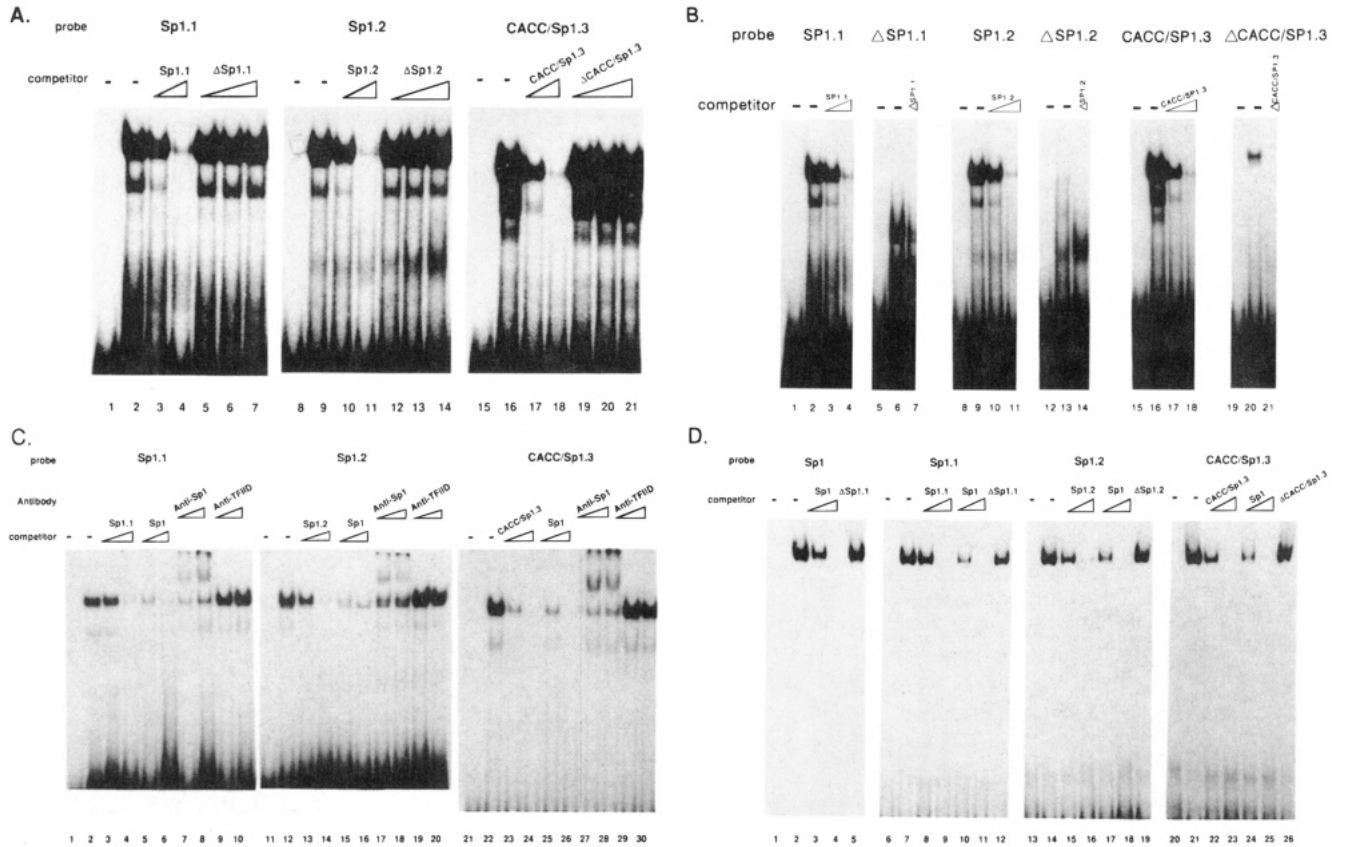

FIGURE 5: Competitive mobility shift analysis of the binding sites in the promoter region of hFR-KB gene. Control reactions (panel A, lanes 1, 8, and 15; panel B, lanes 1, 5, 8, 12, 15, and 19; panel C, lanes 1, 11, and 21; and panel D, lanes 1, 6, 13, and 20) containing radiolabeled probes (10 000 cpm) were incubated in the absence of nuclear extract. Each of the radiolabeled probes (10 000 cpm) including wild type or mutated Sp1.1-, Sp1.2-, and CACC/Sp1.3-oligonucleotides were incubated with 2 μg of KB cell nuclear extract (panels A−C) or 20 ng of purified human Sp1 (panel D) in the absence (panel A, lanes 2, 9, and 16; panel B, lanes 2, 6, 9, 13, 16, and 20; panel C, lanes 2, 12, and 22; and panel D, lanes 2, 7, 14, and 21) or presence of competitors. The competitors are indicated as follows: (panel A) 10- and 100-fold molar excess of unlabeled probe (lanes 3, 4, 10, 11, 17, and 18) and 10-, 100-, and 1000-fold molar excess of the site-specific mutated oligonucleotides ΔSp1.1, ΔSp1.2, and ΔCACC/Sp1.3 (lanes 5−7, 12−14, and 19−21, respectively); (panel B) 10- and 100-fold molar excess of unlabeled probe (lanes 3, 4, 10, 11, 17, and 18) or 1000-fold unlabeled probe (lanes 7, 14, and 21); (panel C) 100- and 1000-fold molar excess of unlabeled probe (lanes 3, 4, 13, 14, 23, and 24) and 100- and 1000-fold molar excess of the commercial Sp1-oligonucleotide (lanes 5, 6, 15, 16, 25, and 26); (panel D) 100- and 1000-fold molar excess of unlabeled probe (lanes 3, 4, 8, 9, 15, 16, 22, and 23) and of the Sp1-oligonucleotide (lanes 10, 11, 17, 18, 24, and 25), and 1000-fold excess molar of the site-specific mutated oligonucleotides ΔSp1.1, ΔSp1.2, and ΔCACC/Sp1.3 (lanes 5 and 12, 19, and 26, respectively). For reactions including antibodies (panel C), aliquots (0.5 and 2 μL) of anti-human Sp1 (lanes 7, 8, 17, 18, 27, and 28) or anti-human TFIID (lanes 9, 10, 19, 20, 29, and 30) antibodies were added to binding reactions as described under Experimental Procedures.

formed with the wild type sequences (Figure 5B). Although a slower migrating complex is formed with the radiolabeled ΔCACC/Sp1.3-oligonucleotide (Figure 5B, lanes 20 and 21), a protein−DNA complex of the same size is not formed with the wild type CACC/Sp1.3-oligonucleotide (Figure 5B, lane 16) indicating that this complex is an artifact. These results demonstrate that the binding of nuclear proteins to these oligonucleotides is specific and suggest that the mutated region in each oligonucleotide is critical for protein binding.

To identify the KB cell nuclear proteins that bind to the Sp1.1, Sp1.2, and CACC/Sp1.3 sequences, the mobility shift assay was performed (a) in the presence of commercial Sp1-oligonucleotide, (b) in the presence of purified Sp1 protein, and (c) in the presence of antibodies to the human Sp1 or TFIID. The results of these mobility shift assays are shown in Figure 5C,D. The protein−DNA complexes formed with each of Sp1.1-, Sp1.2-, or CACC/Sp1.3-oligonucleotide probes are strongly inhibited (Figure 5C, lanes 5−6, 15−16, and 25−26, respectively) by the Sp1-oligonucleotides (100- and 1000-fold molar excess). The major protein−DNA complex with a slower migration was supershifted following incubation with anti-human Sp1 antibodies (lanes 7−8, 17−18, and 27−28), whereas the minor complex with a faster migration appeared to be unshifted. The incubation with anti-human TFIID did not affect the migration of either complex (lanes 9−10, 19−20, and 29−30). The ability of Sp1 to bind to these oligonucleotides was confirmed using purified human Sp1 (Figure 5D). Purified Sp1 formed a single complex with each of Sp1-, Sp1.1-, Sp1.2-, and CACC/Sp1.3-oligonucleotide (lanes 2, 7, 14, and 21, respectively), in contrast to two complexes observed when using KB cell nuclear extract (Figure 5A−C). The specificity of the formation of each complex was confirmed by competition with unlabeled oligonucleotides (100- and 1000-fold molar excess) and site-specific mutated oligonucleotides (1000-fold molar excess) (lanes 3−5, 8−12, 14−19, and 22−26 in Figure 5D). These results suggest that two or more proteins contained in KB cell nuclear extracts bind to each of the oligonucleotide probes (Figure 5A−C), and for each probe, one of these proteins appears to be Sp1 or a factor(s) immunologically related to Sp1. The other nuclear protein-(s) that form relatively minor but specific complexes with these probes are not recognized by the anti-Sp1 (or anti-TFIID) antibody and appear to share the binding site with Sp1 or require Sp1 binding in order to bind to these probes.

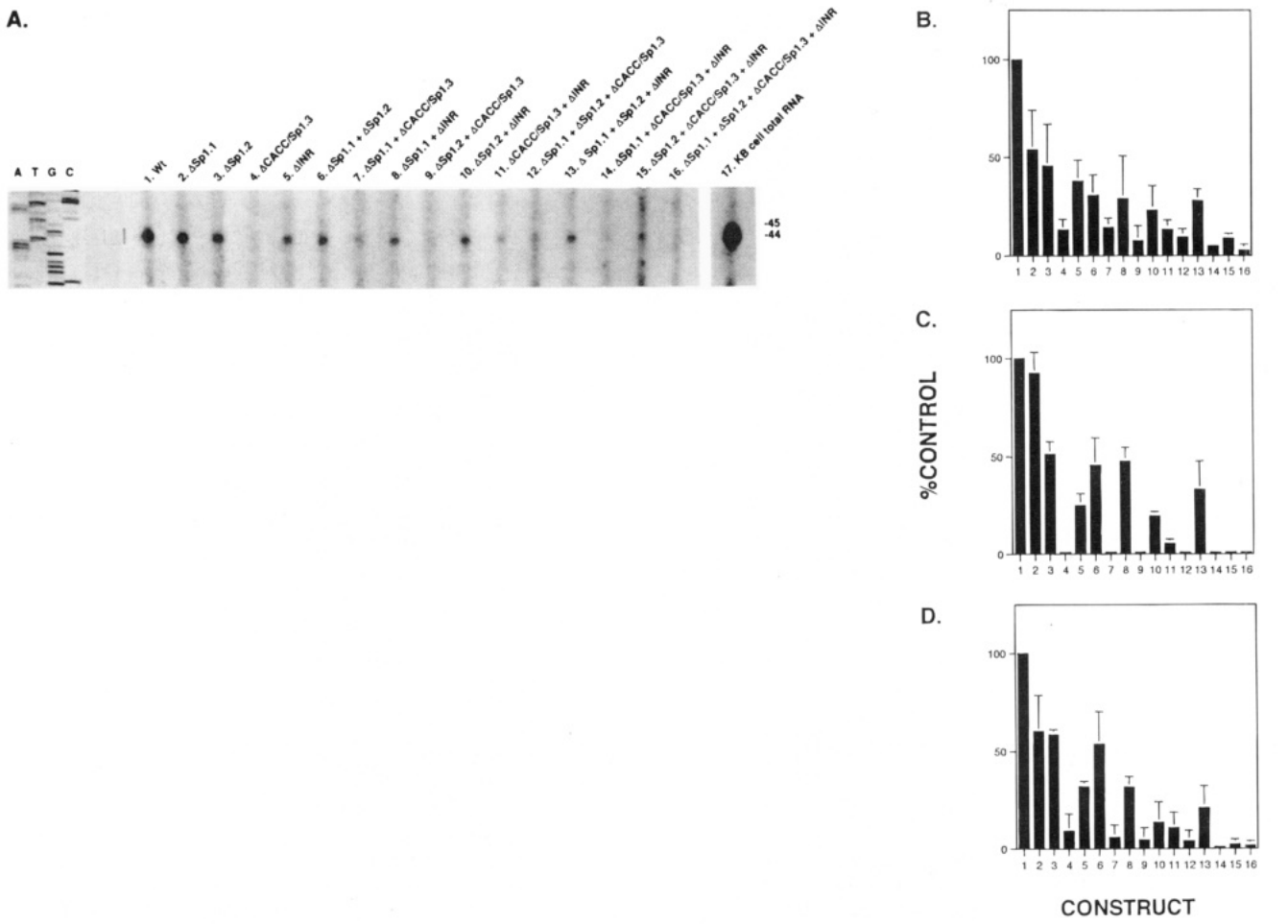

FIGURE 6: *In vitro* (panels A and B) and *in vivo* (panels C and D) analysis of P4 promoter constructs containing site-specific mutations in the Sp1-binding site and INR sequence. (A) *In vitro* transcription assay. The hFR-KB promoter-luciferase reporter gene construct containing −305/+102 promoter fragment (WT), the constructs with the site-specific mutations for the Sp1.1, Sp1.2, CACC/Sp1.3, and INR sequences, and their various combinations were generated by recombinant PCR methods (see Experimental Procedures). *In vitro* transcription reaction with HeLa cell nuclear extract followed by primer extension of each construct (lanes 1−16) was performed as described under Experimental Procedures. The constructs containing site-specific mutations (Δ) are indicated above each lane. Total RNA (20 μg) from KB cells and wheat germ tRNA (20 μ) served as a positive control (lane 17) and a negative control (data not shown), respectively. The extended products in total RNA from KB cells (lane 17) represent the major transcription start sites at nt −45 and nt −44. The sizes of the extended products were determined using the sequencing ladder shown to the left. This figure shows a representative experiment from three experiments. (B) PhosphorImager analysis of the *in vitro* transcription assays. The signal intensities from each experiment were measured with a PhosphorImager (Molecular Dynamics, Inc.) and expressed as the percent control of the signals in WT construct. The number of *x*-axis represents the same constructs as the number in panel A. The bar graph shows the mean of three independent experiments, and the bracket represents one standard deviation from the mean. (C and D) Transient transfection analysis of the P4 promoter-luciferase constructs in HeLa cells (panel C) and KB cells (panel D). Luciferase and β-galactosidase activity were measured by luminometry in the cell extracts following cotransfection with each construct (15 μg) and pCMV-β gal plasmid (3 μg) as described under Experimental Procedures. Luciferase activity was corrected for differences in transfection efficiency as determined by β-galactosidase activities. Luciferase activity is expressed as the percent control of activity in the WT construct. The number of *x*-axis represents the same constructs as described in panels A and B. The bar graph shows the mean of at least three independent transfections, and the bracket represents one standard deviation from the mean.

P4 Promoter Is Activated through Three Clustered Sp1-Binding Sites with Cooperation of Initiator Element in Vitro and in Vivo. The above results suggest that the transcription factor Sp1 and/or a factor related to Sp1 specifically binds to three clustered sites within the proximal upstream sequence of exon 4 (nt −200/−46) and that this region is required for basal P4 promoter activity. We next studied the relative contribution of each site to P4 promoter activity by means of *in vitro* transcription. Furthermore, we sought to determine whether the INR sequence (Figure 1) contributes to this promoter activity and/or specifies accurate localization of transcription. For these experiments, the −305/+102 hFR promoter-luciferase constructs (WT construct) containing various combinations of site-specific mutations for Sp1.1, Sp1.2, CACC/Sp1.3, and INR as shown in Table 1 were generated by recombinant PCR methodology. Following *in vitro* transcription, the size and abundance of RNA transcripts were analyzed by primer extension. The primer extension reaction of the transcription reaction containing only HeLa cell extract, e.g., no plasmid DNA, and wheat germ tRNA served as the negative controls and did not yield detectable signals (data not shown). The primer extension reaction of KB cell RNA, which served as the positive primer extension control, contained a prominent signal (Figure 6A, lane 17). On the basis of the size of extended products and the origin of the primers, the hFR-KB transcripts initiate at nt −44 and nt −45 relative to the translation start site. The WT construct has strong promoter activity (Figure 6A, lane 1) and initiates transcription at nt −44. Single-site mutations of Sp1.1, Sp1.2, or CACC/Sp1.3 (Figure 6A, lanes 2−4, respectively) demonstrated a decrease in transcript abundance, such that promoter activity was 53.9 ± 20.1%, 42.6 ± 21.4%, and

13.1 ± 5.4% of control, respectively (Figure 6B). These results indicate that each single site contributes to the promoter activity and that the more proximal site appears to have more potent promoter activity. The mutation of the INR sequence (lane 5) significantly decreased the promoter activity (37.9 ± 10.6% of control) but did not change the transcription initiation site. The constructs containing two or more mutated sites (lanes 6−16 in Figure 6A,B) exhibited further decreases in promoter activity. The mutations of all four sites (lane 16) resulted in near complete loss of activity (3.0 ± 2.9% of control). These results demonstrate the cooperative interaction of these Sp1 sites and INR sequence in the hFR-KB P4 promoter activity. To determine if these mutations impart a similar effect on promoter activity *in vivo*, we performed transient transfection assays in human KB and HeLa cells using the identical reporter gene constructs. Transfection of the WT construct (Figure 6C,D, lane 1) resulted in a 10-fold increase in luciferase activity compared to transfection with the promoterless plasmid. Relative to the WT construct, transfection with the constructs containing mutations at each site and all combinations thereof (lanes 2−16 in Figure 6C,D) yielded significantly lower promoter activity. These findings are consistent with the results of *in vitro* transcription assays (Figure 6B). Interestingly, all of the constructs containing the mutation of the CACC/Sp1.3 site (Figure 6C,D, lanes 4, 7, 9, 11, 12, and 14−16) showed a dramatic decrease in activity in HeLa cells (<1% of the control value) as well as in KB cells (range 1−9% of the control value) suggesting that this "CACC/Sp1.3 site" plays a critical role for transcriptional activation of the hFR-KB gene. Thus, the data from both *in vitro* and *in vivo* analyses indicate that (i) the region nt −200−nt −38 contains the elements required for basal promoter activity, (ii) three sites (Sp1.1, Sp1.2, and CACC/Sp1.3) are integral elements required for full activity of the P4 promoter and act cooperatively, and (iii) the INR sequence (nt −45−nt −38) in conjunction with upstream Sp1 elements supports the transcriptional activation.

DISCUSSION

We have previously reported the structural and functional organization of the hFR-KB gene (Elwood et al., 1993) in which several observations of interest for regulatory mechanisms of this gene expression are demonstrated. First, alternative splicing involving exons 1−4 generates a number of KB cell cDNA isoforms from a single hFR-KB gene. Second, the hFR-KB gene contains at least two discrete promoters, P1 and P4, located upstream of exons 1 and 4, respectively, and the utilization of these promoters is in a tissue-specific manner in normal and cultured human tissues. Third, transcriptional initiation from the P4 promoter which lacks the TATA sequence is precise in contrast to the multiple transcriptional start sites from the P1 promoter which contains potential TATA and CAAT sequences. In order to further elucidate the molecular basis of these observations, we focused on the functional cis and trans elements contained in the P4 promoter of the hFR-KB gene in this study. The hFR-KB gene P4 promoter (Figure 1) is a "TATA-less promoter" characterized by the following: (i) the P4 promoter does not contain canonical TATA or CAAT boxes; (ii) multiple potential transcription factor-binding sites (e.g., Ap2, Sp1, and E motifs) as well as palindromic sequences are present; (iii) three potential transcription factor Sp1 (or GC box) sites are clustered between residues −185 and −76; (iv) the region containing these three Sp1-binding elements is required for basal promoter activity (Figures 2 and 3); (v) transcription factor Sp1 or the factor(s) immunologically related to Sp1 contained in nuclear extracts from human cells specifically binds to each of these Sp1 sites (Figures 4 and 5); (vi) each of these Sp1 sites is an integral element in supporting full activity of the P4 promoter (Figure 6); and (vii) the INR sequence (Figure 1) cooperates with the upstream elements for transcriptional activation (Figure 6).

The segregation of GC boxes and their cooperative activation of transcription have been reported in several genes with a TATA-less promoter (Blake et al., 1990; Chen et al., 1992; Boisclair et al., 1993). We examined the contribution of each Sp1-binding site to the promoter activity of the P4 hFR-KB gene since the potency of individual GC boxes for the full basal activity appears to be different among reported genes. For example, in the hamster dihydrofolate reductase gene promoter (Blake et al., 1990), two proximal GC boxes are required for efficient initiation of transcription; in the rat insulin-like growth factor-binding protein-2 gene promoter (Boisclair et al., 1993), three GC boxes are equally important for promoter function; and in the rat transforming growth factor-α (TGF-α) gene promoter (Chen et al., 1992), the most proximal GC box confers the maximal effect on the promotion of gene transcription. In the hFR-KB gene promoter P4, the cooperative interaction of three GC boxes was demonstrated by both *in vitro* and *in vivo* assays. Similar to the rat TGF-α gene, mutation of the most proximal Sp1-binding site (CACC/Sp1.3) exhibited the largest reduction in promoter activity compared to mutations of the other two Sp1 sites, indicating the critical function of this site for activation of transcription. The hFR-KB P4 promoter shares several features with a promoter recently described by Sadasivan et al. (1994) upstream from exon 1 of the hFR-P gene (Page et al., 1993). Both promoters lack a canonical TATA box or CAAT box and contain cis elements that bind Sp1 or Sp1-related DNA-binding proteins. The hFR-P promoter also contains three tandemly repeated AGGAAG motifs that bind a GA-binding protein(s) (GABP). The GABP protein is a member of the *ets* gene family which encodes a set of proteins that are involved in tissue-specific transcriptional regulation (Macleod et al., 1992). In contrast, the hFR-KB P4 promoter described herein consists of an array of three Sp1-binding sites juxtaposed to a consensus initiator region sequence (INR) (see Figure 1). On the basis of *in vitro* and *in vivo* assays of promoter activity (Figure 6), each of three elements is necessary to fully support the basal transcriptional activity of the P4 promoter. More importantly, the observed tissue-specific expression (Page et al., 1993; Elwood et al., 1993) of hFR transcripts originating from the hFR-P and hFR-KB promoters is activated in a tissue-specific manner either directly (e.g., by trans activation via DNA-binding proteins such as GABP) or by interaction with other unknown flanking transcriptional regulatory elements, e.g., tissue-specific enhancers and/or suppressors.

The KB cell nuclear extract formed a major and one or more minor specific protein−DNA complexes with each of the probes (Sp1.1, Sp1.2, CACC/Sp1.3, AB-1, AB-2, and AB-3) used in the gel shift assays (see Figures 4 and 5). Complexes of identical sizes were also observed using nuclear extracts prepared from other human cell lines including MCF-7 mammary carcinoma cells, lung carcinoma

(2009) cells, several human renal carcinoma (ACHN, TK-10, and SN12C) cells, and HeLa cells (data not shown). Several lines of evidence support the hypothesis that the moieties in the KB cell nuclear extract that bind to the three GC boxes within the P4 promoter in the formation of the major protein−DNA complexes include Sp1 or Sp1-related factors. First, the three regions protected in DNase footprinting (Figure 3) contain Sp1-binding sequences (Figure 1). Second, nuclear extracts form specific protein−DNA complexes with P4 promoter restriction fragments and synthetic oligonucleotides corresponding to each of the potential Sp1-binding sequences (Sp1.1, Sp1.2, and CACC/Sp1.3). In each case, the major protein complex formation is specifically inhibited by commercially available Sp1 consensus sequences (Figures 4 and 5C). Third, the major nuclear protein−DNA complex containing either Sp1.1-, Sp1.2-, or CACC/Sp1.3-oligonucleotide is supershifted by anti-human Sp1 antibodies in mobility shift assays (Figure 5C). Fourth, purified human Sp1 binds to Sp1.1, Sp1.2, and CACC/Sp1.3 sequences and/or oligonucleotides (Figure 5D) forming DNA−protein complexes of identical sizes. Kriwacki et al. (1992) have demonstrated that the two G's (underlined) at positions 2 and 3 of the first triplet (GGG) and the two G's (underlined) at positions 1 and 3 of the second triplet (GCG) of the canonical Sp1-binding site (GGG;GCG;GGGC) are required to contact with zinc fingers 3 and 2, respectively, of Sp1 for high-affinity binding, while the third subunit (GGGC) contributes less to total binding. Although the core nucleotide sequence of the Sp1 sites contained in the hFR-KB gene P4 promoter differs from that of the Sp1 consensus sequence, the G's at the critical positions are well conserved in each of the three sites, e.g., three of four in Sp1.1 and Sp1.2 sites, and four of four in the CACC/Sp1.3 site. Lastly, oligonucleotides containing substitutions at three or four residues within each of the Sp1 core sequences (including alteration of G's at these critical positions) did not form the same protein−DNA complexes as observed when wild type sequences were used and did not inhibit complex formation between wild type sequences and KB nuclear extracts or purified Sp1 (Figure 5). These results confirm the specificity of Sp1 (or a related protein) binding to these three binding sites and in the formation of the major DNA−protein complexes.

The identity of the nuclear proteins involved in the formation of the minor protein−DNA complexes (Figures 4 and 5) and the relative contribution of each nuclear protein in P4 promoter activation are not clear from these experiments. However, these proteins appear to be related to each other as well as to Sp1 with respect to their DNA-binding properties. For example, the formation of the minor protein−DNA complexes is inhibited by the oligonucleotides that contain each of the putative P4 promoter Sp1-binding sites and by a consensus Sp1-oligonucleotide (Figure 5A). Except for the potential Sp1-binding site, the flanking sequences of these oligonucleotides are unrelated (Table 1). In contrast, oligonucleotides containing mutations within the core Sp1-binding sites do not inhibit formation of the minor (or major) complexes (Figure 5A) and do not form specific complexes that are similar in size relative to the complexes formed with the wild type oligonucleotides (Figure 5B). Others (Hagen et al., 1992) have observed multiple protein−DNA complexes when probes containing Sp1-binding sites were studied by gel shift assays and suggested that the minor complexes might be the result of proteolysis of Sp1. Although it is possible that the faster migrating complexes are due to alteration of Sp1 (e.g., proteolytic degradation) during the preparation of nuclear extracts, it seems unlikely since the anti-Sp1 antibodies are directed toward epitopes (residues 520−538) that are immediately adjacent to the DNA-binding domain of Sp1 (residues 540−620) (Kadonaga et al., 1987; Jackson & Tjian, 1988). Despite these similarities, these proteins appear to be distinct from Sp1 based on their lack of reactivity with anti-human Sp1 antibodies and on the observation that purified human Sp1 forms a single complex with each oligonucleotide probe (Figure 5D). These observations suggest that these proteins share the Sp1-binding site, require Sp1 binding in order to form protein−DNA complexes, or directly interact with the Sp1-related factors. For example, several proteins other than Sp1 may bind to the GC-rich elements, including Ap2 (Faisst & Meyer, 1992), murine factor Krox-24 and Krox-20 (Faisst & Meyer, 1992), and BTEB (basic transcription element-binding factor) (Imataka et al., 1992). Furthermore, novel GT box-binding factors, Sp2 and Sp3 (Kingsley & Winoto, 1992), SPR-1, and SPR-2 (Hagen et al., 1992) have been recently described. These proteins are members of a Sp1 gene family that share biochemical properties with Sp1 including binding to GC and GT boxes and also share nucleotide sequence homology. Since the Sp1.1 and CACC/Sp1.3 sites in the hFR-KB gene promoter also contain GT-rich sequences (see Table 1), it is possible that other members of this gene family are interacting with Sp1 or binding to these sites. Ultimately, the identification of the protein(s) involved in the formation of these minor complexes and the elucidation of the role of each nuclear protein in the activation of the P4 promoter will require further study.

In contrast to the hFR-KB P1 promoter that initiates transcription at multiple sites within exon 1 (Elwood et al., 1993), the hFR-KB gene P4 promoter precisely initiates transcription at nt −45 or nt −44 (Figure 6). Interestingly, we observed that the sequences (from nt −45 to nt −38, CATTCCTT) surrounding the transcriptional start site of the P4 promoter are homologous to previously reported initiator region (INR) sequences. Since the INR sequences have been shown to contribute to promoter activity and/or accurately initiate transcription, we sought to investigate the role of the hFR-KB P4 INR sequence on P4 promoter function. Reported INR consensus sequences [$YA_{+1}YTCYYY$ (Roeder, 1991), $CTCA_{+1}NTCT$ (Smale & Baltimore, 1989), $GN_{1-2}CTCA_{+1}N_3T$ (Conaway et al., 1990; Carcamo et al., 1991), or $YYA_{+1}N^T/_AYY$ (Javahery et al., 1994)] contain a conserved central CA residue where transcription typically initiates at A_{+1} nucleotide. Therefore, we chose to mutate (underlined) the hFR-KB P4 INR sequence to CGTAGCT. Decreased promoter activity in the construct containing ΔINR region was demonstrated by both *in vitro* and *in vivo* assays (Figure 6) indicating that the INR in the P4 promoter is functionally active in supporting P4 promoter activity. In addition, combining the INR mutation with one or more Sp1-binding site mutations resulted in further decreases in promoter activity consistent with a cooperative interaction between these elements. The P4 INR does not appear to be required for accurate transcript initiation since the transcriptional initiation site (nt −44) of transcripts from the ΔINR construct is identical to that from the intact INR construct (Figure 6). As several proteins have been reported to bind to specific initiator regions including YY1, E2F, TFII-I, and USF (Javahery et al., 1994; Wingender, 1993), further

experiments are needed to define the mechanism by which the transcription is initiated precisely by the P4 promoter and to determine which transcription factors are mediating the P4 INR activity.

Methylation of DNA affects the transcriptional activity of many eukaryotic genes (Wingender, 1993) and has been reported to regulate folate-binding protein gene transcription (Hsueh & Dolnick, 1994). The hFR-KB P4 promoter region is relatively rich in GC content and in "GC" doublets (Figure 1) compared to the remainder of the downstream nucleotide sequence. Importantly, intracellular folates (e.g., N^5-methyltetrahydrofolate) serve as the one-carbon donor for all known DNA methylases (Voet & Voet, 1990). Furthermore, folate depletion induces hFR expression (transcript and protein) in tissue culture cells (Kane & Waxman, 1989; Henderson, 1990; Antony, 1992). These observations suggest that the folate status of the cell or organism may influence the overall regulation of the hFR-KB gene or other hFR genes by altering its state of methylation. Experiments are underway to address this hypothesis.

In summary, we have described the structural and functional properties of the hFR-KB P4 promoter upstream from exon 4. The current study establishes the framework to study the regulatory mechanisms resulting in changes in hFR expression related to folate homeostasis, differences in expression of hFR isoforms in cells and human tissues, apparent differences in promoter utilization by various human tissues, and differences in expression following malignant degeneration. The characterization of the other functional promoter, P1, and the determination of regulatory elements such as enhancers or suppressors for each promoter and the interaction of two promoter elements resulting in the observed tissue-specific, differential transcription are on going. Further investigation of the molecular basis for regulation of the hFR-KB gene will advance the understanding of the overall role of hFRs in cellular accumulation of folate and antifolates.

REFERENCES

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

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1991) *Current Protocols in Molecular Biology*, pp 12.1.1-12.1.6, Greene Publishing Assoc. & Wiley-Interscience, New York, NY.

Azizkhan, J. C., Jensen, D. E., Pierce, A. J., & Wade, M. (1993) *Crit. Rev. Eukaryotic Gene Expression 3*, 229−254.

Blackwell, T. K., & Weintraub, H. (1990) *Science 250*, 1104−1110.

Blake, M. C., Jambou, R. C., Swick, A. G., Kahn, J. W., & Azizkhan, J. C. (1990) *Mol. Cell. Biol. 10*, 6632−6641.

Boisclair, Y. R., Brown, A. L., Casola, S., & Rechler, M. M. (1993) *J. Biol. Chem. 268*, 24892−24901.

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

Carcamo, J., Buckbinder, L., & Reinberg, D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 8052−8056.

Chen, X., Azizkhan, J. C., & Lee, D. C. (1992) *Oncogene 7*, 1805−1815.

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

Conaway, J. W., Travis, E., & Conaway, R. C. (1990) *J. Biol. Chem. 265*, 7564−7569.

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

deBoer, E., Antoniou, M., Mignotte, V., Wall, L., & Grosveld, F. (1988) *EMBO J. 7*, 4203−4212.

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

Elwood, P. C., Price, K., Saikawa, Y., Pacheco, P., & Chung, K.-N. (1993) *Blood 82*, 93a.

Faisst, S., & Meyer, S. (1992) *Nucleic Acids Res. 20*, 3−26.

Hagen, G., Muller, S., Beato, M., & Suske, G. (1992) *Nucleic Acids Res. 20*, 5519−5525.

Henderson, G. B. (1990) *Annu. Rev. Nutr. 10*, 319−335.

Higuchi, R. (1990) In *PCR Protocols: A Guide to Methods & Applications* (Innis, M. A., Gelfand, D. H., Sninsky, J. J., & White, T. J., Eds.) 1st ed., pp 177−183, Academic Press, Inc., San Diego, CA.

Hooft van Huijsduijnen, R. A. M. (1992) *Biotechniques 12*, 830−832.

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

Imataka, H., Sogawa, K., Yasumoto, K., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M., & Fujii-Kuriyama, Y. (1992) *EMBO J. 11*, 3663−3671.

Jackson, S. P., & Tijan, R. (1988) *Cell 55*, 125−133.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., & Smale, S. T. (1994) *Mol. Cell. Biol. 14*, 116−127.

Kadonaga, J. T. (1990) *J. Biol. Chem. 265*, 2624−2631.

Kadonaga, J. T., Jones, K. A., & Tjian, R. (1986) *Trends Biochem. Sci. 11*, 20−23.

Kadonaga, J. T., Carner, K. R., Masiarz, F. R., & Tjian, R. (1987) *Cell 51*, 1079−1090.

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

Kingsley, C., & Winoto, A. (1992) *Mol. Cell. Biol. 12*, 4251−4261.

Kriwacki, R. W., Schultz, S. C., Steitz, T. A., & Caradonna, J. P. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9759−9763.

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

Macleod, K., Leprince, D., & Stehelin, D. (1992) *Trends Biochem. Sci. 17*, 251−256.

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

Ragoussis, J., Senger, G., Trowsdale, J., & Campbell, I. G. (1992) *Genomics 14*, 423−430.

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

Roeder, R. G. (1991) *Trends Biochem. Sci. 16*, 402−408.

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

Sadasivan, E., Cedeno, M. M., & Rothenberg, S. P. (1994) *J. Biol. Chem. 289*, 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.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

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

Schule, R., Muller, M., Otsuka-Murakami, H., & Renkawitz, R. (1988) *Nature 332*, 87−90.

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

Smale, S. T., & Baltimore, D. (1989) *Cell 57*, 103−113.

Voet, D., & Voet, J., Eds. (1990) in *Biochemistry*, John Wiley & Sons, Inc., New York, NY.

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

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

Wingender, E., Ed. (1993) in *Gene Regulation in Eukaryotes*, VCH, New York, NY.

Yu, C.-Y., Motamed, K., Chen, J., Bailey, A. D., & Shen, C.-K. J. (1991) *J. Biol. Chem. 266*, 8907−8915.

BI950246U