Articles

Localization of a Transcription Promoter within the Second Exon of the Cytochrome P-450c27/25 Gene for the Expression of the Major Species of Two-Kilobase mRNA[†,‡]

Jayati Mullick, Sankar Addya, Carmen Sucharov, and Narayan G. Avadhani*

Laboratories of Biochemistry, Department of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6046

Received June 19, 1995; Revised Manuscript Received August 31, 1995[⊗]

ABSTRACT: The rat P-450c27/25 (CYP27) gene is expressed as two distinctly sized mRNAs of 2 and 2.3 kb (kilobase). The 2 kb mRNA is the predominant form in the liver with negligible 2.3 kb species. Rat kidney and hepatoma, on the other hand, contain significant levels of the 2.3 kb species. Rat CYP27 gene contains 11 exons of 80−415 nucleotides that are separated by 10 introns of 83 bases to ~10 kb. S1 nuclease protection and primer extension analyses using liver RNA showed a prominent 5′ terminus 86 nucleotides downstream from the start of exon 2. This site, designated as +1, is the start site for the 2 kb mRNA. 5′ RACE analysis of rat kidney and hepatoma RNAs showed the presence of a 5′ extended mRNA with a sequence complementary to the Spi2 mRNA. A cryptic TATA box (TTTAAA) is located 24 nucleotides upstream of the 2 kb mRNA transcription initiation site at +1. A 106 bp DNA fragment (sequence −83 to +23) that houses the putative TATA motif forms three differently migrating complexes with nuclear extract from the murine 3T3 cells. DNAse I footprinting and competition with synthetic DNA showed that complex A represents the bound Sp1 factor and complexes B and C are due to unknown factors binding to the −83 to −71 and −20 to −12 sequences, respectively. *In vivo* transcription analysis using −840/+23 DNA and its 5′ deletions cloned in a CAT reporter plasmid suggests that the basal promoter elements are located within sequence −45 to +23 of the gene. Finally, *in vitro* transcription analysis in HeLa cell nuclear extract showed that intact TTTAAA motif and complex C-forming sequence from this region are essential for transcription initiation at the +1 position of the promoter. Our results demonstrate that the 2 kb mRNA is transcribed as an independent transcript driven by an immediate upstream promoter located within exon 2.

It has long been recognized that cytochrome P-450 isoenzymes located inside the hepatic mitochondrial inner membrane compartment play important roles in the metabolism of cholesterol and also related sterols. The metabolic steps carried out by the hepatic mitochondrial cytochrome P-450 include c-27 hydroxylation of cholesterol or cholestenol. This is believed to be an initial step of oxidation of the sterol side chain leading to bile acid synthesis (Bjorkhem, 1985; Carey, 1982; Wikvall, 1984), and there is increasing evidence suggesting that the c-27 hydroxylation of the sterol side chain is a rate-limiting step in the "acidic" bile synthetic pathway (Princen *et al.*, 1991; Stravitz *et al.*, 1994). It is also believed that the oxysterol, 27-hydroxycholesterol, plays an important regulatory function in modulating cholesterol homeostasis. Mutations that affect either the expression or primary sequence of cholesterol c-27 hydroxylase have been implicated in cerebrotendinous xanthomatosis, an inherited cholesterol storage disease (Cali *et al.*, 1991; Leitersdorf *et al.*, 1993; Kim *et al.*, 1994).

Another important P-450[1] dependent function carried out by the hepatic mitochondrial membrane compartment is the 25-hydroxylation of vitamin D_3, which is the first step in the conversion of inactive vitamin D_3 to biologically active 1α,25-dihydroxycholecalciferol (DeLuca & Schnoes, 1976; Pedersen *et al.*, 1979). This hormonal form of the vitamin plays crucial roles in calcium transport and the maintenance of calcium homeostasis in vertebrates. Studies on cDNA expression in heterologous cell systems, such as COS cells (Su *et al.*, 1990) and yeast (Sakaki *et al.*, 1992), followed by enzyme purification (Bergman & Postlind, 1991; Su *et al.*, 1990; Wikvall, 1993) have demonstrated that the same hepatic mitochondrial enzyme, termed P-450c27, alternatively called P-450c27/25 (Addya *et al.*, 1991), carries out both of these metabolic steps. More recent studies (Guo *et al.*, 1993; Axen *et al.*, 1994) indicate that this multifunctional

[†]This research was supported in part by U.S. Public Health Grant GM-34883.

[‡] The nucleotide sequence reported in this paper has been submitted to GenBank data base with accession number U17363, U17369, U17370-76.

* To whom correspondence should be addressed. Fax: 215-898-9923. Phone: 215-898-8819.

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

[1] Abbreviations: P-450, cytochrome P-450; CYP27, rat gene for cytochrome P-450c27 (variably termed P-450c27/25); CAT, chloramphenicol acetyl transferase; kb, kilobase; bp, base pairs; PCR, polymerase chain reaction; Spi2, a member of serine protease inhibitor; AS, antisense.

0006-2960/95/0434-13729$09.00/0

enzyme may also catalyze vitamin D_3 1α-hydroxylation and 24(*R*)-hydroxylation steps. It has also been shown that P-450c27 plays an important role in the removal of cholesterol from macrophages, suggesting a role for it in atherogenesis (Bjorkhem *et al.*, 1994). In support of this hypothesis, cytochrome P-450c27 is expressed in many extrahepatic tissues (Andersson *et al.*, 1989; Addya *et al.*, 1991), including the epithelial cells of blood vessels (Reiss *et al.*, 1994).

Initial studies using Northern blot hybridization and also immunoblot analysis showed that the steady state level of P-450c27 is under the control of pituitary-regulated steroids and growth hormone (Shayiq & Avadhani, 1992). More recent studies suggest that CYP27 gene expression may also be under the control of glucocorticoids, bile acids, and the diurnal cycle.[2] Additionally, in the rat and mouse hepatic tissues, P-450c27 is expressed as a female predominant enzyme (Saarem & Pedersen, 1987; Addya *et al.*, 1991), making it an attractive system for studies on transcription regulation. In keeping with the initial observation in the rabbit liver (Andersson *et al.*, 1989), Northern blot hybridization detects mRNAs of two distinct size classes in both the rat (Su *et al.*, 1990) and human (Cali & Russell, 1991) systems; the major mRNA species in the rat liver resolves with an apparent size of 1.9−2.0 kb (hereafter referred to as 2 kb mRNA) and a minor species of 2.3 kb. Characterization of the rat cDNA for the larger 2.3 kb mRNA showed sequence identical to the smaller 2 kb RNA in its protein-coding region, except for a ~400-nucleotide extended sequence at its 5′ end (Shayiq & Avadhani, 1992). Surprisingly, a sequence stretch of 291 nucleotides from the 5′ terminus of the 2.3 kb mRNA exhibited 100% complementarity to the 5′ translated region of a member of a family of growth hormone regulated serine protease inhibitor, Spi2 mRNA (Shayiq & Avadhani, 1992). In order to understand the precise nature and the origin of the smaller 2 kb mRNA, we have characterized the rat CYP27 gene encoding the entire 2.3 kb mRNA and identified a distinct transcription promoter within the 5′ terminal region of exon 2, which drives the expression of the shorter 2 kb mRNA. Interestingly, some of the basal promoter elements, including the TTTAAA sequence (a putative TATAA box), Sp1 and AP-2 protein binding motifs, and also a functionally important protein binding motif of unknown nature are located within the first 86-nucleotide sequence of exon 2, which is expressed as part of the 5′ extended 2.3 kb mRNA.

EXPERIMENTAL PROCEDURES

Isolation and Sequencing of Genomic DNA. A rat genomic DNA library in λ Dash vector (Stratagene Inc.) was screened with ^{32}P-labeled 2.0 kb P-450c27/25 cDNA probe (Su *et al.*, 1990) as described (Sambrook *et al.*, 1989). Positive clones were isolated by repeated screening, and DNA from the resultant clones was isolated on a large scale by the plate lysis method, followed by adsorption to Lambdasorb, an affinity matrix (Promega Biotech. Corp., Madison, WI), and phenol/chloroform extraction. The genomic DNA clones were digested with various restriction enzymes, and the DNA inserts were characterized by Southern hybridization to either full length P-450c27/25 cDNA probe (Su *et al.*, 1990) or various synthetic oligonucleotides complementary to selected regions of the rat P-450c27/25 cDNA. The genomic DNA inserts (0.6−14 kb) were subcloned into pGEM 3z, 5z, or 7z plasmids (Promega Biotech. Corp., Madison, WI) and sequenced by the dideoxynucleotide method of Sanger *et al.* (1977) using the Sequenase sequencing kit (U.S. Biochemicals Corp., Cleveland, OH). Sequencing was performed on both of the strands and compared to the full length rat P-450c27/25 2 and 2.3 kb cDNA (Su *et al.*, 1990; Shayiq & Avadhani, 1992) sequence to identify the exons and the exon−intron boundaries.

Probe Preparation and Hybridization Method. DNA probes specific for the 5′, middle, and 3′ terminal regions of the P-450c27/25 cDNA were prepared by digesting the 2 kb cDNA with *Eco*RI, *Bam*HI, and *Bgl*II. The restriction fragments *Eco*RI/*Bgl*II, *Bgl*II/*Bam*HI, and *Bam*HI/*Eco*RI represent the 5′, middle, and 3′ terminal regions of the cDNA, respectively. Double-stranded DNA probes were isolated by restriction digestion and fractionation on an agarose gel. The DNA bands were purified by adsorption to "Geneclean", an affinity resin (Bio-101, La Jolla, CA), and labeled by the random primer oligolabeling method using a kit from Pharmacia LKB Technologies. Ltd., in the presence of added [^{32}P]dCTP (3000 Ci/mmol, ICN Biochemicals), to a specific activity of $>10^9$ cpm/μg. Synthetic primers (20−22mer) were end labeled to a specific activity of about 5×10^5 cpm/ng using [γ-^{32}P]ATP and polynucleotide kinase.

Southern and Northern hybridizations were carried out using moderate stringency conditions as described before (Basu & Avadhani, 1991; Carter & Avadhani, 1991). Total RNA from various tissues and cells was isolated by using the guanidinium thiocyanate solubilization and extraction with phenol−chloroform as described before (Chomczynski & Sacchi, 1987). RNA was resolved by electrophoresis on formaldehyde-containing gels, transblotted to Nytran membranes, and hybridized with the ^{32}P-labeled probes under moderate stringency conditions as described (Carter & Avadhani, 1991).

S1 Nuclease Protection and Primer Extension Analysis. The transcription initiation site of the rat CYP27 gene was mapped by S1 nuclease protection using a 5′ end-labeled single-stranded DNA probe (Berk & Sharp, 1977) and also a reverse transcriptase-based primer extension. The S1 single-stranded DNA probe was prepared by the Taq Polymerase-based extension of 5′ end-labeled synthetic DNA primer, PAS-3 (5′ TCTCATGCGGCTCAACACAG 3′), which is complementary to nucleotides 452−471 of the P-450c27/25 2.3 kb cDNA (Shayiq & Avadhani, 1992; Su *et al.*, 1990). The PAS-3 primer was end-labeled to a specific activity of 5×10^5 cpm/ng using [γ-^{32}P]ATP and polynucleotide kinase (Sambrook *et al.*, 1989). DNA (1 μg) of a genomic subclone containing the 0.5 kb intron a and most of exon 2 was linearized by digestion with *Sac*I and coprecipitated with 50 ng of the labeled primer. The probe was synthesized by the Taq Polymerase (Perkin-Elmer Cetus, Corp.) based unidirectional PCR amplification using the manufacturer's recommended conditions. Twenty cycles of amplification (1 min at 95 °C, 30 s at 55 °C, and 2 min at 72 °C) were carried out using an Ericomp Thermal Cycler. The 289-nucleotide product (sequence −194 to +95 of CYP27 gene in Figure 2) was gel-purified on a 4% urea acrylamide gel. About 2.5×10^5 cpm of the probe was hybridized with 100 μg of rat liver RNA, and the S1 analysis was carried out using 100−300 Vogt units of S1 nuclease at 30 °C for 30 min as described (Basu & Avadhani, 1991).

[2] T. R. Stravitz, and Z. R. Vlahcevic, personal communications.

Primer extension using the 5′ end-labeled PAS-3 primer was carried out essentially as described (Basu & Avadhani, 1991) against 30 μg of total RNA or 3 μg of poly(A) RNA from rat liver. The S1 and primer extension products were resolved on 6% polyacrylamide, 8 M urea-sequencing gels.

Characterization of the 5′ Extended Region of mRNA by 5′ RACE. The presence of cytochrome P-450c27 mRNA with a 5′ extended sequence was further confirmed by using a PCR-based 5′ RACE system. Poly(A) RNAs from rat kidney and rat hepatoma H4II EC3 cells (4 μg each) were used to reverse transcribe the 5′ end using the −8 to −22 AS primer (5′ TTGCGCGACCTGGCG 3′), and the single-stranded DNA was ligated to the 5′ anchor primer using a kit supplied by Clonetech Laboratories, Inc. The single-stranded DNA was amplified by PCR initially using a primer complementary to the 5′ anchor sequence and −8 to −22 AS primer, and subsequently with two nested primers PS-1 (5′ ACAGTATCGGCTGCT 3′) and PAS-7 (5′ AGC-GAGATCCTTTCA 3′) using the manufacturer's recommended conditions. The amplified DNA products were sequenced by the PCR-based dideoxy sequencing and found to be identical to the previously reported sequence for the extended region of the 2.3 kb cDNA sequence (Shayiq & Avadhani, 1992).

In Vivo Transcription Studies. Restriction fragments or PCR-amplified DNA segments surrounding the putative transcription initiation region were cloned in a promoter-less and enhancer-less CAT basic plasmid, pCATb (Promega Biotech. Corp., Madison, WI). We have previously shown that both the 2 and 2.3 kb cDNAs contain a number of in-frame as well as out-of-frame ATGs in the 5′ untranslated region of the 2.3 kb cDNA (Shayiq & Avadhani, 1992). As shown in Figure 2, two of these ATG motifs are located immediately upstream (position −82) and downstream (+20) of the putative transcription start site of the 2 kb mRNA. The −840 to +23 DNA with or without the downstream ATG or both mutated ATGs were generated by PCR amplification of the CYP27 genomic DNA using synthetic primers listed below:

Pr-A (sense) −840 to −823,
5′ caagcttCTAACTCTAGCTGACCTGACCTG 3′

Pr-B (antisense) −83 to −63,
5′ cgtcgacGACAAGCGAGATCCTTTCGTA 3′

Pr-C (sense) −83 to −63,
5′ TACGAAAGGATCTCGCTTGTC 3′

Pr-D (antisense) +4 to +23,
5′ cgtcgacCCATCCAGGCAAGGAGGCAG

Pr-E (antisense) +4 to +23,
5′ cgtcgacCGATCCAGGCAAGGAGGCAG

The primers were synthesized with seven nucleotide overhangs (lowercase letters) that provided either *Hin*dIII or *Sal*I restriction sites for cloning. The −840/+23 wild type DNA was generated using primers A and D, and the −840/+23 mDNA with mutated downstream ATG was synthesized using primers A and E. The −840/+23 mm-DNA with both upstream and downstream ATGs mutated was synthesized by ligating two PCR-generated DNA fragments corresponding to nucleotide positions −840 to −63 and −83 to +23. Primers A and B were used to generate the −840/−63 DNA, and primers C and E were used to generate the −83/+23 DNA. The common *Xho*II restriction site within the −83 to −63 region was used for ligating the two fragments. The fragments were cloned in *Hin*dIII and *Sal*I sites of pCAT basic plasmid. Various 5′ deletions were created by restriction digestion or PCR amplification of −840/+23mCAT plasmid DNA. The −75/+23C$^{\Delta}$mCAT with mutations in the complex C protein binding site was generated by using overlapping PCR.

Transcriptional activities were measured by transient expression of various CAT plasmid constructs in murine 3T3 fibroblasts or, in some cases, SV-40-transformed COS cells and rat hepatoma H4II EC3 cells essentially as described (Basu *et al.*, 1993; Carter *et al.*, 1992).

In Vitro Transcription. In vitro transcription was carried out using a HeLa cell nuclear extract system prepared according to Dignam *et al.* (1983). Reactions were carried out in 20 μL final volumes with 100 μg of nuclear protein and 400 ng of template DNA using conditions essentially as described in a recent paper from this laboratory (Sucharov *et al.*, 1995). RNA from reaction mixtures was recovered by phenol−chloroform extraction and ethanol precipitation and assayed by reverse transcriptase-based primer extension analysis.

^{32}P end-labeled CAT reverse primer (1 ng) (5′ CGGTG-GTATATCCAG 3′) which is complementary to CAT gene sequence 2331−2345 was used for assaying the *in vitro* transcripts by primer extension.

Gel Mobility Shift Assays for DNA-Protein Binding. The −83/+23 double-stranded DNA from exon 2 of the CYP27 gene containing the putative transcription initiation site and the upstream protein binding motifs was used as the probe. Nuclear protein extracts from 3T3 fibroblast cells were prepared by the method of Dignam *et al.* (1983). The 106 bp DNA was released by digesting 4 μg of −83/+23 DNA with *Eco*RI and *Sal*I, gel-purified, and 3′ end-labeled by gap filling with [^{32}P]dATP using Klenow polymerase. The probe was freed of unincorporated radioactivity by chromatography on a Sephadex G50 column. DNA−protein binding reactions were carried out in 20 μL volumes using 0.1−0.2 ng (10 000−15 000 cpm) of labeled DNA, 5−8 μg of nuclear protein extract, and 1 μg of poly(dI:dC) under conditions previously described (Singh *et al.*, 1986). Concentrations of double-stranded oligonucleotides used for competition were determined by ethidium bromide staining on nondenaturing polyacrylamide gels. Competing DNAs were preincubated in binding reactions with added proteins for 10 min before the addition of labeled probe. DNA−protein complexes were resolved by electrophoresis on 4% nondenaturing acrylamide gels using 0.25 × TBE (12.5 mM Tris-borate (pH 7.5) and 0.25 mM EDTA) buffer system.

DNAse I Footprinting. The −45/+23mCAT DNA was digested with *Hin*dIII, and the 5′ end of the sense strand was labeled using polynucleotide kinase and [γ-^{32}P]ATP following dephosphorylation with calf intestinal phosphatase. The −45/+23 insert DNA was released by digestion with *Sal*I and recovered by gel purification. The DNA probe (10^5 cpm) was incubated with 100 μg of 3T3 cell nuclear extract for 30 min at room temperature under conditions described above for the DNA−protein binding assays. The mixture was treated with 0.1 unit of DNAse I (Promega Biotech. Corp., Madison, WI), and the complexes were resolved and analyzed for protection essentially as described (Basu *et al.*, 1993; Carter *et al.*, 1992).

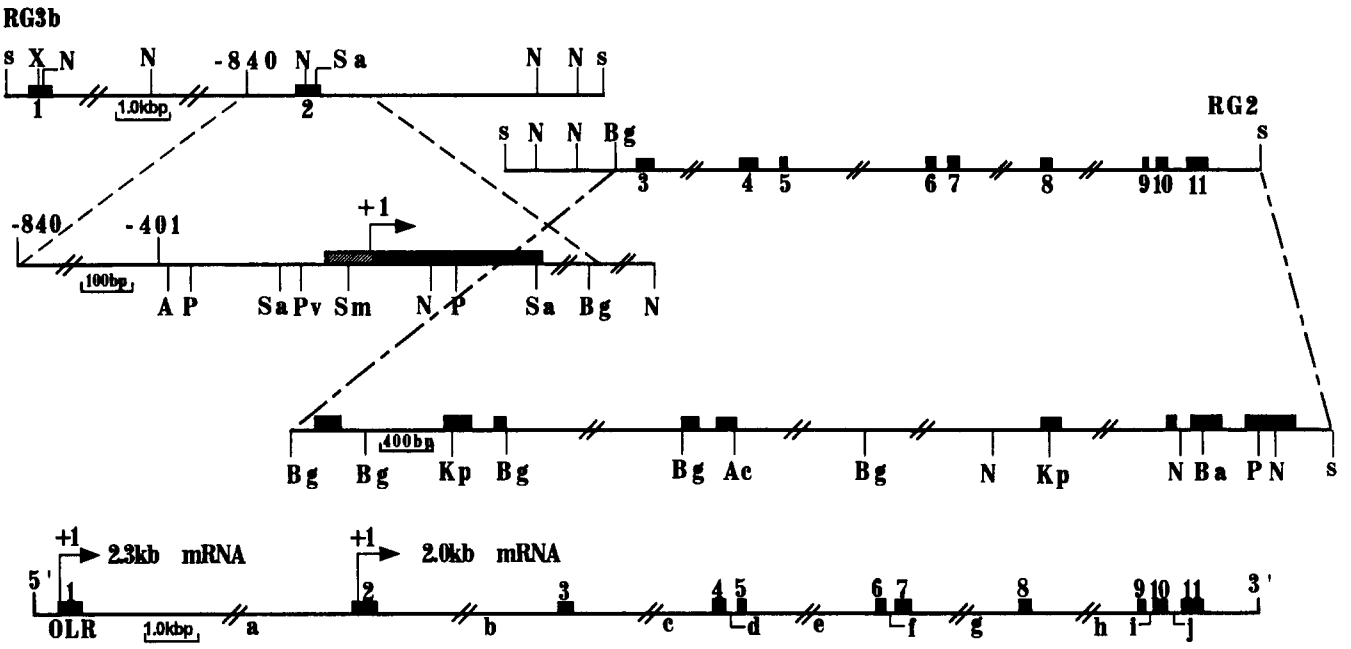

FIGURE 1: Restriction map and intron–exon junctions of the rat CYP27 gene. Two overlapping genomic clones, RG3b and RG2, were isolated and restriction mapped by Southern hybridization using different regions of the P-450c27/25 cDNA probe. The two clones represent a contiguous segment of the rat genomic DNA as tested by the presence of a common 0.6 kb *Nco*I fragment and its flanking sequence. The positions of intron–exon junctions are based on the partial nucleotide sequence presented in Figure 2. The composite gene map is presented at the bottom. The dark boxes represent exons numbered 1–11. The introns are marked as a–j. The hatched box represents the upstream 86 bp promoter region from the 5′ end of exon 2. The position of the putative transcription initiation site at +1 was based on the S1 mapping and primer extension analysis presented in Figure 3. The restriction sites on the genomic map are denoted as follows: A = *Ava*I, Ac = *Acc*I, Bg = *Bgl*II, Ba = *Bam*HI, Kp = *Kpn*I, N = *Nco*I, P = *Pst*I, Pv = *Pvu*II, s = *Sal*I, Sa = *Sac*I, Sm = *Sma*I, and X = *Xba*I. Restriction sites from the vector region are represented by lowercase letters.

RESULTS

Intron–Exon Organization and the Restriction Map of the Rat CYP27 Gene. By screening about 10^6 plaque-forming units of a rat genomic DNA library (Stratagene Inc.), we isolated four positive clones, and the DNA inserts from these clones were mapped by using various restriction enzymes as shown in Figure 1. By Southern blotting and probing with the full length cDNA probe (Su *et al.*, 1990) and also probes corresponding to the 5′ and 3′ termini, two clones designated as RG3b and RG2 were found to contain the entire length of the CYP27 gene (see Figures 1 and 2). DNA fragments hybridizing to the various regions of the cDNA as well as the 3′ and 5′ flanking regions were subcloned in pGEM 3z, 5z, and/or 7z vectors and sequenced. Most of the gene, including the exons, intron–exon junctions, and introns of <2 kb, was sequenced in both directions. The sizes of some of the large introns (>2 kb) were determined by restriction analysis. The restriction map of the entire rat CYP27 gene and the nucleotide sequence of the region encoding the shorter 2 kb mRNA are presented in Figures 1 and 2, respectively. A genomic map presented in Figure 1 shows that the region encoding the 2.3 kb full length mRNA spans over a 30 kb region. The gene is comprised of 11 exons of 290, 415, 195, 198, 80, 121, 171, 167, 83, 212, and 360 nucleotides in length that are intervened by 10 introns of over 10 kb, 5.2 kb, 1.8 kb, 148 bases, 3.0 kb, 112 bases, 3.2 kb, 4.2 kb, 83 bases, and 148 bases, respectively, from the 5′ to the 3′ end. The clone RG3b contained an undetermined portion of the 5′ upstream region, in addition to exons 1 and 2, the >10 kb intron a, as well as a small portion of intron b (see Figure 1). Another clone, RG2, contained exons 3–11, introns b–j, and also the 3′ flanking region of the CYP27 gene. The overlapping region between the two clones was detected by restriction mapping, Southern hybridization, and sequencing.

The exon sequences are identical to the corresponding cDNA sequence (Shayiq & Avadhani, 1992; Su *et al.*, 1990). The 3′ boundary of the first exon separated by a large (over 10 kb) intron corresponds to sequence 290 of the 2.3 kb cDNA reported previously (Shayiq & Avadhani, 1992). Also, the putative 5′ end of the smaller 1.9–2.0 kb mRNA described in our previous study (Shayiq & Avadhani, 1992) lies within the second exon. The intron–exon splice junctions of the rat CYP27 gene (see Figure 2) follow the universal donor–acceptor rule of GT and AG, respectively.

Nature of the 5′ Truncated 2 kb mRNA. Previous S1 analysis of rat liver RNA using a probe from the 5′ terminal region of the 2.3 kb cDNA suggested the occurrence of a shorter 1.9 kb mRNA in addition to the full length 2.3 kb mRNA. The existence of the 5′ truncated RNA and its precise map position on the gene were determined by S1 nuclease protection. The primer PAS-3, containing the antisense sequence of the translated region, was used for preparing the S1 probe as described in Experimental Procedures. As shown in Figure 3A, a 95–96-nucleotide DNA fragment was protected against S1 nuclease digestion after hybridization with total rat liver RNA. The major protected fragment maps to a C residue (marked with an arrow), although a minor component mapping to an A residue at the +2 position is also seen. The +1 position corresponds to nucleotide 377 of the 2.3 kb cDNA sequence [see Shayiq & Avadhani (1992)]. Primer extension analysis using the same 5′ end-labeled PAS-3 primer (Figure 3B) demonstrated the existence of RNA species with a nearly identical 5′ terminus mapping to the A residue immediately downstream of the +1 position. The one-base difference in the mRNA

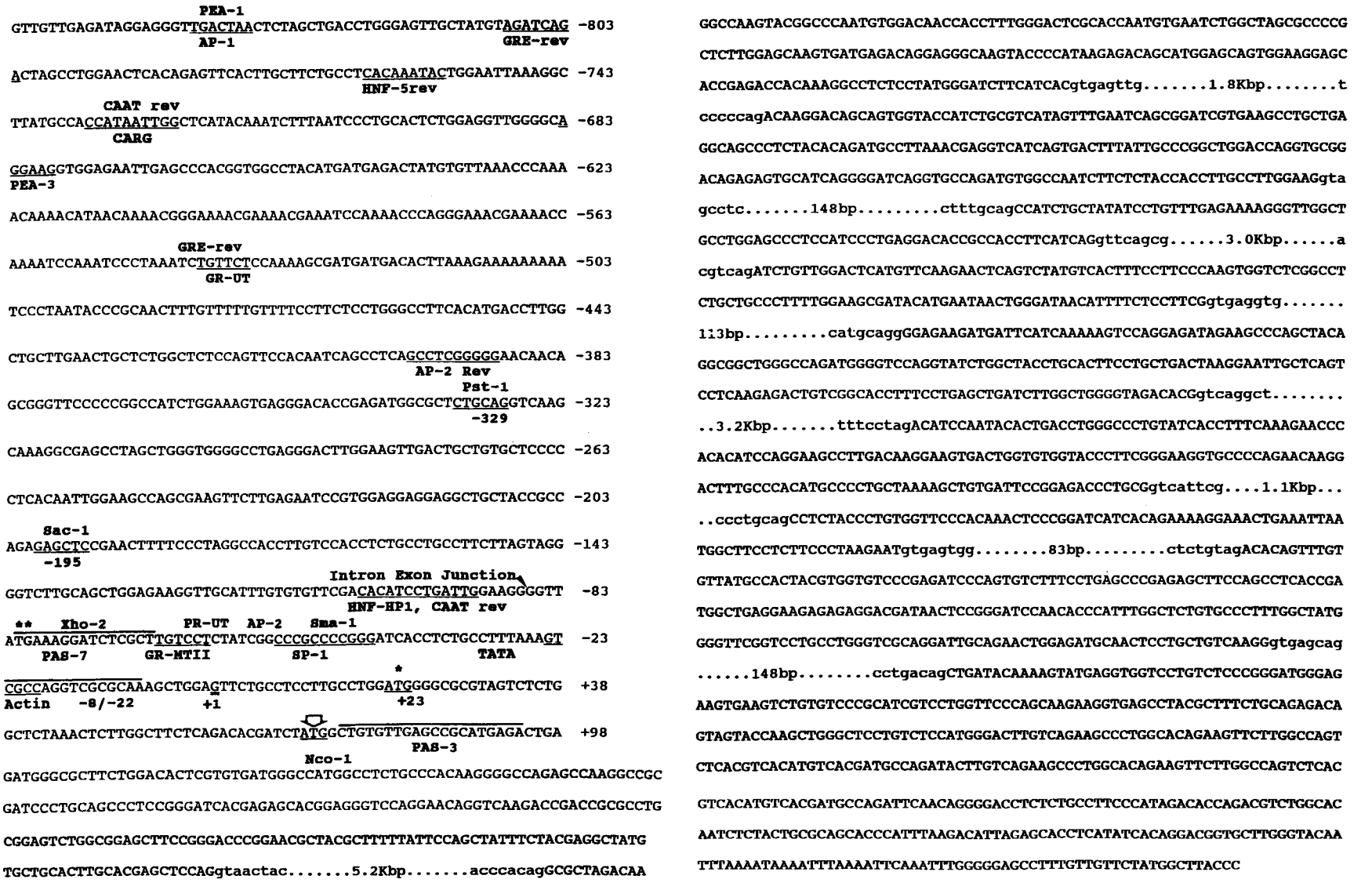

FIGURE 2: Nucleotide sequence of the rat CYP27 gene encoding the 2 kb mRNA and its 5′ flanking region. Restriction fragments of genomic clones RG3b and RG2 were cloned in various sequencing vectors and sequenced as described in Experimental Procedures. The intron sequence is presented in lowercase letters, and the intron sizes are indicated. Except for the introns larger than 2 kb, the genomic region shown was sequenced in both directions. The transcription start site at +1 was based on the S1 and primer extension analyses presented in Figure 3. Various protein binding motifs and hormonal response elements were identified on the basis of a computer search. The junction at −86 is based on the nucleotide sequence of the second exon and its flanking sequences . The translation initiation site at the +71 position is indicated by an arrow. The in-frame upstream ATG at −82 is indicated with two asterisks, and the out-of-frame downstream ATG at +20 is indicated with a single asterisk. These ATGs were also identified in the 2.3 kb cDNA sequence (Shayiq & Avadhani, 1992). The positions of the antisense PAS-3, PAS-7, and −8/−22 primers have been indicated by a line above the sequence.

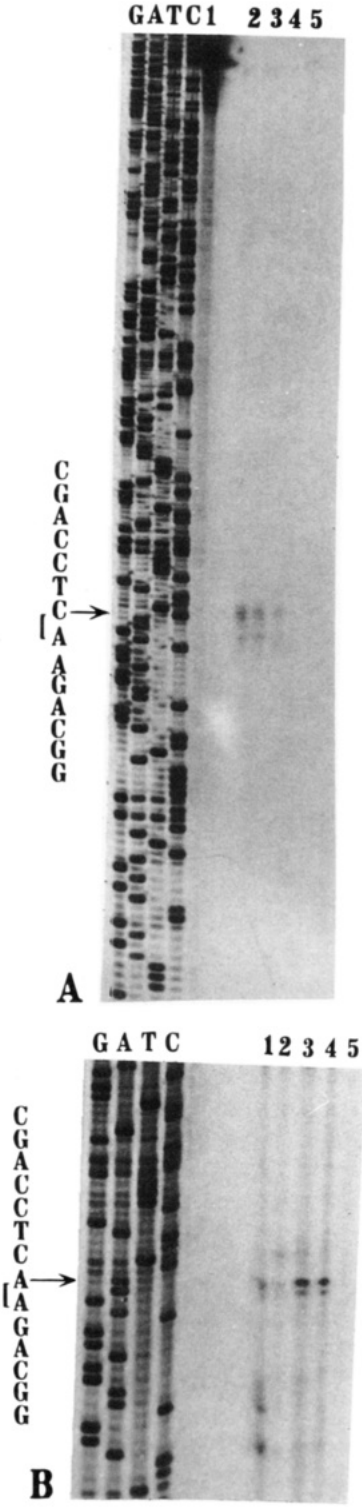

FIGURE 3: S1 nuclease protection and primer extension analyses to map the mRNA 5′ termini. In part A, 5′ end-labeled PAS-3 primer was used to generate a 289-nucleotide single-stranded DNA probe and used for the S1 nuclease protection analysis: lane 1, 25 000 cpm of the untreated DNA probe; lanes 2−4, 100 000 cpm of the probe in each case was annealed with 100 μg of total rat liver RNA and digested with 100, 200, and 300 Vogt units of S1 nuclease, respectively; and lane 5, 100 000 cpm of the probe was annealed with 100 μg of yeast tRNA and digested with 100 Vogt units of S1 nuclease. In part B, 5′ end-labeled PAS-3 primer was used for the reverse transcriptase-based primer extension against RNA templates: lanes 1 and 2, 30 μg of total RNA from rat liver; lanes 3 and 4, 3 μg of rat liver poly(A) RNA; and lane 5, 30 μg of yeast tRNA. Sequencing ladders in both parts A and B were generated using the PAS-3 primer and a genomic DNA subclone from the intron a−exon 2 region of the gene as the template. Both the primer extension and S1 products were resolved under denaturing conditions on 6% polyacrylamide sequencing gels.

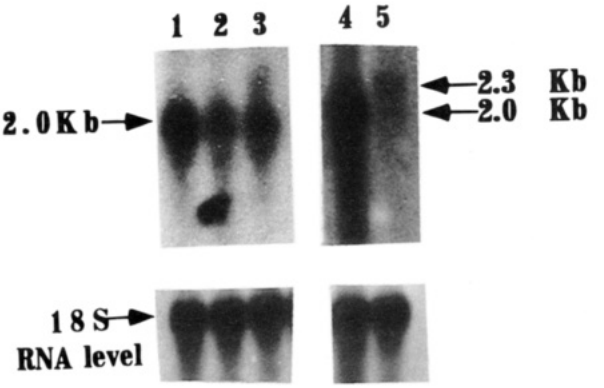

FIGURE 4: Northern blot hybridization of rat liver, kidney, and hepatoma RNA. RNA (30 μg) from tissues or rat hepatoma cells (H4II EC3) were electrophoresed on a formaldehyde-containing agarose gel, transblotted to Nytran membrane, and hybridized with ^{32}P-labeled P-450c27/25 cDNA probe as described in Experimental Procedures: lanes 1 and 4, female liver RNA; lane 2, male liver RNA; lane 3, rat hepatoma RNA; and lane 5, kidney RNA. *Hin*dIII-digested phage λ DNA fragments were run alongside as molecular weight markers and were detected by hybridization with ^{32}P-labeled λ DNA probe. The apparent sizes of mRNAs were determined on the basis of their migration relative to the λ DNA markers. The RNA loading concentrations were monitored by hybridizing the blots with ^{32}P-labeled human 18S DNA probe.

5′ map position obtained with the primer extension method probably represents the stearic restrictions of the 5′ cap structure on the reverse transcriptase. Although not shown, primer extension with rat kidney RNA also yielded larger products, possibly due to the occurrence of the 5′ extended 2.3 kb mRNA species. These results therefore suggest that the putative 2 kb mRNA start site lies 86 nucleotides downstream from the 5′ junction of the second exon (see Figures 1 and 2).

The nature of the putative 2 kb mRNA was further investigated by Northern blot analysis of RNA from the male and female rat livers, cultured rat hepatoma (H4II EC3) cells, and kidney. As shown in Figure 4, hybridization with the P-450c27/25 cDNA probe yielded predominately the 2 kb mRNA with very low 2.3 kb mRNA species in both female and male rat livers (lanes 1 and 2). It should be noted that, in our previous study (Shayiq & Avadhani, 1992), the major species of transcripts in the liver was designated in error as the 2.3 kb species. Rat kidney RNA contains a minor amount of the 2 kb species and a relatively more abundant 2.3 kb mRNA species (lane 5). Results in Figure 4 also show that the rat hepatoma RNA contains mostly the 2 kb species with detectable 2.3 kb species (lane 3). Finally, female liver contains nearly 2−3-fold higher levels of the 2 kb mRNA than male liver.

As shown in Figure 5A, the use of the 5′ RACE analysis to reverse transcribe the extended region of the rat kidney and hepatoma mRNAs, followed by the PCR amplification using two nested primers, yielded a 300 bp DNA fragment. Similarly performed control reactions without the reverse transcriptase step yielded no detectable DNA products (Figure 5A). The 300 bp amplified DNA hybridized with a 172 bp ^{32}P-labeled probe from the 5′ end of the 2.3 kb cDNA sequence reported previously (Figure 5B). Additionally, both a ~400 bp PCR product obtained with the two terminal primers, −8/−22 AS and a sense primer from the anchor region (results not shown), and the 300 bp product generated by using the nested primers (PS-1 and PAS-7) were sequenced and found to contain the 5′ terminal sequence of

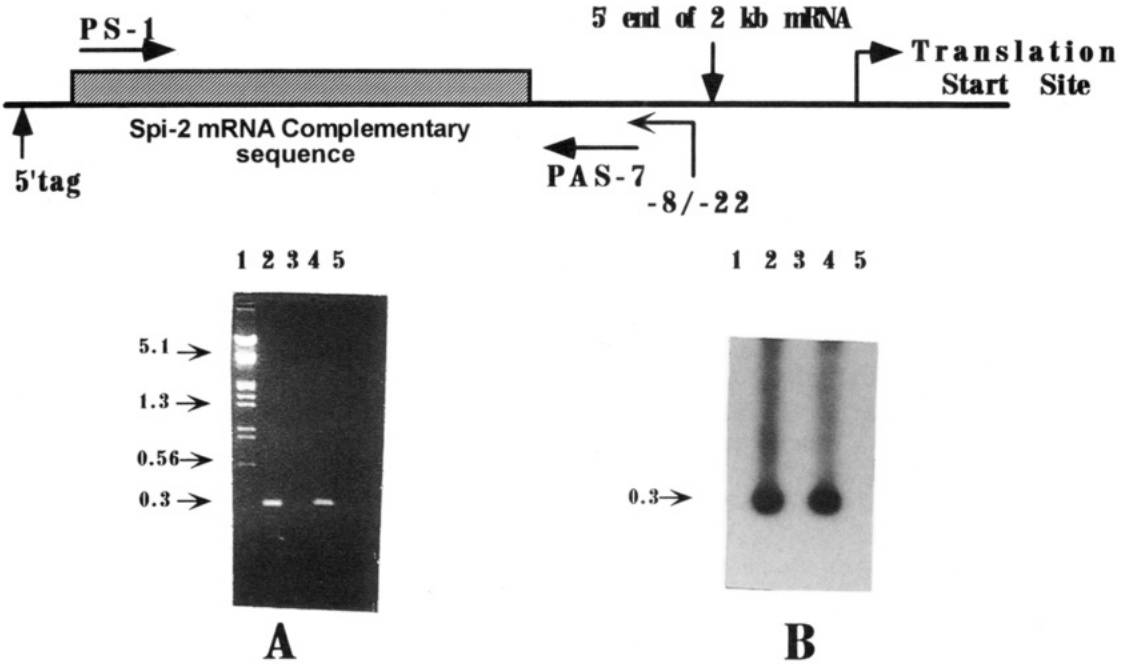

FIGURE 5: 5′ end analysis of the extended 2.3 kb mRNA by 5′ RACE. Poly(A) RNA (4 μg each) from rat kidney and rat hepatoma was used for the reverse transcriptase-based synthesis of antisense strand DNA and subsequent amplifications with the terminal primers (sense 5′ tag and −8/−22 antisense) in the first step and the nested primers (PS-1 and PAS-7) in the second step as described in Experimental Procedures. The illustration at the top of the figure shows the positions of various primers, the Spi2 mRNA complementary region, and the strategies used for the RACE analysis. In part A, 10 μL aliquots of PCR products using the nested primers were resolved on a 1.5% agarose gel: lanes 2 and 4, PCR-amplified RACE products using poly(A) RNA from rat hepatoma and kidney, respectively; lanes 3 and 5, control reactions with rat hepatoma and kidney RNAs, respectively, run without addition of AMV reverse transcriptase; and lane 1, λ DNA digested with *Hind*III and *Eco*RI restriction enzymes run as a marker. Part B represents the blot hybridization pattern of the gel in part A, using the ^{32}P-labeled 172 bp DNA probe from the Spi2 mRNA complementary region.

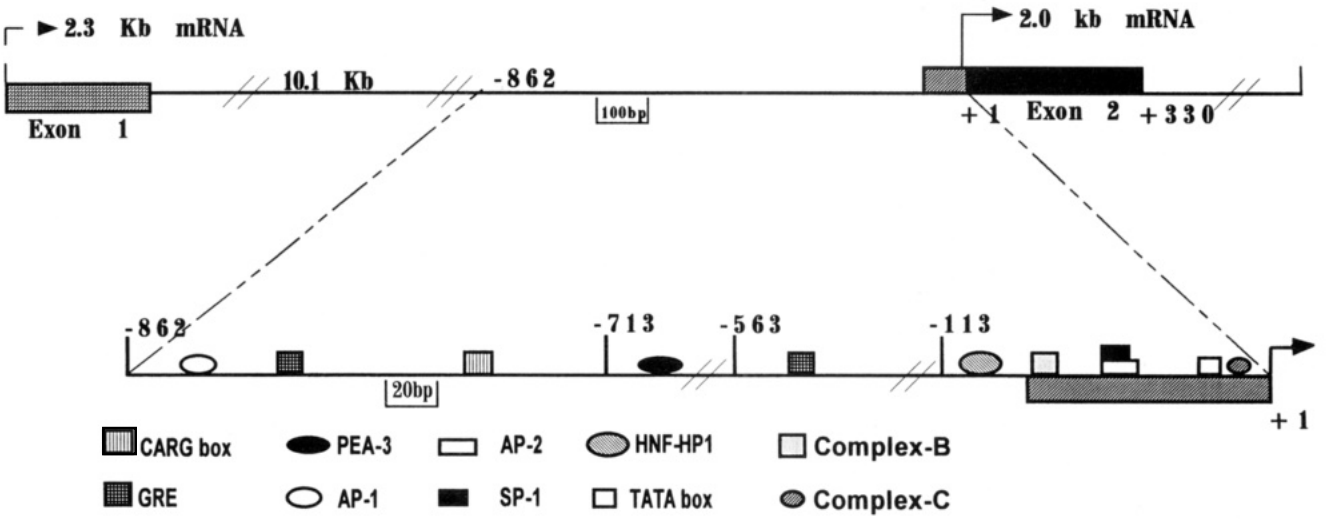

FIGURE 6: Locations of transcription initiation site and various protein binding motifs relative to the exon 1−intron a−exon 2 junctions. The position of transcription initiation site on exon 2 and the various protein binding consensus motifs are indicated. The lightly shaded region upstream of the +1 site represents the 5′ terminal region of exon 2 on the basis of the nucleotide sequence data. The positions of protein binding motifs (Sp1/Ap-2, complexes B and C) within the −86 to +1 exon region are based on DNA−protein binding and DNAse 1 footprinting presented in Figures 8 and 9. The locations of consensus protein binding motifs upstream of the −86 sequence are mainly based on a computer search. The position of exon 1 and the transcription start site for the 2.3 kb mRNA are indicated.

the 2.3 kb cDNA sequence reported previously (Shayiq & Avadhani, 1992). The results of Northern blot analysis and 5′ walking therefore suggest that the second exon region upstream of +1 and also the entire first exon are expressed as part of the 5′ extended 2.3 kb mRNA.

Characteristics of Sequence 5′ to the 2 kb mRNA Start Site. Analysis of the 840 bp region upstream of the +1 site revealed many interesting features as depicted in Figures 2 and 6. An atypical TATA like motif, TTTAAA, is located about 24 nucleotides upstream of the start site at +1. Similar sequences in the human P-450c17 gene (Leonard & Miller, 1987) and the rat tyrosine hydroxylase gene (Cambi *et al*., 1989) have been shown to serve as functional TATAA motifs. Two CAAT boxes in the reverse orientation (TGATTGG, and TAATTGG) were detected at nucleotide positions −91 to −97 and −725 to −731, respectively (see Figure 2). A computer-assisted analysis of the 5′ upstream region revealed consensus sequences for a number of putative cis-acting regulatory elements within the −840 bp flanking region of the rat CYP27 gene. There are two GC rich sites, CCCGCCCCGGG at position −43 to −53 and CCCCCG-GCC at position −367 to −375, that exhibit complete or

partial homology to Sp1 factor binding motifs (Kadonaga *et al.*, 1986). There is no canonical pyrimidine rich "initiator" within the immediate flanking region of the transcription start site at +1, which is a hallmark of many TATA-containing and TATA-less genes (Smale & Baltimore, 1989; Zawel & Reinberg, 1992; Basu *et al.*, 1993). Nevertheless, the transcription initiation region contains a sequence stretch rich in pyrimidine nucleotides, although its role in transcription initiation currently remains unknown. The sequence TGTCCT at position −66 to −61 resembles the glucocorticoid receptor (GR) binding site of the human metallothionein IIA (MT-IIA) gene (Karin *et al.*, 1984). Sequence TGTCCTCT at position −66 to −59 shows over 80% homology to the progesterone receptor (PR) binding site TGTTCACT of the uteroglobin gene and appears to overlap with the putative GR site (Bailly *et al.*, 1986) . Similar overlapping organization of both the GR and the PR binding sites have been reported in the chicken lysozyme gene (Renkawitz *et al.*, 1984) and the mouse mammary tumor virus promoter (Vonder Ahe *et al.*, 1985). Sequence motifs resembling the glucocorticoid responsive element (GRE) (Beato, 1989; Okret *et al.*, 1986) in the reverse orientation are present at positions −809 to −802 and −544 to −537. Overlapping the reverse GRE is a GR binding site of the uteroglobin gene (Bailly *et al.*, 1986) at position −542 to −537. A Pea-3 site at −683 to −677 and two sequence motifs showing partial consensus to AP-2-like binding sites (Roesler *et al.*, 1988), CCGCCCCGGG and GCCTCGGGGG (an inverted motif) at positions −56 to −44 and −390 to −399, respectively, were found within the 5′ upstream region. Similar AP-2-like elements were found in the human MT-IIA (Imagawa *et al.*, 1987) and human GH (DeNoto *et al.*, 1981) promoters. The AP-2 element is a basal transcriptional enhancer which is thought to have a role in both cAMP and phorbol ester-mediated transcription stimulation (Roesler *et al.*, 1988). The 5′ upstream region of the rat CYP27 gene also contains sequence motifs generally found in some muscle specific genes such as the actin binding site GTCGCC (Walsh & Schimmel, 1987) at position −24 to −19 and a CArG box (Miwa *et al.*, 1987) at −734 to −725. Currently, the physiological significance of many of these sequences remains unclear.

Functional Analysis of the 5′ Flanking Region. Initially, a 352 bp (position −329 to +23) fragment of the 5′ flanking region containing the transcription start site was cloned in the expression plasmid pCATb in the sense orientation and used in transient expression assays in NIH-3T3 cells (Figure 7) and also in COS cells and rat hepatoma H4II EC3 cells (results not shown). Surprisingly, the promoter construct, −329/+23CAT, failed to drive CAT expression to a significant level in any of these cell systems. A survey of the nucleotide sequence of the promoter construct showed the presence of two ATG motifs flanking both sides of the putative transcription start site at positions −82 and +20, respectively. As shown in Figure 2, the ATG motif upstream of the start site (at −86) was found to be in-frame, while the downstream ATG at +20 was out-of-frame with respect to CAT reading frame. Mutation of the downstream ATG to ATC (−329/+23mCAT) resulted in significant CAT activity in 3T3 cells, which remained unchanged when the upstream ATG was also mutated to ACG (−329/+23 mmCAT, see Figure 7). For these reasons, all subsequent CAT expression constructs used in this study were designed to carry mutated downstream ATG. The −840/+23mCAT construct yielded the highest activity in 3T3 fibroblasts (indicated in Figure 7 as 100% activity). The 5′ deletion constructs −329/+23mCAT, −194/+23mCAT, and −75/+23mCAT gave consistently declining levels of activities in the range of 8−22. The −45/+23mCAT construct yielded activity comparable to that of the −75/+23mCAT construct (Figure 7). These results suggest that the activity of the −45/+23mCAT may represent the basal promoter activity which is upwardly regulated by the various enhancer elements located at the 5′ upstream positions. Although not shown, similar patterns of CAT activity were obtained in COS and rat hepatoma cells.

Protein Factors Binding to the Basal Promoter Region. The nature of proteins binding to the basal promoter region of the CYP27 gene was studied by gel mobility shift analysis using nuclear protein extract from 3T3 fibroblast cells. Although not shown, nuclear extracts from COS and rat hepatoma cells yielded similarly migrating complexes. The 3′ end-labeled −83 to +23 DNA showing a basal level of transcription activity in transient expression assays forms three protein-bound complexes termed A, B, and C (see lane 2, Figure 8). All three complexes were competed by unlabeled excess (100 and 200 ng) −83 to +23 DNA (lanes 3 and 4). As shown in lane 5, addition of unlabeled Sp1 DNA preferentially eliminated complex A without affecting complex B. Further, NF-E1 (YY-1) DNA which was shown to bind to some transcription initiator motifs (Basu et *al.*, 1993; Seto *et al.*, 1991; Javahery *et al.*, 1994) did not compete with any of the complexes. Both −45/+23 DNA and −33/+23 DNA selectively competed with complex C, suggesting that the protein binding site for this unknown complex resides within sequence −33 to +23 of the promoter. The complex B binding sequence appears to reside within −83 to −71 since this complex formation was not affected by addition of excess unlabeled −45/+23 DNA, Sp1 DNA, or glucocorticoid consensus sequence at −69 to −58 (results not presented). Direct support for this possibility was obtained by showing that a synthetic DNA containing sequence −83 to −68 selectively competed for complex B formation (see lanes 11 and 12).

The complex C binding region was fine mapped by DNAse I footprinting of the sense strand-labeled −45/+23 DNA (Figure 9A). A comparison of DNAse I-digested bands from the unbound (lane 2) and protein-bound complexes (lanes 3−5) indicate that sequence region −12 to −20 is relatively more protected in the bound complex. Although not shown, the same sequence region is protected when the antisense strand-labeled −45/+23 DNA probe is used. That the DNAse I-protected sequence represents the complex C-forming region of the promoter was further verified by the ability of synthetic DNA sequence to compete for complex formation in gel mobility shift assays. As shown in Figure 9B, the fastest migrating protein complex (complex C) with the −83/+23 DNA probe was selectively competed against by the wild type −22 to −8 synthetic DNA sequence (CGCCAGGTCGCGCAA), but not by mutant DNA-containing base substitutions.

On the basis of the results of competition in gel mobility shift assays and DNAse I footprinting of select DNA probes, regions of −83/+23 DNA involved in the formation of the three complexes are indicated in Figure 9C. A computer analysis suggests that both complex B and C binding regions do not represent defined protein binding motifs, and currently, the nature of these complexes remains unknown.

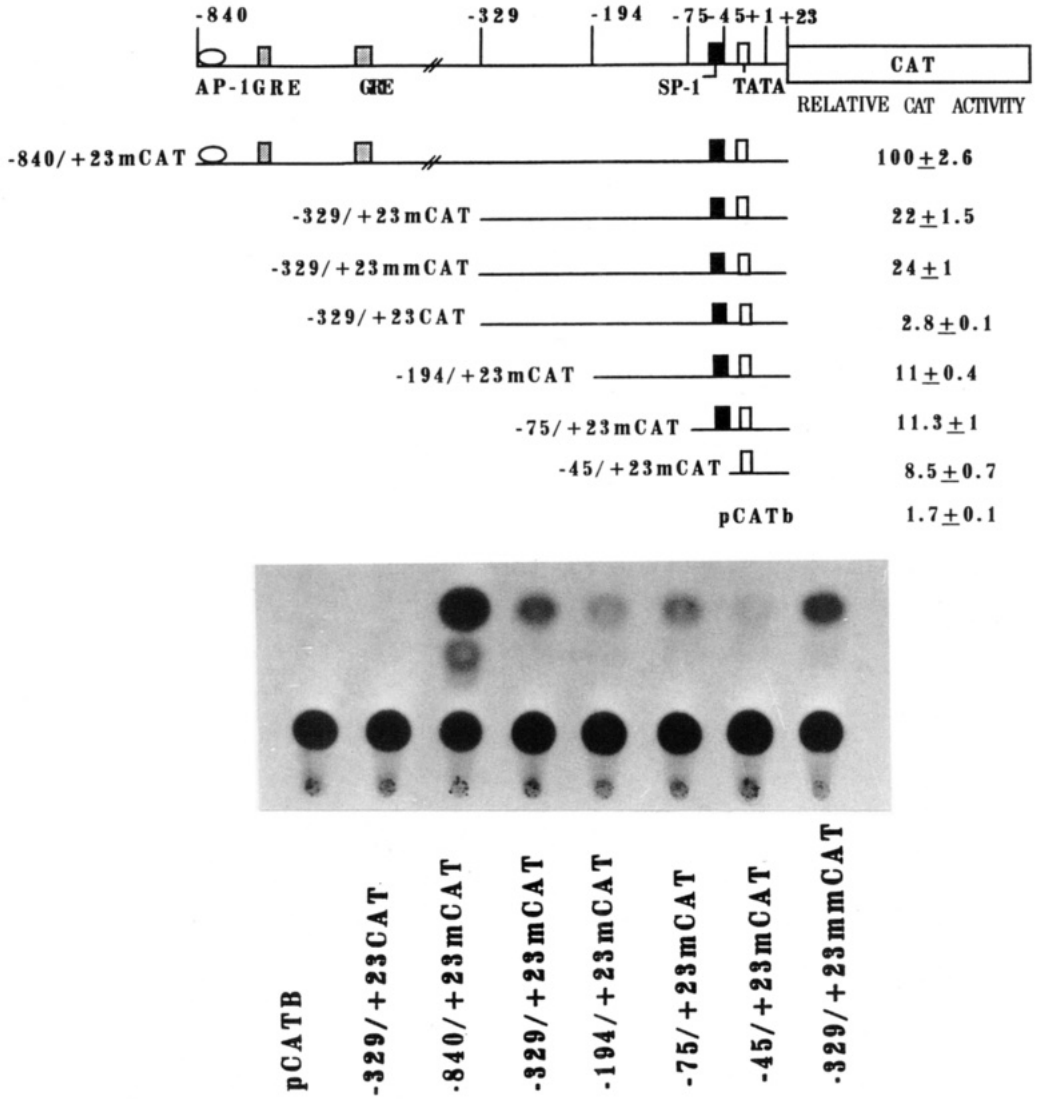

FIGURE 7: *In vivo* transcription analysis of the sequence upstream of the putative transcription start site. The −840/+23mCAT DNA with downstream ATG mutated was generated as described in Experimental Procedures, and various 5′ deletion clones were generated by either restriction digestion or PCR amplification. The intact −840/+23mCAT DNA and also various 5′ deletions were transfected in NIH 3T3 fibroblast cells. Cells were cotransfected with a β-gal expression plasmid, and the CAT activities of cell extracts were normalized for the β-galactosidase activities. The acetylated products were resolved by thin layer chromatography (TLC) and quantitated by scanning through a Vanguard radiometric scanner. CAT activity is expressed as a relative value compared to the activity of −840/+23mCAT DNA which is considered to be 100. Values are an average (± standard deviation) of determinations performed with three independent experiments. A typical TLC pattern has been presented at the bottom.

Role of the TTTAAA and the Complex C-Forming Sequence Motifs in Promoter Function. The basal transcription activity of the 45 bp sequence upstream of the putative transcription start site (−45/+23) and the role of sequence motifs contained within this region in promoter function were investigated using an *in vitro* HeLa nuclear transcription system (Dignam *et al.*, 1983). As shown in Figure 10, the −75/+23mCAT DNA as well as −840/+23mCAT DNA templates supported *in vitro* transcription at comparable rates. The −45/+23mCAT DNA also supported *in vitro* transcription, though at a significantly reduced level, suggesting that it is a weaker template. All three DNA templates yielded transcripts mapping to the +1 position in addition to a less prominent species mapping to the −8 position of the promoter. In contrast, pCAT basic plasmid without any insert failed to support any significant *in vitro* transcription initiation at these positions, suggesting the specificity of the *in vitro* system. The *in vitro* results also showed that the −22/+23mCAT template with deleted TTTAAA motif yields no detectable transcription initiation at either of these positions, indicating the importance of this motif in promoter function. The −75/+23C$^{\Delta}$mCAT DNA construct with intact TTTAAA sequence, but mutated complex C-forming region, yielded no significant level of transcripts mapping to the −8 position and also a vastly reduced transcript mapping to the −3 position of the promoter. These results suggest that complex C-forming motif has an important role in transcription activation.

DISCUSSION

Previous studies from this laboratory indicated that the rat cytochrome P-450c27/25 mRNAs are expressed as a full length 2.3 kb and a 5′ truncated 2 kb species, both of which encode the 52 kDa P-450c27/25 protein targeted to the mitochondrial compartment in heterologous cells (Shayiq & Avadhani, 1992; Su *et al.*, 1990). Results of Northern blot analysis showed that the ovaries from immature rats express only the 2.3 kb species, while treatment of rats with pregnant mare's gonadotropin resulted in the induction of the 2.3 kb RNA and also a distinctly migrating 2 kb RNA (Su *et al.*,

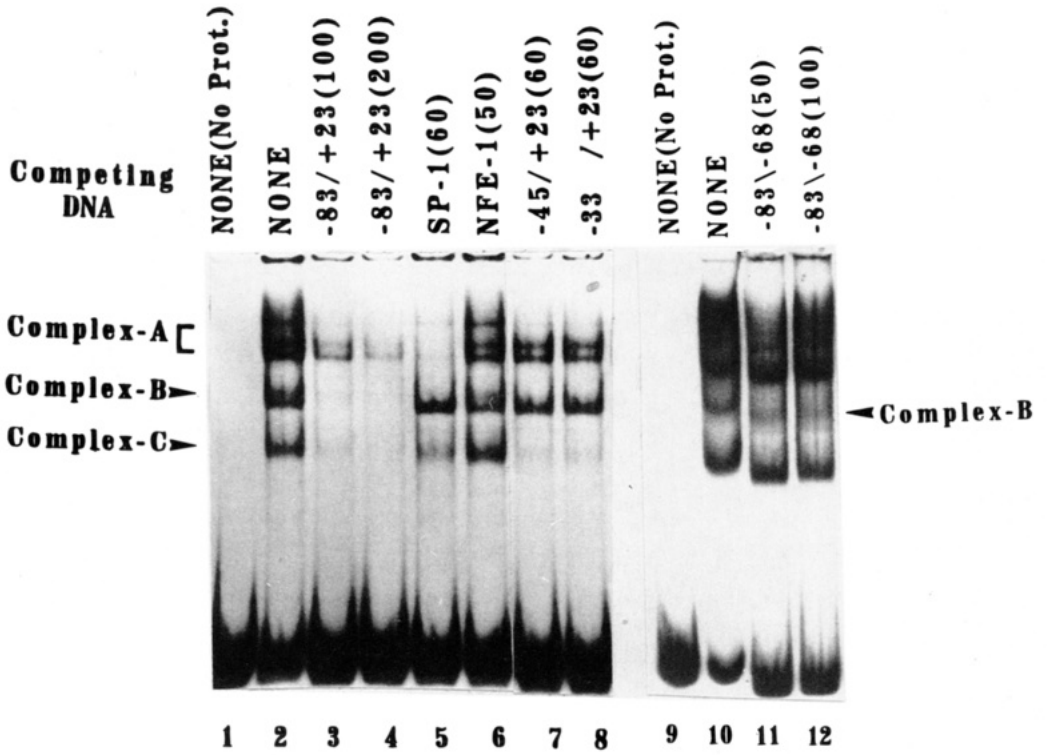

FIGURE 8: Protein binding ability of the exon region immediately upstream of the transcription start site. The −83/+23 DNA representing the exon 2 transcription initiation region was used as the probe, and protein binding was assayed by gel mobility shift assays as described in Experimental Procedures. Competing DNAs were added to the reaction mixture 10 min before the addition of probe DNA. The reaction mixtures were analyzed on a 4% acrylamide gel using 0.25 × TBE buffer. Reactions in lanes 1 and 9 were carried out without added nuclear extract, and those in lanes 2−8 and 10−12 contained 6 μg each of 3T3 nuclear protein. Lanes 3−8, 11, and 12 also contained various competing DNAs. Numbers in parentheses indicate nanogram amounts of competing DNAs added.

1990). In extension of these observations, results presented in this study show that the 2 kb mRNA is the major species in the liver while kidney expresses the longer 2.3 kb mRNA as the major form. These results on the detection of multiple mRNA species, along with the cDNA sequence studies (Shayiq & Avadhani, 1992) suggest that the rat CYP27 gene might be expressed differentially using alternate transcription initiation sites. In this study, we have investigated this possibility by characterizing the rat CYP27 gene and found that the 5′ truncated 2 kb mRNA represents a distinct transcription unit directed by an immediate upstream promoter.

The rat CYP27 gene consists of 11 evenly sized exons, and in this regard, the structural features compare well with those of other steroidogenic P-450 genes such as the mouse adrenal steroid 21-hydroxylase (Parker *et al.*, 1986), human and bovine adrenal steroid 21-hydroxylase (Higashi *et al.*, 1986; Chung *et al.*, 1986), rat side chain cleavage gene (Oonk *et al.*, 1990), human side chain cleavage gene (Morohashi *et al.*, 1987), human and bovine 11-β-hydroxylase gene (Mornet *et al.*, 1989; Hashimoto *et al.*, 1989), and rat vitamin D_3 24(*R*)-hydroxylase genes (Ohyama *et al.*, 1993). As in the case of other mitochondrial P-450 genes, the heme binding domain and the putative adrenodoxin binding domains are encoded by different exons. A distinctive feature appears to be the existence of four very short introns of 148, 112, 83, and 148 bp separating exons 4 and 5, 6 and 7, 9 and 10, and 10 and 11, respectively. Thus, as suggested in the evolutionary branch patterns of various P-450 genes based on nucleotide sequence similarities and subcellular location (Nebert & Nelson, 1991), the rat CYP27 gene appears to be a close relative of the CYP11 family. Overall, the intron−exon organization of the rat CYP27 gene appears to differ significantly from that of the human c27 gene which has been reported to contain nine exons (Leitersdorf *et al.*, 1993).

Sequencing of rat genomic DNA (Figure 2) showed that the first intron−second exon junction corresponds to the cDNA sequence 156 nucleotides upstream of the translation initiation site [see Figure 2 and Shayiq & Avadhani (1992)]. On the basis of the apparent size of the faster migrating 2 kb mRNA, initially we felt that this intron−exon junction may correspond to the start of this transcription unit. Surprisingly, however, both S1 analysis using a 5′ end-labeled probe from this region (nucleotides −194 to +95 of the gene, see Figure 2) and a primer extension analysis failed to show any significant RNA 5′ ends mapping to this intron−exon junction. A prominent S1 protection, as well as a primer extension product, however, corresponded to a position 86 nucleotides downstream of this junction (see Figure 3). In keeping with the RNA-mapping data, the intron−exon junction appears not to represent the transcription initiation site for the following reasons. (1) there is no TATA motif within the 50 nucleotides upstream of this junction region. (2) there is no pyrimidine rich initiator motif around the intron−exon junction which is the hallmark of most TATA-less as well as TATAA-containing genes, and also there is no Sp1 site in the immediate upstream region of the junction. (3) *In vivo* transcriptional assays by transient expression of 5′ deletion CAT constructs show no significant drop in transcription activity when sequence region −194 to −76, which contains the intron−exon junction, is deleted (see Figure 6). In support of our cDNA sequencing data reported previously (Shayiq & Avadhani, 1992), results of Northern blot and 5′ RACE analyses presented here (Figures 4 and 5) provide direct evidence that the −86 to +1 region

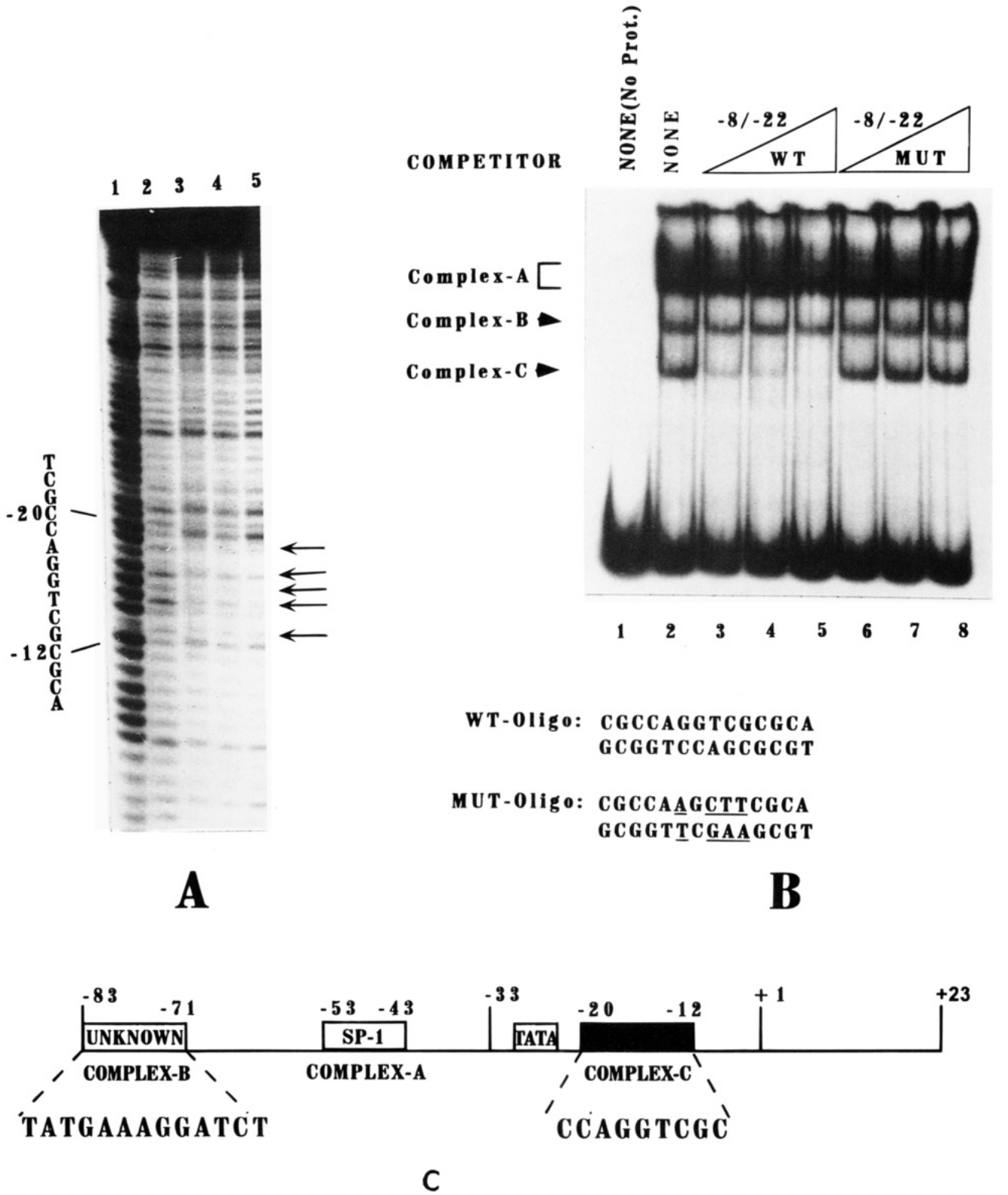

FIGURE 9: Localization of protein binding sites by DNAse 1 footprinting. The sense strand-labeled −45/+23 DNA was used for protein binding with 3T3 nuclear extract and subjected to DNAse 1 digestion as described in Experimental Procedures. In part A, the DNAse 1 digestion patterns of bound complex and also free DNA were analyzed on a 10% sequencing gel: lane 1, G/A sequence ladder; lane 2, free DNA from reaction mixture treated with DNAse for 15 min; and lanes 3−5, digestion patterns of DNA from protein-bound complex subjected to DNAse treatment for 2 and 4 min at room temperature and 15 min on ice, respectively. In part B, the binding specificity of complex C was further verified by gel mobility shift assays using the −83/+23 DNA probe as described in Figure 8: lane 1, probe alone with no extract; and lanes 2−8, probe with added 3T3 nuclear extract. In lanes 3−5, 25, 50, and 100 ng, respectively, of synthetic −22/−8 DNA were added as a competitor. In lanes 6−8, 25, 50, and 100 ng, respectively, of −22/−8Mut DNA were used for competition. Part C indicates the relative locations of complexes A, B, and C on the basis of competition with various DNA sequence motifs and DNAse I footprinting in Figures 8 and 9A,B.

of the first exon is expressed as part of the 5′ extended 2.3 kb mRNA.

As reported previously (Shayiq & Avadhani, 1992), both the 2.3 and 2 kb mRNAs contain additional in-frame as well

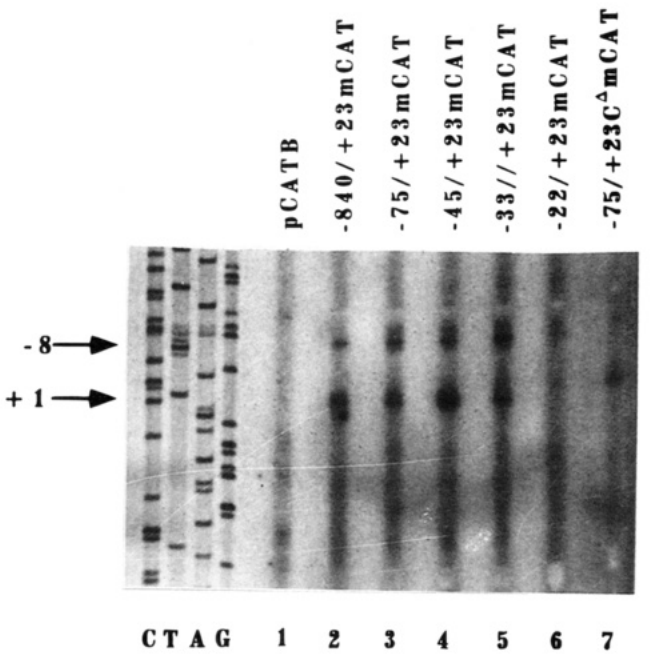

FIGURE 10: Measurement of promoter activity and identification of transcription initiation site by *in vitro* transcription. Various CAT plasmid constructs were used as templates in a HeLa cell *in vitro* transcription system, and the level of *in vitro* transcription was assayed by the reverse transcriptase-based primer extension, using 5′ end-labeled CAT reverse primer as described in Experimental Procedures: lane 1, pCATB; lane 2, −840/+23mCAT DNA; lane 3, −75/+23mCAT DNA; lane 4, −45/+23mCAT DNA; lane 5, −33/+23mCAT DNA; lane 6, −22/+23mCAT DNA; and lane 7, −75/+23C$^{\Delta}$mCAT DNA. The sequencing ladders were generated by using the CAT reverse primer and −840/+23mCAT DNA as the template. The primer extension products were resolved on a 6% sequencing gel.

as out-of-frame ATGs at positions upstream of the translation initiation sites. Preliminary results (R. Shayiq & N. Avadhani, unpublished results) suggest that the translation initiation on these mRNAs at the correct internal ATG codons may involve an unusual ribosome-docking mechanism that enables the ribosomes to bypass the out-of-frame upstream ATGs (Pelletier & Sonenberg, 1988; Macejak & Sarnow, 1991). *In vivo* transcriptional experiments by transfection in 3T3 cells (Figure 7) indicate the need for mutation of the out-of-frame ATG at position +20 to allow a meaningful translation of the CAT mRNA. It should be pointed out that this need stems from the unusual structural features and an intricate translational regulatory controls that feature the mechanism of translation of both 2.3 and 2 kb mRNAs. The CAT reporter constructs described in this paper contain only a part of the 5′ untranslated regions and therefore appear to lack some essential regulatory sequences for bypassing the out-of-frame ATG at the upstream position.

In vivo transcriptional analyses show that the −840/+23mCAT DNA yields maximum activity and deletion of the −840 to −329 region results in a marked reduction (nearly 80%) in activity, suggesting that this region probably contains yet uncharacterized enhancer elements. The region between −75 and +23 shows about 11−12% transcription activity which is consistent with the basal promoter activity of the gene. This stretch of DNA sequence, localized within exon 2, contains a TTTAAA sequence motif, reminiscent of the TATAA sequence at a position 24 nucleotides upstream of the putative transcription initiation site. Additionally, there are upstream Sp1 and Ap-2 sites. Multiple lines of evidence suggest that this is a true transcription promoter. (1) Removal of the TTTAAA sequence by 5′ deletion (−22/+23mCAT clone) causes a near complete loss of *in vitro* transcription activity (Figure 10), suggesting that this motif acts as a TATAA box as in the case of rat tyrosine hydroxylase gene and the human P-450c17 gene (Cambi *et al.*, 1989; Leonard & Miller, 1987). (2). *In vitro* transcription studies (see Figure 10) with intact −840/+23mCAT and also 5′ truncated −45/+23mCAT DNA templates show that transcription initiation occurs very close to or identical to the map position determined by the S1 nuclease protection and primer extension using endogenously expressed RNA. The *in vitro* transcription pattern also shows an additional initiation site not observed in the endogenously expressed CYP27 mRNAs from different sources. This additional site at −8 may be due to activation of a cryptic initiation site under the *in vitro* conditions. Similar activation of cryptic sites downstream of Sp1 specific GC boxes and also some vector sites under *in vitro* conditions have been observed in other promoters (Smale & Baltimore, 1989; O'Shea-Greenfield & Smale, 1992; Zawel & Reinberg, 1992).

The −83 to +23 sequence region of the rat CYP27 gene containing the basal transcription promoter forms at least three different complexes with 3T3 cell nuclear extract. The slowest migrating component, designated as complex A (Figure 8), resolves on polyacrylamide gel as closely migrating multiple bands, all of which are effectively competed with unlabeled excess Sp1 factor binding DNA. It is likely that these multiple components reflect Sp1 factor phosphorylated to different extents as shown by Jackson *et al.* (1990) or other factors known to bind to this GC rich motif (Faisst & Meyer, 1992). On the basis of competition with different restriction fragments and synthetic DNA, the complex B binding sequence appears to reside within the −83 to −68 region of the promoter that shows no consensus to any of the protein binding motifs reported in the literature. However, *in vivo* transcription analysis of 5′ deletion clones by transfection in 3T3 cells shows no detectable stimulatory or repressive effect of this region on promoter activity (Figure 7). Thus, the physiological significance of protein binding to this motif remains unclear. The complex C binding region on the DNA was initially mapped to sequence −20 to −12 of the promoter using DNAse I footprinting (Figure 9A). The binding specificity of complex C was further verified by direct competition with synthetic DNA corresponding to sequence −22 to −8 in DNA−protein binding assays (Figure 9B), and as seen from Figures 2 and 6, this protein binding motif lies between the TTTAAA (putative TATAA) sequence and the initiation site. The complex C forming sequence motif does not compare with any of the known protein binding DNA motifs from the computer data base. Mutations that affect complex C formation nearly abolish transcription originating from the −8 position and cause a markedly reduced and shifted transcription initiation at the +1 position. Thus, the complex C-forming motif may have an important role in both transcription activation and start site positioning. Our results suggest that, in conjunction with the TTTAAA and upstream Sp1 motif, the complex C protein binding sequence functions as part of a minimal promoter. It may therefore act as an unusual initiator that functions from outside the context of the initiation site. As reported for the initiator motifs from TdT and dihydrofolate reductase gene and cytochrome oxidase subunit IV gene (Smale & Baltimore, 1989; O'Shea-Greenfield & Smale, 1992; Carter & Avadhani, 1994; Zawel & Reinberg, 1992), an upstream

TATAA or Sp1 is required for the activity of this motif.

The 5′ heterogeneity of mRNA through transcription initiation at multiple sites has been documented for a number of genes, for example, the *Drosophila* ADH gene (Corbin & Maniatis, 1989), the mouse α-amylase gene (Schibler *et al.*, 1983), the chicken c-*myb* gene (Hahn *et al.*, 1989), the murine cytochrome oxidase subunit Vb gene (Basu *et al.*, 1993), including the bovine adrenodoxin gene (Chen & Waterman, 1992), and the human as well as murine *myc* genes (Marcu *et al.*, 1992). In almost all of the reported cases, the regulatory sequence elements that drive the activity of the downstream transcription units are located within the intron regions immediately upstream of the start sites. The *myc* gene is an exception in that one of the major promoters designed as P_2 is located within exon 1 (Battey *et al.*, 1983; Watt *et al.*, 1983; Bernard *et al.*, 1983). We have therefore characterized a cytochrome P-450 promoter that belongs to a rare family, in which the basal transcription elements are located within the expressed region of the gene. In support of the relative P-450c27/25 levels reported previously (Addya *et al.*, 1991; Saarem & Pedersen, 1987), Northern blot results presented in this paper show that the 2 kb mRNA is detected at 2−3-fold higher levels in the female liver as compared to the male liver (Figure 4). Unpublished results also show that the 2 kb mRNA is induced by pituitary- regulated steroids and growth hormone in hypophysectomized rats. It is therefore likely that the downstream promoter characterized in this study may have an important role in responding to various physiological factors. Further efforts are underway to determine the hormonal responsiveness of this novel promoter.

ACKNOWLEDGMENT

We thank Dr. Byron Kemper for critically reading the manuscript, Drs. A. Basu, C. Vijayasarathy, and R. Carter for helpful suggestions and criticisms reading the manuscript, and Dr. Ji-Kang Fang for synthesizing the DNA primers used in this study. We also gratefully acknowledge the excellent technical help of Vasantha Reddy.

REFERENCES

Addya, S., Zheng, Y.-M., Shayiq, R. M., Fan, J., & Avadhani, N. G. (1991) *Biochemistry 30*, 8323−8330.

Andersson, S., Davis, D. L., Dahlback, H., Jornvall, H., & Russell, D. W. (1989) *J. Biol. Chem. 264*, 8222−8229.

Axen, E., Postlind, H., Sjoberg, H., & Wikvall, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 10014−10018.

Bailly, A., Le-Page, C., Rauch, M., & Milgrom, E. (1986) *EMBO J. 5*, 3235−3241.

Basu, A., & Avadhani, N. G. (1991) *J. Biol. Chem. 266*, 15450−15456.

Basu, A., Park, K., Atchison, M. L., Carter, R. S., & Avadhani, N. G. (1993) *J. Biol. Chem. 268*, 4188−4196.

Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., & Leder, P. (1983) *Cell 34*, 779−787.

Beato, M. (1989) *Cell 56*, 335−344.

Bergman, T., & Postlind, H. (1991) *Biochem. J. 276*, 427−432.

Berk, A. J., & Sharp, P. A. (1977) *Cell 12*, 721−732.

Bernard, O., Cory, S., Gerondakis, S., Webb, E., & Adams, J. M. (1983) *EMBO J. 2*, 2375−2383.

Bjorkhem, I. (1985) in *Sterols and Bile acids* (Danielsson, H., & Sjovall, J., Eds.) pp 231−278, Elsevier, Amsterdam.

Bjorkhem, I., Andersson, O., Diczfalusy, U., Sevastik, B., Xiu, R. J., Duan, C., & Lund, E. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8592−8596.

Cali, J. J., & Russell, D. W. (1991) *J. Biol. Chem. 266*, 7774−7778.

Cali, J. J., Hsieh, C.-L., Francke, U., & Russell, D. W. (1991) *J. Biol. Chem. 266*, 7779−7783.

Cambi, F., Fung, B., & Chikaraishi, D. (1989) *J. Neurochem. 53*, 1656−1659.

Carey, M. C. (1982) in *The Liver: Biology and Pathobiology* (Arias, I. M., Popper, H., Schachter, D., & Shapritz, D. A., Eds.) pp 429−465, Raven Press, New York.

Carter, R. S., & Avadhani, N. G. (1991) *Arch. Biochem. Biophys. 288*, 97−106.

Carter, R. S., & Avadhani, N. G. (1994) *J. Biol. Chem. 269*, 4381−4387.

Carter, R. S., Bhat, N. K., Basu, A., & Avadhani, N. G. (1992) *J. Biol. Chem. 267*, 23418−23426.

Chen, J.-Y., & Waterman, M. R. (1992) *Biochemistry 31*, 2400−2407.

Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156−159.

Chung, B.-C., Matteson, K. J., & Miller, W. L. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4243−4247.

Corbin, V., & Maniatis, T. (1989) *Genes Dev. 3*, 2191−2220.

Deluca, H. F., & Schnoes, H. K. (1976) *Annu. Rev. Biochem. 45*, 631−666.

DeNoto, F. M., Moore, D. D., & Goodman, H. M. (1981) *Nucleic Acids Res. 9*, 3719−3730.

Dignam, J. D., Lebovitz, R. M., & Roeder, R. G. (1983) *Nucleic Acids Res. 11*, 1475−1489.

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

Guo, Y.-D., Strugnell, S., Back, D. W., & Jones, G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8668−8672.

Hahn, S. L., Hahn, M., & Hayward, W. S. (1989) *Mol. Cell. Biol. 9*, 837−843.

Hashimoto, T., Morohashi, K.-I., & Omura, T. (1989) *J. Biochem. 105*, 676−679.

Higashi, Y., Yoshioka, H., Yamane, M., Gotoh, O., & Fujii, K. Y. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2841−2845.

Imagawa, M., Chiu, R., & Karin, M. (1987) *Cell 51*, 251−260.

Jackson, S. P., Mac Donald, J. J., Lees-Miller, S., & Tjian, R. (1990) *Cell 63*, 155−165.

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

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

Karin, M., Haslinger, A., Holtgreve, H., Richards, R. I., Krauter, P., Westphal, H. M., & Beato, M. (1984) *Nature 308*, 513−519.

Kim, K.-S., Kuboto, S., Kuriyama, M., Fujiyama, J., Bjorkhem, I., Eggersten, G., & Seyama, Y. (1994) *J. Lipid Res. 35*, 1031−1039.

Leitersdorf, E., Reshef, A., Meiner, V., Levitzki, R., Schwartz, S. P., Dann, E. J., Berkman, N., Cali, J. J., Klapholz, L., & Berginer, V. M. (1993) *J. Clin. Invest. 91*, 2488−2496.

Leonard, J. P., & Miller, W. L. (1987) *DNA 6*, 439−448.

Macejak, D. G., & Sarnow, P. (1991) *Nature 353*, 90−94.

Marcu, K. B., Bossone, S. A., & Patel, A. J. (1992) *Annu. Rev. Biochem. 61*, 809−860.

Miwa, T., Boxer, L. M., & Kedes, L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6702−6706.

Morohashi, K.-I., Sogawa, K., Omura, T., & Fujii, K. Y. (1987) *J. Biochem. 101*, 879−887.

Mornet, E., Dupont, J., Vitek, A., & White, P. C. (1989) *J. Biol. Chem. 264*, 20961−20967.

Nebert, D. W., & Nelson, D. R. (1991) *Methods Enzymol. 206*, 3−11.

Ohyama, Y., Noshiro, M., Eggersten, G., Gotoh, O., Kato, Y., Bjorkhem, I., & Okuda, K. (1993) *Biochemistry 32*, 76−82.

Okret, S., Poellinger, L., Dong, Y., & Gustafsson, J. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5899−5903.

Oonk, R. B., Parker, K. L., Gibson, J. L., & Richards, J. S. (1990) *J. Biol. Chem. 265*, 22392−22401.

O'Shea-Greenfield, A., & Smale, S. T. (1992) *J. Biol. Chem. 267*, 1391−1402.

Parker, K. L., Schimmer, B. P., Chaplin, D. D., & Seidman, J. G. (1986) *J. Biol. Chem. 261*, 15353−15355.

Pederson, J. I., Bjorkhem, I., & Gustafsson, J. (1979) *J. Biol. Chem. 254*, 6464−6469.

Pelletier, J., & Sonenberg, N. (1988) *Nature 334*, 320−325.

Princen, H. M., Meijer, P., Wolthers, B. G., Vonk, R. J., & Kuipers, F. (1991) *Biochem. J. 275*, 501−505.
Reiss, A. B., Martin, K. O., Javitt, N. B., Martin, D. W., Grossi, E. A., & Galloway, A. C. (1994) *J. Lipid Res. 35*, 1026−1030.
Renkawitz, R., Schutz, G., Von der Ahe, D., & Beato, M. (1984) *Cell 37*, 503−510.
Roesler, W. J., Vandenbark, G. R., & Hanson, R. W. (1988) *J. Biol. Chem. 263*, 9063−9066.
Saarem, K., & Pederson, J. I. (1987) *Biochem. J. 247*, 73−78.
Sakaki, T., Akiyoshi, S. M., Yabusaki, Y., & Ohkawa, H. (1992) *J. Biol. Chem. 267*, 16497−16502.
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.
Seto, E., Shi, Y., & Shenk, T. (1991) *Nature 345*, 241−245.
Shayiq, R. M., & Avadhani, N. G. (1992) *J. Biol. Chem. 267*, 2421−2428.
Singh, H., Sen, R., Baltimore, D., & Sharp, P. A. (1986) *Nature 319*, 154−158.
Smale, S. T., & Baltimore, D. (1989) *Cell 57*, 103−113.
Stravitz, R. T., Vlahcevic, Z. R., Russell, T. L., & Hylemon, P. B. (1994) *Hepatology 20*, 262A.
Su, P., Rennert, H., Shayiq, R. M., Yamamoto, R., Zheng, Y.-M., Addya, S., Strauss, J. F., & Avadhani, N. G. (1990) *DNA Cell Biol. 9*, 657−665.
Sucharov, C., Basu, A., Carter, R. C., & Avadhani, N. G. (1995) *Gene Expression* (in press).
Von der Ahe, D., Janich, S., Scheidereit, C., Renkawitz, R., Schutz, G., & Beato, M. (1985) *Nature 313*, 706−709.
Walsh, K., & Schimmel, P. (1987) *J. Biol. Chem. 262*, 9429−9432.
Watt, R., Nishikura, K., Sorrentino, J., ar-Rushdi, A., Croce, C. M., & Rovera, G. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6307−6311.
Wikvall, K. (1984) *J. Biol. Chem. 259*, 3800−3804.
Wikvall, K. (1993) in *Handbook of Experimental Pharmacology* (Schenkman, J. B. & Greim, H., Eds.) Vol. 105, pp 705−718, Springer-Verlag, New York.
Zawel, L., & Reinberg, D. (1992) *Curr. Opin. Cell Biol. 4*, 488−495.

BI951383U