Vol. 194, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1993 Pages 194-201

GENE STRUCTURE AND UPSTREAM REGULATORY REGIONS OF HUMAN *CYP2C9* AND *CYP2C18*[1]

Sonia M.F. de Morais, Helmut Schweikl[2]*, Joyce Blaisdell and Joyce A. Goldstein*[3]

National Institutes of Environmental Health Sciences, P.O. Box 12233,
Research Triangle Park, NC 27709

Received May 20, 1993

There is a genetic polymorphism in humans in the metabolism of *S*-mephenytoin which has been suggested to be mediated by either CYP2C18 or CYP2C9. We have isolated genomic clones for *CYP2C9* and *CYP2C18* from the liver of an individual phenotyped *in vitro* as an extensive metabolizer of *S*-mephenytoin. Analysis of the genes reveals nine coding exons spanning approximately 55 kb. The intron-exon organization was similar to that of other members of the CYP2C subfamily. Analysis of 2200 bp of 5′ upstream sequence for *CYP2C9* and 1300 bp 5′ upstream sequence for *CYP2C18* reveals canonical TATA boxes situated 57 bp upstream from the first codon, multiple consensus sequences for glucocorticoid regulatory elements, and identification of a 15 base sequence with high homology to a 5′-flanking sequence responsible for barbiturate-inducible expression of $P450_{BM-3}$ in *Bacillus megaterium*. The upstream region for *CYP2C9* was highly homologous (75%) to that of human *CYP2C8* through most of the 2200 bp sequenced, but the upstream region of *CYP2C18* was similar to *CYP2C8* and *CYP2C9* for only the first 200 bases. The availability of the sequences of the upstream regions and intron-exon junctions of *CYP2C9* and *CYP2C18* will allow future analysis of these genes in humans which differ in their ability to metabolize *S*-mephenytoin and other drugs.

The human cytochrome P4502C family metabolizes drugs such as mephenytoin, tolbutamide and barbiturates (1,2). There is a genetic polymorphism in the 4′-hydroxylation of the *S*-enantiomer of mephenytoin in which approximately 3% of Caucasians and 20% of Orientals are deficient in the ability to metabolize this drug (3,4). This reaction is thought to

[1]The sequences reported in this paper have been submitted to GenEMBL/Genbank, under accession numbers: L16869 through L16876 for *CYP2C18*, and L16877 through L16883 for *CYP2C9*.

[2]Present Address: Universität Regensburg, Franz-Josef-Strauss-Allee, W-8400 Regensburg, Germany.

[3]To whom correspondence should be addressed. Fax number: 919-541-3647.

0006-291X/93 $4.00

be metabolized by a member of the CYP2C subfamily. Both recombinant CYP2C9 (5) and CYP2C18 (6) have been reported to 4′-hydroxylate *S*-mephenytoin. However, other conflicting reports suggesting little or no metabolism of *S*-mephenytoin by recombinant CYP2C9 when expressed in a variety of cDNA expression systems (6-8) and the intermediate turnover numbers of recombinant CYP2C18 for 4′-hydroxylation of *S*-mephenytoin compared to human liver microsomes make the identity of the *CYP2C* gene responsible for this polymorphism still uncertain.

To date, the only information available on the gene structure of the human CYP2C subfamily is the sequence of the 5′ regulatory region for *CYP2C8* (9) and a partial gene structure for *CYP2C18* covering 932 bases of 5′-flanking region and the intron/exon junctions of exons 1-4 (10). As a first step in the analysis of the molecular basis for the polymorphism in *S*-mephenytoin metabolism, we have isolated and sequenced genomic clones for the *CYP2C9* and *2C18* genes from an individual with high hepatic *S*-mephenytoin 4′-hydroxylase activity.

MATERIALS AND METHODS

Preparation and Screening of Genomic Library: A genomic library was constructed in bacteriophage EMBL3 (Stratagene) from liver DNA of an individual phenotyped *in vitro* as an extensive metabolizer of *S*-mephenytoin (0.8 nmol/min/nmol P450). The library was screened by *in situ* plaque hybridization (11) using *CYP2C18* or *CYP2C9* cDNA probes (12,13). Additional probes used included a 5′-fragment of *CYP2C18* (from -199 to +127 bp of exon 1) and a PCR fragment containing exons 5 and 6 of *CYP2C18*. Fragments of positive clones derived from restriction analysis were subcloned into pBluescript SK$^-$ vectors (Stratagene) and double-stranded DNA templates were isolated and sequenced using exon-specific synthetic oligonucleotides as primers, and Sequenase kits (Stratagene) or an automated sequencer (Applied Biosystems, Inc.). Additional primers were synthesized to completely sequence the 5′-flanking regions in both directions.

RESULTS AND DISCUSSION

A map of the genomic clones and the gene organization for *CYP2C9* and *CYP2C18* is shown in Figure 1. We obtained a complete set of overlapping clones for *CYP2C18*. Exon 6 of *CYP2C9* was missing, but the nucleotide sequences of the remaining exons were identical to those of the Leu^{359} variant of *CYP2C9* cDNA (12). Both genes are much larger than their rabbit and rat counterparts (14-17), with 2C18 spanning approximately 55 kb. The intron-exon junctions and polyadenylation sites for *CYP2C9* and *2C18* are shown in Table 1. The intron-exon structure appears to be conserved in the *CYP2* family, comparing the human genes to the rabbit (14) and rat *CYP2* genes (15-17). All *CYP2* genes have nine exons, and there is a very high homology in the amino acids at the exon-intron junctions. A variant cDNA clone for *CYP2C9* has been reported which has a six-base deletion at positions 820-826 (18). The missing

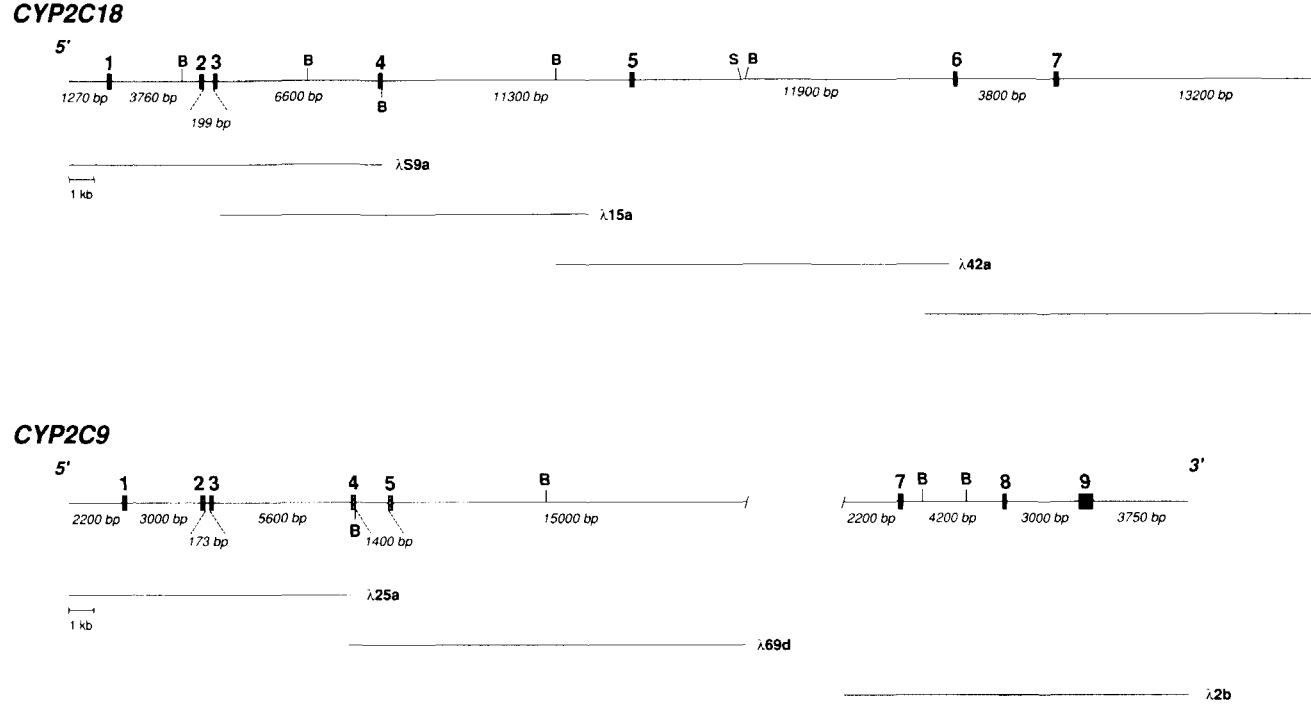

Fig. 1. Representation of the structure of *CYP2C9* and *CYP2C18* genes, and alignment of λ-EMBL 3 clones. Clones were isolated and characterized by restriction analysis, partial sequencing, and by hybridization studies to determine overlapping regions. DNA fragments were subcloned into pBluescript vectors for sequence analysis. Exons were located by restriction analysis and by PCR analysis using exon specific primers and primers located in the vector cloning sites. Overlapping areas are indicated by the positions of the clones on the diagram.

TABLE 1. LOCATION OF INTRON-EXON BOUNDARIES IN *CYP2C9* AND *CYP2C18*

Intron	Gene	Donor	Intron size	Acceptor
		N56		L57
1	2C9	CTTAACCAATgtaagtatgc	3000	actttcctagCTCTCAAAGG
	2C18	CTTAACCAATgtaagtatgc	3750	ctatgtttagTTCTCAAAAG
				F57
		G111		G111
2	2C9	AGAGGATTTGgtaggtgtgc	173	ttcctgttagGAATTGTTTT
	2C18	AAAGGACTTGgtaaatgtgc	191	gttctgctagGAATCCTTTT
		A161		A161
3	2C9	AAAACCAAGGgtgggtgacc	5600	aattatttagCCTCACCCTG
	2C18	AAAACCAATGgtgggtgact	6600	aatcctttagCCTCACCCTG
		Q214		I215
4	2C9	CTGGATCCAGgtaaggccaa	1400	tttctcttagATCTGCAATA
	2C18	ATGGATCCAGgtgagatcaa	11300	aatctttaagGTCTGCAATA
				V215
		K273		E274
5	2C9	AATGGAGAAGgtaaaatgta	>17200	GAAAAGCACA
	2C18	AATGGAACAGgtaaaatgtt	11900	aatttttagGAAAAGCACA
		Q273		
		A321		A321
6	2C9	GAGGTCACAG	>2200	gtcttatcagCTAAAGTCCA
	2C18	GAGGTCACAGgtatgatgat	3800	gtcttatcagCTAAAGTCCA
		K383		G384
7	2C9	CATTCCCAAGgtaagtttgt	4200	ttgtttctagGGCACAACCA
	2C18	CATCCCCAAGgtaagcttgt	13200	ctattttcagGGCATGACCA
		G431		G431
8	2C9	TTCTCAGCAGgtaatataaa	3000	ctattttcagGAAAACGGAT
	2C18	TTCTCAGCAGgtaatagata	1750	ttattttcagGAAAACGGAT
Poly(A) Site	2C9	TTCTAAATAAAAAGCATTATTATTTGCTGA▾gtcagtttat		
	2C18(1976[a])	TGATTAATAAATGACAATTCAGAGCC▾ATTTATTCTCT		
	2C18(2197[a])	GTCTAAATATATGCTTTCATATGGC•TAATCATgtgttaatga		

Exons are shown in upper case and introns in lower case. Codons at junctions and polyadenylation sites are underlined and the amino acids indicated above the sequences. ▾Location of polyA tail or •longest cDNA clone. [a]Refers to positions in the *CYP2C18* (6b allele) cDNA.

bases are located at the intron 5-exon 6 junction and contain a possible alternative splice acceptor site (GAAAAG), explaining the probable mechanism by which this aberrant 2C9 cDNA was generated. The *CYP2C18* gene isolated in the present study was identical to the 6b allelic variant (Met^{385}) (12,13). An additional clone was isolated which contained exons 8 and 9 of the 29c allelic variant of *CYP2C18* (not shown). Both clones contained the additional 420 bases found in the 3′-noncoding region of the cDNA for the 6b variant but not found in the cDNA for the shorter 29c variant. Exon 9 of both *CYP2C18* clones contained two alternative polyadenylation signals, making this exon longer ($\geq$925 bases) in *CYP2C18* than in *CYP2C9* (545 bases).

The nucleotide sequences for the 5′-flanking regions of *CYP2C9* and *CYP2C18* are shown in Figure 2. *CYP2C9* contains several consensus sequences for glucocorticoid response elements

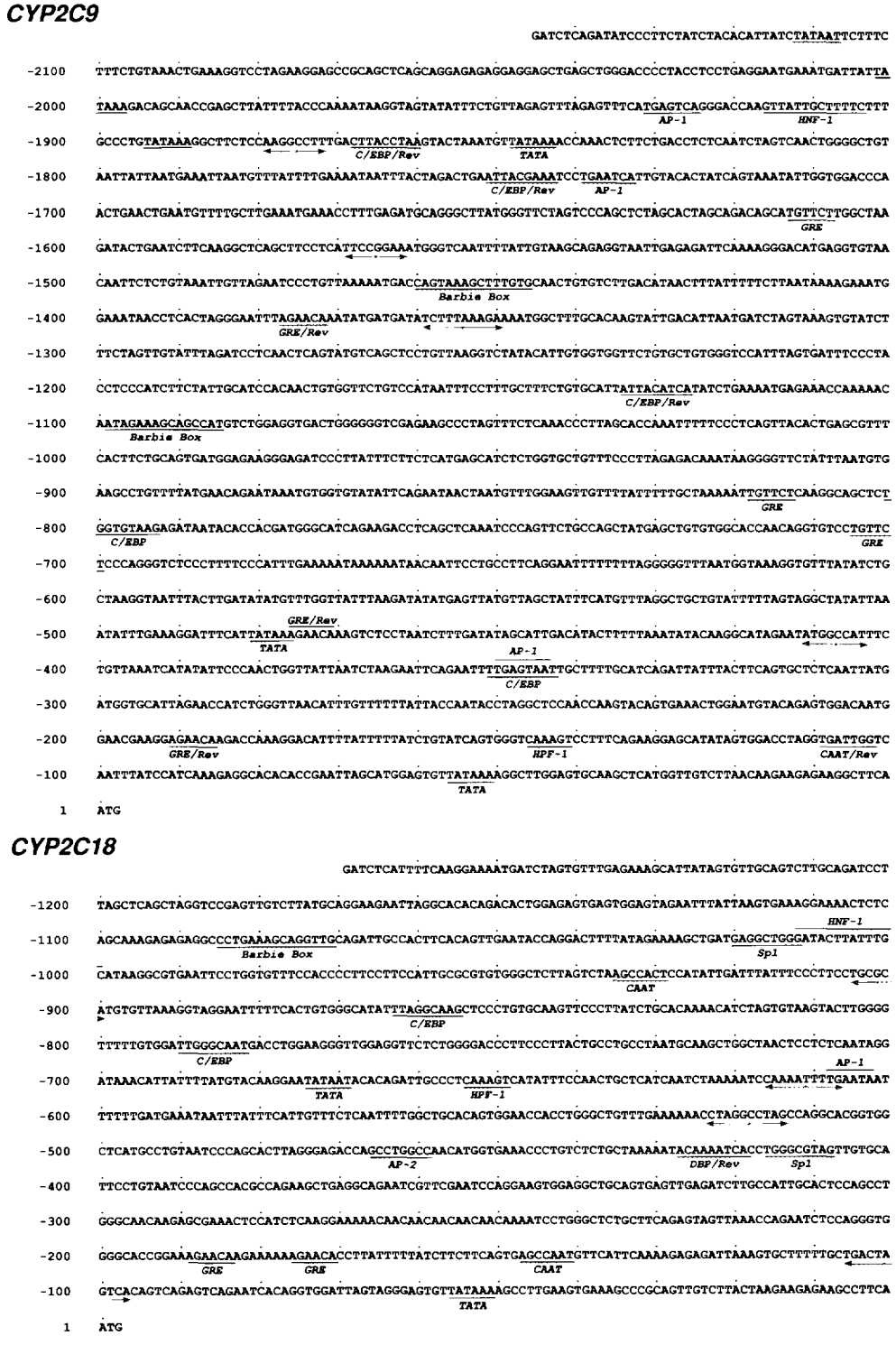

Fig. 2. Nucleotide sequence of the 5'-flanking region of the *CYP2C9* and *CYP2C18* genes. Nucleotides are arbitrarily numbered in negative numbers from the ATG coding for the initiation codon. The sequences of putative binding sites for transcription factors are underlined, as are presumed TATA and CAAT boxes. Arrows indicate palindromic sequences.

(GREs) within the 2200 bp of 5′-upstream sequences analyzed. GRE consensus sequences also occur in the upstream region of *CYP2C18* and human *CYP2C8* (9,10). A 15 base sequence was identified in the upstream regions of both *CYP2C18* and *CYP2C9* which had high homology (7-8 of 15 bases including the prerequisite AAAG at bases 5 to 8) to a 5′-flanking sequence responsible for barbiturate-inducible expression of $P450_{BM-3}$ in *Bacillus megaterium* (19). A similar sequence has been identified in the 5′-flanking region of a number of phenobarbital inducible genes including *CYP2B1* and *CYP2B2* (20). The presence of these sequences in the upstream regions of human *CYP2C18* and *CYP2C9* suggests the possibility that these cytochromes may be inducible by glucocorticoids and phenobarbital.

A comparison of the regulatory regions of the three human P4502C genes using the COMPARE/DOTPLOT program (Genetics Computer Group, Madison, WI) shows that the 5′-flanking region of *CYP2C9* is highly similar (75%) to that of *CYP2C8* through most of the available 2200 bp sequence, with only one stretch of *CYP2C8* between -684 and -739 showing no homology to *CYP2C9* (Fig. 3). We reexamined the sequence of 2C9 in this area and found no comparable sequence in this area. However, comparison of the sequence of *CYP2C8* with itself in the reverse orientation indicated that this area was almost an exact reverse repeat of bases -54 to -112, suggesting that this sequence may have been accidentally inserted twice in *CYP2C8* during the sequence compilation. In contrast, the upstream region of *CYP2C18* showed homology with *CYP2C9* and *CYP2C8* for only the first 200 bases. The position of one TATA box (57 bp upstream from the ATG start codon) is conserved in all three human 2C genes, but

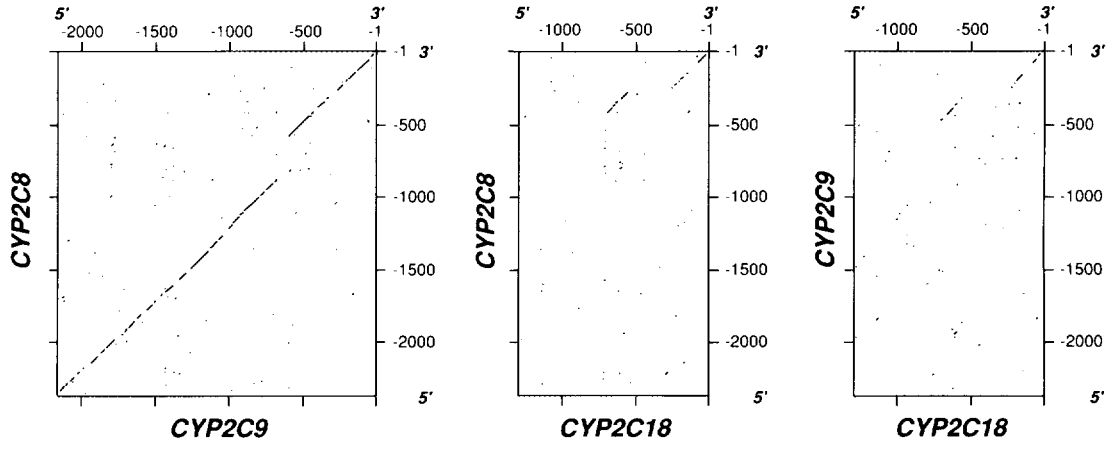

Fig. 3. Comparison of the 5′-flanking regions of human *CYP2C* genes. Sequences were compared by dot matrix analysis (COMPARE/DOTPLOT, GCG) in which the best match is plotted at each position for a sequence window of 21 nucleotides. The number -1 denotes the first nucleotide prior to the ATG coding for the initiation codon.

CAAT boxes are located at different positions: at -147 bp for *CYP2C18*, at -110 bp for *CYP2C9* (in reverse orientation) and at -113, -118 and -123 bp for *CYP2C8*. The transcription start site for *CYP2C8* (9) has been reported to be located 23 bp upstream of the ATG, consistent with those of the rabbit and rat CYP2C genes. Interestingly, *CYP2C18* contains an additional TATA box, located at -674 bp and an additional CAAT box at -935 bp. This may indicate two start sites for *CYP2C18* transcription, and could explain the isolation of a *CYP2C18* cDNA clone containing 200 bases of 5′-noncoding leader sequence (12,13).

Promotor regions of *CYP2C18* and *CYP2C9* contained a motif which agreed with the consensus sequence for the HepG2-specific factor 1 (HPF-1)(21) at positions -650 and -150 bp, respectively. This sequence, which is similar to the consensus sequence for HNF-4, has been found to be essential for the expression of the rabbit CYP2C genes in HepG2 cells but not for expression in COS-1 cells. It appears to be conserved in the *CYP2A*, *CYP2C*, and *CYP2D* genes, and it may be important for the liver-specific expression of these genes. Sites for other liver-enriched transcription factors common to the two genes include: C/EBP, HNF-1, and AP-1 (Figure 2). *CYP2C18* also contains sites for AP-2, Sp1, and DBP, that are not found in *CYP2C9* (Figure 2). DBP is a liver transcription factor which has been shown to play a role in the developmental regulation of expression of the rat CYP2C6 (22). Additional studies will define the role of these regulatory elements in the expression of the human *CYP2C* genes. The determination of the sequence of the 5′-upstream regions and the exon-intron junctions of *CYP2C9* and *CYP2C18* will allow analysis of these regions in humans differing in their ability to express *S*-mephenytoin 4′-hydroxylase.

Acknowledgments: We are grateful to Dr. Judy Raucy, University of New Mexico, for the sample of human liver phenotyped in her laboratory as having high liver microsomal *S*-mephenytoin hydroxylase activity *in vitro*, and to Dr. Jonathan Wiest, NIEHS, for assistance with the Applied Biosystems automatic sequencer.

REFERENCES

1. Wilkinson, G. R., Guengerich, F.P., and Branch, R. A. (1989) *Pharm. Ther.* 43, 53-76.
2. Wrighton, S. A., and Stevens, J. C. (1992) *Crit. Rev. Toxicol.* 22, 1-21.
3. Kupfer, A. and Preisig, R. (1984) *Eur. J. Clin. Pharmacol.* 26, 753-759.
4. Wedlund, P.J., Aslanian, W.S., McAllister, C.B., Wilkinson, G.R. and Branch, R.A. (1984) *Clin. Pharmacol. Ther.* 36, 773-780.
5. Yasumori, T., Yamazoe, Y. and Kato, R. (1991) *J. Biochem.* 109, 711-717.
6. Faletto, M. B., Romkes-Sparks, M., Raucy, J. L., Lasker, J. M., and Goldstein, J. A. (1992) Expression and metabolism studies of the human P4502C subfamily in yeast. *FASEB J.* 6, 322 (Abstr.)

7. Relling, M.V., Aoyama, T., Gonzalez, F.J. and Meyer, U.A. (1990) *J. Pharmacol. Exp. Ther.* 252, 442-447.
8. Srivastava, P. K., Yun, C.H., Beaune, P., Ged, C. and Guenguerich, F.P. (1991) *Mol. Pharmacol.* 40, 69-79.
9. Ged, C. and Beaune, P. (1991) *Biochem. Biophys. Acta* 1088, 433-435.
10. Ged, C. and Beaune, P. (1992) *Pharmacogenetics* 2, 109-115.
11. Sambrook, J., Fritsch, E.J. and Maniatis, T. (1989) Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
12. Romkes, M., Faletto M.B., Blaisdell, J.A., Raucy, J.L. and Goldstein, J.A. (1991) *Biochemistry* 30, 3247-3255.
13. Romkes, M., Faletto, M. B., Blaisdell, J. A., Raucy, J. L, and Goldstein, J. A. (1992) *Biochemistry* 32, 1390.
14. Zhao, J., Chan, G., Govind, S., Bell, P. and Kemper, B. (1990) *DNA Cell Biol.* 9, 37-48.
15. Morishima, N., Yoshioka, H., Higashi, Y., Sogawa, K. and Fujii-Kuriyama, Y. (1987) *Biochemistry* 26, 8279-8285.
16. Eguchi, H., Westin, S., Strom, A., Gustafsson, J.-A. and Zaphiropoulos, P.G. (1991) *Biochemistry* 30, 10844-10849.
17. Zaphiropoulos, P.G., Westin, S., Strom, A., Mode, A. and Gustafsson, J.-A. (1990) *DNA Cell Biol.* 9, 49-56.
18. Ohgiya, S., Komori, M., Ohi, H., Shiramatsu, K., Shinriki, N. and Kamataki, T. (1992) *Biochem. Internat.* 27, 1073-1081.
19. Shaw, G-C. and Fulco, A.J. (1992) *J. Biol. Chem.* 267, 5515-5526.
20. He, J.-S. and Fulco, A. (1991) *J. Biol. Chem.* 266, 7864-7869.
21. Venepally, P. Chen, D. and Kemper, B. (1992) *J. Biol. Chem.* 267, 17333-17338.
22. Yano, M., Falvey, E., and Gonzalez, F. J. (1992) *Mol. Cell Biol.*12, 2847-2854.