Vol. 166, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1990 Pages 365-372

STRUCTURAL CHARACTERIZATION OF THE HUMAN ESTROGEN SYNTHETASE (AROMATASE) GENE

Nobuhiro Harada, Kazuyo Yamada, Kuniaki Saito, Naomi Kibe, Sumiko Dohmae, and Yasuyuki Takagi

Division of Molecular Genetics, Institute for Comprehensive Medical Science, School of Medicine, Fujita Health University, Toyoake, Aichi 470-11, Japan

Received December 4, 1989

The estrogen synthetase (aromatase, cytochrome P-450AROM) gene has been isolated from human genomic libraries and characterized. The restriction map of 43 positive clones obtained indicated that this enzyme is present as a single copy gene. The aromatase gene is unexpectedly large compared with other forms of the cytochrome P-450 superfamily, spanning at least 70 kilobases. The gene consists of 10 exons and its 5'-untranslated region is divided into 2 exons by an intron of more than 35 kilobases long. This organization of the first exon in the aromatase gene is unique in the cytochrome P-450 superfamily. All the exon-intron junctional sequences conform to the canonical GT/AG rule. The sequences of a TATA box and a CAAT box are present 27 and 83 base pairs upstream from the transcriptional initiation site. Within 3 kilobases upstream from the initiation site, there are no typical consensus sequences of responsive elements for glucocorticoid and c-AMP, which regulate aromatase expression.

Aromatase (estrogen synthetase; EC 1.14.14.1), so-called cytochrome P-450AROM, catalyzes the aromatization of androgens to estrogens, which is a rate-limiting step in their biosynthesis. Aromatase is known to be involved in estrogen production in the ovary and placenta of gonadal tissues. But it has also been found in extra-gonadal tissues (1,2) such as brain, adipose tissue, muscle and skin. These findings raise questions about its functions and regulation. As aromatase is an unstable membrane-bound protein, it has been difficult to purify it to homogeneity and to prepare a specific antibody. Therefore, extensive studies on these questions have been done by the conventional method of measuring aromatase activity in various culture cells. Many kinds of factors, including cAMP, phorbor ester (TPA), dexamethasone, insulin, glucose, androgen, epidermal growth factor, prolactin, and insulin-like growth factor-I, have been found to influence the expression of aromatase, although some of these factors have different effects in different types of cells(3-5).

Recently, aromatase was purified to homogeneity from human placenta (6-8) and cDNA encoding the complete molecule was isolated(9,10). Characterization of the purified aromatase and its cDNA revealed that the enzyme differed in several properties from other members of the cytochrome P-450 superfamily and so should be classified in a new family of cytochromes P-450(9,10). Here we report the isolation of the human aromatase gene and show that its exon-intron organization differs from those of other members of the cytochrome P-450

0006-291X/90 $1.50

superfamily. We also report the isolation of the 5'-promotor region of the aromatase gene and its nucleotide sequencing as a first step in determining the responsive elements for various factors that regulate aromatase expression.

MATERIALS AND METHODS

Isolation and Characterization of Aromatase Genomic Clones. A λ Charon 4A library, obtained by partial *Eco*RI digestion of human placental genomic DNA, was generously provided by Drs. C.Setoyama and K.Shimada (Kumamoto Univ.). The λ EMBL 3 library, constructed by partial digestion of human genomic DNA from peripheral blood cells with *Sau*3AI by Drs.S.Tomatsu and Y.Sakaki (Kyushu Univ.), was obtained from J.C.R.B. (Japanese Cancer Research Resources Bank). These libraries were screened with human aromatase cDNA(9) as a probe. Recombinant λ phage DNAs containing an aromatase gene were prepared and mapped by the method of Rackwitz et al.(11). All *Eco*RI restriction fragments of aromatase genomic clones were subcloned and further characterized by detailed restriction endonuclease mapping and Southern blot analysis. DNA sequencing of specific restriction fragments was performed by the dideoxynucleotide chain-termination method with *Taq* DNA polymerase. The nucleotide sequence of the 5'-promotor region of the aromatase gene was determined by exonuclease III deletion, as described by Henikoff(12).

Primer-Extension Analysis. Total RNA was prepared from human placenta at term by the method of Chirgwin et al(13). A 22-base oligonucleotide primer (5'-CCGAGCACAGGAC CTTCCGTCC-3'), complementary to the sequence expected to be near the 5'-end of human aromatase m-RNA, was synthesized and labeled at the 5'-end. The primer (1×10^5 cpm) and 70 μg of total RNA were hybridized and the primer extension reaction was carried out as described by Maniatis(14). The products were analyzed by 5% polyacrylamide/6M urea gel electrophoresis. As size markers, *Hae*III-digested fragments of ϕχ-174 were also 5'-end-labeled and applied to the same gel. For determination of the exact initiation site for transcription, ^{32}P-labeled primer was annealed with the single stranded DNA containing the aromatase 5'-promoter region, and the nucleotide sequence from the primer site was determined.

S1 Nuclease Protection Analysis. A 439 base pair (bp) *Nco*I-*Ava*I restriction fragment containing the 5'-end of human aromatase cDNA was prepared from λAROM-G and labeled at the 5'-end. The *Nco*I-*Ava*I fragment was hybridized at 65°C for 3 hrs with nylon membrane-immobilized, single stranded DNA which was inserted by a DNA fragment complementary to human aromatase mRNA. The membrane was washed extensively and single stranded DNA of the *Nco*I-*Ava*I fragment was recovered with water from the membrane by heating at 100°C for 5 min. This single stranded DNA contains the sequence of a 22 base oligonucleotide primer used for primer-extension analysis at the 5'-end. The ^{32}P-labeled, 439 bp single stranded DNA (6×10^4cpm) and human placental total RNA (70 μg) were hybridized. S1 nuclease digestion was performed as described by Maniatis(14).

RESULTS AND DISCUSSION

Screening of Human Aromatase Gene. Genomic clones of the human aromatase gene from a human placental genomic library were screened in Charon 4A with human aromatase cDNA as a probe. Of 3×10^6 clones screened, 27 positive clones were isolated and characterized. All genomic DNA fragments were roughly classified with *Eco*RI restriction fragments and further mapped with several restriction endonucleases by the method of Rackwitz et al.(11). As shown in Fig. 1, nine different clones were obtained by these analyses and the map of the human aromatase gene was constructed by connecting genomic DNA fragments with overlapping portions, which were determined by Southern blot analysis of their restriction endonuclease fragments. As there was one missing link between the λAROM-F and

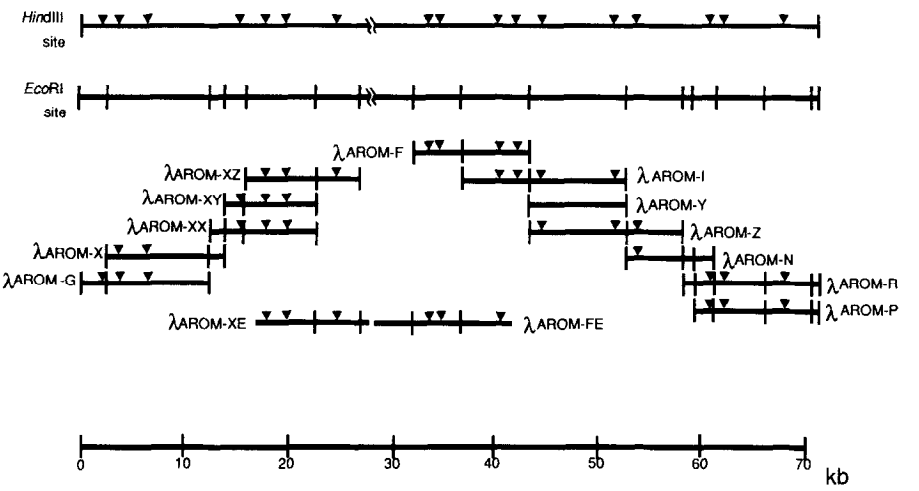

Fig. 1. Restriction endonuclease map of genomic clones of human aromatase. Nine clones (λ AROM-G, -X, -F, -I, -Y, -Z, -N, -R, and -P) were isolated from a λ Charon 4A library, which was constructed with a partial *Eco*RI-digest of human genomic DNA. Three clones (λ AROM-XX, -XY, and -XZ) were isolated by successive gene walking from the same library and two clones (λ AROM-XE, and -FE) were isolated from another library of λ EMBL3 obtained with a partial *Sau*3AI-digest of genomic DNA. The restriction sites of representive restriction enzymes, *Hin*dIII and *Eco*RI, are shown on composite maps and on DNA fragments of the clones. The interrupted line on maps indicates a missing intron region.

λAROM-X clones on the map, gene walking was performed with λ Charon 4A and λEMBL 3 genomic libraries using DNA fragments of the 5'-end and 3'-end portions, respectively, as probes. Of 3×10^6 clones screened by successive gene walking, 16 positive clones were isolated and five different clones were connected with overlapping portions on the map. However, the isolated clones still did not cover the unidentified region. Another genomic library in λ Charon 4A, constructed from a partial digest of genomic DNA with *Hae*III and *Alu*I was also screened, but still this missing region could not be isolated. Similar difficulties in isolation of genomic clones containing specific intractable sequences have been reported and discussed(15,16). Interestingly, a missimg link in the first intron of a human P-450scc gene has also been difficult to isolate(17). All 43 genomic clones isolated from the libraries could be located on the same map, indicating that the aromatase gene exists in the human haploid genome as a single copy. Southern blot analysis of restriction fragments of human total genomic DNA (data not shown) also supported this conclusion, although it is not consistent with the report by Chen et al.(18).

Structural Organization of Human Aromatase Gene. For determination of the locations of individual exons on the map of the aromatase gene, all the genomic DNAs isolated were subjected to Southern blot analysis with human aromatase cDNA after digestion with various restriction endonucleases, and DNA fragments containing exons and their adjacent regions were sequenced. The human aromatase gene consists of 10 exons and 9 introns and is more than 70 kb as shown in Fig. 2. The exon-intron organization of the aromatase gene differs from those of other genes of the cytochrome P-450 superfamily. The 3' half portion (1579 bp) of aromatase mRNA (cDNA) was encoded by exon 10, whereas exons 1-9 contained

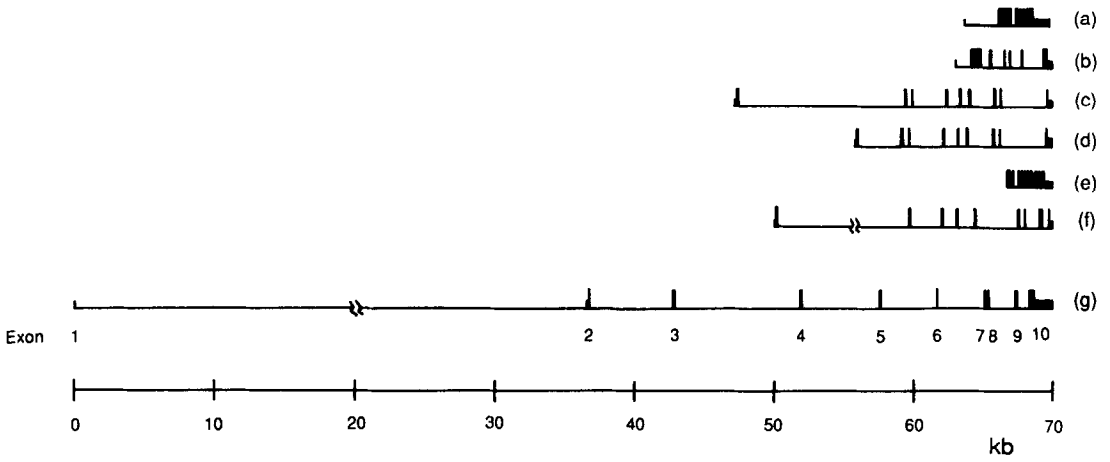

Fig. 2. Exon-intron structures of the aromatase gene and six other genes of the cytochrome P-450 superfamily. P-450c (a), P-450d (b), P-450b (c), P-450e (d), P-450c-21 (e), P-450scc (f), and aromatase (g) are classified as cytochrome P-450 (sub) family IA, IA, IIB, IIB, XXI, XXII, and XIX, respectively. Solid columns indicate exons, coding regions being shown as higher columns and noncoding regions as low ones. The length of the gene is shown in kb below.

small pieces (103 bp-242 bp) of the 5' half portion. Exons 2-10 are located within approximately 35 kb of the 3' end of the aromatase gene. By contrast, exon 1, encoding only a 5' flanking region of aromatase mRNA, is separated by a large intron of more than 35 kb from exon 2. Figure 2 shows the exon-intron organizations of several other genes of the cytochrome P-450 superfamily besides the aromatase gene. Although the aromatase gene has the same number of exons and introns as P-450C21, it is larger than any other genes of the cytochrome P-450 superfamily reported so far and its exon-intron organization, especially the location of exon 1 far from the other exons, is very different from those of other genes of this family. Methylcholanthrene-inducible cytochrome P-450(19,20), P-450c and P-450d, and phenobarbital-inducible cytochrome P-450(21), P-450b and P-450e, are classified in the P-450 subfamilies IA and IIB, respectively, according to the nomenclature recommended by the Committee for Standardized Nomenclature of P-450 Genes(22). Similarly, aromatase, P-450C-21(23), and P-450SCC(17) are classified in the P-450 families XIX, XXI, and XXII, respectively. The exon-intron organization of genes is generally conserved within each of the subfamilies, but differs from one another among the P-450 families. The sequence of all introns of the human aromatase gene begins with GTA at the 5' end and terminates with C(T)AG at the 3' end, indicating that the aromatase gene follows the canonical GT/AG rule at all splice junctions (Table 1).

Identification of Transcriptional Initiation Site. The transcriptional initiation site(s) was identified by primer-extension and S1 nuclease protection analyses. The human aromatase cDNA isolated in this laboratory contained a 5'-noncoding region of 124 bp. So, a 22-base sequence upstream of the *AvaI* site in this 5'-noncoding region was used as primer and a primer-extension experiment was carried out with this primer and human placental RNA as a

Table 1. Exon-Intron Junctions of Human Aromatase Gene

Exon	3'-Exon junction	Intron			5'-Exon junction	Exon	Exon size bp
					GGAGTT	1	103
1	CGTCGC	GTATGT	>30kb --------	CCACAG	GACTCT	2	183
2	TACCAG	GTAAGT	6kb --------	TCCTAG	GTCCTG	3	151
3	CAGCAA	GTAAGT	9kb --------	TTTCAG	GTCCTC	4	155
4	TGAAAG	GTAAGC	5.8kb --------	CTGCAG	CTCTGT	5	177
5	TGGACG	GTACTG	4kb --------	CTACAG	AAAGTG	6	115
6	GTCTGT	GTAAGT	3.5kb --------	TTCCAG	CAAGGA	7	115
7	GCAGAG	GTACTG	0.4kb --------	GTTCAG	AAACGT	8	163
8	TTATTG	GTAAGA	2kb --------	CCACAG	GTGAGA	9	242
9	AAGAAT	GTAAGA	1kb --------	TCATAG	GTTCCT	10	1579
10	ATAGGT						

Sequences at splice junctions of the aromatase gene were determined. The exon sites and the estimated intron sizes are indicated.

template. The *Eco*RI-*Hin*dIII fragment (1.7 kb) containing the region upstream of the primer site was isolated from λAROM-G and this fragment was also sequenced using the same primer. The primer-extension reaction produced a major DNA fragment of 86 bp and two minor fragments of 85 and 87 bp (Fig. 3), indicating that the major transcriptional initiation site is located 86 bp upstream from the primer site. This indicates initiation at a guanine on the aromatase gene (Figs. 3 and 4). This result was confirmed by S1 nuclease protection assay. The *Eco*RI-*Hin*dIII fragment from λAROM-G was further digested with *Nco*I and *Ava*I. The human placental RNA mainly protected a 86 bp length of single stranded *Nco*I-*Ava*I fragment from S1 nuclease digestion(Fig. 3). This length of protected DNA strongly supported the conclusion that transcription of the aromatase gene was initiated at a guanine 86 bp upstream of the primer site.

Analysis of 5' Promoter Region of Aromatase Gene. The DNA sequence on the 5'region upstream of exon 1 in λAROM-G was determined. The consensus sequences of a "TATA box" and a "CAAT box" are located close to the transcriptional initiation site as shown in Fig. 4, the sequence ATAAAA is located at positions -27 to -22 as a "TATA box"and the sequence CCAAT is located at positions -83 to -79 as a "CAAT box". Expression of the aromatase gene is known to be regulated by various factors, such as glucocorticoid, cAMP, phorbor ester, and androgen. However, there are no typical consensus sequences

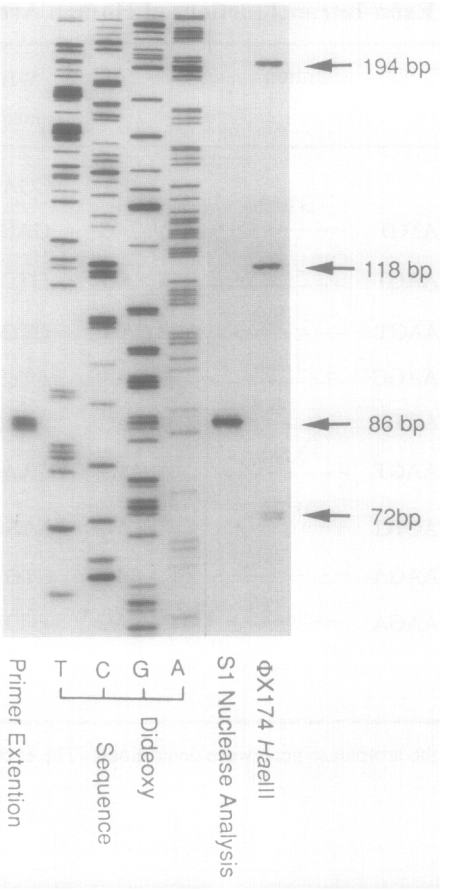

Fig. 3. Mapping of the transcriptional initiation site on the human aromatase gene by primer extension and S1 nuclease protection analyses. The primer extension and S1 nuclease protection analyses were carried out as described in "Materials and Methods". *Hae*III fragments of $\phi\chi$–174 were used as size markers. For determination of the exact transcriptional initiation site, a DNA sequence ladder was produced by dideoxynucleotide sequencing reactions with the same 22 base primer as for primer extension assay.

corresponding to responsive elements for any of these factors in the 5' promoter region within about 3 kb upstream from the transcriptional initiation site. Expression of P-450SCC is known to be regulated by adrenocorticotrophic hormone (ACTH) through production of cAMP, but there is also no consensus sequence of a cAMP responsive element in the 5' promoter region of the P-450SCC gene(17). These findings imply that unknown *cis*- or *trans*-elements are involved in the regulation of gene expression of aromatase by various factors. Possible TPA (phorbor ester) responsive elements, are located at -604 to -598 (TGACTGA in Fig. 4) and at -1094 to -1088 (TGAGTCA, data not shown). The consensus sequence of a *cis*-element widely found in cytochrome P-450 genes is also located at -160 to -153 (CCCTGGAGGAA GGAAT).

The secondary structure of the 5' promoter region of the aromatase gene was examined. Potential sites for stem-loop structure formation were found to be concentrated within 1200 bp upstream of the transcriptional initiation site (data not shown). There were two kinds of repeated sequence, 11 bp of TGAATAAATGA and 10 bp of TTT/CTTGTCCA, within the region 500 bp upstream of the initiation site. The significances of these DNA sequences are

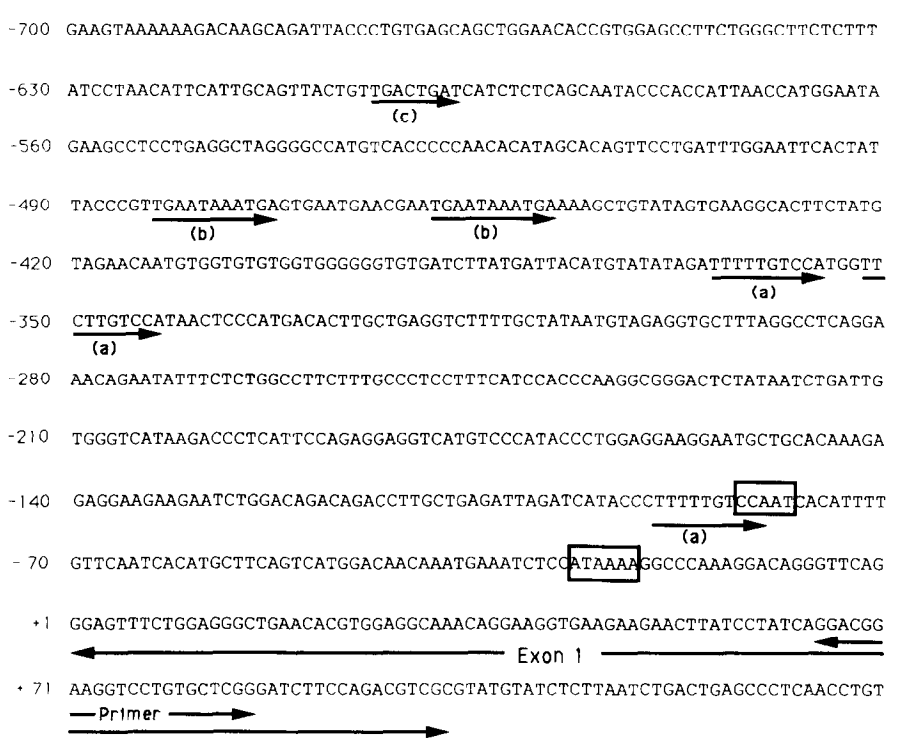

Fig. 4. DNA sequence of the 5'-promoter region of the human aromatase gene. The transcriptional initiation site is designated as +1. The "TATA box" (-27 to -22) and the "CAAT box" (-83 to -79) are boxed. One of two putative motives of the TPA (phorbor ester) responsive element is underlined (c). Repeated sequences of 10 bp (a) and 11bp(b) are underlined.

unknown. We are now trying to identify responsive elements other than those known already and demonstrate the bindings of regulatory proteins to the DNA to clarify the regulation mechanism of tissue specific and hormone (factor) specific gene expression of aromatase.

ACKNOWLEDGMENT

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

REFERENCES

1. Naftolin, F., Ryan, K. J., Davis, I. J., Reddy, V. V., Flores, F., Petro, Z., Kuhn, M., White, R. J., Takosha, T. & Wolin, L. (1975) Recent Prog. Horm. Res. *31*, 295-319
2. Simpson,E.R., Cleland,W.H. & Mendelson,C.R. (1983) J.Steroid Biochem. *19*, 707-713
3. Garzo, V. G. & Dorrington, J. H. (1984) Am. J. Obstet. Gynecol. *148*, 657- 662
4. Mendelson, C.R. & Simpson, E. R. (1987) Mol. Cell. Endocrinol. *52*, 169-176
5. Nestler, J. E. & Williams, T. (1987) Endocrinology *121*, 1845-1852
6. Nakajin, S., Shinoda, M. & Hall, P.F. (1986) Biochem. Biophys. Res. Commun. *134*, 704-710
7. Tan, J. & Muto, N. (1986) Eur. J. Biochem. *156*, 243-250
8. Harada, N. (1988) J. Biochem. *103*, 106-113
9. Harada, N. (1988) Biochem. Biophys. Res. Commun. *156*, 725-732

10. JoCorbin, C., Graham-Lorence, S., McPhaul, M., Mason, J. I., Mendelson, C. R. & Simpson, E. R. (1988) Proc. Natl. Acad. Sci. USA *85*, 8948-8952
11. Rackwitz, H.-R., Zehetner, G., Frischauf, A. -M. & Lehrach, H. (1984) Gene *30*, 195-200
12. Henikoff, S. (1984) Gene *28*, 351-359
13. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) Biochemistry *18*, 5294-5299
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning : A Laboratory Manual (Cold Spring Habor Labs., Cold Spring Habor, NY).
15. H. Morioka, Magnuson, M. A., Mitsuhashi, T., Dong, M. -K. H., Rall, J. E. & Nikodem, V. M. (1989) Proc. Natl. Acad. Sci. USA *86*, 4912-4916
16. Leach, D.R.F. & Stahl, F. W. (1983) Nature *305*, 448-451
17. Morohashi, K., Sogawa, K., Omura, T.& Fujii-Kuriyama, Y. (1987) J.Biochem. *101*, 879-887
18. Chen, S., Besman, M. J., Sparkes, R. S., Zollman, S.,Klisak, I., Mohandas,T., Hall, P. F. (1988) DNA *7*, 27-38
19. Sogawa, K., Gotoh, O., Kawajiri, K., Harada, T. & Fujii-Kuriyama, Y. (1985) J. Biol. Chem. *260*, 5026-5032
20. Hines, R. N., Levy, J. B., Shen, M. -L.,Renli, A. M. & Bresnik, E. (1985) Arch. Biochem. Biophys. *237*, 465-476
21. Suwa, Y., Mizukami, Y., Sogawa, K. & Fujii-Kuriyama, Y. (1985) J.Biol.Chem. *260*, 7980-7984
22. Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P.,Gunsalus, I.C.,Johnson, E.F.,Kemper, B.,Levin, W.,Phillips, I.R., Sato, R. & Waterman, M.R. (1987) DNA 6, 1-11
23. White, P. C., New, M. I. & Dupont, B. (1986) Proc. Natl. Acad. Sci. USA *83*, 5111-5115