Structural Analysis of the Gene Encoding Human Aromatase Cytochrome P-450, the Enzyme Responsible for Estrogen Biosynthesis*

(Gary D. Means†, Mala S. Mahendroo‡, C. Jo Corbin, J. Michael Mathis§, Frances E. Powell‖, Carole R. Mendelson, and Evan R. Simpson¶
From the Departments of Obstetrics and Gynecology and Biochemistry, Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9051

The structural gene encoding aromatase cytochrome P-450 (P-450AROM) was isolated from human genomic DNA. The gene spans at least 52 kilobases and is composed of 10 exons, the first of which is untranslated. Analysis of the transcription initiation site of human P-450AROM mRNA reveals the differential use of 1 of 3 consecutive G residues at the cap site. DNA sequence analysis indicates that the gene has a putative TATA (ATAGAAA) sequence at -23 base pairs (bp) and putative CAAT binding sequences beginning at -41, -67, and -83 bp. The 5'-flanking region contains sequences similar to consensus sequences of cis-acting elements defined as regulators of aromatase gene expression. These putative sequences include a cAMP regulatory element at -211 bp, an AP1 (protein kinase C) site at -54 bp, and glucocorticoid regulatory elements at -352 bp and within the first intron at +346 bp. There appears to be only one gene encoding P-450AROM in the human genome. Two major species of human P-450AROM mRNA (3.4 and 2.9 kilobases) are derived from the use of two polyadenylation signals.

The biosynthesis of estrogens from androgens is catalyzed by an enzyme complex termed aromatase, which is present in the endoplasmic reticulum of cells in which it is expressed and is composed of a specific form of cytochrome P-450, aromatase cytochrome P-450 (P450AROM, P450XIX; CYP19).

* This work was supported in part by United States Public Health Service Grants AG 08174 and AM 31208. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J05105.

†Supported in part by United States Public Health Service Training Grant 5-T32 HD 07190.
‡Present address: Eukaryotic Regulatory Biology Program, UCSD School of Medicine, M-013, La Jolla, CA 92039.
§Present address: Dept. of Physiol., 460 Medical Science Research Bldg., University of North Carolina, Chapel Hill, NC 27514.
‖To whom correspondence should be addressed: Dept. of Obstetrics and Gynecology and Biochemistry, Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-9051.
¶To whom correspondence should be addressed: Depts. of Obstetrics and Gynecology and Biochemistry, Cecil H. and Ida Green Center for Reproductive Biology Sciences, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-9051.

The biosynthesis of estrogens from androgens is catalyzed by an enzyme complex termed aromatase, which is present in the endoplasmic reticulum of cells in which it is expressed and is composed of a specific form of cytochrome P-450, aromatase cytochrome P-450 (P450AROM, P450XIX; CYP19). This enzyme utilizes molecular oxygen and reducing equivalents provided by a ubiquitous NADPH-cytochrome P-450 reductase to catalyze a series of three sequential hydroxylations that results in loss of the angular methyl group at carbon 19 and phenolization of the A ring of the steroid (6-9).

In addition to ovarian granulosa cells, aromatase is expressed in several tissue sites including the placenta, Sertoli (10) and Leydig cells (11, 12) in the male, and adipose tissue (13) and several sites in the brain including the hypothalamus, hippocampus, and amygdala (14, 15) in both sexes. Aromatase expression in adipose tissue has been implicated in the development of endometrial cancer (16) as well as estrogen-dependent breast cancer (17). In adipose tissue, the principal estrogen formed is estrone, whereas in granulosa cells, the primary estrogen formed is estradiol-17β and in the placenta, estriol. Expression of a full-length cDNA encoding human P-450AROM suggests that a single enzyme is capable of catalyzing the aromatization of all three classes of androgen substrates, namely androstenedione, testosterone, and 16α-hydroxylated androgens (18). It is likely therefore that the formation of different estrogen products in the various sites of expression reflects the presentation of these different substrates to the same enzyme, rather than the presence of different forms of aromatase in each tissue site.

In previous studies (19, 20), we have shown that aromatase activity of human adipose and ovarian granulosa cells is subject to complex and multifactorial regulation which is correlated with changes in the levels of mRNA encoding this protein. In order to examine, in greater detail, the regulation of aromatase expression in the human, we have isolated and characterized the gene encoding P-450AROM. Utilizing the full-length cDNA (18) and a primer-extended cDNA insert as hybridization probes, four genomic clones encoding the entire P-450AROM structural gene have been isolated from two different human genomic libraries. The gene spans at least 52 kb and contains an untranslated first exon as well as two polyadenylation signals. Characterization of the regulatory sequences of this gene should pave the way to understanding the multifactorial regulation and tissue-specific expression of human P-450AROM.

P450XXIB or CYP21B gene (5)); P-45017α, 17α-hydroxylase cytochrome P-450 (the product of the P450XVII or CYP17 gene (5)); P-450C17, cholesterol side-chain cleavage cytochrome P-450 (the product of the P450XIA or CYP11A gene (5)); P-450C21, 21-hydroxylase cytochrome P-450 (the product of the
**Experimental Procedures and Results**

**Discussion**

Structure of Human Aromatase Cytochrome P-450 Gene—In this study is presented an analysis of the gene encoding human P-450 arom, the enzyme responsible for the conversion of androgens to estrogens. The gene is similar to that of other cytochrome P-450 species in that the structural gene comprises 10 exons (most cytochrome P-450 genes contain between 8 and 10 exons); the heme-binding region and the entire 3' untranslated region are encoded by the last exon (29). However, the gene is much larger than those of other steroidogenic cytochrome P-450 species and may in fact be the largest cytochrome P-450 gene analyzed at this time. The entire gene spans at least 52 kb; and since there are two regions where the clones do not overlap, the actual size is unknown. By comparison, the genes for the other two microsomal steroidogenic cytochrome P-450 species, namely P-450c17 (30) and P-450, (31), span 3.7 and 6.5 kb, respectively. Of the mitochondrial steroidogenic cytochrome P-450 species, the gene for bovine 11β-hydroxylase cytochrome P-450 (P-450,β) is 8 kb (32) long, whereas that for human cholesterol side-chain cleavage cytochrome P-450 (P-450scC) is at least 20 kb long (33), and this also has an intron in which the clones do not overlap.

Analysis of the intron/exon boundaries of human P-450 arom reveals rather poor correlation with other microsomal steroidogenic cytochrome P-450 species. Poor intron/exon boundary alignment between the various genes is a characteristic feature of the cytochrome P-450 superfamily, leading to speculation that the ancestral common gene had many more exons than the modern counterparts (29). Number of Aromatase Cytochrome P-450 Genes Present in Human Genomes—The issue of whether there is more than one aromatase enzyme in the human is an important one for several reasons. In the first place, it has been suggested that different aromatase enzymes exist in the placenta, ovary, and adipose tissue since the major estrogen produced in each of these tissues is different, namely estriol in the placenta, estradiol in the ovary, and estrone in the adipose tissue. Equally, however, this could be due to the presentation to the same enzyme of different substrates, namely 16α-hydroxylated androgens in the case of the placenta, testosterone in the case of the ovary, and androstenedione in the case of adipose tissue. The issue is also of importance clinically. At present, there is much interest in the development of more effective inhibitors of aromatase for use clinically in the management of patients with breast cancer (35). However, the only source of aromatase available for the testing of such inhibitors is that derived from human placenta. Moreover, there is evidence to suggest that the estrogen which may be of consequence in the development of breast tumors is that produced in breast adipose tissue surrounding the developing tumor. If the enzyme responsible for such estrogen in adipose tissue were different from that in the placenta, then clearly, inhibitors developed against the placental enzyme might be less efficacious toward the enzyme present in adipose tissue.

Our previous work (18) on the expression of the cDNA encoding P-450 arom together with previous work (3, 4) on the purified enzyme suggest that a single enzyme is capable of metabolizing all three categories of C19-steroid substrate and that it is not necessary to postulate the presence of different enzymes in the different tissues which synthesize estrogens. This work on the characterization of the gene encoding P-450 arom is consistent with this view. In all of our restriction mapping and Southern analyses, we have obtained no evidence to suggest that there is more than one P-450 arom gene within the human genome. Our conclusions in this context therefore differ from those of Chen et al. (36), who suggested, on the basis of Southern mapping, that there exist two human P-450 arom genes.

Characterization of 5'-Untranslated Exon—Although DNA sequence upstream from the bases encoding the start of translation contained putative CAAT and TATA boxes, primer extension failed to reveal any start of transcription associated with these. Moreover, an oligonucleotide prepared against a region commencing 39 bp downstream from this putative TATA box failed to hybridize in Northern analysis of poly(A) RNA from human placenta (Fig. 3, lane 7). We conclude therefore that an intron is present in the DNA 5' of the exon encoding the start site of translation and that the intron/exon boundary occurs at the point where the genomic sequence and those of the cDNAs all become identical. A sequence similar to the splice junction consensus sequence is also present at this site (Ref. 24 and Fig. 8). The region of the gene that we have called exon I therefore fulfills the criteria of a 5' untranslated exon, although the clone that contains it does not overlap with the clones containing the exons encoding the translated gene product. This conclusion is also supported by the following criteria. First, the genomic clone contains sequence identical to that of the 5'-end of a cDNA (clone cDNA-2) isolated from a primer-extended human placental library and is identical to the human P-450 arom cDNA sequence published by Harada (28). Second, an oligonucleotide prepared complementary to this region hybridizes to poly(A) RNA from human placenta to give bands identical to those observed using the full-length cDNA as a hybridization probe (Fig. 3, lane 5). Third, primer extension products of human placental poly(A) RNA initiated using an oligonucleotide complementary to exon II also contain DNA corresponding to exon I (Fig. 7). These findings indicate that exon I is present on the same mRNA species as is the first translated exon.

Putative Regulatory Sequences within Human Aromatase Cytochrome P-450 Gene—5'-Flanking sequence comprising 918 bp upstream of exon I has been sequenced. Analysis of 5'-flanking region of exon I indicates the presence of a putative TATA (ATAAA) box. Downstream from this (23 bp) is the site of transcription initiation as revealed by primer extension. Sequences similar to putative CAAT binding elements are present at -41, -67, and -83 bp. P-450 arom expression in vivo and in vitro is under the control of a number of hormonal regulators including factors that act via cAMP-dependent protein kinases (e.g. gonadotropins) and protein kinase C as well as growth factors such as epidermal growth factor, basic fibroblast growth factor, transforming growth factor β, tumor necrosis factor (37), and glucocorticoids, which act via specific receptors that bind responsive cis-acting elements (38). The 5'-flanking sequence of P-450 arom was therefore evaluated for sequences that have been shown to confer responsiveness to the above factors or their second messengers. A putative cAMP-responsive element was found at -211 bp. This sequence (TGTGCTCA) is identical to the published consensus sequence for cAMP-responsive elements (TGACGTCA) (39) at six of eight of the positions. A putative AP1-like cis-regulatory sequence (TCAGTCA) that would confer protein kinase C responsive-
ness was found at -54 bp. A single deviation from the consensus sequence occurs: the C at position 2 is a T/G in the poly(A+)
RNA by primer extension using oligonucleotides.

The Aromatase Cytochrome P-450 Gene-Although there appears to be only one P-450AROM gene within the human genome, Northern analysis of RNA from tissues in which the gene is expressed reveals the presence of two hybridizable mRNA species, one of 3.4 kb and one of 2.9 kb (Fig. 3). We conclude therefore that although this sequence is present within the 5'-region of a human P-450AROM cDNA upstream from a splice junction, it does not appear to be a component of the normal pattern of expression of the P-
450AROM gene. Its presence within the cDNA therefore may result either from an unlikely artifact of the original cloning or from an extremely rare polymorphism, perhaps involving an extra piece of DNA inserted 3' of exon 1.

Nature of Transcribed RNA Message-Although there appears to be only one P-450AROM gene within the human genome, Northern analysis of RNA from tissues in which the gene is expressed reveals the presence of two hybridizable mRNA species, one of 3.4 kb and one of 2.9 kb (Fig. 3). We believe that this is due to the use of alternative polyadenylation signals for the following reasons. In the first place, there are two putative polyadenylation signals within the 3'-untranslated region of the human P-450AROM gene, and we have isolated cDNAs that contain polyadenylated tails corresponding to the use of each of these polyadenylation signals (18, 42). In all other aspects, the sequences of the cDNAs are identical, including the sequence of the coding region. Second, when an oligonucleotide complementary to the area of the 5'-untranslated region between these polyadenylation signals is used as a hybridization probe in Northern analysis of human placental poly(A+) RNA, only of these mRNA species hybridizes, namely the 3.4-kb band (Fig. 3, lane 3). This proves convincingly that the RNA species of 2.9 kb does not contain the 3'-untranslated region between the polyadenylation signals. Third, an estimate of the size of the mRNA that would be expected based on the length of these cDNAs is in good agreement since sizes of the corresponding cDNAs are 3.0 and 2.7 kb, and most polyadenylated tails are 200-400 bases long.

Conclusions—Understanding of the regulation of the bio-
synthesis of estrogens in the human is of great importance for a number of reasons. First, the ratio of androgen to estrogen is responsible for a number of important physiological parameters such as the expression of the appropriate sexually dimorphic phenotype as well as reproductive capacity. The expression of the gene within the hypothalamus and other regions of the brain is probably required for the imprinting of sex-related behavior as well as the pattern of gonadotropin secretion by the hypothalamic-pituitary axis. The expression of aromatase in the preimplantation blastocyst may well provide a signal for implantation, and this could account for the observation that, at present, no mutations resulting in a loss of aromatase activity have been characterized. Last, a number of common human cancers, in particular endometrial and breast cancer, are estrogen-dependent. Therefore, for these reasons, understanding the differential regulation of aromatase in the various tissues in which it is expressed in the human as well as understanding the developmental and tissue-specific expression of the enzyme are of great importance and interest. Characterization of the gene encoding human P-450AROM should pave the way for studies addressing the issue of the different modes of regulation of this gene.

Acknowledgments—We gratefully acknowledge the assistance of Dr. Michael McPhaul (Department of Internal Medicine, University of Texas Southwestern) in the preparation of the primer-extended cDNA library and the expert advice and help of Dr. Sandra Graham-Lorence (Green Center, University of Texas Southwestern) as well as the expert editorial assistance of Sandra Finley.

Addendum—Subsequent to submission of this manuscript, we became aware of the publication of Toda et al. (43). These authors used a similar strategy to ours to characterize the two P-450AROM mRNA species. They also stated, without showing substantiating data, that they had isolated a second cDNA differing from their published insert in terms of the sequence of the 5'-untranslated region.

REFERENCES
1. Mendelson, C. R., Wright, E. E., Evans, C. T., Porter, J. C., and Simpson, E. R. (1982) Proc. Natl. Acad. Sci. U. S. A. 88, 6950-6953
2. Nakajin, S., Shimoda, M., and Hall, P. F. (1981) Biochem. Biophys. Res. Commun. 104, 794-799
3. Kattwinkel, J., Reiss, M., and Vickers, L. E. (1987) J. Biol. Chem. 262, 4413-4420
4. Osawa, Y., Yoshida, N., Frankowski, M., and Kitawaki, J. (1987) Steroids 52, 1-26
5. Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guggerich, P. F., Guzelian, I. C., Johnson, E. F., Kemper, B., Levits, W. Philps, J. J., Sato, R., and Waterman, M. R. (1987) DNA (N. Y.) 6, 1-11
6. Thompson, E. A., Jr., and Sierter, P. K. (1974) J. Biol. Chem. 249, 5927-5938
7. Akhtar, M., Calder, M. R., Corina, D. L., and Wright, J. N. (1982) Biochem. J. 201, 569-580
8. Casag. E., Wicha, I., Arnesahcham, T., Nelson, P., and Spitteller, G. (1984) J. Am. Chem. Soc. 106, 7282-7285
9. Cole, P. A., and Robinson, C. H. (1984) J. Am. Chem. Soc. 110, 1284-1285
10. Fritz, B. H., Griswold, M. D., Louis, B. F., and Darrington, J. H. (1976) Mol. Cell. Endocrinol. 6, 298-294
11. Valadas, L. R., and Arnaiz, A. (1979) Endocrinology 109, 431-436
12. Tsai-Morris, C. H., Aquilano, D. R., and Dufau, M. L. (1984) Ann. N. Y. Acad. Sci. 438, 666-668
13. Grodin, J. M., Siteri, P. K., and MacDonald, P. C. (1973) J. Clin. Endocrinol. Metab. 36, 207-214
14. Nakajin, S., and Kato, J., Davis, I. J., Reddy, V. V., Flores, F., Zerto, K., Kohn, M., White, R. J., Takosha, Y., and Woltan, L. (1975) Recent Prog. Horm. Res. 31, 295-315
15. Rossel, C. E., Horton, L. E., and Resko, J. A. (1985) Endocrinology 117, 2471-2476
16. Hemsell, D. L., Grodin, J., Brenner, P. F., Sitter, P. K., and MacDonald, P. C. (1974) J. Clin. Endocrinol. Metab. 34, 474-479
17. O'Neill, J. S., Elton, R. A., and Miller, W. R. (1986) Br. Med. J. 296, 741-743
18. Corbin, C. J., Graham-Lorence, S., McPhaul, M., Mason, J. C., and Simpson, E. R. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 6948-6953
19. Evans, C. T., Corbin, C. J., Saunders, C. T., Merrill, J. C., Simpson, E. R., and Mendelson, C. R. (1987) J. Biol. Chem. 262, 6914-6920
20. Steinke, M., Mendelson, C. R., and Simpson, E. R. (1988) Mol. Cell. Endocrinol. 59, 93-99
21. Marazk, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
22. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517
23. Sanger, T., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
24. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472
25. Salk, R. K., Geltman, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Lisk, M. J., Erlich, H. A., and Squier, C. K. (1983) Nature 304, 519-523
26. Chinhyun, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299
27. Doebere, H., Apostol, B., Lin, R., Newman, A., Brody, E., and Abeles, J. (1984) Cell 39, 611-621
28. Hardes, N. (1986) Biochem. Biophys. Res. Commun. 158, 725-730
29. Deeb, D. W., and Gonzalez, F. G. (1987) Annu. Rev. Biochem. 56, 945-993
30. White, P. C., New, M. I., and DuPont, R. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 511-515
31. Picado-Leonard, J., and Miller, W. L. (1987) DNA (N.Y.) 6, 439-448
32. Hashimoto, T., Morohoshi, K., and Omura, T. (1987) J. Biochem. (Tokyo) 101, 879-887
33. Osawa, Y., and Higashiyama, Y. (1984) in Microsomes, Drug Oxidations and Chemical Carcinogenesis (Cson, M. J., Conney, A. H., Estropook, R. W., Gelbain, H. V., Gillette, J. R., and O'Brien, P. J., eds) Vol. I, 215-238, Academic Press
34. Brodie, A. M., Coombes, R. C., and Dowsett, M. (1987) J. Steroid Biochem. 27, 899-903
35. Cheng, J. S., Redman, M. J., Sparks, R., Zollman, S., Kliska, I., Mohandes, T., Hall, P. F., and Shively, J. E. (1988) DNA (N. Y.) 7, 27-38
36. Simpson, E. R., Johnson, C., Hollah, A. J., and Mendelson, R. E. (1989) Endocrinology 130, 136-148
37. Simpson, E. R., Ackerman, G. E., Smith, M. E., and Mendelson, C. R. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 1570-1574
38. Rosser, W., Vandenberg, G. R., and Hanson, R. W. (1988) J. Biol. Chem. 263, 9603-9609
39. Jones, N. C., Rigby, P. W. J., and Ziff, E. B. (1988) Genes Dev. 2, 267-281
40. Schmitz, U., Klock, G., and Schütz, G. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 7671-7675
41. Simpson, E. R., and Mendelson, C. R. (1987) Mol. Cell. Endocrinol. 52, 169-177
42. Toda, K., Terashima, M., Mitsuuchi, Y., Yamakawa, Y., Yokoyama, Y., Nojima, S., Ushiro, H., Maeda, T., Yamamoto, Y., Sagarra, Y., and Shizuka, Y. (1988) PNAS Lett. 247, 371-376

SUPPLEMENTAL MATERIAL TO
Structural Analysis of the Gene Encoding Human Aromatase Cyp21a1 in P. pastoris
The Enzyme Responsible for Estrogen Biosynthesis

by
Gopi D. Rama, Mani S. Murthy, Carolina M. Coria, John Michael Mathis, Frances A. Pochett, Geno R. Michaudon, and Donna R. Sampson

EXPERIMENTAL PROCEDURES

Cloning and Nucleotide Sequencing Analysis of the CYP450 Gene and 5'- flanking Region

The human aromatase gene was PCR amplified from genomic DNA isolated from human breast cancer cell line MCF-7 (ATCC CCL-23) and is the 3.5 kb fragment containing the full-length human aromatase CYP21 gene (CYP21, Exon 8) cloned into pBluescript KS- vector (Stratagene, La Jolla, CA) using PCR amplification kit

RESULTS

Polymorphism of the Sequence Containing the Structural Gene Encoding Human P450

Two different human genomic libraries were screened with a full-length DNA probe specific for human P450. The DNA probe (DNA1) excorates the entire structural P450 sequence and is generated from a polyclonal human spleen cDNA probe. Further analyses indicated that the DNA is the same size as the DNA probe (DNA1) and is consistent with the DNA probe (DNA1) being the same size as the DNA probe (DNA1)

NIH Partial Digest of Specific DNA - Specific DNA fragments were simultaneously amplified and labeled using a modification of the worker's polymerase elongation. The reactions were performed as described [37] using a DNA polymerase elongation kit from New England Biolabs. The amplified DNA was then separated by electrophoresis on 6% polyacrylamide gels containing 0.5% agarose, and the gel was stained with ethidium bromide and visualized under UV light

Results Analysis - Total human DNA (labeled at excision point) was isolated with minor modifications according to the methods of Chiang, et al. (12) using the DNA probe (DNA1) labeling and detection

Figure 1: Schematic of the gene encoding human aromatase CYP21A1. Restriction sites for EcoRI (1) and Hind III (2) are shown, as are the transcription initiation site (TIS) and TSS (transcription start site). The +1 position is the site of transcription initiation. The 3'-UTR is shown as a solid line, while the 5'-UTR is shown as a dashed line. The arrow indicates the direction of transcription. The positions of the EcoRI and Hind III restriction sites are also indicated.
**Analysis of the Aromatase Cytochrome P-450 Gene**

![Diagram](image)

**Fig. 3** Northern analysis of human placental poly(A) RNA. Poly(A) RNA was prepared as described in Experimental Procedures and 15 µg was applied to each lane. The RNA on each lane was hybridized using the following probes: lane 1 - full-length cDNA (cDNA-1); lane 2 - oligonucleotide complementary to the codon region of exon X; lane 3 - oligonucleotide complementary to the portion of the 3'-untranslated region between the polyadenylation signal; lane 4 - oligonucleotide complementary to the 3'-end of intron IX; lane 5 - oligonucleotide complementary to exon I; lane 6 - oligonucleotide complementary to the corresponding region of cDNA-2; lane 7 - oligonucleotide complementary to a region between the polyadenyl cap site and the 5'-end of the identified cDNA sequence of exon II. The site designations were based on use of an RNA denatured weight ladder.

**Fig. 4** Diagram of 5'-end of original cDNA clone (cDNA-1), two primer-extended cDNA clones (cDNA-2 and cDNA-3) and the corresponding genomic sequence. Exon/intron junctions and spliced sequences are shown. The cDNA-1 and cDNA-2 and cDNA-3 are different. The start of cDNA-1 is downstream from this junction.

---

**Fig. 5** Sequence of exons I of the gene encoding cytochrome P-450 arom. Exonic sequence is indicated by bold caps, spurious sequence and intronic sequence by lower case letters. The exon/intron boundary between exon I and intron I is indicated by underline. The putative transcriptional start site is indicated by +1, and upstream from this, enclosed by boxes, are a putative TATA box and three putative CAAT boxes. Other sequences of potential regulatory significance are indicated in bold with underlining and their possible roles so designated. GRE - glucocorticoid regulatory element; CRE - cAMP regulatory element.
Analysis of the Aromatase Cytochrome P-450 Gene

Fig. 6 Primer extension analysis to determine the transcription start site with exon 1. The analysis was conducted using human placental poly A RNA as described in Experimental Procedures. Lanes A, C, G, T denote the appropriate lanes of the sequencing ladder, while P denotes the lane containing the primer-extended product, indicated by the arrow, aligned with the nucleotide of the top of a triplet of Gs. The position of the putative TATA box (TATAAAA) on the sequencing ladder is also indicated.

Fig. 7 Amplification of exon 1 from human placental poly A RNA and P-450aromDNA. First strand synthesis was initiated using placental poly A RNA as a template and MMLV-RT with oligonucleotide primer comprising either to the 3′-end of exon 1 or to the 3′-end of exon II. Selective primers and substrate was accomplished by PCR using a primer corresponding to the 5′-end of exon 1 (5′-3′) in all incubations. The products were separated on 2% agarose gels, visualized using ethidium bromide staining (left panel), transferred to Zetablot, and probed with a labeled 3′ bp oligonucleotide hybridizing to exon 1 (80-1660 bp right panel).

Fig. 8 Splice junction boundaries of the exon-intron and intron-exon junctions of the human aromatase P-450arom gene. Exon sequence is indicated by bold caps, intron by lower case letters. The consensus sequences of each of the boundaries according to Mount (26) are also included for comparison.

Fig. 9 Comparison of the positions of introns in the coding regions of human P-450arom, P-450a, and P-450f. Arrows are denoted by their single letter designation. Solid arrowheads between amino acids indicate amino acids belonging to codons, hatched arrowheads pointing to amino acids would indicate that the amino acid between the first and second bases of the corresponding codon, open arrowheads pointing to an amino acid indicate that the amino acid between the second and third bases of the corresponding codon. The data for human P-450arom, P-450a, and P-450f were derived from ref. 11.