Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 6582–6586, September 1989
Biochemistry

Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus

(21-hydroxylase deficiency/congenital adrenal hyperplasia/overlapping genes/HLA class III genes/*CYP21* locus)

YVES MOREL, JAMES BRISTOW, STEPHEN E. GITELMAN, AND WALTER L. MILLER*

Department of Pediatrics and The Metabolic Research Unit, University of California, San Francisco, CA 94143-0434

Communicated by Seymour Lieberman, June 12, 1989 (received for review March 30, 1989)

ABSTRACT **The gene encoding human adrenal steroid 21-hydroxylase (P450c21) and its highly similar pseudogene are duplicated in tandem with the two genes encoding the fourth component of human serum hemolytic complement (C4). This 60-kilobase gene complex, which lies within the major histocompatibility complex on the short arm of human chromosome 6, has been studied in considerable detail because genetic disorders in steroid 21-hydroxylation and in C4 are common. We have cloned a cDNA encoded by a previously unidentified gene in this region. This gene lies on the strand of DNA opposite from the strand containing the P450c21 and C4 genes, and it overlaps the last exon of P450c21. The newly identified gene encodes mRNAs of 3.5 and 1.8 kilobases that are expressed in the adrenal and in a Leydig cell tumor but are not expressed in nonsteroidogenic tissues. The sequence of the longest cDNA (2.7 kilobases) shows no similarity to known sequences available in two computerized data bases. The 5′ end of this sequence is characterized by three repeats, each encoding about 100 amino acids flanked by potential sites for proteolytic cleavage. Although numerous studies have shown that gene deletions causing congenital adrenal hyperplasia occur in this region, none of these gene deletions extends into this newly identified gene, suggesting that it encodes an essential function.**

The class III region of the human leukocyte antigen (HLA) locus on chromosome 6p21 contains tandemly duplicated genes for the fourth component of serum complement (C4) and for the P450 enzyme (P450c21) steroid 21-hydroxylase [steroid,hydrogen-donor:oxygen oxidoreductase (21-hydroxylating), EC 1.4.99.10] (reviewed in refs. 1 and 2). This locus has attracted considerable attention, in part because 1 in 7000 persons has deficient 21-hydroxylase activity causing congenital adrenal hyperplasia (CAH), a life-threatening disease (reviewed in ref. 1). The C4 genes, *C4A* and *C4B*, are transcribed in the same orientation as the P450c21B gene [also termed P450XXIA2 (3) or *CYP21B* (4)]. The P450c21A locus (*CYP21A*) also has the same transcriptional orientation but has several mutations rendering it a nonfunctional pseudogene (5, 6). A map of this region is shown in Fig. 1*A*. Genetic rearrangements in this locus are common (12). Deletions extending from the middle of the P450c21A gene to the middle of the P450c21B gene occur in 11% of patients with CAH (7); however, no deletion has been described extending to the 3′ end of the P450c21B gene. We present evidence for the existence of a gene lying on the DNA strand opposite from that encoding the C4 and P450c21 genes. This gene overlaps exon 10 of the P450c21 gene(s) and extends 3′ to the P450c21 gene(s). This gene(s) encodes 3.5- and 1.8-kb mRNAs expressed in the adrenal and in a Leydig cell tumor but not in nonsteroidogenic tissues. The sequence of the available cDNA shows no similarity to known sequences.† Because no deletion of this gene has been found in numerous studies of this region of DNA, we propose that it encodes an essential function.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

RNA was prepared from tissues either by guanidinium isothiocyanate extraction followed by CsCl ultracentrifugation (8) or by LiCl precipitation (13); RNA was prepared from cultured cells as described (14). Electrophoresis of total RNA in 1% agarose gel, its transfer to nylon membranes, and the hybridization and washing of such membranes were done as described (14).

RNA was reverse-transcribed into cDNA, cloned into λgt11, and screened with a human P450c21 cDNA (10), all as described previously (15). Positively hybridizing plaques were purified and the sizes of the inserts were determined by agarose gel electrophoresis followed by blotting and hybridization with the P450c21 cDNA. The longest insert obtained (2.7 kb) was subcloned in M13 and pUC18 and sequenced by the dideoxy chain-termination technique using either single-stranded M13 templates or double-stranded sequencing (16). The cDNA sequence and the encoded protein sequence were compared to the GenBank and National Biomedical Research Foundation data bases on April 28, 1989.

Southern blots of normal human genomic DNA were prepared and probed with cDNA fragments isolated from their cloning vectors and labeled with ^{32}P with random primers, as described (7).

RESULTS AND DISCUSSION

Cloning of cDNA Encoded by this Gene. To determine the genetic lesion causing one form of 21-hydroxylase deficiency, we obtained adrenal tissue from a homozygously affected human fetus aborted at week 20. The diagnosis had been established hormonally and by HLA identity to an affected sibling (1). The adrenal cDNA library of 9×10^5 clones prepared from this tissue was screened with human P450c21 cDNA (10), identifying 36 cDNA clones of 1.0–2.7 kb. As human P450c21 mRNA is about 2.0 kb long (17), we considered that the cDNAs longer than 2.0 kb might be copies of incorrectly spliced mRNA precursors and that such splicing error(s) might account for the 21-hydroxylase deficiency. Therefore, we studied the 2.7-kb cDNA insert further.

Location of the Gene Encoding the 2.7-kb cDNA. Restriction endonuclease mapping of the 2.7-kb cDNA suggested that only a portion of it corresponded to the P450c21 gene. When

Abbreviations: CAH, congenital adrenal hyperplasia; C4, fourth component of complement; P450c21, cytochrome P450 specific for steroid 21-hydroxylation.

*To whom reprint requests should be addressed at: Room 677-S Pediatrics, University of California, San Francisco, CA 94143-0434.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M25813).

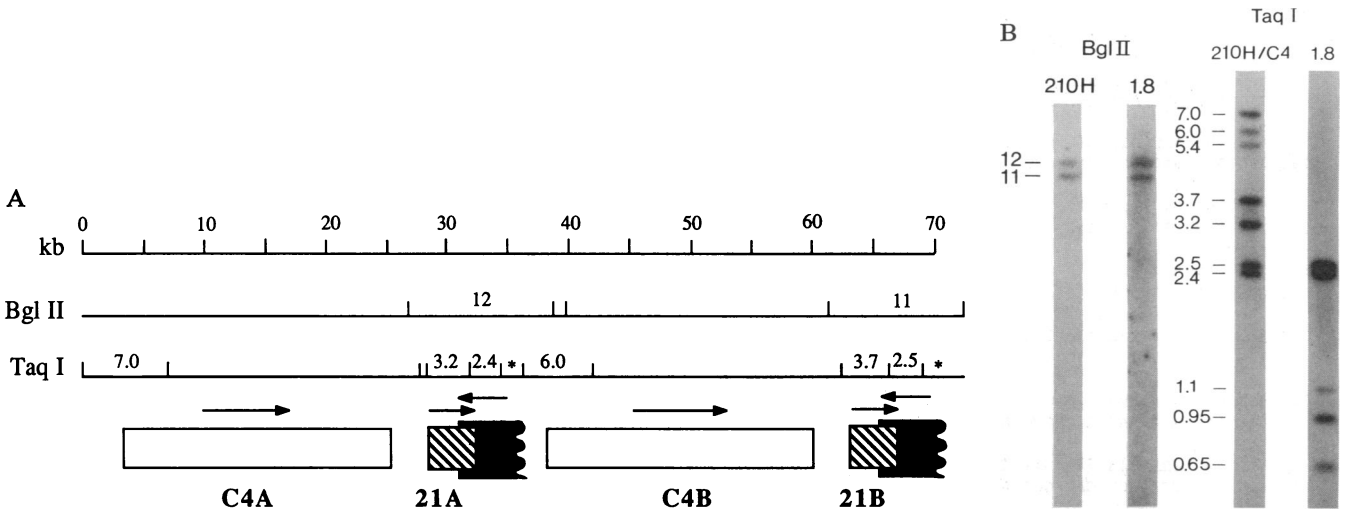

FIG. 1. (*A*) Map of the P450c21/C4 gene locus in the class III region of the HLA locus on human chromosome 6p21. The top line shows the scale in kilobases (kb). Boxes show the extent of *C4A* and *C4B* (open boxes), the P450c21A and -B genes [designated 21A and 21B (hatched boxes)], and the approximate extent of the detected portions of the opposite-strand gene(s) described in this report (solid boxes). Some of the known loci cleaved by restriction endonucleases *Bgl* II and *Taq* I, derived from published studies (5–7, 9, 12), are shown. Numbers indicate sizes (in kilobases) of DNA fragments hybridizing to the 1.8-kb *Bam*HI–*Eco*RI fragment of the 2.7-kb cDNA. The asterisks 3′ to the 2.5- and 2.4-kb *Taq* I fragments indicate that several small fragments lie in those regions. (*B*) Southern blots of genomic DNA from a normal person. DNA was cleaved with *Bgl* II or *Taq* I and probed with the indicated ^{32}P-labeled cDNAs: 21OH, a human P450c21 cDNA fragment comprising the 1229 bases at the 3′ end of P450c21 mRNA (10); 21OH/C4, a mixture of the 476-base-pair *Bam*HI–*Kpn* I fragment from the 5′ end of the full-length human C4 cDNA clone pAT-A (11) and the 21OH probe; and 1.8, the 1.8-kb *Bam*HI–*Eco*RI fragment of the 2.7-kb cDNA excluding all P450c21 sequences. The sizes of the hybridizing bands seen in these blots were determined by comparison to migration of *Hin*dIII-cleaved DNA from bacteriophage PM2. The map locations of the hybridizing bands are shown in *A*. The 12- and 11-kb *Bgl* II bands, respectively, correspond to the P450c21A and -B genes (and their 3′ flanking DNA). The 7.0-kb *Taq* I band corresponds to the 5′ end of *C4A*, whereas the 6.0- and 5.4-kb *Taq* I bands correspond to the 5′ ends of the long and short alleles of the *C4B*, respectively (11). The 3.2- and 2.4-kb *Taq* I fragments correspond to the P450c21A pseudogene and its 3′ flanking DNA, respectively, whereas the 3.7- and 2.5-kb *Taq* I fragments correspond to the functional P450c21B gene and its 3′ flanking DNA, respectively (7). The 1.1-, 0.95-, and 0.65-kb *Taq* I fragments correspond to the map region(s) designated by the asterisk(s).

Southern blots of human genomic DNA cleaved with *Bgl* II and *Taq* I were probed with the 2.7-kb insert, DNA fragments corresponding to the P450c21 genes and their 3′ flanking DNA hybridized (data not shown). When these blots were probed with a 1.8-kb *Bam*HI–*Eco*RI fragment of the 2.7-kb cDNA that lacks P450c21 sequences, only the 3′ flanking DNA hybridized (Fig. 1*B*). This suggested that the cDNA did not correspond to P450c21 or C4 mRNA, but to another transcript encoded by a gene lying downstream (3′) from the P450c21 gene. Such a transcript could have two possible origins. First, it might have the same transcriptional orientation as the P450c21 gene and be initiated within that gene and extend downstream from it. Second, it might be initiated downstream from the P450c21 gene and be transcribed in the opposite orientation (i.e., from the opposite strand) and extend back to the P450c21 gene. Inspection of the sequence of the P450c21A and -B genes (5, 6) showed no sequences similar to transcriptional initiation complexes within either P450c21 gene; however, an AATAAA sequence, generally implicated as a polyadenylylation signal (18), was found on the antisense strand 40 bases 3′ to the TGA translational stop codon of P450c21. These observations were consistent with the second possible origin of the 2.7-kb cDNA. To confirm this hypothesis, we determined the nucleotide sequence of the 2.7-kb cDNA. The 3′ end of the 2.7-kb cDNA was identified by a 13-base poly(A) tail; the adjacent DNA had a nucleotide sequence identical to that of the antisense strand of the 3′ untranslated region of the P450c21 gene and contained the AATAAA sequence in the predicted location. Therefore, the 2.7-kb cDNA was encoded by a gene on the opposite DNA strand from the P450c21 and C4 genes that had a transcriptional initiation site downstream from the P450c21 locus and whose 3′ end overlapped P450c21 exon 10 (Fig. 2).

To determine whether the 2.7-kb cDNA arose from the opposite strand of the P450c21A or P450c21B gene, we compared its 3′ sequence to the known P450c21A and -B gene sequences. The P450c21A pseudogene and P450c21B gene differ in only 87 of the 3400 bases comprising each of these two genes (5). Within the P450c21 3′ untranslated region

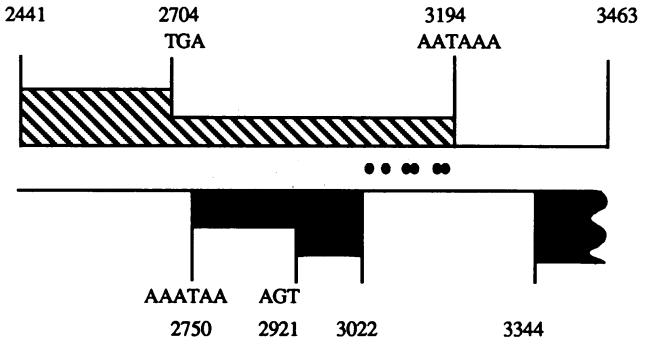

FIG. 2. Detailed map of the overlapping portions of the P450c21 gene(s) and the gene(s) encoding the 2.7-kb cDNA. Upper line: exon 10 of the P450c21B gene; the broader region of the bar indicates sequences encoding P450c21 protein and the narrower region indicates the 3′ untranslated region of P450c21 mRNA; the thin line indicates 3′ flanking DNA. Lower line: presumed structure of the 3′ end of the gene encoding the 2.7-kb cDNA, inferred from partial cDNA sequencing. The narrow region of the bar shows the presumed 3′ untranslated portion of the corresponding mRNA. The broad bar corresponds to protein-coding sequences, and the thin lines correspond to 3′ flanking DNA and the last intron of this gene. Nucleotide numbers correspond to the P450c21B gene sequenced by Higashi *et al.* (5). The six dots between the two diagrams indicate the positions of the only nucleotides differing between the P450c21A and -B genes in exon 10. The diagram only extends to the last available base of the genomic DNA sequence of the 3′ flanking region of P450c21B (5). Only those regions of the cDNA whose genomic locations could be confirmed from the genomic DNA sequence (5) are shown.

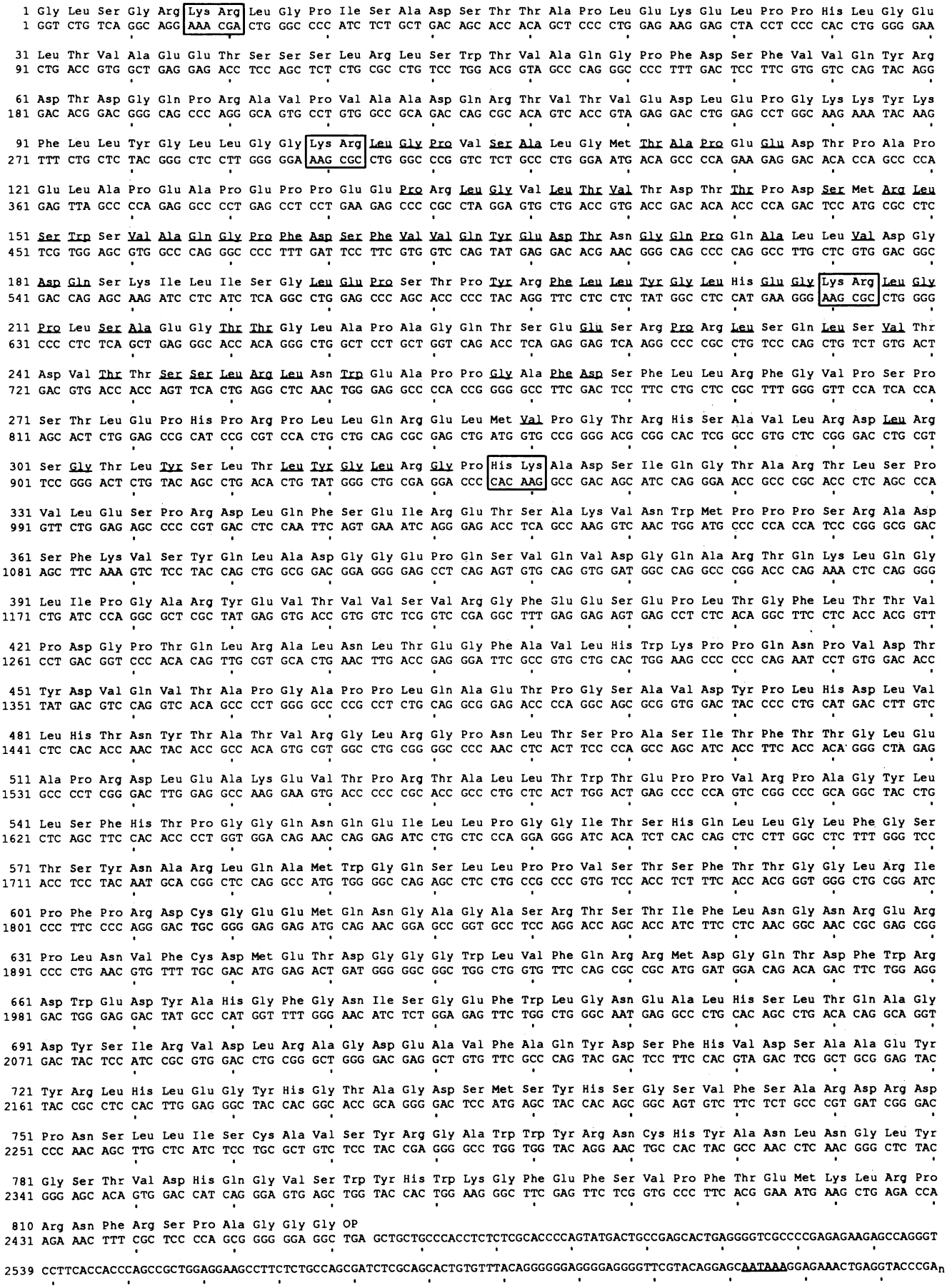

FIG. 3. (*Legend appears at the bottom of the opposite page.*)

there are 6 nucleotide differences, all lying between bases 3074 and 3180 of the P450c21B gene, according to the numbering of Higashi *et al.* (5). However, DNA corresponding to bases 3022–3343 of the P450c21B gene, encompassing all 6 nucleotide differences with P450c21A in this region, were absent in the 2.7-kb cDNA (Fig. 2). The limits of this absent region have canonical intron/exon splice-junction sequences (19), consistent with P450c21B bases 3022–3343, constituting an intron in this newly described gene. Because all nucleotides distinguishing the P450c21A pseudogene from the P450c21B gene in this region are in the spliced intronic region, we cannot determine whether the 2.7-kb cDNA arises from the opposite strand of the P450c21A pseudogene or from the opposite strand of the P450c21B gene.

Structure of the 2.7-kb cDNA. The complete sequence of the 2.7-kb cDNA is shown in Fig. 3. It contains a single open reading frame of 2457 bases encoding 819 amino acids, followed by a TGA translational termination signal. It has a 193-base 3′ untranslated region containing a typical AATAAA polyadenylylation signal followed by a poly(A) tail 20 bases further downstream. The cDNA is clearly not full-length, as the first in-phase ATG codon is 500 bases from the 5′ end (amino acid 148 in Fig. 3) and lacks a consensus translational initiation sequence (20). Furthermore, transfer blots of adrenal and other RNAs showed that a full-length cDNA should be at least 3.5 kb (see Fig. 4). Searching of several other cDNA libraries, including an Okayama–Berg human adrenal library that has yielded 5-kb cDNAs (21), did not yield a cDNA for this gene longer than 2.7 kb. The 5′ end of the sequence is characterized by three repeated units lying between potential sites for proteolytic cleavage: Lys-Arg at amino acids 5 and 6, Lys-Arg at 100 and 101, Lys-Arg at 207 and 208, and His-Lys at 316 and 317. Of the 91 amino acids in the first repeat, 54 are identical in the second and 25 in the third. Sequencing of some genomic DNA fragments has revealed three additional homologous repeats lying 5′ of the region encoding the 2.7-kb transcript (22). This structure suggests that the protein encoded by this gene may undergo proteolytic cleavage into a series of related peptides, plus a carboxyl-terminal peptide of unrelated sequence.

Expression of the Gene. We considered that this transcript from the opposite strand of the P450c21 locus might represent a fetal-specific transcript or might be an aberrant transcript associated with 21-hydroxylase deficiency. Therefore, we probed transfer blots of RNA prepared from a variety of human fetal tissues and from normal adult adrenals (Fig. 4). The 2.7-kb cDNA hybridized to mRNAs of 3.5 and 1.8 kb in normal adrenal RNA and in RNA from a human Leydig cell adenoma (35) but did not hybridize detectably to such mRNAs in heart, liver, brain, spleen, placenta, or fibroblasts (Fig. 4). Furthermore, transcription of the gene encoding the 2.7-kb cDNA is not specific to the fetus and is unrelated to CAH. However, the lack of significant amounts of this transcript in nonsteroidogenic tissues suggests it might be associated with steroidogenesis.

Role of this Gene. Overlapping genes encoded on opposite strands of DNA occur commonly in prokaryotes, bacteriophages, and viruses, but only a few examples have been reported in higher eukaryotes. These include a gene on the opposite strand of the gene for rat gonadotropin-releasing hormone (23), two *Drosophila* loci (24, 25), a mouse locus of unknown function (26), and a newly described member of the c-*erbA* gene family on the opposite strand from the gene encoding a rat (27) and human (28) thyroid hormone receptor. While the functional relationship, if any, between P450c21

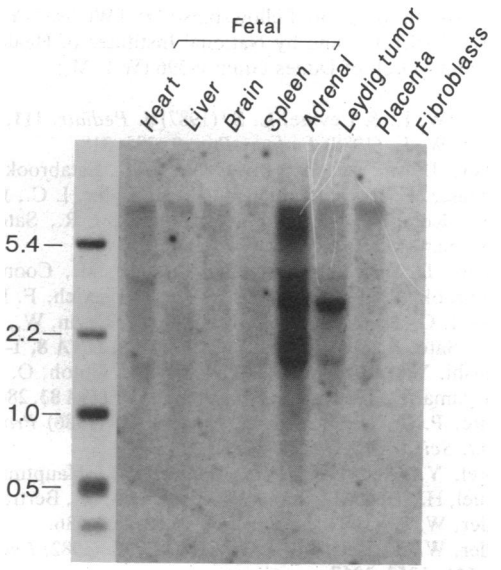

FIG. 4. Transfer blots of human RNA probed with the 1.8-kb *Bam*HI–*Eco*RI fragment of the 2.7-kb cDNA; this fragment lacks all P450c21 sequences. The tissue source of each 20-μg RNA sample is shown. Molecular size (kb) markers are *Hin*dIII-cleaved bacteriophage PM2 DNA.

and the transcript from its opposite strand is unknown, it is obviously of interest that both mRNAs are found in the adrenal, suggesting a role in steroidogenesis. Potential roles in cholesterol transport or as a trans-acting regulatory protein cannot be evaluated at present, but the opposite-strand transcript we have described probably cannot encode another steroidogenic enzyme, as the genes for these are largely known (2), the available protein sequence lacks a steroid binding site (29), and only deficiency of 21-hydroxylase is linked to *HLA* (1, 2). P450c21 and the mRNA encoding the 2.7-kb cDNA are both found in the adrenal, but we do not know whether they are expressed in the same cells. If expressed in the same cell, the overlapping 3′ sequences might regulate transcription or mRNA processing, translation, or half-life (30, 31).

Deletions extending from the P450c21A pseudogene to *C4B* are found in 14% of normal chromosomes (7), and a larger deletion encompassing the 5′ half of the P450c21B gene occurs in 11% of patients with CAH (7, 12, 32). However, among 453 chromosomes bearing CAH alleles (reviewed in ref. 12), no deletion has been described extending into the 3′ portion of the P450c21B gene. Therefore, we speculate that the 2.7-kb cDNA arises from a gene on the opposite strand of the 3′ end of the P450c21B gene, and that it encodes a transcript whose absence causes early embryonic or fetal demise. Examination of the published sequences of the 3′ ends of the murine (33) and bovine (34) P450c21 genes shows that these other mammals retain polyadenylylation sites, open reading frames, translational stop signals, and canonical splice-junction sites in the same loci as shown in Fig. 2. The encoded amino acid sequence of the last exon of the opposite-strand transcript is very similar in all three species. Thus, it appears that this gene and its relationship to P450c21 may be a general mammalian feature.

We thank Synthia H. Mellon for productive discussions and Carol Dahlstrom for typing the manuscript. This work was supported by

Fig. 3 (*on opposite page*). Sequence of the 2.7-kb cDNA. The longest open reading frame is shown. The potential proteolytic cleavage sites between amino acids 6 and 7, 100 and 101, 207 and 208, and 316 and 317 are boxed; the amino acids found between the first pair of potential cleavage sites that are conserved between the other pairs of potential cleavage sites are underlined. Note the AATAAA sequence 20 bases from the end of the cDNA. The poly(A) tail following the last base is designated A_n.

National Institutes of Health Fellowships FO5-TWO3935 (Y.M.) and T32-DK07161 (S.E.G.) and by National Institutes of Health Grant DK37922 and March of Dimes Grant 6-396 (W.L.M.).

1. Miller, W. L. & Levine, L. S. (1987) *J. Pediatr.* **111,** 1–17.
2. Miller, W. L. (1988) *Endocr. Rev.* **9,** 295–318.
3. 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.
4. Nebert, D. W., Nelson, D. R., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. G., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R. & Waterman, M. R. (1989) *DNA* **8,** 1–13.
5. Higashi, Y., Yoshioka, H., Yamane, M., Gotoh, O. & Fujii-Kuriyama, Y. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2841–2845.
6. White, P. C., New, M. I. & Dupont, B. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5111–5115.
7. Morel, Y., Andre, J., Uring-Lambert, B., Hauptmann, G., Betuel, H., Tosi, M., Forest, M. G., David, M., Bertrand, J. & Miller, W. L. (1989) *J. Clin. Invest.* **83,** 527–536.
8. Miller, W. L., Leisti, S. & Johnson, L. K. (1982) *Endocrinology* **111,** 1358–1367.
9. Prentice, H. L., Schneider, P. M. & Strominger, J. L. (1986) *Immunogenetics* **23,** 274–276.
10. Matteson, K. J., Phillips, J. A., III, Miller, W. L., Chung, B., Orlando, P. J., Frisch, H., Ferrandez, A. & Burr, I. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5858–5862.
11. Belt, K. T., Carroll, M. C. & Porter, R. R. (1984) *Cell* **36,** 904–914.
12. Miller, W. L. & Morel, Y. (1989) *Annu. Rev. Genet.*, in press.
13. Cathala, G., Savouret, J.-F., Mendez, B., West, B. L., Karin, M., Martial, J. & Baxter, J. D. (1983) *DNA* **2,** 329–335.
14. Voutilainen, R. & Miller, W. L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1590–1594.
15. Matteson, K. J., Chung, B., Urdea, M. S. & Miller, W. L. (1986) *Endocrinology* **118,** 1296–1305.
16. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
17. Voutilainen, R. & Miller, W. L. (1986) *J. Clin. Endocrinol. Metab.* **63,** 1145–1150.
18. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
19. Brethnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
20. Kozak, M. (1986) *Cell* **44,** 283–292.
21. Yamamoto, T., Davis, C. G., Brown, M. S., Schneider, W. J., Casey, M. L., Goldstein, J. L. & Russell, D. W. (1984) *Cell* **39,** 27–38.
22. Bristow, J., Gitelman, S. E., Morel, Y., Shi, Y. & Miller, W. L. (1989) *Pediatr. Res.* **25,** 81A.
23. Adelman, J. P., Bond, C. J., Douglass, J. & Herbert, E. (1987) *Science* **235,** 1514–1517.
24. Henikoff, S., Keene, M. A., Fechtel, K. & Fristrom, J. W. (1986) *Cell* **44,** 33–42.
25. Spencer, C. A., Giriz, R. D. & Hodgetta, R. B. (1986) *Nature (London)* **322,** 279–281.
26. Williams, T. & Fried, M. (1986) *Nature (London)* **322,** 275–279.
27. Lazar, M. A., Hodin, R. A., Darling, D. S. & Chin, W. W. (1989) *Mol. Cell. Biol.* **9,** 1128–1136.
28. Miyajima, N., Horiuchi, R., Shibuya, Y., Fukushige, S., Matsubara, K., Toyoshima, K. & Yamamoto, T. (1989) *Cell* **57,** 31–39.
29. Picado-Leonard, J. & Miller, W. L. (1988) *Mol. Endocrinol.* **2,** 1145–1150.
30. Strickland, S., Huarte, J., Belin, D., Vassalli, A., Rickles, R. J. & Vassalli, J.-D. (1988) *Science* **241,** 680–684.
31. van der Krol, A. R., Mol, J. M. M. & Stuitje, A. R. (1988) *Biotechniques* **6,** 958–976.
32. Higashi, Y., Tanae, A., Inoue, H., Hiromasa, T. & Fujii-Kuriyama, Y. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7486–7490.
33. Chaplin, D. D., Galbreath, L. G., Seidman, J. G., White, P. C. & Parker, K. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9601–9605.
34. Chung, B., Matteson, K. J. & Miller, W. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4243–4247.
35. Solish, S. B., Goldsmith, M. A., Voutilainen, R. & Miller, W. L. (1989) *J. Clin. Endocrinol. Metab.*, in press.