Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 4453–4457, July 1985
Genetics

Liver mRNA probes disclose two cytochrome P-450 genes duplicated in tandem with the complement *C4* loci of the mouse *H-2S* region

(steroid 21-hydroxylase/gene family/DNA sequence)

MOUNIRA AMOR*, MARIO TOSI*, CHRISTIANE DUPONCHEL*, MICHAEL STEINMETZ†, AND TOMMASO MEO*

*Unité d'Immunogénétique and Institut National de la Santé et de la Recherche Médicale U.276, Institut Pasteur, 75724 Paris Cédex 15, France; and †Basel Institute for Immunology, CH-4005 Basel, Switzerland

Communicated by R. R. Porter, March 4, 1985

ABSTRACT **A search for uncharacterized genes of the *S* region of the murine *H-2* major histocompatibility complex was undertaken; a series of cosmid clones previously aligned by overlap hybridizations were used as radiolabeled probes. Sequences hybridizing with liver poly(A)$^+$ RNA were found within a cosmid covering a region 3′ to the *C4-Slp* gene (the gene encoding the hemolytically inactive isoform of the fourth component of serum complement). Radiolabeled, short cDNA complementary to liver poly(A)$^+$ RNA was used to establish the transcriptional polarity of the newly detected gene and to define fragments containing its 3′ end. DNA sequence analyses and comparisons with porcine peptides established that the gene encodes the enzyme steroid 21-hydroxylase (EC 1.14.99.10), a cytochrome P-450 often referred to as P-450(C21), whose major site of expression is the adrenal gland. Two copies of the *P-450(C21)* gene, very similar yet distinguishable by restriction endonuclease analysis, were found individually associated with *C4* and *C4-Slp*, genes that encode isoforms of mouse fourth component of complement. One of the *P-450(C21)* genes is coamplified with *C4-Slp* in *H-2^{w7}*, a haplotype carrying a rare elongation of the *S* region. Comparisons with other members of the P-450 gene family show that the *P-450(C21)* genes encode peptides of extraordinary evolutionary conservation. The detection of a liver transcript of P-450(C21) raises the issue of the specific metabolic role of this enzyme in this organ and may have implications for the interpretation of human congenital adrenal hyperplasia.**

The *S* region of the murine *H-2* major histocompatibility complex is commonly viewed as composed of loci related to the complement system, or class III genes. Although it represents the largest genetic segment (0.4 centimorgan) of the *H-2* complex, surprisingly few marker genes have been definitely mapped in this region.

To extend the functional characterization of this *H-2* segment, we sought coding DNA sequences in cloned genomic fragments by using mRNA hybridization and overlapping cosmid clones. We now report the finding of DNA sequences that hybridize with a 20S liver mRNA and that lie immediately adjacent to the gene (*C4*) encoding the fourth component of complement (C4) and to the gene (*C4-Slp*) encoding the nonhemolytic isoform of C4, sex-limited protein (Slp). By sequencing genomic DNA, we have established that these sequences correspond to two copies of the gene encoding steroid 21-hydroxylase [21-OHase; steroid 21-monooxygenase; steroid,hydrogen-donor:oxygen oxidoreductase (21-hydroxylating), EC 1.14.99.10], a member of the cytochrome P-450 isozyme superfamily (1).

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

Preparation of Liver DNA and RNA. Genomic DNA was isolated (2) from liver nuclei prepared by the citric acid method (3). RNA was isolated essentially as described (4) and precipitated in 2.7 M lithium chloride. The polyadenylylated RNA fraction was bound to poly(U)-Sephadex (Bethesda Research Laboratories) according to the instructions of the manufacturer but was eluted at 50°C in 50% formamide/10 mM Na Hepes, pH 7.0/1 mM EDTA/0.4% $NaDodSO_4$. For some experiments, poly(A)$^+$ was size-fractionated by denaturation with 15 mM methylmercury hydroxide followed by sedimentation in 10–33% sucrose exponential gradients prepared in 25 mM NaCl/5 mM EDTA/10 mM Na Hepes, pH 7.0.

Nucleic Acid Blotting and Hybridizations. Total genomic DNA or cloned DNA was digested with the restriction endonuclease specified in the figure legends, and the restriction fragments were separated by agarose gel electrophoresis, denatured, and blotted onto a GeneScreen membrane (New England Nuclear) in 10× NaCl/Cit (1× is 150 mM NaCl/15 mM sodium citrate, pH 7) (5). RNA was denatured for 15 min at room temperature with 10 mM methylmercury hydroxide in borate buffer (25 mM boric acid/2.5 mM sodium tetraborate/4 mM sodium sulfate/0.2 mM EDTA) and then adjusted to 3% (wt/vol) formaldehyde. RNAs were separated in 1.2% agarose gels in borate buffer containing 3% formaldehyde (running buffer) (2). After electrophoresis, the gels were washed with distilled water for 1 hr and with 0.2 M ammonium acetate for 30 min and then stained with ethidium bromide in 0.2 M ammonium acetate. Before blotting to a GeneScreen membrane, the gel was reequilibrated for 1 hr in the running buffer.

After a 2-hr baking in a vacuum oven at 80°C, DNA or RNA filters were prehybridized for 2–6 hr at 43°C in 50% formamide/5× Denhardt's solution (1× is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin)/5× NaCl/Cit/50 mM sodium phosphate buffer, pH 6.5/0.2% $NaDodSO_4$/yeast RNA (250 μg/ml)/depurinated and denatured salmon sperm DNA (100 μg/ml). Hybridization was for 40 hr at 43°C in the same solution but supplemented with poly(A) (50 μg/ml), poly(C) (25 μg/ml), and probe (≈10^6 dpm/ml). Filters were washed extensively at room temperature and twice for 30 min at 67°C in 1× NaCl/Cit/0.2% $NaDodSO_4$/0.05% sodium pyrophosphate/15 mM sodium phosphate buffer, pH 7.3, and then twice for 30 min at 67°C in 0.1× NaCl/Cit/0.2% $NaDodSO_4$. Dried filters were ex-

Abbreviations: C4 and C4-Slp, hemolytically active and inactive isoforms, respectively, of the fourth component of murine complement; *H-2*, murine major histocompatibility complex; *HLA*, human major histocompatibility complex; 21-OHase, steroid 21-hydroxylase [also referred to as P-450(C21)]; kb, kilobase(s); bp, base pair(s).

posed to Kodak XAR5 films for 1–6 days at −70°C with a Cronex Lightning Plus intensifying screen (DuPont).

Identification of Coding Sequences on Genomic Clones. The construction and screening of the AKR mouse cosmid library and the alignment of cosmid clones are described elsewhere (6). DNA from cosmid clones was ^{32}P-labeled by nick-translation (7) and hybridized to blots of electrophoretically separated liver poly(A)$^+$ RNA. To localize the hybridizing sequences, restriction endonuclease fragments of cosmid clones were separated on agarose gels, blotted as described above, and hybridized with oligo(dT)-primed, ^{32}P-labeled cDNA (2) synthesized against a sucrose-gradient-enriched fraction of liver poly(A)$^+$ RNA. In parallel hybridizations, repetitive DNA sequences were detected by use of ^{32}P-labeled total mouse DNA. [The average length of these cDNA probes was varied as indicated in the figure legends by using final concentrations of either 10 μM or 3 μM limiting ^{32}P-labeled deoxynucleoside triphosphates and a constant concentration (200 μM) of the remaining deoxynucleoside triphosphates.] These hybridizations were for 48 hr at 60°C, with constant agitation, in the hybridization solution described above, except that formamide was omitted.

DNA Sequence Analysis. DNA fragments were made blunt-ended with the Klenow fragment of DNA polymerase I and inserted into the *Sma* I site of bacteriophage vector M13mp8. Complementary strands were sequenced by the chain-termination method of Sanger *et al.* (8). The sequence was confirmed by subcloning and sequencing internal *Hae* III fragments.

RESULTS

Detection of a Previously Unknown *H-2S* Region Gene. We have previously described two cosmid clones, isolated from an AKR/J (*H-2^k*) mouse library and containing the genes encoding the two isoforms C4 and C4-Slp of the fourth component of murine complement (6, 9, 10). Starting from these clones, we isolated and linked by overlap-hybridization methods a number of *S*-region cosmids. Several new genes could be expected on these cosmids, primarily those anticipated from analogy with the major histocompatibility complex of other mammals (11). However, the structural genes for the complement components C2 and factor B (see ref. 12 for their molecular map in the human genome) were stronger candidates also because their location in the *H-2S* region had been inferred directly by formal genetic studies in the mouse (13, 14). As these genes are primarily expressed in the liver, we chose hepatic poly(A)$^+$ RNA for detecting coding DNA sequences possibly carried by the newly isolated cosmids. Whereas a 150-kilobase (kb) cluster of clones contiguous to the 3′ end side of the *I*-region gene E_α^k failed to hybridize to blots of electrophoretically fractionated mouse liver poly(A)$^+$ RNA, clone 3.3, which overlaps a cosmid containing the *C4-Slp* gene (Fig. 1) yielded a strong hybridization signal.

The sequences responsible for these hybridizations were localized by use of radiolabeled cDNA. A positive 2.7-kb *Eco*RI fragment of cosmid 9.2, surprisingly close to the 3′ end of the *C4-Slp* gene, was selected for subcloning on the grounds of its lack of repetitive DNA sequences. Fig. 1*B* shows the detection of a 20S liver RNA species with this 2.7-kb *Eco*RI fragment as probe. The size of this mRNA (2.1 kb) and the location of the probe with respect to *C4-Slp* ruled out a possible identification of this mRNA as a transcript of the gene for complement component C2 or factor B. The physical location of the latter gene has meanwhile been established 5′ to the *C4-Slp* locus (15). These data, therefore, strongly suggest the presence, within 7 kb from the 3′ end of *C4-Slp*, of a hitherto unknown gene. The direction of transcription of this putative gene was deduced from differential

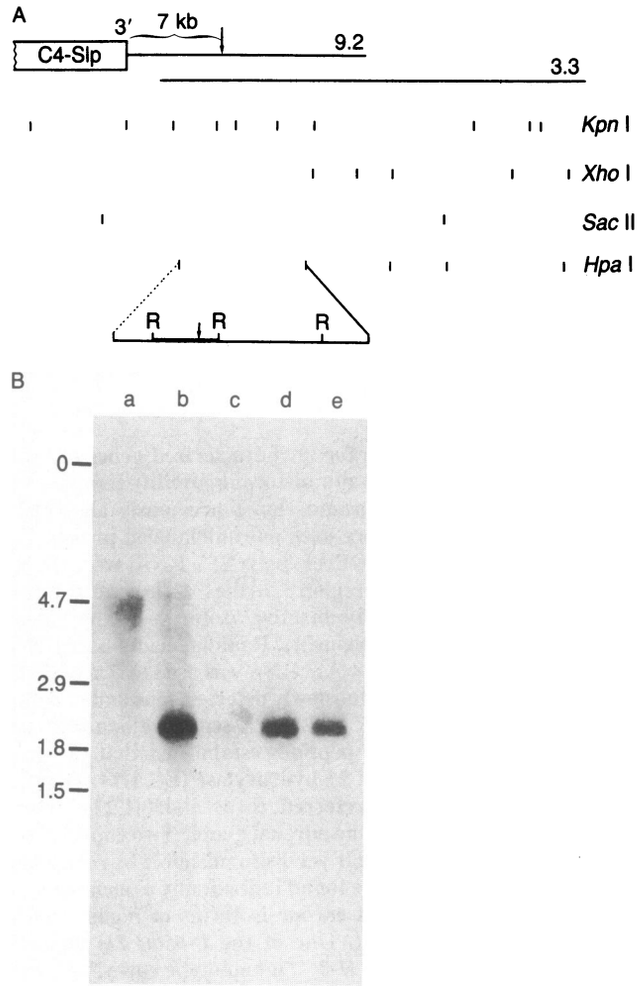

FIG. 1. Location within cosmid clones of sequences hybridizing with liver mRNA. (*A*) Cleavage map of overlapping AKR/J (*H-2^k*) cosmid clones. Lines indicate the entire insertion of clone 3.3 and part of the insertion of clone 9.2, for which the 3′ portion of the *C4-Slp* gene is boxed. An arrow marks the 3′ end of a *21-OHase* gene. The 10.6-kb *Hpa* I fragment (expanded) contains a 2.7-kb *Eco*RI fragment (thick line) which is devoid of repetitive sequences and hybridizes with ^{32}P-labeled cDNA synthesized against total poly(A)$^+$ liver RNA. The left boundary of the enlarged segment is dotted to indicate that no other *Eco*RI (R) site is present in clone 9.2. This made it convenient to isolate the 2.7-kb *Eco*RI fragment from clone 9.2 instead of clone 3.3 for subcloning in the vector pBR322 (2). (*B*) Hybridization of mouse liver RNA with the 2.7-kb *Eco*RI fragment. Lane a: 10 μg of poly(A)$^-$ RNA. Lane b: 4 μg of total poly(A)$^+$ RNA. Lanes c–e: 0.8 μg of contiguous sucrose gradient fractions of poly(A)$^+$ RNA that sedimented around the 18S ribosomal RNA marker. The side markers indicate the origin of migration (o) and the size in kb of *Escherichia coli* and mouse rRNAs used as standards. Autoradiographic exposure was for 6 days.

hybridization of short and longer oligo(dT)-primed cDNA probes corresponding to sequences within the 2.7-kb *Eco*RI fragment (Fig. 2). The orientation of this fragment with respect to the cosmid map of Fig. 1 then was determined by hybridizing appropriate restriction fragments of cosmid 9.2 either with the 547-bp *Msp* I–*Taq* I probe or with the 281-bp *Taq* I–*Msp* I probe (data not shown). The resulting orientation of this *Eco*RI fragment demonstrates that the gene has the same transcriptional polarity as *C4-Slp*.

Demonstration that the Gene Encodes 21-OHase. The nucleotide sequence of the 828-bp *Msp* I fragment, which hybridizes with short liver cDNA and thus covers the 3′ end of the putative gene, was then determined. The sequence (Fig. 3*A*) contains, within the stretch corresponding to the 281-bp *Taq*

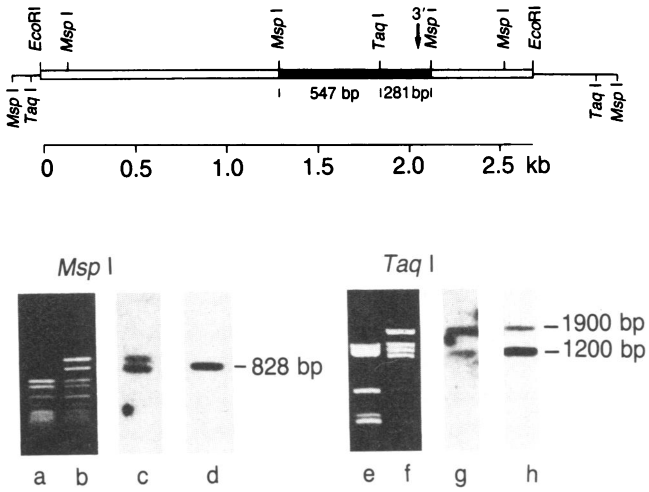

FIG. 2. Fine mapping of cosmid sequences hybridizing with liver mRNA and deduction of their transcriptional polarity. *Taq* I and *Msp* I sites were mapped by double digestions of a pBR322 subclone of the 2.7-kb *Eco*RI fragment from cosmid 9.2; plasmid pBR322 sequences are shown as a thin line. Poly(A)$^+$ liver RNA that sedimented in the 20S region in sucrose gradients was used to prepare radiolabeled, oligo(dT)-primed cDNA. This probe revealed a strongly hybridizing 828-base-pair (bp) *Msp* I fragment (shaded bar), and an adjacent 1100-bp *Msp* I fragment hybridizing with lower intensity (lane c). Within the 828 bp, the hybridizing sequences encompass the *Taq* I site (lane g). Shorter cDNA probes, obtained by lowering the concentration of the limiting deoxynucleoside triphosphate, hybridize to the 828-bp *Msp* I fragment but not to the adjacent 1100-bp *Msp* I fragment (lane d) and are particularly homologous to the 1200-bp *Taq* I fragment (lane h), indicating a left-to-right polarity of transcription. Lanes: a and e, pBR322 digested with *Msp* I and *Taq* I, respectively; b and f, pBR322 subclone of the 2.7-kb *Eco*RI fragment from cosmid 9.2 digested with *Msp* I and *Taq* I, respectively; c and d, hybridizations of lane b with the longer and the shorter oligo(dT)-primed liver cDNA probes, respectively; g and h, hybridizations of lane f with longer and shorter cDNA probes, respectively. Lanes a, b, e, and f show ethidium bromide staining; lanes c, d, g, and h are autoradiograms.

I–*Msp* I fragment, a single potential polyadenylylation site preceded by several stop codons in all three reading frames, suggesting the presence of a rather long 3′ noncoding region. Indeed, long open reading frames, which could be compared with the available data on proteins possibly encoded in this portion of the major histocompatibility complex, were found in the sequence corresponding to the 547-bp *Msp* I–*Taq* I fragment.

In an attempt to identify these putative coding sequences, a number of genetic markers were considered because of their association with the *S* region of the *H-2* complex or with its analogue in man, the *HLA-B:HLA-D* segment (listed in refs. 11 and 24). In particular, amino acid (18) and nucleotide (25) sequences have been reported for the enzyme 21-OHase, whose deficiency is suspected to cause a congenital form of adrenal hyperplasia previously linked to the complement region of *HLA* (26). Thus, the amino acid sequence predicted from the nucleotide sequence in Fig. 3 was compared with the sequenced peptides from the porcine adrenal gland 21-OHase (18). Surprisingly, one of the open reading frames of the gene revealed a high degree of similarity with the sequence of the cysteine-containing peptides T-37 and Tsa-23 and with a recently determined extension of the known sequence of peptide Tsa-23 toward the carboxyl terminus (19). The overall matching of 47 out of 62 residues leaves no doubt as to the identification of this gene.

Our data obviously support a direct link between peptides T-37 and Tsa-23 and thus contradict the previously reported 10-residues extension at the amino terminus of peptide Tsa-23

A

```
MspI                                                     (T-37
 argproleuleuglyglnthrserleualaleuhisleuleuprothraspArgPheLeu
 CGGCCTCTGCTGGGACAAACTTCTCTAGCTCTCCATCTACTACCCACAGATCGCTTCCTG  60

 Ala      Ala      SerAlaLeuAla                 )(Tsa-23
 GluProGlyLysAsnProArgThrProSerPheGlyCysGlyAlaArgValCysLeuGly
 GAACCTGGCAAGAATCCCAGAACACCATCCTTTGGCTGTGGGGCACGCGTGTGCCTGGGA 120

                                 ValGln
 GluProLeuAlaArgLeuGluLeuPheValValLeuAlaArgLeuLeuGlnAlaPheThr
 GAGCCTCTGGCACGGCTGGAGCTCTTTGTGGTGCTGGCTCGTCTGCTGCAGGCCTTCACT 180

       ---ThrGlu    Ala                 Val     X Leu
 LeuLeuProProProAspGlyThrLeuProSerLeuGlnProGlnProTyrAlaGlyIle
 CTGCTGCCTCCACCAGATGGAACCCTGCCTTCCCTGCAGCCCCAGCCTTATGCTGGCATC 240

 AsnLeuProIleProProPheGlnValArgLeuGlnProArgAsnLeuAlaProGlnAsp
 AATCTCCCGATTCCTCCTTTCCAGGTGCGGCTGCAGCCCAGAAACCTGGCGCCCCAAGAC 300

 GlnGLyGluArgPro
 CAGGGTGAGCGTCCT[TGA]CAGGATAGGACGAGTCTCTTTAAAGTTTCTCCTTTATTGCTC 360

 AGTTCCCCCCCCCCCCCCCCCGTAAACATGGTGCTGTGAGATTGTGGCAGAGAAGGCTTC 420

 CTCGGTGAGGCACGGTGGCTCTTCTCTACTGGGGTGTGAGTGAGGTAGGCAGCAGGCTCA 480

 GGTGCCCTGGTGGGGGCCTGGAAGTTTCTGGGTCGGAGCTTCATTTCCGTGAAGGGCACA 540

       TaqI
 GAAAACTCGAAGCCCTTCCAGTGGTACCAGCTCACTCCCTGAGGAGAGGGTTGTCAAGGG 600

 AAGAGAGTCAAGACAGCGCCATATCAATGCCACCTAATCCAGCCCCACCCTGCTCCTGTA 660

 GGTGACTCCCAGGGTCCAGTCATCTCCCCCATGATCCCTGAGCTGGTGCCTCTGCCCAGT 720

 [AATAAA]TGGTCTCTTCAGATTTCCTCAGCAAGGTGGGCATGAGTGTCTCATTTCACAGTT 780

                                                 MspI
 GAGCTTCTGAGGCATGAGTTCAGCCCTTCATGTGAGAGGACTTCAGCCCGG          830
```

B

P-450(C21)S (mouse)	F	G	C	G	A	R	V	C	L	G	E	P	L	A	R	L	E	L	F	V	V
$P\text{-}450_e$ (rat)	F	S	T	G	K	R	I	C	L	G	E	G	I	A	R	N	E	L	F	L	F
$P\text{-}450_c$ (rat)	F	G	L	G	K	R	K	C	I	G	E	T	I	G	R	L	E	V	F	L	F
P-450 (SCC) (bovine)	F	G	W	G	V	R	Q	C	V	G	R	R	I	A	E	L	E	M	T	L	F
$P\text{-}450_{cam}$ (P. putida)	F	G	H	G	S	H	L	C	L	G	Q	S	L	A	R	R	E	I	I	V	T

FIG. 3. Identification of a murine *21-OHase* gene by nucleotide sequence analysis. (*A*) The genomic DNA sequence shown covers the 828-bp *Msp* I fragment represented as a solid bar in Fig. 2. The protein stop codon and a potential polyadenylylation site (16) are boxed. The first 17 amino acids are represented with lower-case initial letters to indicate the possibility that the corresponding nucleotide sequence represents the 3′ portion of an intron ending at a potential splice acceptor site (17) (underlined AG). Amino acid residues are compared with the sequence of two porcine 21-OHase peptides (18), T-37 and Tsa-23, which are delimited by parentheses, and with a recent carboxyl-terminal extension of the sequence of the latter peptide (19). Underlined residues in the deduced sequence indicate identity, amino acids that differ in the porcine sequence are indicated above the deduced murine sequence. A dashed line marks a gap introduced to maximize matches. (*B*) Amino acid sequence conservation of cytochrome P-450 proteins around the heme-binding cysteine. The single-letter amino acid notation is used. Boxed residues indicate identity between the protein predicted from the nucleotide sequence of *P-450(C21)S*, the murine *21-OHase* gene adjacent to *C4-Slp*, and proteins representing different classes of cytochromes P-450. Conservative replacements are underlined. The sequences flanking this highly conserved region are essentially specific for each class of P-450. An arrow points to the glycine codon interrupted by an intron in the phenobarbital-inducible rat cytochrome $P\text{-}450_e$ gene. For the bovine cytochrome P-450(SCC), no genomic sequence is available for comparison. $P\text{-}450_c$ (20) and $P\text{-}450_e$ (21) are the major rat liver isozymes inducible by methylcholanthrene and phenobarbital, respectively. P-450(SCC) (22) is a steroidogenic isozyme found in the adrenal cortex mitochondria, and $P\text{-}450_{cam}$ is a prokaryotic enzyme from *Pseudomonas putida* (23). In the region amino-terminal to the heme-binding cysteine, the residues recently proposed for the porcine P-450(C21) (19) have the more divergent sequence G A A G S R V.

(19). Formally, we cannot exclude that this discrepancy reflects a species difference. Moreover, the cDNA sequence recently reported for the bovine adrenal gland 21-OHase (25) is uninformative in this context because it does not cover this carboxyl-terminal region of the enzyme. Peptide Tsa-23 lies within the most conserved sequence of all cytochromes P-450, as shown by previous comparisons around the cysteine (bold print in Fig. 3) believed to bind the heme iron atom in these isozymes (22). The new murine sequence allows extension of such comparisons to a region of 21-OHase amino-terminal to this heme-binding cysteine (Fig. 3*B*). Interestingly, these protein alignments reveal that, at variance with the aforementioned porcine sequence (19), the sequence deduced from the murine gene more faithfully complies with the canonical matches traced for this region from all other cytochromes P-450, including the prokaryotic camphor hydroxylase.

Tandem Duplications of *C4* and *P-450(C21)* Genes. A direct implication of our findings is that a gene encoding 21-OHase, typically expressed in the adrenal gland, is also transcribed in the liver. This observation raises the question of whether there are multiple copies of genes encoding 21-OHase, which could be expressed in a tissue-specific fashion. Most restriction endonuclease analyses were uninformative on this point. Essentially a single DNA fragment was detected (data not shown) in each case, when Southern blots of genomic DNA from various mouse strains were probed with the 2.7-kb *Eco*RI fragment subcloned from cosmid 9.2, after digestion of the DNA with *Bam*HI (4.0 kb; strains BALB/c, B10.W7R, and B10.HTT), *Hin*dIII (16.5 kb; strains BALB/c, AKR/J, DBA/1, and C57BL/6), *Pvu* II (2 kb; strains BALB/c, DBA/1, and C57BL/6), or *Eco*RI (2.7 kb; strains BALB/c, B10.W7R, and B10.HTT). Although single hybridization bands could hide multiple conserved *21-OHase* gene copies, a DNA blot analysis of hamster–BALB/c somatic cell hybrids showed that all hybridizing sequences reside on mouse chromosome 17 (Fig. 4). However, the endonuclease *Taq* I permitted the detection of two nonallelic gene copies in the BALB.K ($H\text{-}2^k$) strain (lane e) as well as in all strains mentioned above. As one would have suspected from the very close proximity of a 21-OHase-encoding gene to the *C4-Slp* gene on cosmid 9.2 and from the existence of two nonallelic *C4* genes in the mouse (10, 15), the other copy is located on the cosmid containing the second *C4* gene (refs. 6 and 10; see Fig. 4). To distinguish the *21-OHase* gene associated with the *C4* gene from that associated with the *C4-Slp* gene, we here refer to the two gene copies as *P-450(C21)C* and *P-450(C21)S*, respectively, in line with the designation more commonly used for this gene family. The close association of *P-450(C21)* and *C4* genes is further underscored by a genetic peculiarity of strain B10.W7R, which carries an amplification of the *H-2S* region correlating with a testosterone-independent form of expression of the C4-Slp protein (10). As seen in lane f of Fig. 4, B10.W7R carries one *P-450(C21)C* but multiple (probably four) copies of the *P-450(C21)S* gene, in line with our finding that only the *C4-Slp* isotype is amplified in this strain (35).

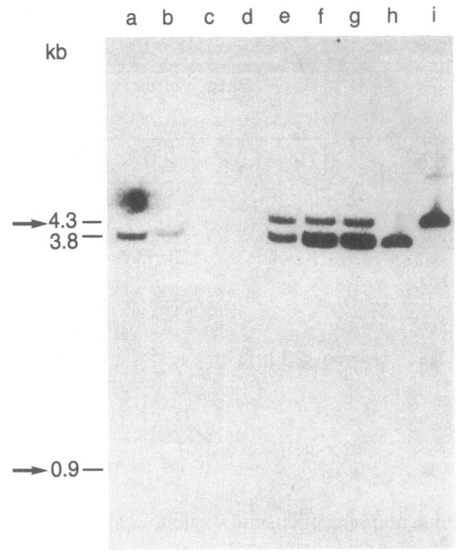

FIG. 4. Duplications of the *P-450(C21)* gene within *H-2* of different strains. Uncloned genomic DNA from cell lines (lanes b–d; 30 μg), mouse liver (a, e, f, and g; 15 μg) or cloned genomic DNA (lanes h and i; 1 ng) was digested with *Taq* I, electrophoresed in a 0.8% agarose gel, blotted, and hybridized with the M13mp8 subclone of the 547-bp *Msp* I–*Taq* I fragment (Fig. 2) labeled in its vector portion by use of a M13 probe primer. All *P-450(C21)* sequences lie on chromosome 17, as shown by the comparison of the genomic DNAs of BALB/c mouse liver (lane a), of a hamster–BALB/c hybrid cell line carrying the mouse chromosome 17 (lane b), and of its sister line, which lacks mouse chromosome 17 but carries at least one copy of each of the other mouse chromosomes (10) (lane c). The hamster *P-450(C21)* sequences are not detectable under these conditions in the parental hamster cell line (lane d). In the $H\text{-}2^k$ haplotype (BALB.K; lane e) the probe reveals two *Taq* I fragments of 3.8 kb and 4.3 kb, which correspond to the *P-450(C21)S* gene (cosmid 9.2 containing *C4-Slp*; lane h) and to the *P-450(C21)C* gene (cosmid 10.8, containing *C4*; lane i), respectively. Strains B10.W7R (lane f) and C3H.W7 (lane g), possess multiple copies of the *P-450(C21)S* gene. Note that a *Taq* I polymorphism distinguishes the *P-450(C21)C* gene of BALB/c mouse $H\text{-}2^d$; lane a) from the corresponding gene of the $H\text{-}2^k$ and of the $H\text{-}2^{w7}$ haplotypes, because the hybridization band of 4.3 kb is replaced in BALB/c by a 0.9-kb *Taq* I fragment (see arrows).

DISCUSSION

A most surprising outcome of our search for new *H-2S* region genes that might be expressed in the liver was the detection of *P-450(C21)* sequences tandemly duplicated with the genes encoding isoforms (C4 and Slp) of the fourth component of complement. The enzyme 21-OHase is typically produced in the adrenal gland, and a deficiency of this form of the enzyme has long been correlated with an *HLA*-linked disorder of steroidogenesis (27).

Our data clearly establish that at least one *P-450(C21)* gene copy is transcribed in the liver. In fact, it could be argued that the mRNA detected in this tissue does not encode 21-OHase but another isozyme species whose mRNA expression in the liver is sufficiently high to offset a weak sequence homology with *P-450(C21)*. Two observations contradict this hypothesis. First, no other member of the *P-450* gene family was found to hybridize with the short *P-450(C21)* probe used in the Southern blot shown in Fig. 4, even after longer autoradiographic exposure. This probe includes the nucleotides encoding the conserved region around the heme-binding cysteine, which is essentially shared by all P-450 proteins (Fig. 3*B*). Second, radiolabeled liver cDNA hybridizes with a 1.2-kb *Taq* I fragment (Fig. 2*B*) containing exclusively the 3′ noncoding sequences of mouse *P-450(C21)S*. Significantly, the 3′ noncoding regions of the known major P-450 mRNAs show a high degree of length and sequence heterogeneity among (20, 21, 28) or even within (29) subfamilies. Accordingly, probes that include the 3′ noncoding region of *P-450(C21)S*, such as the subcloned 2.7-kb *Eco*RI fragment (Fig. 2), also failed to reveal crosshybridizing genes.

A 21-OHase activity in organs other than the adrenal gland has been previously reported in a wide variety of species, including rabbit (30), ox (31), and even spiny lobster (32). However, due to the biochemical complexity of cytochromes P-450, this activity was explained as cross-metabolization by nonsteroidogenic isozymes (18, 31). Thus, our detection of a mRNA homologous in sequence to the gene encoding 21-OHase directly raises the issue of the metabolic role that such

a typically adrenal gland isozyme might play in the liver. Perhaps its role will be found not (or not exclusively) in steroidogenesis but rather in detoxification of endogenous or "exogenous" steroid hormones. It is worth noting that in the adult mouse, the 21-OHase mRNA represents a less abundant species in the liver than in the adrenal gland, where we estimate that it accumulates at least at a 50-fold higher level. In this context, it is tempting to speculate that the liver *P-450(C21)* plays a role during fetal development.

The finding of two gene copies in the common mouse strains raises the issue of whether the 21-OHase mRNA found in the liver is specifically transcribed from one of the two genes. A direct answer to this question will be hard to acquire, in view of the apparently strong sequence conservation between these genes. We have proposed elsewhere that the *C4* gene duplications occurred recently and independently in several ancestors of mammals (10). If the syrian hamster, which characteristically possesses a single *C4* gene (unpublished data), is found to possess also a single *P-450(C21)*, a search of liver 21-OHase transcripts in this species might be revealing.

Finally, it should be noticed that the glycine marked by an arrow in Fig. 3*B* is the site where an intron occurs in the rat sequence of the phenobarbital-inducible P-450$_e$ form (20). This intron is obviously missing from the corresponding sequence of murine *P-450(C21)* (Fig. 3*A*), thus likening this gene to that encoding rat P-450$_e$, a member of the methylcholanthrene-inducible class (21). However, limited sequence data at the 5′ boundary of the *Eco*RI fragment expanded in Fig. 2 revealed the presence of introns that have no precise counterpart either in the methylcholanthrene-inducible or in the phenobarbital-inducible cytochrome P-450 genes cloned from the rat. A complete genomic sequence of mouse *P-450(C21)* might contribute important clues as to the origin of introns or the evolutionary stability of their location.

Note. P. White and collaborators have localized *21-OHase* genes within *H-2* by using their adrenal gland cDNA probe from the ox (33). Meanwhile, *21-OHase* genes have also been physically mapped at homologous positions in the *HLA* by virtue of their expression in the adrenal gland (ref. 34 and work cited in ref. 33). If 21-OHase has a hepatic form also in man, a better understanding of the regulation of 21-OHase in the adrenal gland and in the liver might contribute significantly to the molecular interpretation of the different syndromes associated with congenital deficiency of this enzyme (27).

We thank Pierre Bourgarel for skillful help in several phases of this work and Jacqueline Badella for her continual editorial assistance. We are very grateful to Harvey Eisen for critically reading the manuscript. The Ministère de l'Industrie et de la Recherche has supported this work under Grant 83LO855. A preliminary report of this work was given at the Third *H-2, HLA* Cloning Meeting, Strasbourg, France, May 2–4, 1984.

1. Nebert, D. W. & Negishi, M. (1982) *Biochem. Pharmacol.* **31,** 2311–2317.
2. Maniatis, T., Fritsch, F. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
3. Higashi, K., Shankarnarayanan, K., Adams, H. R. & Busch, H. (1966) *Cancer Res.* **26,** 1582–1590.
4. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
5. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
6. Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R. A., Dastoornikoo, G. R., Stephen, D., Gibb, E. & Romanink, R. (1985) *EMBO J.* **3,** 2995–3003.
7. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
8. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. (1980) *J. Mol. Biol.* **143,** 161–178.
9. Tosi, M., Lévi-Strauss, M., Duponchel, C. & Meo, T. (1984) *Philos. Trans. R. Soc. London Ser. B* **306,** 389–394.
10. Lévi-Strauss, M., Tosi, M., Steinmetz, M., Klein, J. & Meo, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1746–1750.
11. Robson, E. B. & Lamm, L. U. (1983) *Cytogenet. Cell Genet.* **37,** Suppls. 1–4, 47–70.
12. Carroll, M. C., Campbell, R. D., Bentley, D. R. & Porter, R. R. (1984) *Nature (London)* **307,** 237–241.
13. Gorman, J. C., Jackson, R., Desantola, J. R., Shreffler, D. C. & Atkinson, J. P. (1980) *J. Immunol.* **125,** 344–351.
14. Paolucci, E. & Shreffler, D. C. (1983) *Immunogenetics* **17,** 67–78.
15. Chaplin, D. D., Woods, D. E., Whitehead, A. S., Goldberger, G., Colten, H. R. & Seidman, J. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6947–6951.
16. Proudfoot, N. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
17. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–473.
18. Yuan, P. M., Nakajin, S., Hanju, M., Shinoda, M., Hall, P. F. & Shively, J. E. (1983) *Biochemistry* **22,** 143–149.
19. Hanju, M., Shively, J. E. & Hall, P. F. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **42,** 1774.
20. Mizukami, Y., Sogawa, K., Suwa, Y., Muramatsu, M. & Fujii-Kuriyama, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3958–3962.
21. Sogawa, K., Gotoh, O., Kawajiri, K. & Fujii-Kuriyama, Y. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5066–5070.
22. Morohashi, K., Fujii-Kuriyama, Y., Okada, Y., Sogawa, K., Hirose, T., Inayama, S. & Omura, T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4647–4651.
23. Hanju, M., Armes, L. G., Tanaka, M., Yasunobu, K. T., Shastry, B. S., Wagner, G. C. & Gunsalus, I. C. (1982) *Biochem. Biophys. Res. Commun.* **105,** 889–894.
24. Klein, J., Figueroa, F. & Klein, D. (1982) *Immunogenetics* **16,** 285–317.
25. White, P. C., New, M. I. & Dupont, B. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1986–1990.
26. O'Neill, G. J., Dupont, B., Pollack, M. S., Levine, L. S. & New, M. (1982) *Clin. Immunol. Immunopathol.* **23,** 312–322.
27. New, M. I., Dupont, B., Grumbach, K. & Levine, L. S. (1982) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Wyngaarden, J. B., Frederikson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York), pp. 973–1000.
28. Yabusaki, Y., Shimizu, M., Murakami, H., Nakamura, K., Oeda, K. & Ohkawa, H. (1984) *Nucleic Acids Res.* **12,** 2930–2938.
29. Kimura, S., Gonzalez, F. J. & Webert, D. W. (1984) *J. Biol. Chem.* **17,** 10705–10713.
30. Dieter, H. H., Muller-Eberhard, V. & Johnson, E. F. (1982) *Biochem. Biophys. Res. Commun.* **105,** 515–520.
31. Hiwatashi, A. & Ichikawa, Y. (1981) *Biochim. Biophys. Acta* **664,** 33–48.
32. James, M. O. & Shiverick, K. T. (1984) *Arch. Biochem. Biophys.* **233,** 1–9.
33. White, P. C., Chaplin, D. D., Weis, J. H., Dupont, B., New, M. I. & Seidman, J. G. (1984) *Nature (London)* **312,** 465–467.
34. Carroll, M. C., Campbell, R. D. & Porter, R. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 521–525.
35. Lévi-Strauss, M., Georgatsu, E., Tosi, M. & Meo, T. (1985) *Immunogenetics*, in press.