MOLECULAR AND CELLULAR BIOLOGY, May 1992, p. 2124–2134
0270-7306/92/052124-11$02.00/0

Vol. 12, No. 5

Mechanism and Consequences of the Duplication of the Human C4/P450c21/Gene X Locus

STEPHEN E. GITELMAN,[1] JAMES BRISTOW,[1] AND WALTER L. MILLER[2]*

Department of Pediatrics[1] and the Metabolic Research Unit,[2] University of California, San Francisco, California 94143

Received 11 December 1991/Accepted 17 February 1992

The adjacent C4 and P450c21 genes encode the fourth component of serum complement and steroid 21-hydroxylase respectively, and are tandemly duplicated in the human, murine, and bovine genomes. We recently cloned a cDNA for another duplicated gene, operationally termed X, which overlaps the 3′ end of human P450c21 and has the opposite transcriptional orientation. Thus, the organization of the locus is 5′-C4A-21A-XA-C4B-21B-XB-3′ (Y. Morel, J. Bristow, S. E. Gitelman, and W. L. Miller, Proc. Natl. Acad. Sci. USA 86:6582–6586, 1989). To determine how this locus was duplicated, we sequenced the DNA at the duplication boundaries and the 7 kb between P450c21A and C4B comprising the XA locus. The sequences located the duplication boundaries precisely and indicate that the duplication occurred by nonhomologous recombination. The boundaries are substantially different from those of the corresponding duplication in the mouse genome, suggesting that similar gene duplications may have occurred independently in ancestors of rodents and primates after mammalian speciation. Compared with XB, the XA gene is truncated at its 5′ end and bears a 121-bp intragenic deletion causing a frameshift and premature translational stop signal. Nevertheless, XA is transcribed into a stable 2.6-kb polyadenylated RNA that is expressed uniquely in the adrenal gland.

Gene duplications are especially common in the human leukocyte antigen (HLA) compatibility locus on chromosome 6p21. The class III HLA region of about 1,000 kb contains many duplicated genes with diverse functions, including genes encoding tumor necrosis factors α and β (9, 15, 54), heat shock proteins HSP70 (51), complement cascade factors B, C2, C4A, and C4B (8), and the adrenal steroid 21-hydroxylase genes P450c21A and -B (21, 46, 58). Many other genes of unknown function have also been identified in this region (28, 50, 53).

The class III region is widely studied because of its relationship to human disease, and the tandemly duplicated C4/P450c21 gene cluster has been the subject of intense scrutiny. The highly polymorphic C4A and C4B genes encode different forms of the fourth component of serum complement, and particular haplotypes may account for HLA-linked autoimmune diseases and systemic lupus erythematosus (for a review, see reference 20). The P450c21 genes (formally termed CYP21 [37]) encode adrenal steroid 21-hydroxylase. P450c21 gene lesions cause 21-hydroxylase deficiency, a common, potentially life-threatening disease affecting about 1 in 12,000 newborns. P450c21 gene conversions account for about 85% of 21-hydroxylase deficiency (including apparent point mutations), indicating frequent genetic recombination in this locus (for reviews, see references 32 and 36).

The duplication of the C4/P450c21 gene cluster has been thought to predate mammalian speciation, as mice (2) and cattle (12, 52) also have duplicated C4 and P450c21 genes linked to the leukocyte antigen loci. However, there are substantial differences among various mammals. The human P450c21B gene is functional, but the P450c21A gene has several mutations rendering it a nonfunctional pseudogene (21, 46, 58). In mice, the P450c21A gene is functional whereas P450c21B is a pseudogene bearing a single large deletion (10, 40). In cattle, both P450c21 genes are functional (12, 13, 61). Although both human C4 genes encode functional C4 protein, only one murine C4 gene is functional; the other gene, termed sex-limited protein (*Slp*), has structural changes that render it inactive in the complement cascade (39), and it has a 5′ flanking DNA insertion that confers androgen responsiveness (56). Also, 40 kb of DNA separates the 3′ end of the mouse P450c21A gene from the 5′ end of the mouse C4B gene (57), while only 6.5 kb separates the corresponding human genes. Finally, pigs (18), whales (55), Syrian hamsters, and possibly guinea pigs (29) appear to have only one P450c21 gene, suggesting that this locus may have duplicated independently in cattle, mice, and humans after mammalian speciation.

We recently discovered an additional tandemly duplicated gene pair in the human C4/P450c21 locus (34). These genes, operationally termed XA and XB, overlap the last exon of the P450c21A and -B genes, respectively, and are encoded by the opposite strand of DNA. Cloning and sequencing of a 2.7-kb X cDNA fragment (34) showed that the encoded protein has juxtaposed fibrinogen and fibronectin type III-like domains (7, 59), thus resembling the extracellular matrix protein tenascin (reviewed in reference 16). P450c21 gene deletions causing 21-hydroxylase deficiency never extend into the XB gene, but deletions of XA are found in 14% of human chromosomes (32, 36). This finding suggested that the cDNA arose from the XB gene and that the XB gene was essential for life, while the XA gene appeared to be a pseudogene (34).

We have now determined that the human C4/P450c21/X locus was duplicated by a nonhomologous recombination. This truncated the 5′ end of the XA gene. Compared with the XB gene, the XA gene also has a 91-bp exonic deletion causing a frameshift and premature termination signal. These changes suggest that the XA locus is a residual pseudogene fragment; however, the XA gene is expressed,

* Corresponding author.

and it encodes a transcript substantially different from XB in size, encoded sequence, and tissue-specific expression. It is not clear whether the XA transcript has a function; the XA locus may be on its way to becoming a pseudogene, despite abundant tissue-specific expression.

MATERIALS AND METHODS

Library screening. Two human genomic bacteriophage libraries were screened with both a 1.8-kb *Bam*HI-*Eco*RI X cDNA fragment (34) and a 476-bp *Bam*HI-*Kpn*I 5′ C4 cDNA fragment (4) under stringent conditions (42). Positively hybridizing plaques were purified through subsequent rounds of screening. Phage inserts were mapped by digestion with several restriction endonucleases followed by agarose gel electrophoresis. The authenticity of these clones was verified by Southern blotting and hybridization with the original probes. An Okayama-Berg fetal adrenal cDNA library (60) was screened as described above with the X cDNA fragment. Probes were isolated from their cloning vectors, ^{32}P labeled by the random primer method, and used to probe Southern blots, all as described previously (35).

Sequencing. Fragments of the bacteriophage inserts were subconed into Bluescript vectors and sequenced by the dideoxy technique as described previously (42). Templates were prepared either by the double-stranded method (11) or by single-stranded rescue with VCS helper phage. All sequencing was done on both strands. Sequences were entered into the DNA Inspector software program (TEXTO, West Lebanon, N.H.) for initial alignment, editing, and restriction enzyme site analysis. Further analysis was conducted in Eugene sequence analysis software through the Department of Biochemistry, University of California, San Francisco. The XA cDNA sequence and the encoded protein sequence were compared with entries in the GenBank and National Biomedical Research Foundation data bases.

RNA preparation. Human fetal tissues were obtained from the International Institute for the Advancement of Medicine (Philadelphia, Pa.). U937 cells obtained from the UCSF Tissue Culture facility were grown under standard conditions prior to harvesting for RNA preparation. RNA was prepared by tissue or cell homogenization in 4 M guanidinium thiocyanate and then ultracentrifugation over a cushion of 5.7 M CsCl (33). Pellets were resuspended in 10 mM Tris-HCl (pH 7.6)–1 mM EDTA–1% sodium dodecyl sulfate, extracted with phenol-$CHCl_3$, and precipitated with ethanol.

RNase protection assays. RNase protection assays were done essentially as described previously (33). DNA fragments were subcloned into the polylinker of a Bluescript vector (Stratagene, La Jolla, Calif.). The plasmid was linearized, and antisense RNA was transcribed with T3 or T7 RNA polymerase (Promega, Madison, Wis.). Hybridizations were performed with 20-μg samples of total RNA mixed with 5×10^5 cpm of antisense ^{32}P-riboprobe and incubated for 16 h at 45°C. For RNase digestions, a total of 0.1 U of DNase-free RNase A and 50 U of RNase T_1 (both from Boehringer Mannheim, Indianapolis, Ind.) were added to each sample, and the samples were then incubated at 37°C for 45 min. The samples were electrophoresed on a denaturing 6% polyacrylamide–7 M urea gel and autoradiographed for variable time periods.

PCR amplifications. Genomic DNA was prepared as described previously (35), and samples of genomic DNA or subcloned genomic DNA fragments were amplified by the polymerase chain reaction (PCR) as described previously (30). For subcloning of PCR products, the primers were first

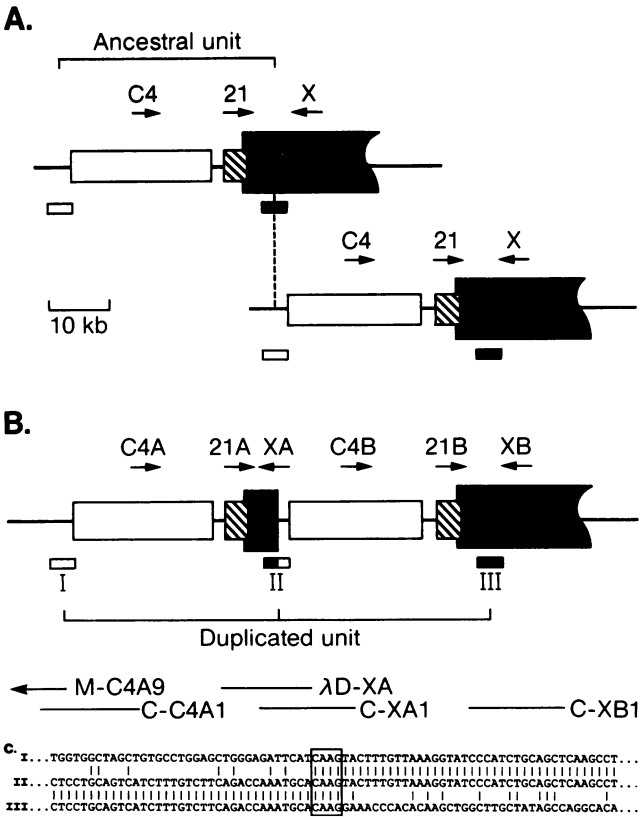

FIG. 1. (A) Model for duplication of the C4/P450c21/X locus, with unequal crossover between ancestral chromosomes. The C4 gene is represented by the open boxes, the P450c21 gene is represented by the hatched boxes, and the X gene is represented by the black boxes; the arrows denote transcriptional orientation. The smaller open and closed bars beneath the C4 and X genes indicate that the approximate sites for recombination contain no sequence homology. The dotted vertical line designates the site of nonhomologous recombination. (B) Current map of the C4/P450c21/X duplicated gene cluster. I is defined arbitrarily as the upstream boundary of the duplication, II is the junctional boundary, and III is the downstream boundary. A map of the genomic phage clones is shown under this diagram; M refers to clones obtained from the Maniatis library, C refers to those from the Clonetech library, and λD refers to a clone from the λ-Dash library from B. Chung. (C) Alignment of the sequences spanning each duplication boundary. I, II, and III indicate the boundaries identified above. More than 1 kb of sequence from I, II, and III was aligned, although only the relevant region is shown. Vertical lines between nucleotides indicate nucleotide identity. The boxed region around the four central nucleotides denotes the overlap in the duplication boundary that is found at all three boundary sites.

treated with T4 polynucleotide kinase (Boehringer Mannheim). Following amplification, the PCR products were purified on a Centricon-100 filter (Amicon, Beverly, Mass.), blunt ended with the Klenow fragment of DNA polymerase (Pharmacia), and subcloned into pBluescript vectors. For cDNA amplification, 1 μg of total RNA from fetal adrenal glands was resuspended in 20 μl of 10 mM Tris-HCl (pH 8.3)–50 mM KCl–5 mM $MgCl_2$–1 μM random hexamers–1 mM deoxyribonucleotide triphosphates–1 μl of RNase inhibitor (1 U/μl; Perkin Elmer/Cetus). The mixture was heated to 65°C for 2 min and cooled to room temperature, and 2.5 U of Moloney murine leukemia virus reverse transcriptase was

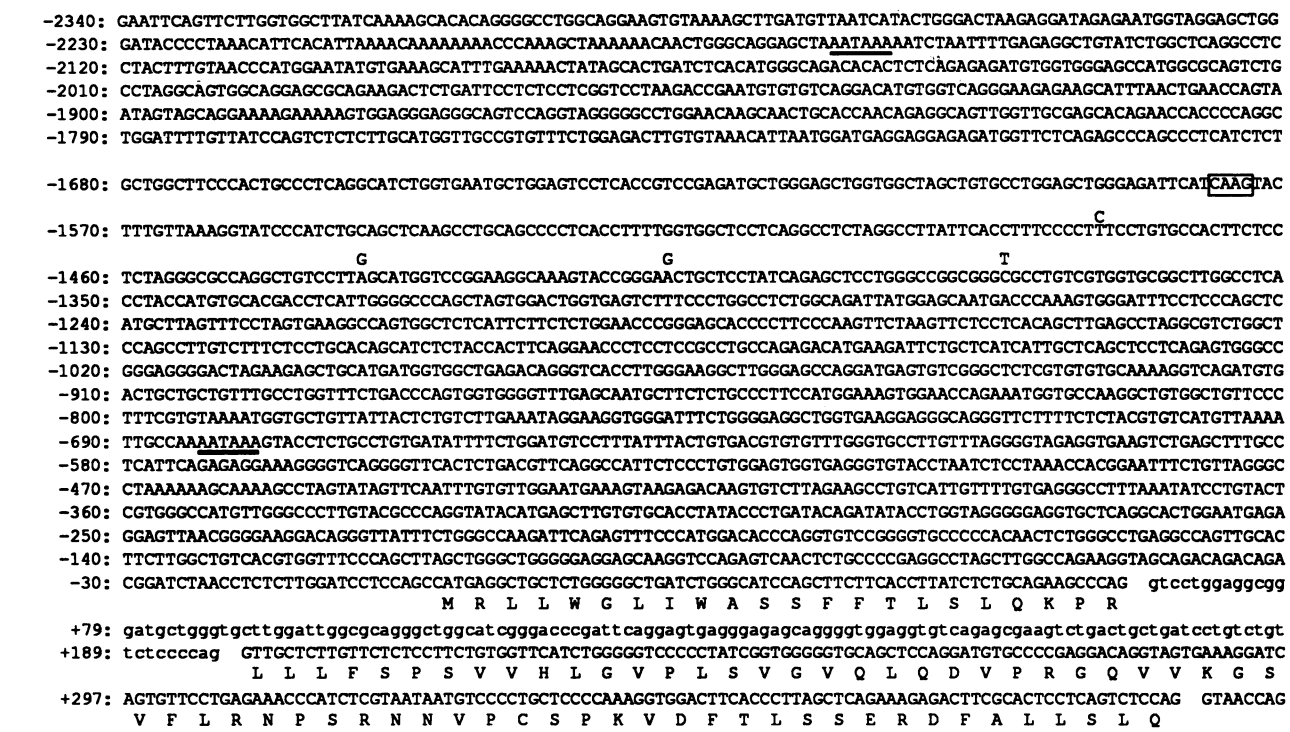

FIG. 2. 5′ flanking DNA and first two exons of the C4A gene. Nucleotides are numbered such that the first nucleotide of the translational initiation codon is +1. The nucleotide sequence from −1577 to +405 is identical in both the C4A and C4B sequences, with the exception of four nucleotide differences in C4B that are shown above the corresponding C4A nucleotides. The four boxed nucleotides constitute the upstream duplication boundary (I in Fig. 1). Amino acids are aligned below the first nucleotide of each codon. Introns are represented in lowercase letters. The potential polyadenylation signals discussed in the text are underlined.

added. After incubation for 10 min at 20°C and then at 42°C for 1 h, the reaction was terminated by heating to 95°C for 5 min. A 5-μl sample of this cDNA synthesis product was used for subsequent PCR amplification as described above.

RESULTS

Duplication of the C4/P450c21/X gene cluster. The arrangement of the duplicated C4/P450c21/X gene cluster indicates that an unequal crossover event occurred between ancestral chromosomes (Fig. 1). To determine whether this unequal crossover occurred at homologous or nonhomologous sequences, we obtained genomic phage clones spanning this locus (Fig. 1). The approximate limits of the C4/P450c21/X duplication were known from the restriction map of this region, suggesting a duplication of about 30 kb (14, 45). Two genomic libraries were screened with a 1.8-kb *Bam*HI-*Eco*RI fragment from the 5′ end of the X cDNA (34) that contains no overlapping P450c21 sequences and with a 476-bp *Bam*HI-*Kpn*I fragment from the 5′ end of the C4 cDNA (4). XA genomic clones hybridized to both probes, while XB clones hybridized only to the X probe. The XA clones also had the predicted restriction fragment length variations that differentiate this region from the XB gene (34, 45). The C4A phage clones were identified by hybridization only to the C4 cDNA probe or to a 1-kb *Apa*I fragment of genomic DNA that lies in the 5′ flanking DNA of both C4 genes.

To determine the precise duplication boundaries, we sequenced portions of each of the phage clones. For clarity, we refer to C4A as lying 5′ to XB. Region I, the upstream duplication boundary, is where the sequences of the 5′ flanking DNA of the C4A and C4B genes become dissimilar (Fig. 1B). Region II, the junctional boundary of the duplication, is where XA sequences are joined to C4B sequences. Region III, the downstream duplication boundary, is where the XB gene sequence diverges from that of the XA gene. Homology matrix analysis of 1 kb of DNA from regions I and III showed no regions of substantial similarity, suggesting that the duplication did not arise by homologous recombination at a repetitive element. Alignments of about 1 kb of sequence from all three regions revealed the precise site of a nonhomologous recombination (Fig. 1C). As predicted from the model in Fig. 1, the 5′ flanking DNAs of C4A and C4B (3′ halves of I and II) align exactly. Similarly, the corresponding portions of XA and XB (5′ halves of II and III) align exactly. There is a central block of four nucleotides that are identical in all three sequences, defining the duplication boundary. Similar short regions of sequence identity have been described with other nonhomologous recombinations (31, 47).

C4 genes. While characterizing regions I and II (Fig. 1), we determined the sequence of the 5′ flanking DNA and first two exons of the C4A and C4B genes. We found only four nucleotide differences between these two genes from nucleotides −1577 to +405, all of which were within the 5′ flanking DNA. As expected from the data in Fig. 1, there is no significant sequence homology between C4A and C4B upstream from the duplication boundary at −1574 to −1577. Figure 2 shows 2,340 bp of C4A 5′ flanking DNA, highlighting the CAAG duplication boundary. This also corrects the reported human C4 cDNA sequence (4) wherein the first two cDNA bases were not present in the genomic sequence, cDNA bases +7 to 78 (bases −49 to +22 in Fig. 2) were

inverted, and the cDNA had a 15-bp deletion between bases 78 and 79 (bases +23 to +37 in Fig. 2). This sequence places a Kozak translational initiation signal (26) at the start of the C4 protein-coding region, whereas the previously reported cDNA lacked such a sequence.

Sargent et al. (50) described a gene termed G11 that lies adjacent to the C4A gene, in the same transcriptional orientation. G11 appears to be a housekeeping gene of unknown function expressed in all cell lines examined; its RNA abundance was greatest in monocyte U937 cells. We found two potential polyadenylation signals in the 5′ flanking DNA of the C4A gene (Fig. 2); the one at −2158 to −2163 lies upstream from the duplication boundary (5′ of region I in Fig. 1), and the one at −678 to −683 lies downstream from the duplication boundary (3′ of region I in Fig. 1). To determine whether either of these sites corresponded to the 3′ end of the G11 gene, and to determine whether a portion of the G11 gene was duplicated with the C4/P450c21/X locus, we probed Northern (RNA) blots of U937 RNA with two genomic fragments, one lying upstream and one lying downstream from the duplication boundary. Neither of these probes hybridized to the reported 1.4-kb G11 mRNA; the upstream probe hybridized at approximately 10 kb, and the downstream probe hybridized at 4.8 kb (data not shown). The origins of these signals are not yet known. These data suggest that the 3′ end of G11 must lie further upstream and that G11 was not affected by the duplication of the C4/P450c21/X gene cluster.

Structure of the XA gene. To determine which X gene encoded the 2.7-kb X cDNA, we determined the genomic sequence from the 3′ end of the P450c21A pseudogene to the 5′ end of the C4B gene (Fig. 3). Comparing this genomic sequence with that of the 2.7-kb X cDNA (34) reveals 13 apparent exons flanked by canonical intron/exon splice sites (Fig. 3). However, the 5′ end of the XA gene is truncated by the duplication of the C4B/P450c21B/XB locus, so that there are no XA gene sequences corresponding to the first 332 bp of the 2.7-kb cDNA fragment. Furthermore, the middle of the cDNA contains 91 bp not found in exon 4 of the XA gene. As a result of this deletion in the XA gene, the predicted XA open reading frame terminates 30 bp downstream in exon 5, rather than maintaining the long continuous open reading frame found in the 2.7-kb X cDNA.

Because this deletion within the XA gene closes the reading frame, we carefully ruled out the possibility that it was an artifact. First, we isolated an additional XA genomic clone from an independent library; sequencing showed that this clone contained an identical deletion. Second, we used PCR to amplify this region of the XA and XB genes simultaneously from genomic DNA (Fig. 4). Amplification of genomic DNA revealed two bands of approximately equal intensity (Fig. 4A, lane 4). The exonic sequence of the larger band (Fig. 4B) corresponds to that predicted from our 2.7-kb X cDNA sequence; there are two exons that are separated by a 182-bp intron and that are flanked by canonical splice junction sequences. Therefore, this larger fragment corresponds to the XB gene. The sequence of the smaller fragment revealed a 121-bp deletion, encompassing the 91 bp from the upstream exon and the first 30 bp of the adjacent intron; thus, this smaller fragment derived from the XA gene. Both the truncation at its 5′ end and the 121-bp deletion show that the XA gene cannot encode the 2.7-kb X cDNA (which we now term the 2.7-kb XB cDNA). Furthermore, these structural changes suggested that XA, like the unexpressed P450c21A gene that it overlaps, should be a pseudogene.

The XA gene is expressed. To determine whether the XA gene is expressed, we performed RNA-based PCR with the same set of primers used for Fig. 4 that distinguishes the XA and XB genes. Using human fetal adrenal RNA, the tissue from which the 2.7-kb XB cDNA was cloned (34), two PCR products of 269 and 178 bp result; sequence analysis confirmed that the two bands differ by the expected 91 bp (Fig. 4A, lanes 5 to 7). Thus, the XA gene is transcribed, and the XA RNA is stable.

To obtain an XA cDNA, we screened 10^6 clones from an Okayama-Berg fetal adrenal cDNA library (60) with the 1.8-kb *Bam*HI-*Eco*RI XB cDNA fragment, identifying eight hybridizing clones. We used the same PCR tactic shown in Fig. 4 to characterize each cDNA, identifying one clone as an XA cDNA. The complete nucleotide sequence of this 2,450-bp cDNA corresponded with the XA gene sequence (Fig. 3), with all of the splice junctions shown and a poly(A) tail. This sequence contained the expected 91-bp deletion of exon 4, and the 5′ end contained 216 bp found in the XA gene but not in the XB cDNA. Thus, this 2.45-kb cDNA derived from the XA gene.

A translational initiation site is not readily apparent in the XA RNA. There are six potential methionine initiation codons in the first 600 nucleotides (at nucleotides 2729, 2738, 2835, 2890, 2940, and 2949 in Fig. 3), of which only those at 2729 and 2890 match the Kozak consensus sequence well (A/GCCATGG) (26). However, these sites initiate only short open reading frames that are different from the reading frame of the 2.7-kb XB cDNA. A more likely translational start site might be the leucine codon at nucleotides 2713 to 2715; this sequence, AGCCTGG, closely resembles the canonical consensus sequence; such leucine translational initiation sites have been described for several genes (1, 5, 17, 19, 44). There are three in-frame stop codons upstream from this leucine codon, and its downstream reading frame corresponds to the open reading frame of the XB cDNA. This XA reading frame remains open for 933 bases but terminates 30 bases downstream from the 91-bp deletion in exon 5 (Fig. 3). Thus, the deletion and subsequent frameshift eliminates over half of the open reading frame found in the 2.7-kb XB cDNA. The resulting 3′ downstream region is 1,344 bp long, is encoded by eight spliced exons, and is followed by a polyadenylation signal (at nucleotides 6808 to 6813 in Fig. 3) and a poly(A) tail.

However, if the XA cDNA had been modified with a single-base deletion or two-base addition near the 91-bp deletion, then the downstream sequence would retain the same long open reading frame found in the 2.7-kb XB cDNA. To rule out a sequencing error, an artifact in cDNA synthesis, or a posttranscriptional mRNA modification (e.g., RNA editing) that might restore the open reading frame, we examined additional XA cDNA clones. First, we used the tactic shown in Fig. 4 to examine the 36 X cDNA phage clones identified during our initial screening of a fetal adrenal cDNA library (34). Of these, 30 were XA cDNAs, but none was longer than 2.2 kb and hence they were not full length. Five unique clones were characterized further by cloning and sequencing the PCR products encompassing the deletion site; the corresponding regions of all five clones contained precisely the same nucleotide sequence, with the 91-bp deletion and frameshift as identified previously. Second, we performed three independent rounds of RNA-based PCR from fetal adrenal RNA with use of these same primers and found identical sequences in the subcloned products from each XA cDNA. Thus, the frameshift and subsequent down-

A
C4B
XA
P450
C21A

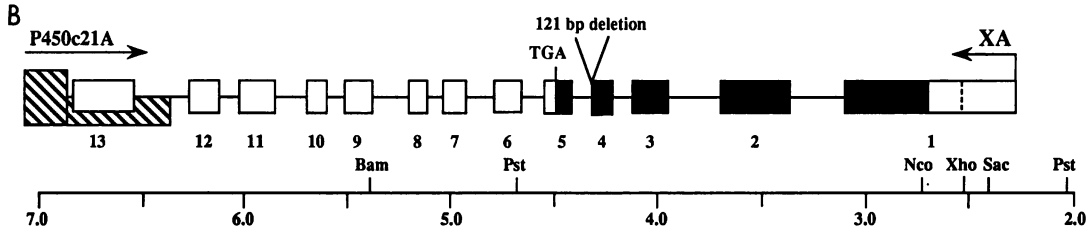

FIG. 3. Structure of the XA gene. (A) Nucleotide sequence of the XA gene, extending from the 5′ end of the C4B gene to the 3′ end of the P450c21A gene. The sequence is in the sense orientation with respect to XA, and thus the translated amino acid sequences for C4B and P450c21A arise from the opposite strand of DNA as represented by the inverted amino acid letter code for these two genes. The four boxed nucleotides at 1982 to 1985 are the junctional duplication boundary (II in Fig. 1); note that this sequence is antisense to that presented in Fig. 2. The most commonly used XA cap sites are denoted with dots over the corresponding bases; the predominant cap site for the XA gene is marked by the arrow preceding nucleotide 2232. The single-letter amino acids code is aligned below the first nucleotide of each codon. The arrowhead following nucleotide 4267 denotes the site of the 121-bp deletion. XA introns are denoted by lowercase letters. The polyadenylation signal of the P450c21A pseudogene is lightly underlined at bases 6363 to 6368. The polyadenylation signal for XA is boldly underlined in the last line of sequence; the space 21 bases downstream indicates the poly(A) addition site for the XA gene. (B) Map of the XA gene, with transcriptional orientation from right to left. Black boxes indicate coding regions, and open boxes represent 5′ and 3′ untranslated regions. The dotted line in exon 1 designates the 5′ end of the XA cDNA. The site of the 121-bp deletion is marked, as is the stop codon. The hatched box represents the last exon of the P450c21A gene. The base numbers in the scale bar shown below the diagram correspond to the base numbers in panel A. Some restriction sites used in making various probes are shown: B, *Bam*HI; P, *Pst*I; N, *Nco*I; X, *Xho*I; S, *Sac*I.

stream termination signal are bona fide features of XA transcripts.

Identification of the XA cap site. Northern blots of fetal adrenal RNA probed with the 1.8-kb *Eco*RI-*Bam*HI XB cDNA fragment hybridizing to XA and XB (but lacking P450c21 sequences) revealed bands at 12, 8, 3.5, 2.6, and 1.8 kb. The origin of these multiple mRNA species was unclear. The XA and XB genes each contribute at least one mRNA species, and additional bands may represent alternate splicing, exon skipping, alternate polyadenylation signals, the use of alternate promoters, or even additional overlapping genes in this region. To identify the length of the XA transcript(s), we probed a series of Northern blots with genomic fragments extending from the 3′ end of P450c21A to the 5′ end of P450c21B. A 711-bp *Pst*I-*Bam*HI probe that contains exons common to both the XA and XB genes but lacks P450c21

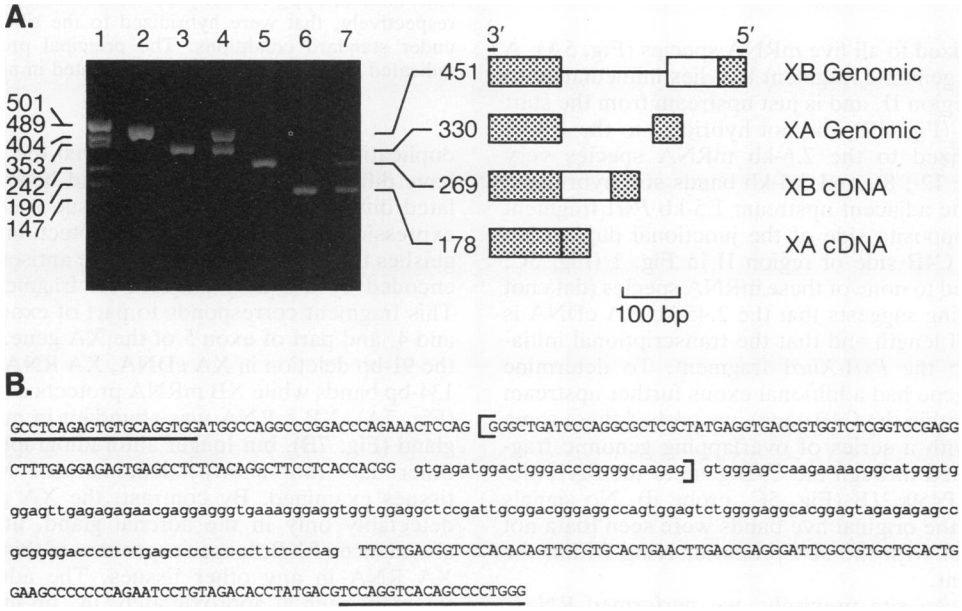

FIG. 4. PCR amplification of the XA and XB regions containing the 121-bp deletion. The 5′ sense PCR primer hybridized 49 bp upstream from the deletion and within the same exon, and the 3′ antisense PCR primer hybridized to the 3′ end of the next exon. (A) 2.5% agarose gel of PCR amplifications of genomic DNA and cDNA. The gel was stained with ethidium bromide and photographed under UV light. Lane 1 contains molecular size markers of pUC18 digested with *Hpa*II. Lanes 2 to 4 are amplifications performed with genomic DNA; lane 2 is from the XB genomic phage, lane 3 is from the XA genomic phage, and lane 4 is from genomic DNA of a normal individual. Lanes 5 to 7 are amplifications performed with cDNA; lane 5 is from the 2.7-kb XB cDNA (34), lane 6 is from the XA cDNA clone, and lane 7 is from fetal adrenal RNA that was reverse transcribed into cDNA. The diagrams at the right illustrate the structures of the amplified products. Exons common to XA and XB are represented by stippled boxes, the open box shows exonic sequence present only in the XB gene, and the thin lines denote introns. The predicted size (in base pairs) of each fragment is shown. (B) Nucleotide sequence from the XB genomic amplification. The primers are underlined, and the intron is represented in lowercase letters. The extent of the XA deletion is denoted by brackets.

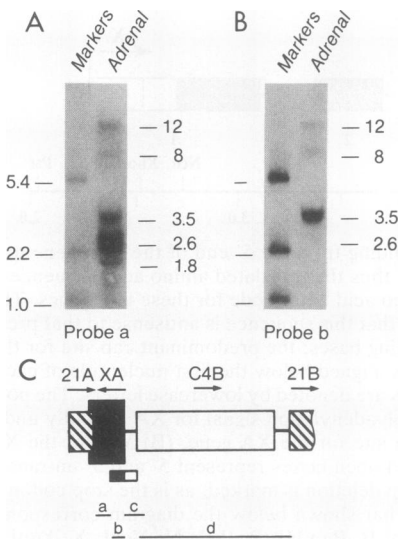

FIG. 5. Northern blot analysis. (A) Total RNA from a fetal adrenal gland probed with a 711-bp *Pst*I-*Bam*HI genomic probe near the 3′ end of the XA gene and 929 bp 3′ of the P450c21A gene (probe a; nucleotides 4678 to 5389 in Fig. 3). Approximate sizes (in kilobases) of the hybridizing bands are indicated. (B) A duplicate blot probed with a 483-bp *Pst*I-*Xho*I XA genomic fragment (probe b; nucleotides 2073 to 2556 in Fig. 3), localized just on the C4A side of the junctional boundary (region II in Fig. 1) and just upstream from the start of the XA cDNA. Markers are bacteriophage PM2 digested with *Hin*dIII. (C) Map of the probes used in this experiment. Probes c and d did not hybridize to any of the mRNA species detected with probe a.

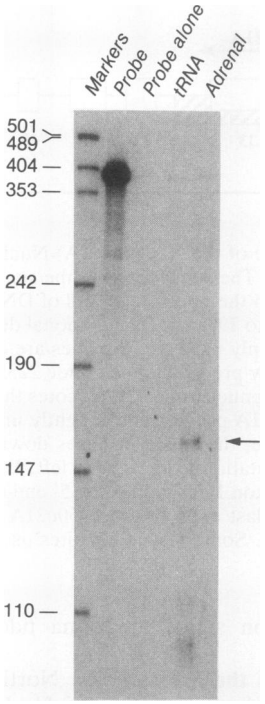

FIG. 6. RNase protection assay mapping the cap site of the XA gene. The riboprobe includes a 320-bp *Pst*I-*Sac*I genomic fragment plus 48 bp of additional vector sequence. Markers are pUC18 digested with *Hpa*II. Lanes: Probe, undigested riboprobe alone; Probe alone, riboprobe that had not been hybridized with any RNA and was then subjected to the standard digestion conditions; tRNA and Adrenal, 10 μg each of yeast tRNA and total fetal adrenal RNA, respectively, that were hybridized to the riboprobe and digested under standard conditions. The principal protected fragment is indicated by an arrow. Sizes are indicated in nucleotides.

sequences hybridized to all five mRNA species (Fig. 5A). A 483-bp *Pst*I-*Xho*I genomic fragment that lies immediately to the C4A side of region II, and is just upstream from the start of the XA cDNA (Fig. 3B), did not hybridize to the 1.8-kb band and hybridized to the 2.6-kb mRNA species very weakly, while the 12-, 8-, and 3.5-kb bands still hybridized well (Fig. 5B). The adjacent upstream 1.5-kb *Pst*I fragment that lies on the opposite side of the junctional duplication boundary on the C4B side of region II in Fig. 1 (Fig. 5C, probe c) hybridized to none of these mRNA species (data not shown). This finding suggests that the 2.45-kb XA cDNA is approximately full length and that the transcriptional initiation site maps to the *Pst*I-*Xho*I fragment. To determine whether the XA gene had additional exons further upstream (perhaps buried within the C4B gene), we probed these same Northern blots with a series of overlapping genomic fragments that extended through the C4B gene to the *Bgl*II site at the 5′ end of P450c21B (Fig. 5C, probe d). No signals corresponding to the original five bands were seen (data not shown), consistent with the XA promoter lying within the *Pst*I-*Xho*I fragment.

To locate the cap site precisely, we performed RNase protection assays. A series of overlapping probes extending from the *Nco*I site at 2736 to the *Sac*I site at 2392 (Fig. 3B) were fully protected by fetal adrenal RNA (data not shown). However, the upstream *Pst*I-*Sac*I riboprobe (nucleotides 2073 to 2392) yielded a protected fragment of 161 bp and minor species at 106 to 107 and 103 to 104 bp (Fig. 6). The exact cap sites were determined from sequencing ladders run in parallel with the protected fragments; these sites are marked in Fig. 3A.

Expression of XA and XB genes. One consequence of the duplication of this gene cluster is that the XA and XB genes have different 5′ flanking DNA and hence should be regulated differently. To study the tissue specificity of gene X expression, we used an RNase protection assay that distinguishes the XA and XB RNAs. The antisense riboprobe was encoded by a 342-bp *Hin*cII-*Pst*I fragment of XB cDNA. This fragment corresponds to part of exon 2, all of exons 3 and 4, and part of exon 5 of the XA gene. Because it spans the 91-bp deletion in XA cDNA, XA RNA protects 117- and 134-bp bands while XB mRNA protects a single 342-bp band (Fig. 7A). XB mRNA was abundant in muscle and adrenal gland (Fig. 7B), but longer autoradiographic exposures and other data (not shown) demonstrated XB mRNA in all tissues examined. By contrast, the XA gene is expressed detectably only in the adrenal gland, at about twice the abundance of XB; longer exposures of this gel did not reveal XA RNA in any other tissues. The additional protected fragment seen at approximately 147 bp in the adrenal samples disappears entirely at more stringent RNase digestion conditions, but the specific XA bands become somewhat broader and less distinct (data not shown). This band does not represent an alternately spliced mRNA, as such an RNA would have also been detected in the RNA-based PCR experiment shown in Fig. 4. Other RNase protection assays demonstrate that XA RNA is also found in adult adrenal mRNA at about twice the abundance of XB mRNA (data not shown).

Function of XA RNA? The XA RNA may function as an

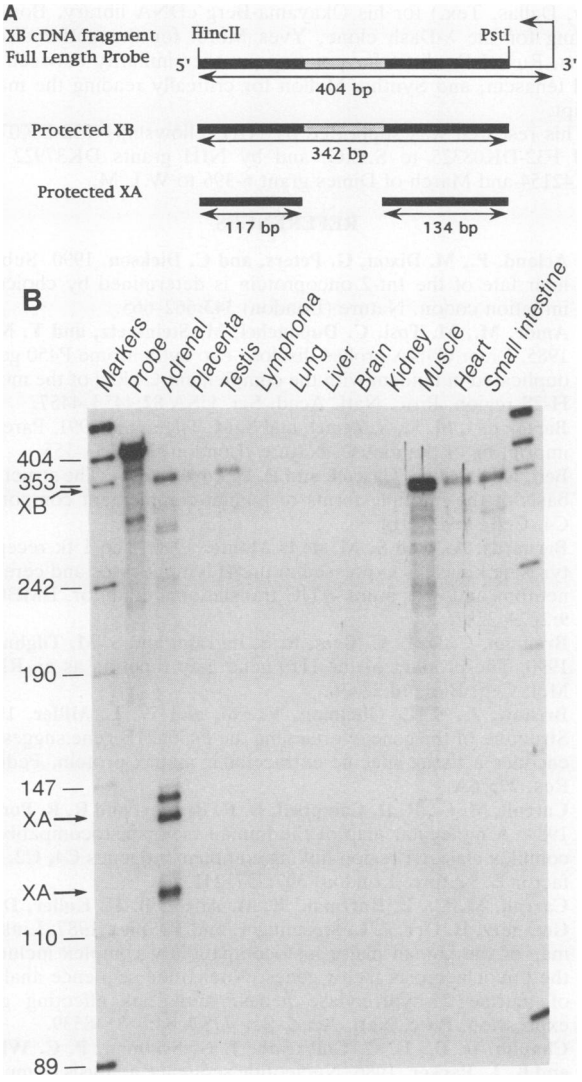

FIG. 7. Expression of XA and XB genes as detected by a solution hybridization-RNase protection assay. (A) Probes. The antisense riboprobe is a 342-bp *Hin*cII-*Pst*I XB cDNA fragment that spans the deletion site present in the XA cDNA and three intron/exon boundaries in either XA or XB. The full-length probe contains an additional 82 bp of vector sequence. The XB mRNA protects a single fragment, while the XA mRNA, lacking the central 91 bp (gray shading), protects two smaller fragments at 117 and 134 bp. (B) RNase protection assay performed with a variety of tissues. The lanes on the extreme right and left contain markers of pUC18 digested with *Hpa*II. The other lanes contain fetal tissues of approximately 20 weeks' gestation, with the exception of the placenta (obtained at term) and lymphoma samples. Simultaneous experiments with probe alone and probe hybridized with tRNA contained no signal following RNase digestion (data not shown).

mRNA, or it may have another untranslated role, e.g., in nucleoprotein formation. The high abundance and tissue specificity of its expression suggest that it is not a gratuitous transcript. If XA encodes a protein, it would be 310 amino acids long and encode 2.5 fibronectin type III repeats, each of approximately 100 amino acids. Searches of protein data bases (59) show that the putative XA protein is similar to human fibronectin (25) and chicken tenascin (22); counting both conserved and identical amino acids, XA has 30%

```
FnIII      NVSPPRRARVTDATETTITISWRTKTETITGFQVDAVPANGQTPIQRTIKPDVRSYTIT
            *  ||     *  ||   * |          | |    |*|  |    *  * |   *
XA         VLESPRDLQFSEIRETSAKVNWMPPPSRADSFKVSYQLADGGEPQSVQVDGQARTQKLQ
           |* ||* * ||*| | ||*| | || || ||**|||    || |  | |||     ||
Tenascin   VVGSPKGISFSDITENSARVSWTPPRSRVDSYRVSYVPITGGTPNVVTVDGSKTRTKLV
```

FIG. 8. Amino acid alignment of a portion of the deduced XA protein (amino acids 243 to 302 in Fig. 3; identical to amino acids 331 to 389 of the partial XB sequence in reference 34) with fragments of human fibronectin type III (FnIII; amino acids 1783 to 1841) (25) and chicken tenascin (amino acids 1285 to 1346) (22). Vertical lines between amino acids indicate identity between proteins, and asterisks denote amino acid homology. The conserved tryptophan residue found in fibronectin type III repeats is underlined.

homology to fibronectin and 60% homology to tenascin in one 59-residue region (Fig. 8). However, unlike fibronectin and tenascin, the deduced XA protein lacks the RGD sequence implicated in binding to the integrin receptor (reviewed in reference 49).

DISCUSSION

The C4/P450c21/X locus was duplicated by a nonhomologous recombination; such events can occur with chromosomal translocations, duplications, deletions, and viral integrations and excisions and with immunoglobulin and T-cell receptor rearrangements (reviewed in references 31 and 47). Although the mechanisms for nonhomologous recombination remain unknown, over half of such recombinations have a very short region of homology at the crossover site (24). Experimental recombination of simian virus 40 digested with two different restriction endonucleases and transfected into CV1 cells occurs at short identical sequences rather than at noncomplementary restriction sites, suggesting a primary role for such regions of patchy homology (48). The nature of the enzymes mediating ligation of DNA at such sites is unknown, although such an activity has been described in *Xenopus laevis* (41) and in HeLa cells (47).

The sequences of the 3′ untranslated and 3′ flanking sequences of murine, human, and bovine P450c21 suggest that all three species have X genes (34), but only human X has been characterized to date. The duplication of the human and murine C4/P450c21/X loci probably occurred independently. Studies with bovine (12, 13, 61), murine (2, 38), and human (8, 21, 46, 58) DNAs show that the C4/P450c21 locus was duplicated in three different mammalian families (ungulates, rodents, and primates), suggesting that the locus duplicated before mammalian speciation. However, the upstream duplication boundary of the murine *Slp* and C4 genes (38) is about 400 bp further upstream than the corresponding human boundary (region I in Fig. 1) and has no sequence similarity with that surrounding the human boundary. This suggests that the human and murine duplications arose independently, postdating mammalian speciation. Independent duplication of the C4/P450c21/X locus in various mammalian genomes might explain the differences in the spacing between C4 and P450c21 in the murine, bovine, and human genomes and the apparent presence of single C4/P450c21/X clusters in some other incompetely studied mammalian species (18, 29, 55). However, the data cannot distinguish postspeciation duplications in some mammalian lineages from a prespeciation duplication followed by postspeciation genetic losses in some lineages. Because of the very high frequency of genetic recombination that occurs in this region, both models are possible.

Although the structure of the XA gene suggests that it is a pseudogene, three lines of evidence show that the XA gene

is expressed and gives rise to a stable RNA. First, XA RNA is readily amplified by RNA-based PCR, yielding a product with greater abundance than XB. Second, we have cloned and sequenced six different XA cDNAs. Third, RNase protection experiments show that XA is expressed in the adult and fetal adrenal glands more abundantly than is XB. However, presence of XA RNA only suggests, but does not demonstrate, that this RNA has a function. One possible function for XA RNA would be to encode protein; however, if XA RNA is an mRNA, then it is most unusual. First, it appears to use CUG rather than AUG as its translational initiation codon; such CUG codons do initiate translation of basic fibroblast growth factor (17, 44), *int-2* (1), c-*myc* (19), and the *ltk* receptor (5). Second, its 3′ untranslated region would contain seven introns. Editing of the mRNA for human apolipoprotein B-100 creates the mRNA for B-48, which then has spliced exons in its 3′ untranslated region (43); however, other such 3′ untranslated regions have not been described. We carefully characterized several XA cDNAs to rule out such RNA editing (23), errors in reverse transcription, or errors in sequencing. It is also conceivable that this spliced 3′ untranslated region could constitute another translation unit, as it contains a second open reading frame of 1,470 bp, and ribosomes can initiate from a second downstream AUG if it occurs 3′ of the termination codon for the first translated product (reviewed in reference 27). Alternatively, XA RNA could function without translation. Traditional untranslated RNAs such as ribosomal, transfer, and small nuclear RNAs transcribed by RNA polymerases I and III are not spliced and polyadenylated, as are polymerase II mRNA transcripts. However, murine H19 is an example of a functional, processed, polyadenylated, and untranslated polymerase II transcript (3, 6).

The duplication of this gene cluster has placed different DNA sequences in the 5′ flanking regions of the XA and XB genes. Because transcriptional regulatory elements are usually located in 5′ flanking DNA, it is not surprising that the XA and XB genes exhibit differences in tissue-specific expression; the XB gene enjoys widespread expression, whereas expression of the XA gene appears limited to the adrenal gland. It is not clear whether the adrenal gland-specific expression of XA is related to its proximity to the P450c21A pseudogene or the more distant adrenal gland-specific P450c21B gene.

The duplication of this gene locus has also affected the function of the X genes. The XB gene appears to encode an extracellular matrix protein: its carboxy-terminal region resembles fibrinogen, and the upstream portion encodes several repeating peptide units homologous to fibronectin type III repeats (7, 59); this arrangement of domains closely resembles that of tenascin (reviewed in reference 16). By contrast, if XA RNA is translated, the predicted XA protein would lack the fibrinogenlike domain and consist of only 2.5 fibronectin type III repeats. While many known proteins contain fibronectin type III domains, none consists solely of such domains. Type III repeats often have binding sites for heparin and other proteoglycans and may contain RGD or other sequences mediating binding to integrin receptors (49). Thus, if synthesized and secreted, the XA protein might compete with other extracellular matrix proteins for proteoglycans or for cell membrane receptors.

ACKNOWLEDGMENTS

We thank Michael Carroll (Harvard University, Boston, Mass.) for providing the C4 probe, David Russell (Southwestern University, Dallas, Tex.) for his Okayama-Berg cDNA library, Bon-chu Chung for the λ-Dash clone, Yves Morel for thoughtful discussions, Russel Doolittle for pointing out the similarity between XB and tenascin, and Synthia Mellon for critically reading the manuscript.

This research was supported by NIH fellowships T32-DK07161 and F32-DK08325 to S.E.G. and by NIH grants DK37922 and DK42154 and March of Dimes grant 6-396 to W.L.M.

REFERENCES

1. **Acland, P., M. Dixon, G. Peters, and C. Dickson.** 1990. Subcellular fate of the Int-2 oncoprotein is determined by choice of initiation codon. Nature (London) **343**:662–665.
2. **Amor, M., M. Tosi, C. Duponchel, M. Steinmetz, and T. Meo.** 1985. Liver cDNA probes disclose two cytochrome P450 genes duplicated in tandem with the complement C4 loci of the mouse H-2S region. Proc. Natl. Acad. Sci. USA **82**:4453–4457.
3. **Bartolomei, M. S., S. Zemel, and S. M. Tilghman.** 1991. Parental imprinting of the mouse. Nature (London) **351**:153–155.
4. **Belt, K. T., M. C. Carroll, and R. R. Porter.** 1984. The structural basis of the multiple forms of human complement component C4. Cell **36**:907–914.
5. **Bernards, A., and S. M. de la Monte.** 1990. The 1 tk receptor tyrosine kinase is expressed in pre-B lymphocytes and cerebral neurons and uses a non-AUG translational inhibitor. EMBO J. **9**:2279–2287.
6. **Brannan, C. I., E. C. Dees, R. S. Ingram, and S. M. Tilghman.** 1990. The product of the H19 gene may function as an RNA. Mol. Cell. Biol. **10**:28–36.
7. **Bristow, J., S. E. Gitelman, Y. Shi, and W. L. Miller.** 1990. Structure of the gene overlapping the P450c21B gene suggests it encodes a tissue-specific extracellular matrix protein. Pediatr. Res. **27**:76A.
8. **Carroll, M. C., R. D. Campbell, D. R. Bentley, and R. R. Porter.** 1984. A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2, and factor B. Nature (London) **307**:237–241.
9. **Carroll, M. C., P. Katzman, E. M. Alicot, B. H. Koller, D. E. Geraghty, H. Orr, J. L. Strominger, and T. Spies.** 1987. Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. Nucleotide sequence analysis of murine 21-hydroxylase genes: mutations affecting gene expression. Proc. Natl. Acad. Sci. USA **84**:8535–8539.
10. **Chaplin, D. D., L. G. Galbreath, J. G. Seidman, P. C. White, and K. L. Parker.** 1986. Nucleotide sequence analysis of murine 21-hydroxylase genes: mutations affecting gene expression. Proc. Natl. Acad. Sci. USA **83**:9601–9605.
11. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
12. **Chung, B., K. J. Matteson, and W. L. Miller.** 1985. Cloning and characterization of the bovine gene for steroid 21-hydroxylase. DNA **4**:211–219.
13. **Chung, B., K. J. Matteson, and W. L. Miller.** 1986. Structure of a bovine gene for P450c21 (steroid 21-hydroxylase) defines a novel cytochrome P450 gene family. Proc. Natl. Acad. Sci. USA **83**:4243–4247.
14. **Collier, S., P. J. Sinnott, P. A. Dyer, D. A. Price, R. Harris, and T. Strachan.** 1989. Pulse field gel electrophoresis identifies a high degree of variability in the number of tandem 21-hydroxylase and complement C4 gene repeats in 21-hydroxylase deficiency haplotypes. EMBO J. **8**:1393–1402.
15. **Dunham, I., C. A. Sargent, J. Trowsdale, and R. D. Campbell.** 1987. Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. Proc. Natl. Acad. Sci. USA **84**:7237–7241.
16. **Erickson, H. P., and M. A. Bourdon.** 1989. Tenascin: an extracellular matrix protein prominent in specialized embryonic tissues and tumors. Annu. Rev. Cell Biol. **5**:71–92.
17. **Florkiewicz, R. Z., and A. Sommer.** 1989. Human basic fibroblast growth factor gene encodes four polypeptides: three initiate translation from non-AUG codons. Proc. Natl. Acad. Sci. USA **86**:3978–3981.

18. **Geffrotin, C., P. Chardon, D. F. DeAndres-Cara, R. Feil, C. Renard, and M. Vaiman.** 1990. The swine steroid 21-hydroxylase gene (CYP21): cloning and mapping within the swine leukocyte antigen locus. Anim. Genet. **21:**1–13.
19. **Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman.** 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell **52:**185–195.
20. **Hauptmann, G., G. Tappeiner, and J. A. Schifferli.** 1988. Inherited deficiency of the fourth component of human complement. Immunodefic. Rev. **1:**3–22.
21. **Higashi, Y., H. Yoshioka, M. Yamane, O. Gotoh, and Y. Fujii-Kuriyama.** 1986. Human chromosome 6: a pseudogene and a genuine gene. Proc. Natl. Acad. Sci. USA **83:**2841–2845.
22. **Jones, F. S., S. Hoffman, B. A. Cunningham, and G. M. Edelman.** 1989. A detailed structural model of cytotactin: protein homologies, alternative RNA splicing, and binding regions. Proc. Natl. Acad. Sci. USA **86:**1905–1909.
23. **Kimelman, D., and M. W. Kirschner.** 1989. An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. Cell **59:**687–696.
24. **Konopka, A. K.** 1988. Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. Nucleic Acids Res. **16:**1739–1758.
25. **Kornblihtt, A. R., K. Umezawa, K. Vibe-Pedersen, and F. E. Baralle.** 1985. Primary structure of human fibronectin: differential splicing may generate at least 10 polypeptides from a single gene. EMBO J. **4:**1755–1759.
26. **Kozak, M.** 1984. Point mutations that define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44:**283–292.
27. **Kozak, M.** 1989. The scanning model for translation: an update. J. Cell Biol. **108:**229–241.
28. **Levi-Strauss, M., M. C. Carroll, M. Steinmetz, and T. Meo.** 1988. A previously undetected MHC gene with an unusual periodic structure. Science **240:**201–204.
29. **Levi-Strauss, M., M. Tosi, M. Steinmetz, J. Klein, and T. Meo.** 1985. Multiple duplications of complement C4 gene correlate with H-2 controlled testosterone-independent expression of its sex-limited isoform, C4-Slp. Proc. Natl. Acad. Sci. USA **82:**1746–1750.
30. **Lin, D., S. E. Gitelman, P. Saenger, and W. L. Miller.** 1991. Normal genes for the cholesterol side-chain cleavage enzyme, P450scc, in congenital lipoid adrenal hyperplasia. J. Clin. Invest. **88:**1955–1962.
31. **Meuth, M.** 1989. Illegitimate recombination in mammalian cells, p. 833–859. *In* D. E. Berg and M. H. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
32. **Miller, W. L., and Y. Morel.** 1989. The molecular genetics of 21-hydroxylase deficiency. Annu. Rev. Genet. **23:**371–393.
33. **Moore, C. C. D., S. T. Brentano, and W. L. Miller.** 1990. Human P450scc gene is induced by cyclic AMP and repressed by TPA/A23187 by independent *cis* elements. Mol. Cell. Biol. **10:**6013–6023.
34. **Morel, Y., J. Bristow, S. E. Gitelman, and W. L. Miller.** 1989. Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus. Proc. Natl. Acad. Sci. USA **86:**6582–6586.
35. **Morel, Y., J. Andre, B. Uring-Lambert, G. Hauptmann, H. Beteuel, M. Tosi, M. Forest, M. David, J. Bertrand, and W. L. Miller.** 1989. Rearrangements and point mutations of P450c21 genes are distinguished by five restriction endonuclease haplotypes identified by a new probing strategy in 57 families with congenital adrenal hyperplasia. J. Clin. Invest. **83:**527–536.
36. **Morel, Y., and W. L. Miller.** 1991. Clinical and molecular genetics of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Adv. Hum. Genet. **20:**1–68.
37. **Nebert, D., D. Nelson, M. Coon, R. Estabrook, R. Feyereisenn, Y. Fujii-Kuriyama, F. Gonzalez, F. Guengerich, I. Gunsalus, E. Johnson, J. Loper, R. Sato, M. Waterman, and D. Waxman.** 1991. The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. DNA Cell Biol. **10:**1–14.
38. **Nonaka, M., H. Kimura, Y. D. Yeul, S. Yokoyama, K. Nakayama, and M. Takahashi.** 1986. Identification of the 5′ flanking regulatory region responsible for the difference in transcriptional control between mouse complement C4 and Slp genes. Proc. Natl. Acad. Sci. USA **83:**7883–7887.
39. **Ogata, R. T., and D. S. Sepich.** 1985. Murine sex-limited protein: complete cDNA sequence and comparison with murine fourth complement component. J. Immunol. **135:**4239–4244.
40. **Parker, K. L., D. D. Chaplin, M. Wong, J. G. Seidman, J. A. Smith, and B. P. Schimmer.** 1985. Expression of murine 21-hydroxylase in mouse adrenal glands and in transfected Y1 adrenocortical tumor cells. Proc. Natl. Acad. Sci. USA **82:**7860–7864.
41. **Pfeiffer, P., and W. Vielmetter.** 1988. Joining of nonhomologous DNA double strand breaks in vitro. Nucleic Acids Res. **16:**907–924.
42. **Picado-Leonard, J., and W. L. Miller.** 1987. Cloning and sequence of the human gene for P450c17 (steroid 17-hydroxylase/17,20 lyase): similarity with the gene for P450c21. DNA **6:**439–448.
43. **Powell, L. M., S. C. Wallis, R. J. Pease, Y. H. Edwards, T. J. Knott, and J. Scott.** 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell **50:**831–840.
44. **Prats, H., M. Kaghad, A. Prats, M. Klagsbrun, J. Lelias, P. Liauzun, P. Chalon, J. P. Tauber, F. Amalric, J. A. Smith, and D. Caput.** 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. Proc. Natl. Acad. Sci. USA **86:**1836–1840.
45. **Prentice, H. L., P. M. Schneider, and J. L. Strominger.** 1986. C4B gene polymorphism detected in a human cosmid clone. Immunogenetics **23:**274–276.
46. **Rodrigues, N. R., I. Dunham, C. Y. Yu, M. C. Carroll, R. R. Porter, and R. D. Campbell.** 1987. Molecular characterization of the HLA-linked steroid 21-hydroxylase B-gene from an individual with congenital adrenal hyperplasia. EMBO J. **6:**1653–1661.
47. **Roth, D. B., and J. H. Wilson.** 1986. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol. Cell. Biol. **6:**4295–4304.
48. **Roth, D., and J. Wilson.** 1988. Illegitimate recombination in mammalian cells, p. 621–653. *In* R. Kucherlapati and R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
49. **Ruoslahti, E.** 1988. Fibronectin and its receptors. Annu. Rev. Biochem. **57:**375–413.
50. **Sargent, C. A., I. Dunham, and R. D. Campbell.** 1989. Identification of multiple HTF-island associated genes in the human major histocompatibility complex class III region. EMBO J. **8:**2305–2312.
51. **Sargent, C. A., I. Dunham, J. Trowsdale, and R. D. Campbell.** 1989. Human major histocompatibility complex contains genes for the major heat shock protein HSP 70. Proc. Natl. Acad. Sci. USA **86:**1968–1972.
52. **Skow, L. E., J. E. Womack, J. M. Petresh, and W. L. Miller.** 1988. Synteny mapping of the genes for steroid 21-hydroxylase, alpha-A-crystallin, and class I bovine leukocyte (BoLA) in cattle. DNA **7:**143–149.
53. **Spies, T., G. Blanck, M. Bresnahan, J. Sands, and J. L. Strominger.** 1989. A new cluster of genes within the human major histocompatibility complex. Science **243:**214–217.
54. **Spies, T., C. C. Morton, S. A. Nedospasov, W. Fiers, D. Pious, and J. Strominger.** 1986. Genes for the tumor necrosis factors α and β are linked to the human major histocompatibility complex. Proc. Natl. Acad. Sci. USA **83:**8699–8702.
55. **Spilliaert, R., A. Palsdottir, and A. Arnason.** 1990. Analysis of the C4 genes in baleen whales using a human cDNA probe. Immunogenetics **32:**73–76.
56. **Stavenhagen, J. B., and D. M. Robins.** 1988. An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. Cell **55:**247–254.
57. **White, P. C., D. D. Chaplin, J. H. Weis, B. Dupont, M. I. New,**

and J. G. Seidman. 1984. The steroid 21-hydroxylase genes are located in the murine S region. Nature (London) **312:**465–467.
58. **White, P. C., M. I. New, and B. Dupont.** 1986. Structure of the human steroid 21-hydroxylase genes. Proc. Natl. Acad. Sci. USA **83:**5111–5115.
59. **Xu, X., and R. Doolittle.** 1990. Presence of a vertebrate fibrinogen-like sequence in an echinoderm. Proc. Natl. Acad. Sci. USA **87:**2097–2101.
60. **Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell.** 1984. The human LDL receptor: a cysteine-rich protein with multiple alu sequences in its mRNA. Cell **39:**27–38.
61. **Yoshioka, H., K. Morohashi, K. Sogawa, M. Yamane, S. Kominami, S. Takemori, Y. Okada, T. Omura, and Y. Fujii-Kuriyama.** 1986. Structural analysis of cloned cDNA for mRNA of microsomal cytochrome P-450(c21) which catalyses steroid 21-hydroxylation in bovine adrenal cortex. J. Biol. Chem. **261:**4106–4109.